



HAL
open science

**Proceedings MOSIM 2018 - 12e Conférence
Internationale de Modélisation, Optimisation et
SIMulation: “ L’essor des systèmes connectés dans
l’industrie et les services ”. Toulouse, France, 27-29 juin
2018**

Élise Vareilles, Alain Haït, Bernard Grabot, Gilles Savard

► **To cite this version:**

Élise Vareilles, Alain Haït, Bernard Grabot, Gilles Savard. Proceedings MOSIM 2018 - 12e Conférence Internationale de Modélisation, Optimisation et SIMulation: “ L’essor des systèmes connectés dans l’industrie et les services ”. Toulouse, France, 27-29 juin 2018. IMT Mines Albi, 692 p., 2018, 979-10-91526-06-7. hal-01989427

HAL Id: hal-01989427

<https://hal.science/hal-01989427>

Submitted on 22 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Toulouse, France, 27-29 Juin 2018

MOSIM2018

**12^e Conférence Internationale
de Modélisation, Optimisation et Simulation**

***L'essor des systèmes connectés
dans l'industrie et les services***

**12^e Conférence Internationale
de Modélisation, Optimisation et SIMulation**

MOSIM2018

27-29 Juin 2018 – Toulouse, France

***« L'essor des systèmes connectés dans l'industrie
et les services »***

**12th International Conference
on Modelling, Optimization and SIMulation**

MOSIM'18

June 27-29, 2018 – Toulouse, France

***“The rise of connected systems in industry and
services”***

Rédacteurs en chef : Élise Vareilles, Alain Haït, Bernard Grabot, Gilles Savard
Juin 2018

ISBN 979-10-91526-06-7 IMT Mines Albi

Sommaire

Sommaire	v
Éditorial	ix
Comité d'organisation 2018	xi
Comité scientifique 2018	xiii
Organisateurs	xvii
Sponsors	xix
1 – <i>A bi-objective model for clearance chain design</i> by Rihab Mechmech, Slim Harbi, Atidel B.Hadj Alouane and Sami Sboui	1
2 – <i>A continuous time inventory routing problem with energy minimization</i> by Yun He, Tolga Bektas, Cyril Briand and Nicolas Jozefowicz	9
3 – <i>A framework for supply chain risk analysis</i> by Saleh Eddine Ben Jbara, Pierre David and Gülgün Alpan	17
4 – <i>A new class of spectral risk measures</i> by Mohammed Berkouch, Ghizlane Lakhnati and Marcelo Righi	25
5 – <i>Algorithme d'approximation pour un problème d'affectation d'une commande multi-produits multi-fournisseurs</i> par Azeddine Cheref, Christian Artigues, Vincent Charvillat, Emmanuel Hébrard, Marie-José Huguet, Pierre Lopez, Sandra Ulrich Ngueveu et Yves Régent	31
6 – <i>Amélioration de l'efficacité énergétique dans un réseau électrique en appliquant les méthodes d'apprentissage et de clustering</i> par Mohamed Hascuri, Mostafa Derrhi et Mustapha Ait Rami ..	39
7 – <i>An approach to solve inventive problems based on the substance-field analysis and systems dynamics</i> by Jesús Delgado-Maciél, Guillermo Cortés-Robles, Stéphane Negny and Aidé Maldonado-Macias	45
8 – <i>An effective column generation algorithm for the inventory-routing problem with transshipment</i> by Wouter Lefever, El-Houssaine Aghezzaf and Khaled Hadj-Hamou	53
9 – <i>Analysing eco-system of ilities for resilience evaluation in system of systems</i> by Behrang Moradi Koutchi, Nicolas Daclin and Vincent Chapurlat	61
10 – <i>Application of simulation and lean techniques for improvement of glutam production : a case study</i> par Marzieh Ghiyasinassab, Nadia Lehoux, Sylvain Ménard and Caroline Cloutier	69
11 – <i>Apport d'agilité aux bioraffineries : vers une démarche collaborative et une servicisation des procédés</i> par Michelle Houngbé, Anne-Marie Barthe-Delanoë et Stéphane Negny	79
12 – <i>Apports des paradigmes sociaux dans les interactions entre objets communicants industriels : pour un social industriel iot</i> par Eddy Bajic et Oussama Hajlaoui	87
13 – <i>Approche comportementale pour une sylviculture communautaire durable : application à la forêt de la Faya (Mali)</i> par Souleymane Dicko, Daniel Roy, Sophie Hennequin et Abdel Kader Keita	95
14 – <i>Approche systémique pour le développement de systèmes embarqués connectés – conception et production d'un MOOC</i> par Claude Baron et Brigitte Daniel-Allegro	101
15 – <i>Approche probabiliste et multiobjectif pour la sélection de scénarios dans un graphe de projet</i> par Majda Lachhab, Thierry Coudert et Cédric Béler	109
16 – <i>Avez-vous identifié toutes les parties prenantes ?</i> par Karla Itzel Gomez Sotelo, Xin Yi, Claude Baron, Philippe Esteban, Citlalih Yollohtli Alejandra Gutierrez Estrada et Luis De Jesus Laredo Velazquez	117

17 – <i>Bi-objective mip formulation for the optimization of maintenance planning on french military aircraft operations</i> by Franco Peschiera, Alain Hait, Olga Battaia and Nicolas Dupin	125
18 – <i>Caractérisation de services industriels en réponse à appels d’offres</i> par Delphine Guillon, Elise Vareilles, Michel Aldanondo, Eric Villeneuve et Christophe Merlo	133
19 – <i>Centres d’appel d’urgence français (SAMU) : une démarche dirigée par les modèles pour le diagnostic de leur fonctionnement</i> par Éva Petitdemange, Franck Fontanili, Matthieu Lauras et Elyes Lamine	141
20 – <i>Column generation for the kidney exchange problem</i> by Lucie Pansart, Hadrien Cambazard, Nicolas Catusse and Gautier Stauffer	149
21 – <i>Conception d’un système de vérification de la diagnosticabilité par model checking à partir du modèle du système</i> par Mehdi Chankate, Alexandre Philippot, Pascale Marange et Véronique Carre-Menetrier	157
22 – <i>Contribution des architectures de contrôle holoniques à l’industrie 4.0</i> par Olivier Cardin, Damien Trentesaux et William Derigent	165
23 – <i>Couplage de simulateur de procédé et de modélisation ERTN pour la planification de systèmes énergétiques</i> par Alexandre Leclercq, Gilles Hétreux, Raphaële Hétreux et Pascal Floquet	173
24 – <i>Data visualization in conceptual design: developing a prototype to support decision making</i> by Xin Yi, Claude Baron, Rob Vingerhoeds, Alessandro Bertoni and Phillipe Esteban	181
25 – <i>Definition of states and modes as general concepts for system design and validation</i> by Ronan Baduel, Jean-Michel Bruel, Iulian Ober and Eddy Doba	189
26 – <i>Development of a framework for optimal design of closed-loop supply chains: application to battery recycling</i> by Magno Angel GonzÁlez-Rodríguez, Catherine Azzaro-Pantel, Jesus Ochoa Robles and Alberto Aguilar-Lasserre	197
27 – <i>Dynamiques des territoires et de l’organisation de l’enseignement supérieur et de la recherche</i> par Emmanuel Caillaud	205
28 – <i>Early checking of sysml models applied to protocols</i> by Pierre de Saqui-Sannes, Rob Vingerhoeds and Ludovic Aprville	213
29 – <i>Educational challenges for cyber-physical systems modelling</i> by Rob Vingerhoeds, Ahlem Mifdaoui and Pierre de Saqui-Sannes	221
30 – <i>Élaboration d’offres en relation client/fournisseur : vers une exploitation des connaissances pour l’ingénierie des risques</i> par Rania Ayachi, Delphine Guillon, François Marmier, Élise Vareilles, Michel Aldanondo, Thierry Coudert, Laurent Geneste et Yvan Beauregard	229
31 – <i>Enhancing operational fault diagnosis by assessing multiple operational modes</i> by Juan José Montero Jiménez and Rob Vingerhoeds	237
32 – <i>Enjeux financiers de DDMRP : une approche simulateur</i> par Stephanie Bayard et Frederic Grimaud	245
33 – <i>Entrepôt du futur : formalisation des caractéristiques clefs</i> par Dmitry Kucharavy, David Damand, Marc Barth et Ridha Derrouiche	253
34 – <i>Gestion de la demande en puissance en milieu industriel : état de l’art et stratégies pour la production de bois d’œuvre</i> par Laurence Dumont, Nadia Lehoux, Louis Gosselin et Hugues Fortin	261
35 – <i>Génération d’un modèle de situation en temps réel : application à la gestion de crise</i> par Audrey Fertier, Anne-Marie Barthe-Delanoë, Aurélie Montarnal, Sébastien Truptil et Frédérick Bénaben	269
36 – <i>Identification des risques, opportunités et facteurs critiques de succès de l’industrie 4.0 pour la performance industrielle des PME</i> par Alexandre Moeuf, Robert Pellerin, Samir Lamouri, Adrien Géraldes et Estefania Tobon-Valencia	277
37 – <i>Implementing scrum to develop a connected robot</i> by Diego Armando Díaz Vargas, Rui Xue, Claude Baron, Philippe Esteban, Rob Vingerhoeds, Citlalih Y. A. Gutierrez Estrada and Chao Liu	285
38 – <i>Inferring extended deterministic labelled transition system models of reactive system behaviours</i> by Luigi Novella, Manuela Tufo and Giovanni Fiengo	293
39 – <i>Ingénierie système agile par modèles simulables</i> par Philippe Gicquel	301

40 – <i>Intégration tactique-opérationnelle d'une planification hiérarchique avec fonction d'anticipation : étude de cas dans le secteur forestier au Québec</i> par Laureline Estivalet, Mustapha Ouhimmou et Amin Chaabane	309
41 – <i>Inventory location routing problem: issues, models and methods</i> by Sanaa Lerhlaly, Maria Lebbar, Sohaib Affi, Hamid Allaoui and Driss Ouazar	319
42 – <i>IWODEMO : un nouvel algorithme d'évolution différentielle hybride pour l'optimisation multi-objectifs</i> par Caroline Gagné, Mathieu Fournier, Aymen Sioud et Marc Gravel	327
43 – <i>Le pilotage des approches intégrées dans la planification et le contrôle de la production</i> par Angela Patricia Velasco Acosta et Christian Mascle	335
44 – <i>L'indispensable évolution des « transport management systems » pour le pilotage des flux de produits sensibles</i> par Quentin Schoen, Matthieu Lauras, Franck Fontanili, Sébastien Truptil and André Machado Alves	343
45 – <i>L'innovation organisationnelle au service de l'autonomisation en santé</i> par Pierre Yves Traynard, Mariana Dorsa, Muriel Londres et Naiditch Michel	351
46 – <i>Mise en œuvre d'un simulateur DEVS utilisant des réseaux de petri temporisés sur FPGA</i> par Clément Foucher et Vincent Albert	359
47 – <i>Mixed-integer and constraint programming for a multi-skill project scheduling problem with partial preemption</i> by Oliver Polo Mejia, Marie-Christine Anselmet, Christian Artigues and Pierre Lopez	367
48 – <i>Modélisation d'une chaîne logistique hybride par la simulation à événements discrets</i> par Najat Bara, Frédéric Gautier et Vincent Giard	375
49 – <i>Modèle de planification pour les activités de recommerce de produits électroniques</i> par Antoine Jeanjean et Nabil Absi	385
50 – <i>Multi level uncapacitated lot sizing problem considering working capital requirement</i> by Yuan Bian, Fabrice Gayraud, David Lemoine, Thomas Yeung and Nathalie Bostel	393
51 – <i>On the forecasting of traffic accidents based on real data</i> by Fernando Crespo, Elyn L. Solano-Charris and Jairo R. Montoya-Torres	401
52 – <i>Optimisation de l'affectation des patients aux médecins spécialistes avec prise en compte de la priorité clinique : cas du CRDS montréalais</i> par Vincent Nauze, Tasseda Boukherroub, Marc Paquet et Claude Olivier	407
53 – <i>Optimisation d'un plan de production multi-périodes avec une demande dynamique et des délais stochastiques</i> par Oussama Ben Ammar, B. Bettayeb et Alexandre Dolgui	415
54 – <i>Optimization of a hydrogen supply chain network design: application to an airport ecosystem</i> by Marialis Giraud Billoud, Jesus Ochoa Robles, Catherine Azzaro-Pantel and Alberto Alfonso Aguilar Lasserre	423
55 – <i>Optimization of analyzers configuration in a clinical laboratory: a mathematical model</i> by Sohrab Faramarzi Oghani, El-Ghazali Talbi, Martin Bué and Eric Varlet	431
56 – <i>Planification d'un complexe minier, de la mine à l'usine, sous contraintes de teneur en éléments du minerai</i> par Guillaume Battaia, Romain Guillaume et Caroline Thierry	437
57 – <i>Planification simultanée de la production, de la remise à neuf et du setup pour un système hybride non fiable</i> par Morad Assid, Ali Gharbi et Adnène Hajji	445
58 – <i>Planification à court terme dans un système de production complexe géré en DDMRP</i> par Pierre Baptiste et Robert Pellerin	453
59 – <i>Predictive cross-management of disaster plans in big data supply chains: fuzzy cognitive map approach</i> by Samia Chehbi-Gamoura, Ridha Derrouiche, Manisha Malhotra and David Damand	463
60 – <i>Processus de fabrication d'éléments préfabriqués : état de l'art et analyse d'un cas d'études</i> par Pierre Ultré, Nadia Lehoux et Sylvain Ménard	471
61 – <i>Proposition d'une modélisation pour l'amélioration des performances du secteur public</i> par Gautier Aubourg, François Galasso, Lamothe Jacques et Bernard Grabot	479
62 – <i>Proposition pour une démarche de modélisation des processus</i> par Sophie Lemoussu, Jean-Charles Chaudemar et Rob A. Vingerhoeds	487

63 – <i>Public construction projects in tunisia: current practices, existing problems and development of a decision support tool</i> by Sirine Touati, Safa Bhar Layeb and Jouhaina Siala Chaouachi	495
64 – <i>Questioning integration of verification in model-based systems engineering: a industrial perspective</i> by Christopher Laing, Pierre David, Eric Blanco and Xavier Dorel	503
65 – <i>Reconfigurable transfer line balancing problem: a new mip approach and approximation hybrid algorithm.</i> by Youssef Lahrichi, Sylvie Norre, Laurent Deroussi and Nathalie Grangeon	511
66 – <i>Revisiting digital mock-up for pme involved in systems engineering deployment</i> by Chapurlat Vincent, Lafon Maxence and Nastov Blazo	519
67 – <i>Robustesse, résilience : une brève synthèse des définitions au travers d’une analyse structurée de la littérature</i> par Antoine Clement, François Marmier, Daouda Kamissoko, Didier Gourc, Lien Wioland, Virginie Govaere et Julien Cegarra	527
68 – <i>RREFlex, an innovative tool for the synthesis of flexible heat exchanger network</i> by Lucille Payet, Gilles Hétreux, Raphaële Hétreux and Pascal Floquet	535
69 – <i>Safety analysis of pilot-system interaction</i> by Vatsal Pant, Jean-Charles Chaudemar and Hamid Demmou	543
70 – <i>Scheduling in additive manufacturing</i> by Anne-Lise Antomarchi, Séverine Durieux and Emmanuel Duc	549
71 – <i>Scoring risk of investment projects using nominal classification</i> by Ayeley Tchangani	557
72 – <i>Semigreedy heuristics for the flexible jobshop scheduling problem with multiple process plans and sequence dependent setup times</i> by Nicolau Santos, Gonçalo Figueira and João Pedro Pedroso	565
73 – <i>Simulation of communication based train control for robustness assessment of a mixed traffic management system</i> by Sélim Cornet, Paul Bouvarel, Christine Buisson, François Ramond and Joaquin Rodriguez	573
74 – <i>Simulation of value for environmentally adaptable packaging solutions architecture</i> by Carlo Leardi	581
75 – <i>Sustainable development in serious games: rethinking game-based learning strategies for master’s degree engineers</i> by Marios Stanitsas, Élise Vareilles, Konstantinos Kirytopoulos and Michel Aldanondo	589
76 – <i>Système de systèmes dans les architectures PHM</i> par Alexandre Sarazin, Sébastien Truptil, Aurélie Montarnal et Jacques Lamothe	597
77 – <i>The digital twin at the service of storage zone optimisation and assembly line simulation</i> by Ariane Piel, Mohamed Benazouz and Charles Brunet	605
78 – <i>The use of internal cross-docking in just-in-time plants</i> by Mohammed Hichame Benbitour and Evren Sahin	613
79 – <i>Towards a new methodology to improve maintenance performance based on the process approach and lean thinking</i> by Frakso Btissame, Mokhlis Ahmed, Elfezazi Said and Elhachadi Ahmed	621
80 – <i>Towards the design of a distributed aircraft flight control system connected to simulation components</i> by Jean-Baptiste Chaudron and David Saussie	629
81 – <i>Une approche centrée usage pour accompagner l’introduction d’un outil numérique pour la prise en charge à domicile</i> par Christophe Merlo et Katarzyna Borgiel	637
82 – <i>Urban freight by rail: an algorithm genetic for optimizing the transport of goods</i> by Abderrahim Sahli, Sana Belmokhtar-Berraf, Doina Uritu and Walid Behiri	645
83 – <i>Validation de la fiabilité de composants au stade de la conception</i> par Daoud Ait-Kadi, Abdelhak El Khaïr El Idrissi et Nathalie Gouget	653
84 – <i>Vers une cartographie de processus explicite pour le modèle Demand Driven Adaptive Enterprise</i> par Guillaume Martin, Pierre Baptiste, Jacques Lamothe, Romain Miclo et Matthieu Lauras	661
85 – <i>Vers une innovation dirigée par les modèles – application en télémédecine</i> par Yosra Chaher, Jean-Pierre Belaud et Hervé Pingaud	675
86 – <i>Vers une méthode de gestion intégrée risques – processus pour la sécurisation du circuit du médicament</i> par Rafika Thabet, Amine Boufaied, Elyes Lamine, Ouajdi Korbaa et Hervé Pingaud	685

Éditorial

Créée en 1997, la conférence MOSIM est un lieu d'échange entre chercheurs, enseignants et industriels actifs dans le domaine de la Modélisation, de l'Optimisation et de la SIMulation des systèmes.

Après des éditions récentes à Hammamet, Bordeaux, Nancy et Montréal, c'est à Toulouse que s'est déroulée la 12^e édition de la conférence, plus particulièrement orientée sur le thème de l'essor des systèmes connectés dans l'industrie et les services. Il a en effet semblé important aux organisateurs de proposer un espace de réflexion prospective et de débat sur la révolution actuelle provoquée par les technologies connectées sur les systèmes de production de biens et de service, comme d'ailleurs dans de nombreux autres domaines. Par les nouvelles possibilités offertes au niveau de la production et du traitement de données, ces technologies sont en effet une composante majeure de l'Industrie et du Service 4.0. Elles impactent ainsi l'ensemble du cycle de vie des systèmes et remettent en question toutes les catégories de processus organisationnels (décisionnels, opérationnels et supports), ainsi que les modèles stratégiques des acteurs économiques.

En quelques chiffres, MOSIM 2018 s'est traduit par 15 sessions spéciales, 104 soumissions et 91 communications acceptées et présentées pour un total de 358 auteurs appartenant à 20 nationalités... Face aux multiples sollicitations de notre communauté, ce résultat peut être considéré comme un succès, mais nous invite néanmoins à réfléchir à l'évolution du congrès pour assurer sa pérennisation.

Le Prix de jeunes chercheurs a fait l'objet de 42 soumissions, 6 ayant été retenues pour des présentations détaillées. À l'issue de ces présentations, le premier prix a été décerné à Wouter Lefever de l'université de Gand pour un travail en logistique sur l'optimisation de tournées de véhicules avec transbordement. Le second prix a été décerné à Rafika Thabet de l'université de Sousse pour un travail sur la minimisation des risques dans les processus de prise en charge médicamenteuse des patients dans les établissements hospitaliers. De nouveau, toutes nos félicitations à ces jeunes chercheurs !

Nous espérons que les textes rassemblés ici des communications présentées lors du congrès vous seront utiles.

Élise, Alain, Bernard, Gilles

Comité d'organisation 2018

Présidents



Alain HAÏT

Institut Supérieur de l'Aéronautique et de l'Espace
ISAE-SUPAERO



Élise VAREILLES

IMT Mines Albi-Carmaux
ETS Montréal

Membres du comité

Anne-Marie BARTHE-DELANOË	INPT-ENSIACET, Toulouse
Olga BATAÏA	ISAE-SUPAERO, Toulouse
Jean-Pierre BELAUD	INPT-ENSIACET, Toulouse
Cédric BELER	ENI de Tarbes
Cyril BRIAND	Université Toulouse 3 Paul Sabatier
Vincent CHAPURLAT	IMT Mines Alès
Paul GABORIT	IMT Mines Albi-Carmaux
François GALASSO	Université Toulouse 1 Capitole, IUT QLIO
Romain GUILLAUME	Université Toulouse 2 Jean Jaurès
Marie-José HUGUET	INSA Toulouse
Jacques LAMOTHE	IMT Mines Albi-Carmaux
Catherine MANCEL	ENAC, Toulouse
Jacky MONTMAIN	IMT Mines Alès
Stéphane NEGNY	INPT-ENSIACET, Toulouse
Sandra Ulrich NGUEVEU	INPT-ENSEEIH, Toulouse
Nicolas PERRY	ENSAM, Bordeaux
Hervé PINGAUD	Institut National Universitaire Champollion, Castres
Emmanuel RACHELSON	ISAE-SUPAERO, Toulouse
Caroline THIERRY	Université Toulouse 2 Jean Jaurès
Rob VINGERHOEDS	ISAE-SUPAERO, Toulouse
Delphine GUILLO	IMT Mines Albi-Carmaux (Doct.)
Abdourahim SYLLA	IMT Mines Albi-Carmaux (Doct.)
Rania AYACHI	IMT Mines Albi-Carmaux (Doct.)
Hamidreza POURZAREI	IMT Mines Albi-Carmaux (Doct.)
Luis Ignacio GARCÉS MONGE	IMT Mines Albi-Carmaux (Doct.)
Pierre-Antoine MORIN	ISAE-SUPAERO, Toulouse (Doct.)
Luca MOSSINA	ISAE-SUPAERO, Toulouse (Doct.)
Franco PESCHIERA	ISAE-SUPAERO, Toulouse (Doct.)
Marios STANITSAS	NTUA, Grèce et IMT Mines Albi-Carmaux, France (Doct.)

Comité scientifique 2018

Présidents



Prof. Bernard Grabot

École Nationale d'Ingénieurs de Tarbes

Bernard Grabot est Professeur à l'ENI de Tarbes, où ses enseignements concernent principalement la gestion de production, les systèmes d'information et les chaînes logistiques. Ses activités de recherche portent sur l'ingénierie des connaissances, avec des applications à la gestion des chaînes logistiques et l'aide à la décision. Bernard Grabot est membre des groupes de travail de l'IFAC 3.2 « Computational Intelligence in Control » et 5.1 « Manufacturing Plant Control » et du groupe de travail de l'IFIP 5.7 « Advances in Production Management ». Il a été impliqué dans plusieurs projets européens (CRAFT, INTERREG, etc.) et appartient aux comités éditoriaux de « International Journal of Computational Intelligence Research, International Journal of Production Research » et « Computers in Industry ».



Prof. Gilles Savard

École Polytechnique de Montréal
et directeur général, IVADO

Le Professeur Gilles Savard a obtenu un doctorat en génie électrique avec spécialisation en recherche opérationnelle de l'École Polytechnique en 1989. Il est devenu professeur au Département de mathématiques appliquées de Polytechnique Montréal en 1993 puis directeur du Département de mathématiques et génie industriel de 2001 à 2007 et directeur de la recherche, de l'innovation et des affaires internationales de Polytechnique de 2007 à 2016. Il occupe depuis le poste de directeur général de l'Institut de valorisation des données (IVADO).

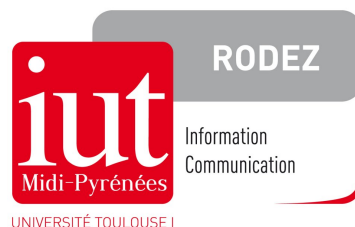
Membres du comité

Bruno AGARD	École Polytechnique de Montreal, Canada
El-Houssaine AGHEZZAF	Université de Gand, Belgique
Daoud AÏT-KADI	Université Laval, Québec, Canada
Michel ALDANONDO	IMT Mines Albi, France
Giua ALLESSANDRO	University of Cagliari, Italy
Beghi ALLESSANDRO	University of Padova, Italy
Hassane ALLA	Université Joseph Fourier, France
Frank ALLGÖWER	Université de Stuttgart, Allemagne
Gülgin ALPAN-GAUJAL	Grenoble INP, France
Abdelhakim ARTIBA	Université de Valenciennes, France
Pierre BAPTISTE	École Polytechnique de Montréal, Canada
Claude BARON	INSA Toulouse, France
Fernando BARROS	Université de Coimbra, Portugal
Anne-Marie BARTHE-DELANOË	Laboratoire de Génie Chimique, France
Francesco BASILE	Université de Salerne, Italy
Yvan BEAUREGARD	École de Technologie Supérieure, ETS, Canada
Jean-Pierre BELAUD	Laboratoire de Génie Chimique, France
Alain BERNARD	École Centrale de Nantes, France
Jean-Charles BILLAUT	Université François-Rabelais de Tours, France
Éric BONJOUR	ENSGSI Nancy et AFIS, France
Valérie BOTTA-GENOULAZ	Institut National des Sciences Appliquées, Lyon, France
Hafida BOULOIZ	ENSA Agadir, Maroc
Jean-Paul BOURRIÈRES	Université de Bordeaux 1, France

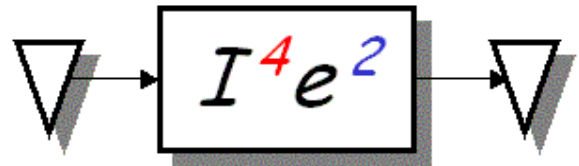
Emmanuel CAILLAUD	Université de Strasbourg et AFIS, France
Simona CARAMIHAI	Université Politechnique de Bucharest, Romania
Jacques CARLIER	Université Technologique de Compiègne, France
Leardi CARLO	TETRA PAK Packaging Solutions S.p.A., Italy
Pierre CASTAGNA	Université de Nantes, France
Vincent CHAPURLAT	IMT Mines Alès, France
Youness CHATER	École Nationale des Sciences Appliquées de Tanger, Maroc
Samia CHEHBI-GAMOURA	Institut national des Sciences appliquées de Lyon, France
Anis CHELBI	École Supérieure des Sciences et Techniques de Tunis, Tunisie
Philippe CHRÉTIENNE	Université Paris 6, France
Thierry COUDERT	École Nationale d'Ingénieurs de Tarbes / LGP, France
Nicolas DACLIN	IMT Mines Alès, France
Stéphane DAUZÈRE-PÉRÈS	IMT Mines Saint-Etienne, France
Isabel DEMONGODIN	Aix-Marseille Université, France
Laurent DEROUSSI	LIMOS, France
Ridna DERROUCHE	EM Strasbourg, France
Maria DI MASCOLO	G-SCOP Grenoble, France
Georgi DIMIROVSKI	Université Dogus, Istanbul, Turquie
Alexandre DOLGUI	IMT Atlantique, Nantes, France
Maria Grazia DOTOLI	Politecnico di Bari, Italy
Antonio DOURADO	Université de Coimbra, Portugal
Yves DUCQ	Université de Bordeaux 1, France
Florin-Gheorghe FILIP	Académie Roumaine, Roumanie
Clément FOUCHER	LAAS, CNRS, France
Franck FONTANILI	IMT Mines Albi, France
Jose FRAMINAN	Université de Séville, Spain
Yannick FREIN	Grenoble INP, France
Georg FREY	Université de Sarre, Allemagne
Claudia FRYDMAN	Aix-Marseille Université, France
Maryam GALLAB	École Mohammadia des Ingénieurs, Maroc
Thierry GARAIX	IMT Mines Saint-Etienne, France
Laurent GENESTE	École Nationale d'Ingénieurs de Tarbes / LGP, France
Ali GHARBI	École de Technologie Supérieure, Montréal, Canada
Vincent GIARD	Université Paris-Dauphine, France
Dave GOLDSMAN	Georgia Tech, Atlanta, USA
Gilles GONCALVES	Université d'Artois, France
Alain HAÏT	ISAE SUPAERO, France
Yannick KERGOSIEN	École Polytechnique de l'Université de Tours, France
Konstantinos KIRYTOPOULOS	University of South Australia, Australia
Jus KOCIJAN	Université de Ljubljana, Slovenie
Elyes LAMINE	Centre universitaire Jean François Champollion, France
Jacques LAMOTHE	IMT Mines Albi, France
Samir LAMOURI	Arts et Métiers Paristech, France
Matthieu LAURAS	IMT Mines Albi, France
Pierre LOPEZ	LAAS-CNRS, Toulouse, France
Mateo MANEL	Université Polytechnique de Catalogne, Espagne
Nadine MESKENS	Université Catholique de Louvain (UCL), Belgique
Stéphane NEGNY	Laboratoire de Génie Chimique, France
Mustapha NOUR EL FATH	Université Laval, Québec, Canada
Mustapha OUHIMMOU	École de Technologie Supérieure, Montréal, Canada
Hervé PANETTO	Université Henri Poincaré, Nancy, France
Robert PELLERIN	École Polytechnique de Montréal, Canada
Henri PIERREVAL	SIGMA Clermont, France
Hervé PINGAUD	Université Champollion, France

Raul POLER	Université Polytechnique de Valence, Espagne
Christian PRINS	Université de Technologie de Troyes, France
Goran PUTNIK	Université de Minho, Portugal
Tatiana REYES	Université de technologie de Troyes, France
Fouad RIANE	École Centrale Casablanca, Maroc
Bertrand ROSE	Université de Strasbourg, France
Angel RUIZ	Université de Laval, Canada
Evren SAHIN	Centrale-Supélec, France
Nathalie SAUER	Université Paul Verlaine, Metz, France
Mamadou SECK	Delft University of Technology, Pays-Bas
Patrick SIARRY	Université Paris Est Créteil, France
Caroline THIERRY	Université Toulouse Jean Jaurès, France
André THOMAS	Université de Lorraine, Nancy, France
Damien TRENTESAUX	Université de Valenciennes, France
Antonios TSOURDOS	Université de Cranfield, Royaume Uni
D.A. (Bert) VAN BEEK	Université Technologique d'Eindhoven, Pays-Bas
Élise VAREILLES	IMT Mines Albi, France
	ETS Montréal, Canada
François VERNADAT	Université de Lorraine, France
Rob VINGERHOEDS	ISAE SUPAERO, France
Gabriel WAINER	Université de Carleton, Canada
Marino WIDMER	Université de Fribourg, Suisse
Enver YUCESAN	INSEAD, Fontainebleau, France
Janan ZAYTOON	Université de Reims, France
Marc ZOLGHADRI	LISMMA, France
Noureddine ZERHOUNI	ENSMM Besançon, France

Organisateurs



Sponsors



Forward Thinking



Institut d'Études des
Systèmes Complexes
de Toulouse



A BI-OBJECTIVE MODEL FOR CLEARANCE CHAIN DESIGN

MECHMECH R., HARBI S., B.HADJ-ALOUANE A.

OASIS Laboratory - ENIT
University of Tunis El Manar
Tunis, Tunisia
slim.harbi@enicarthage.rnu.tn
atidel.hadj@enit.utm.tn
rihab.mechmech@utm.tn

SBOUI S.

SQLI services
Manouba, Tunisia
JASSP SAS.
Champs sur Marne, France
sbouis@yahoo.fr

ABSTRACT: *Unsold inventory represents a significant issue for many companies. Some clearance strategies have been proposed in previous researches to drain this undesirable inventory. The goal of this paper is to help companies in the choice of the most appropriate clearance strategies to improve the efficiency of the clearance chain. In this research, we develop a bi-objective mathematical model that captures the existence of conflicting objectives in the clearance chain. The two objectives used in this study are the total performance and the total risk associated with the clearance chain while allocating the adequate quantity to each strategy. The bi-objective model has been solved using the ϵ -constraint method. The effectiveness of the mathematical model and the proposed solution method is demonstrated in a numerical example from an illustrative case study. Finally, we have conducted a sensitivity analysis where results give an important insight for future researches.*

KEYWORDS: *clearance chain, strategies' selection, MOO, ϵ -constraint method*

1 INTRODUCTION

The management of returned inventory is not a new research topic. Indeed, such a goal is one of the major concerns of the reverse logistics; by collecting and sending them to recovery plants, returned products begin their second life. Also, based on [Sasikumar, P. and Kannan, G. , 2008], we reveal that the managing of the end-of-life (EoL) products has become an area of rapidly growing importance in reverse supply chain management (RSCM). Whether it's a surplus, a used item, a returned product, an end-of-life or an end-of-use one, an unsold product must have the most suitable and cost-efficient recovery process and management since the difference between a product's success and failure, according to [Marisa P. de Brito and Dekker, R., 2002], depends often on how the end-of-life is managed.

[Sboui, S., 2006] has underlined the importance of the management of unsold inventory since it represents a challenge that should not be skipped throughout the supply chain. Having such inventory, the main objective for a company, whether it is a retailer, a manufacturer or a wholesaler, is to drain the total amount as it can provide benefits, while realizing the best value from it.

According to [S. Rogers, D. et al., 2012] companies should focus into two main factors: to liquidate the maximum quantity of its unsold items in the shortest possible period since, the products' value is forfeited over time. These two objectives represent the primary goals of the clearance chain, which has been introduced by [Mechmech, R. et al., 2017] as a number of combined

clearance strategies, having each its advantageous and drawbacks. Within this situation, the decision maker is facing several conflicting goals, such as to maximize benefits (in term of profit), to minimize the delay, to decrease the adverse effects that can be caused, to use the most suitable strategies regarding the nature of the products, the initial amount of unsold item and to allocate the adequate quantity to each clearance strategy. They have also proposed a framework based on the AHP method in order to have a helpful ranking of the different clearance strategies. The evaluation of these strategies is based on two attributes: the performance and the risk of each one. In addition, to liquidate the total amount of unsold inventory, a company has to combine many strategies making the clearance channel. Therefore, the next major step for the company is to optimize its clearance channel by choosing the most suitable combination of clearance strategies that maximizes the total performance and minimizes the channel risk. Once the strategies' selection has been done, the optimization of the liquidated quantity by each strategy in the clearance chain represents, also, a crucial goal. Consequently, the main objective of this research is to help the decision maker, based on an appraisal of the diverse strategies, to select and choose the most adequate ones, that optimize the entire clearance chain, and to allocate to each strategy the optimal quantity that optimizes the clearance network.

Clearance channel evaluation and selection is a multi-criteria decision-making problem that includes both quantitative and qualitative factors. The problem can be decomposed into two interrelated sub-problems:

- Which strategies should be selected?
- How much/many should be allocated to each selected strategy?

In order to deal with our problem, we will rely also on the multi-objective optimization (MOO) which has been used in several topics, such as in the closed loop supply chain (CLSC) network design [Ramezani, M. et al., 2013], in the partner selection problem [Dash Wu, D. et al., 2010] [Kannan, D. et al., 2013] [Yeh, W. and Chuang, M., 2011] in the forward and green supply chain and in many other researches and topics.

The remainder of this paper is structured as follows: section 2 reviews the current literature on clearance strategies and different MOO models used in strategies selection, partners' evaluation and order allocation problems. Section 3 proposes a formulation of the problem as a bi-objective (risk, performance) optimization model. Then, section 4 develops the solving approach using the ϵ -constraint method. An illustration is presented in section 5 that details the computational results. Finally, section 6 concludes the paper by pointing out the main results and provides direction for future research and expansions.

2 LITERATURE REVIEW

2.1 Clearance chain frameworks

According to [S. Rogers, D. and Tibben-Lembke R, 2001], the reverse logistics has been defined as “the movement product or material in the opposite direction for the purpose of creating or recapturing value, or for proper disposal”. In other words, this system has been proposed in order to manage returned products. We have noted that a large number of studies, through the literature analysis, have introduced a four-main-steps based reverse system [Lambert, S. et al., 2011]: gatekeeping, collection, sorting, and disposal. The three first steps, according to [Lambert, S. et al., 2011] are focusing on the choice of the product that can be considered into the reverse system, a grouping or a categorization of the different chosen products and finally, the fate and the firm's decisions about each product/category. The disposal represents the choice of the returned product destination which, according to [Giuntini, R. and Andel, T., 1995a], involves two alternatives: renewal or removal. The first one covers the reselling (to a second market), recycling, repackaging, repairing, refurbishing or remanufacturing, while the second possibility, the removal from the reverse process, is based mainly on the disposal to landfill or incineration [Giuntini, R. and Andel, T., 1995a].

[Agrawal, S. et al., 2015] has bring out the importance, of the secondary market in a reverse logistics network research area. Indeed, he has classified the literature on the reverse logistics into five categories among which we find the category “reverse logistics network from secondary market perspective” with an approximately rate

of 40 % of the examined publications which are focusing on the opportunities for introducing the secondary market into the reverse logistic network design. He also has mentioned that a significant flow is directed to this market thanks to the profit created.

[S. Rogers, D. et al., 2012] has highlighted the complexity of the management of the reverse physical flow and the decision choice. The authors presented several disposal options that a retailer can choose. To make the most appropriate choice, a retailer (or a manufacture) faces significant number of uncertainties.

Very little research has explored how companies dispose of their unsold product. According to [Mechmech et al., 2017], no company can escape from unsold inventory. Therefore, having a global strategy for unsold inventory management brings certainly much more benefits than only returned products management.

We can see that for a company having unsold inventory to liquidate, the set of operation and decisions to be taken is very similar to the one adopted in order to manage the reverse flow. Thus, relying on [S. Rogers, D. et al., 2012] and [Mechmech, R. et al., 2017] we have expanded this idea by proposing a framework for the unsold inventory flow management to be considered by the manufacturer, the retailer and the wholesaler (Figure 1). This unsold inventory includes returned products.

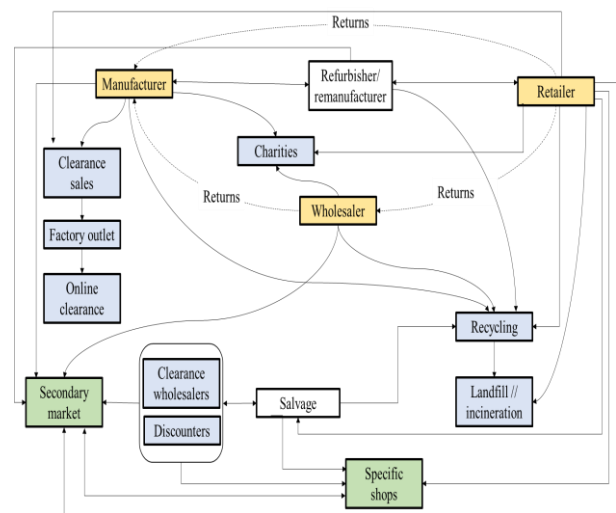


Figure 1: Overview of the clearance process

An exhaustive list of options to liquidate the unsold inventory (Figure1, Table 1) has been identified through the analysis of the business practices in diverse sectors (textile, IT, cosmetics, household appliances). The proposed framework is composed of eight main clearance strategies that have been classified into two categories: the cash recovery strategies (S1..S5) and the disposal ones (S6..S8), as mentioned in Table1.

Strategies	Comments
S1: Clearance sales	A solution that drains much quantity of unsold items.
S2: Factory outlet	A retail store that sells discounted price items.
S3: Online clearance	Company's specific clearance websites.
S4: Discounters	Liquidators that buy unsold products and resale them on <i>BtoC</i> business.
S5: Clearance wholesalers	Similar to discounters but resell products in bulk (<i>BtoB</i> business).
S6: Recycling	Generally used in electronics components.
S7: Destruction	Preserve the brand image in some cases.
S8: Donation	To promote enterprise social responsibility

Table 1 : Clearance strategies [Mechmech, R. et al., 2017]

The cash recovery strategies are divided into two classes: in-house (S1..S3) and external (S4..S5). In-house strategies represent the bunch of adopted clearance solutions by the firm using its own resources; the others constitute the strategies that use an external way.

In the same study, authors have proposed a performance-risk AHP process that supports decision makers to rank the clearance strategies based on two main criteria and six sub-criteria:

- Performance (P): Drained volume, Flow speed, Benefit/Price.
- Risks (R): Deterioration of the brand image, Cannibalization, Change in the consumer buying behavior.

Based on the company's vision and goals, the decision maker can assess the performance (P_i) and the risk (R_i) of each strategy (S_i). The second step is to evaluate the clearance channel as a combination of two or more clearance strategies since, generally, only one strategy is not able to liquidate the total amount of unsold inventory.

Once a company has chosen a strategy, the next step is to deal with new partners, such as discounters, recycling providers, charities, etc. Since the problem of the selection of the clearance strategies is not well studied in the academic literature [S. Rogers, D. et al., 2012], in the next section, we will focus on similar research topics.

2.2 MOO models in supply chain management

It can be noted that a wide range of studies, on the closed loop supply chain (CLSC) network, uses the multi-objective optimization. [Amin, S.H. and Zhang, G., 2013] presents a bi-objective model aiming to minimize the cost while maximizing the use of clean technology. They have used two solving approaches, the weighted

sum method and the ϵ -constraint one. They have shown that the second method leads to better results. [Ramezani, M. et al., R.,2013] has introduced a multi-objective model for forward/reverse logistics, aiming to maximize profit, customer responsiveness and quality. The authors have exploited the ϵ -constraint method. [Feng, Du. and W. Evans, G., 2008] has elaborated a bi-objective model that aims to minimize the total cost and the tardiness of the cycle time. In order to solve their model, they have used a combination of three algorithms: scatter search, the dual simplex method and the constraint method.

In the green supply chain, MOO models are also presented in a significant number of studies. [Kadziński, M. et al., 2017,] focused on three objectives, costs, CO₂, and fine dust emissions. To fix the Pareto front, the authors have adopted four approaches including the ϵ -constraint method. Many researches have used the MOO models to deal with suppliers and /or partners selection and optimal order allocation [Devika Kannan et al., 2013] [Yeh, W. and Chuang, M., 2011] [Narasimhan, R. et al., 2006] [Wadhwa, V. and Ravindran, A.R., 2007]. [Xia, W. and Wu, Z., 2007.] has used an integrated approach that calculates the performance scores of different suppliers using the AHP method and then incorporates them into a multi-objective mixed integer-programming model to select the best one.

MOO has been largely used in the routing problem. [Zhao, J. and Verter, V., 2015] have presented a bi-objective model for the Location-Routing Problem (LRP) that seeks to minimize the total environmental risk and the total cost. [Ghezavati, V.R. and Beigi, M., 2016] have introduced a bi-objective model for the Location-Routing Problem with time window (LRPTW) in reverse logistics. The authors have used the non-dominated sorting genetic algorithm II (NSGA-II) and the ϵ -constraint method. [Keshavarz, E. and Toloo, M., 2015], based on the traditional minimum cost flow problem, have presented a bi-objective minimum cost-time flow problem that aims to minimize the total shipping cost and the total shipping time.

Generally, when dealing with a multi-objective problem, we cannot talk about optimal solution for all objectives but rather we speak about the pareto optimality. The set of solution is, thus, called the pareto optimal solution, and is defined according to [Demir, E. et al., 2014] as the set of solutions that "cannot be improved in one objective function without deteriorating the performance of at least one of the rest". For an actual issue, decision makers, faced with the set of pareto solutions, should choose the most preferred and adequate solution in order to solve their problem.

To solve multi-objective problem, many approaches more or less sophisticated can be adopted such as the iterative method [Mostert, M. et al., 2017,], the goal programming presented by [Charnes, A. and Cooper, W.W., 1977], the evolutionary algorithm [Konaka, A. et al.,

2006] and other approaches. In our case, we have applied the well-known one, used in almost all presented studies, the ϵ -constraint method. It aims to transform the original multi-objective problem into a single-objective one and it can be applied to non-convex multi-objective problems.

Based on the literature, and in order to select the most appropriate clearance strategy, we will incorporate the AHP results, mined from [Mechmech, R. et al., 2017], into a multi-objective optimization model to choose the most suitable strategies and to decide the optimal liquidated quantity for each one.

3 PROBLEM FORMULATION

3.1 Clearance chain specifications

In this section, the clearance strategies selection problem is formulated as a network problem (Figure 2). The ten nodes represent the eight clearance strategies, a source and a sink.

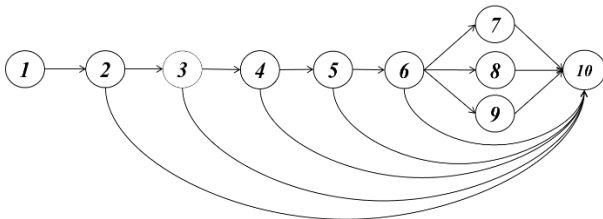


Figure 2: Clearance network flow

Let G be a directed network, having n nodes, with a capacity C_{ij} , a performance P_{ij} and a risk R_{ij} associated to every arc (i,j) . An arc (i,j) represents, if $j=10$, the amount of the closed-out quantity by the strategy i . It reflects, in other words, the amount which has been removed from the initial unsold inventory by selling it using strategy i . In this case, the different values of the triplet (P_{ij}, R_{ij}, C_{ij}) are strictly positive (Figure 3). The weights of the criteria risk (R_{ij}) and performance (P_{ij}) are generated by the AHP method based on the expert's judgements and the (C_{ij}) represents the full capacity of the strategy i . Thereafter, in order to be able to sell off the residual quantity from a node i , we must transmit them to another option/strategy j . Consequently, if $j \neq 10$, the capacity of the arc (i,j) must be unlimited to be able to pass the integrality of the remaining unsold products and the values of P_{ij} and R_{ij} are null since the arc does not represent a strategy use (Figure 3).

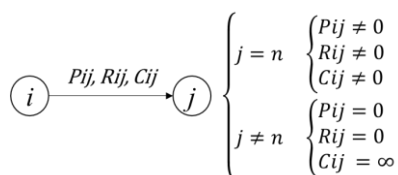


Figure 3: Condition in the triplet (P_{ij}, R_{ij}, C_{ij})

The problem may be stated as follows: define the flow of unsold inventory through the network to liquidate the maximum quantity at minimal risk and maximal performance. To deal with this problem, we propose in the next section a bi-objective mode.

3.2 Bi-objective optimization model

Since each strategy (S_i) has a specific performance (P_i) and risk (R_i), a global optimization of the clearance chain will aim to maximize the global clearance chain performance while minimizing the global clearance chain risk. In our case, we consider that the clearance chain performance depends on the liquidated quantity using each strategy. However, the risk depends whether or not the strategy in question is used.

Parameters:

- N: set of clearance strategies.
- Q: the total amount of unsold inventory.
- P_i : the performance of the strategy i .
- R_i : the risk caused by the strategy i .
- C_i : the capacity of the strategy i .

Decision variables:

- x_i : the percentage of the liquidated quantity using the strategy i .
- Y_i : a binary variable to consider the risk of a strategy i once it is used to liquidate a portion of the unsold products, that is, $x_i > 0$.

To deal with this problem, we propose the following model:

$$\text{Max } \sum_{i=1}^n P_i * x_i \quad (1)$$

$$\text{Min } \sum_{i=1}^n R_i * Y_i \quad (2)$$

s. t

$$0 \leq x_i * Q \leq C_i \quad (3)$$

$$\sum_{i=1}^n x_i = 1 \quad (4)$$

$$Y_i = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i = 0 \end{cases} \quad (5)$$

$$Y_i \in \{0,1\} \quad (6)$$

$$x_i \geq 0; i \in [1..N] \quad (7)$$

As x_i are real valued variables, our model is a bi-objective mixed-integer linear program (MILP).

Equations (1) and (2) represent respectively the total performance and the total risk of the clearance chain.

Constraint (3) represents the capacity strategy. Indeed, each strategy has a known capacity to liquidate unsold product. Constraint (4) forces the liquidation of the entire initial volume of unsold items. Constraints (5) represents a binary variable which, allows to take into account the strategy's risk once it has been chosen in the clearance channel, else it takes the null value. Constraints (6) and (7) are restrictions on the decisions variables. Constraint (5), which reflects a correlation between x_i and Y_i , adds a new challenge for our model.

Therefore, we will replace it with two constraints:

$$\begin{cases} xi \leq M * Yi \\ Yi \leq M * xi \end{cases} \quad (8)$$

Where M is a very large positive number.

4 SOLUTION METHOD: ϵ -CONSTRAINT METHOD

The ϵ -constraint method has been introduced by [Haimes, Y. Y. et al., 1971] as a scalarization method that aims to transform an m-objective problem into a single-objective problem with (m-1) constraints. In our research, since we have a bi-objective model, we can consider two options, (Table 2). In the first formulation, we have considered the performance as our objective function, thus the risk will be necessarily switched to a constraint. The insight of such a model is to maximize the performance while having a defined upper bound and a maximum risk value (ϵ_1) fixed beforehand by the manager. In the second formulation, the risk will be presented as an objective function and the performance as a constraint. In this case, the principle idea is to minimize the risk while maintaining a minimal performance level. Therefore, the performance constraint will be measured by a lower bound instead of an upper bound. The set of pareto-optimal solutions is generated by solving this problem several times, while varying the ϵ_j 's values.

1 st formulation	2 nd formulation
$\begin{cases} \text{Max } \sum_{i=1}^n Pi * xi \\ \text{s.t } \sum_{i=1}^n Ri * Yi \leq \epsilon_1 \end{cases}$ <p>Where: ϵ_1 represents an upper bound to the risk objective.</p>	$\begin{cases} \text{Min } \sum_{i=1}^n Ri * Yi \\ \text{s.t } \sum_{i=1}^n Pi * xi \geq \epsilon_2 \end{cases}$ <p>Where: ϵ_2 represents a lower bound to the performance objective.</p>

Table 2 : Problem formulations with ϵ -constraint method

In this paper, we consider the first formulation that aims to maximize the performance while minimizing the risk. We propose the following model:

$$\text{Maximize } (\sum_{i=1}^n Pi * xi) \quad (9)$$

$$\text{s.t } \sum_{i=1}^n Ri * Yi \leq \epsilon_1 \quad (10)$$

$$0 \leq xi * Q \leq Ci \quad (11)$$

$$\sum_{i=1}^n xi = 1 \quad (12)$$

$$\begin{cases} xi \leq M * Yi \\ Yi \leq M * xi \end{cases} \quad (13)$$

$$Yi \in \{0,1\} \quad (14)$$

$$xi \geq 0; \forall i \in [1..N] \quad (15)$$

5 COMPUTATIONAL RESULTS

5.1 Data setting

The different values of P_i and R_i , are extracted from the illustration of the performance-risk based AHP method where the different judgement have been fixed by the experts (Table3).

Strategies	P_i	R_i
S1	0.211	0.058
S2	0.147	0.161
S3	0.087	0.174
S4	0.093	0.220
S5	0.085	0.220
S6	0.126	0.031
S7	0.120	0.031
S8	0.133	0.106

Table 3 : P_i and R_i values [Mechmech et al., 2017]

The capacity of each strategy can be fixed either by the usual liquidated quantities, for example for the clearance sales and factory outlet, or by a proposed volume, like the discounters and wholesalers. Indeed, for the clearance sales and factory outlet strategies, the firm based on its sales historical data, can estimate a clear insight and a sufficiently accurate approximation of the liquidated quantities by these two strategies. Concerning, the online clearance, its capacity value is closely linked to the unsold product storage space capacity of the company. For the external strategies, the discounters or the wholesalers generally propose their maximum and minimum capacity's values. Finally, for the disposal strategies, their capacities are generally unlimited, but, for those that are costly, like the destruction and recycling, their capacities' values can be fixed from an early stage according to their allocated cost from the total company budget. In this illustration, we have supposed, for the disposal strategies, that they have an unlimited capacity and we have fixed approximated capacities for the others strategies (Table 4). The considered amount of unsold items is $Q=100000$ units.

Strategies	Capacities	C_i
S1 : Clearance sales	50 %	50000
S2 : Factory outlet	30 %	30000
S3 : Online clearance	17 %	17000
S4 : Discounters	15 %	15000
S5 : Clearance wholesalers	14 %	14000
S6 : Recycling	∞	∞
S7 : Destruction	∞	∞
S8 : Donation	∞	∞

Table 4 : Strategies' capacities

5.2 Ideal solutions

The ϵ -constraint method will consider, as we have mentioned, only one objective as the objective function and the others as constraints. As we can notice from the Table 3, the total sum of the strategies' risks values is equal to 1. Such a result is expected since they were extracted from the AHP method. Consequently, the maximum total risk, R_{\max} , that can be reached, if we use all strategies in one channel, is 1, and the minimum one is the lowest risk value, R_{\min} , that corresponds to 0.031 in our example. The ϵ_1 value will be rounded off to three decimal places as the risks' values have been considered.

5.2.1 1st case: unlimited capacities

In this section, we will try to solve our model without any capacity constraint. Results are presented in Table 5. Until the value of $\epsilon_1=0.058$, the clearance sales, which has the most significant performance represents the used strategy. For a value that is lower than the clearance sales risk one, which is 0.058, two solutions are possible S6 or S7, however, the proposed one is the S6, between [0.057,0.031], since it has a higher performance. For an $\epsilon_1 < 0.031$, we have not found any solution because all-risks' values exceed this limit.

ϵ_1	Used strategies / (xi)
[1,0.058]	S1 (1)
[0.057,0.031]	S6 (1)
[0.03, 0]	--

Table 5 : Solutions with unlimited-capacities strategies

5.2.2 2nd case: capacities' constraints

For a value of $\epsilon_1=1$, all strategies can be exploited, but results, in Table 6, show that only S1, S2 and S8 have been selected. This is because these three strategies have the most important performance values, in other words, they satisfy the risk condition while providing the maximum possible performance value. By reducing the value of ϵ_1 , some strategies will be excluded; consequently, we note the switching from the S8, to another one with a smaller risk value, the S6. We also note that the number of combined strategies becomes increasingly smaller in the combination of strategies that, achieves the maximum performance value and satisfies the constraint capacity and the risk limitation.

ϵ_1	Used strategies / (xi)		
[1; 0.325]	S1 (0,5)	S2 (0,3)	S8 (0,2)
[0.324; 0.25]	S1 (0,5)	S2 (0,3)	S6 (0,2)
[0.249; 0.164]	S1 (0,5)		S8 (0,5)
[0.163; 0.089]	S1 (0,5)		S6 (0,5)
[0.088, 0.031]	S6 (1)		

Table 6 : Solutions while considering the capacities' constraints

The disposal strategies facilitate the resolution of the problem since they have unlimited capacities. Therefore,

we propose an application with only the cash recovery strategies ($N=5$) and we observe capacity effects on the global solution; allocated quantities and used strategies. Results are presented in Table 7.

ϵ_1	Used strategies / (xi)			
[1 ; 0.586]	S1 (0.5)	S2 (0.3)	S3 (0.05)	S4 (0.15)

Table 7 : solutions for only in-house strategies' use while considering the capacities' constraints

Only one solution has been found which incorporates four strategies. From the value 0.586, no combination of strategies can satisfy the liquidation of the total unsold inventory while respecting the ϵ_1 value.

5.3 Sensitivity analysis

The significant difference between results regarding the two cases, the use of all strategies or only cash recovery ones, has prompted us to enhance our illustration by conducting a sensitivity analysis based on the capacities constraints. The main objective is to highlight the importance and the effects of this constraint on the global solution. Thus, we will consider the effects of an uncertainty on the clearance sales' capacity, C_1 , in two ways:

- a reducing from $0.5*Q$ to $0.25*Q$
- a raise from $0.5*Q$ to $0.75*Q$

This analysis can be expanded by considering a change in all strategies' capacities.

5.3.1 1st case: $C_1=0.25*Q$

Results are presented in Table 8. We note that the number of solutions and the solution's intervals are the same as in the original application. However, the difference lays only on the percentage of allocated quantities.

ϵ_1	Used strategies / (xi)		
[1; 0.325]	S1 (0,25)	S2 (0,3)	S8 (0,45)
[0.324; 0.25]	S1 (0,25)	S2 (0,3)	S6 (0,45)
[0.249; 0.164]	S1 (0,25)		S8 (0,75)
[0.163; 0.089]	S1 (0,25)		S6 (0,75)
[0.088, 0.031]	S6 (1)		

Table 8 : The effect of reduced capacity C_1

5.3.2 2nd case: $C_1=0.75*Q$

In this section, the capacity C_1 , has been increased. The main observation is that the number of used strategies has been reduced. Solutions that do not incorporate the strategy S1 remain the same. Results are presented in Table 9.

ϵ_1	Used strategies / (xi)	
[1; 0.219]	S1 (0,75)	S2 (0,25)
[0.218; 0.164]	S1 (0,75)	S8 (0,25)
[0.163; 0.089]	S1 (0,75)	S6 (0,25)
[0.088, 0.031]	S6 (1)	

Table 9 : The effect of increased capacity C_1

5.4 Results' interpretations and concluding remarks

5.4.1 Results' interpretations

In this part, we will analyze the impact of the capacity change on the global solution. To simplify the discussion, we will consider the first solution for each case, (Tables 6, 8 and 9) and the same analysis could be made for the other solutions.

Since the main objective is to maximize the total performance, the solving approach has only used strategies which satisfy the risk constraint and have the highest performance values which are; S1, S2 and S8. As the $P_1 > P_2 > P_8$, according to Table 3, the first step is the use of the clearance sales strategy (S1), once it is saturated, the model sways to the factory outlet one (S2) and finally the destruction (S8). Consequently, for $C_1 = 0.25 * Q$, the S1 and S2 are saturated and the remaining items are transmitted to the S8. In the second case, where $C_1 = 0.5 * Q$, the same process has been adopted. The difference lays only in the allocated quantities of S1 and S8, indeed, once S1 and S2 are saturated the remaining percentage of the quantity is only 0.2 (instead of 0.45) which will be affected to S8. Finally, in the third case, since $C_1 = 0.75 * Q$, the two strategies S1 and S2 can liquidate the total amount while maintaining the saturation of the most performant strategy, which is S1. Thus, the S8 has been excluded.

5.4.2 Concluding remarks

In the preceding section, we have varied only one and unique strategy capacity, and the solution has been changed in every time. We have even noted the removal of a strategy, but in practice, all strategies' capacities fluctuate and can be uncertain which may impact the global solution. Based on this analysis, we can conclude that solutions and feasible choices are closely linked to the unsold inventory amount and the estimated capacity of each strategy. Therefore, for a manager, having a clear idea on these two parameters makes the overall view of his problem more lucid. Such an observation can bring a great insight for further researches based on the forecasting of the unsold products and the estimation of strategy capacity.

6 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have proposed a new support tool for decision makers to design their clearance chain. It is based on a bi-objective optimization model to balance the risk and performance of the chosen clearance chain. The ϵ -constraint method has been used in order to solve the bi-objective model. Based on an illustrative case study, we have also solved the model firstly with unlimited capacities and with capacity constraint which have increased the number of solutions. The increase of the number of solutions regarding the capacity constraint has led us to conduct a sensitivity analysis. This analysis

brings a great insight for future researches. Thus, further proposed analysis can be by considering the uncertainty of the quantities of unsold inventory, of the capacities and respectively of the liquidated quantity by each one. The robust MOO represents a significant research area for our problem.

REFERENCES

- Agrawal S., Singh R.K, Murtaza Q., 2015, A literature review and perspectives in reverse logistics, *Resources, Conservation and Recycling*, vol.97, p.76–92
- Amin, S.H., Zhang, G., 2013, A multi-objective facility location model for closed-loop supply chain network under uncertain demand and return, *Applied Mathematical Modelling*, p. 4165-4176.
- Charnes, A., Cooper, W.W., 1977, Goal programming and multiple objective optimizations: Part 1, *European Journal of Operational Research*, vol. 1, p. 39-54.
- Dash Wu, D., Zhang, Y., Wu, D., L. Olson, D., 2010, Fuzzy multi-objective programming for supplier selection and risk modeling: A possibility approach, *European Journal of Operational Research*, vol.200, p.774-787.
- Demir, E., Bektas, T., Laporte, G., 2014, The bi-objective Pollution-Routing Problem, *European Journal of Operational Research*, vol. 232, p.464-478.
- Feng, Du., W. Evans, G., 2008, A bi-objective reverse logistics network analysis for post-sale service, *computers & operations research*, vol. 35, p.2617-2634.
- Ghezavati, V.R., Beigi, M., 2016, Solving a bi-objective mathematical model for location-routing problem with time windows in multi-echelon reverse logistics using metaheuristic procedure, *Journal of Industrial Engineering International*, December 2016, Vol. 12, p. 469–483.
- Giuntini, R., Andel, T., 1995a, Advance with reverse logistics – Part 1. *Transportation and Distribution*, vol. 36, p. 73–77.
- Haimes, Y. Y., Wismer, D. A., and Lasdon, L. S., 1971, On a Bicriterion Formulation of the Problems of Integrated System Identification and System Optimization, *IEEE SMC-1*, pp. 296-97.
- Kadziński, M., Tervonen, T., K.Tomczyk, M., Dekker, R., 2017, Evaluation of multi-objective optimization approaches for solving green supply chain design problems, *OMEGA*, Vol.68, p. 168-184
- Kannan, D., Khodaverdi, R., Olfat, L., Jafarian, A., Diabat, A., 2013, Integrated fuzzy multi criteria decision making method and multiobjective programming approach for supplier selection and order allocation in a green supply chain, *Journal of Cleaner Production*, Vol. 47, p. 355-367.
- Keshavarz, E., Toloo, M., 2015, A Bi-Objective Minimum Cost-Time Network Flow Problem, *Procedia Economics and Finance*, Vol. 23, p.3-8.
- Konaka, A., W. Coit, D., Alice E. Smith, 2006, Multi-objective optimization using genetic algorithms: A tu-

- torial, *Reliability Engineering and System Safety*, vol. 91, p.992–1007.
- Lambert, S., Riopel, D., Abdul-Kader, W., 2011, A reverse logistics decisions conceptual framework, *Computers & Industrial Engineering*, vol.61, p.561–581.
- Marisa P. de Brito, Dekker, R., 2002, Reverse Logistics – a framework, *Econometric Institute Report EI*.
- Mechmech, R., Harbi, S., B. Hadj-Alouane, A., Sboui, S., 2017, Identifying and ranking of clearance chain management (CCM) strategies, *7th IESM Conference, Saarbrücken, Germany*.
- Mostert, M., Caris, A., Limbourg, S., 2017, Intermodal network design: a three mode, bi-objective model applied to the case of Belgium, *Flexible Services and Manufacturing Journal Media*, New York, p.1-24.
- Narasimhan, R., Talluri, S., Mahapatra, S.K., 2006. Multiproduct, multicriteria model for supplier selection with product life-cycle considerations, *Decision Sciences*, vol. 37, p. 577–603.
- Ramezani, M.; Bashiri, M.; Tavakkoli-Moghaddam, R., 2013, A new multi-objective stochastic model for a forward/reverse logistic network design with responsiveness and quality level, *Applied Mathematical Modelling*, vol.37, p. 328-344.
- Sasikumar, P., Kannan, G. , 2008, Issues in reverse supply chains, part I: end-of-life product recovery and inventory management – an overview, *international journal of sustainable engineering*, vol.3, p.154-172.
- S. Rogers, D., Tibben-Lembke R, 2001, An examination of reverse logistics practices, *Journal of Business Logistics*, vol 22, p.129–148.
- S. Rogers, D., Melamed, B., and S. Lembke, R., 2012, Modeling and Analysis of Reverse Logistics, *Journal of Business Logistics*, vol.3, p.107–117.
- Sboui, S., 2006, Unsold and Excess Inventory: between optimization and management: A New challenge for the Supply Chain Management, *International Conference on Service System and Service Management*, vol.2.
- Wadhwa, V., Ravindran, A.R., 2007. Vendor selection in outsourcing, *Computers and Operations Research*, vol. 34, p. 3725–3737.
- Xia, W., Wu, Z., 2007. Supplier selection with multiple criteria in volume discount environments. *OMEGA – International Journal of Management Science*, vol. 35, p.494–504.
- Yeh, W., Chuang, M., 2011, Using multi-objective genetic algorithm for partner selection in green supply chain problems, *Expert Systems with Applications*, vol. 38, p.4244-4253.
- Zhao, J., Verter, V. ,2015, A bi-objective model for the used oil location-routing problem, Canada, *Computers & Operations Research*, vol.62, p.157–168.

A Continuous Time Inventory Routing Problem with Energy Minimization

Yun HE

Tolga BEKTAŞ

Cyril BRIAND

Nicolas JOZEFOWIEZ

LS2N / IMT-Atlantique

CORMSIS

LAAS-CNRS, UPS

LCOMS

4 Rue Alfred Kastler

University of Southampton

Université de Toulouse

Université de Lorraine

44000 Nantes - France

Southampton SO17 1BJ - UK

31000 Toulouse - France

57000 Metz - France

yun.he@imt-atlantique.fr

T.Bektas@soton.ac.uk

briand@laas.fr

nicolas.jozefowicz@univ-lorraine.fr

ABSTRACT: *In this paper, an original continuous time Inventory Routing Problem with Energy Minimization is presented. The problem is defined on a road network with a set of customers under the Vendor Managed Inventory. Each customer has a series of consumption event with a quantity at a certain time instants (hours, minutes or seconds) instead of periods. The road network is represented by a multi-graph, which allows the choice of different paths with different energetic costs and travel durations between the same pair of stations. The total energy consumption is estimated according to the vehicle dynamics, which is linearly dependent on the transported mass on each segment of road. In this way, the combined decision of inventory and routing is completely integrated with the energy criteria by the timing of visits and by the decisions on the mass flows. A decomposition method based on Lagrangian relaxation is proposed for solving the problem. Some preliminary results are given.* **KEYWORDS:** *Inventory Routing Problem, Lagrangian relaxation, Energy minimization, Multi-graph.*

1 INTRODUCTION

The real-life inventory routing challenge (Air Liquide, EURO & ROADEF 2016) reveals three major drawbacks of the assumptions used in the classical academic models for the deterministic Inventory Routing Problem (IRP). Namely, (1) the planning horizon is discretized into periods, assuming that each supply tour can be performed in one period; (2) customer demands can be aggregated into periods, assuming that the demand frequency of each customer is regular; and (3) the travel time and costs are roughly estimated. These assumptions may not necessarily hold in practice.

To contribute to bridging the gap between the practical inventory routing problems and the idealized academic models, as well as to account for energy considerations, this paper describes a new inventory routing problem with an explicit consideration of energy that we name as Energy-Minimizing Inventory Routing Problem (EMIRP). The problem is defined on a real road network that is represented by a multi-graph. This allows the choice of different paths, each of which may have a different energy cost and travel duration, between a given pair of customer points. At the same time, the inventory level is monitored in finer time granularity, which is expressed in any relevant time units (e.g., hours, minutes or seconds) instead of periods.

This paper is organized as follows: in Section 2 a brief review on the recent advance in the energy related IRP is given. Then, in Section 3, the multi-graph representation of real road networks is explained. In Section 4, the problem is defined on the multi-graph. In Section 5, the mathematical model is introduced with a decomposition method based on Lagrangian relaxation. The solution method is explained in details in Section 6. And finally, some preliminary results are given in Section 7.

2 Literature

The Inventory Routing Problem (IRP) is an integration of inventory management and vehicle routing and scheduling. Both exact and heuristic methods exist for the solution of the problem. For detailed literature review of a variety of IRPs and their solution methods, please refer to (Coelho, Cordeau & Laporte 2013). For the industrial aspects and applications, please refer to (Andersson, Hoff, Christiansen, Hasle & Løkketangen 2010).

Under the new sustainable development requirement, the “Green-Vehicle Routing Problem (Green-VRP)” is being studied (see (Lin, Choy, Ho, Chung & Lam 2014) for a survey). In this category of VRP, we can find the energy minimizing VRP (Kara, Kara & Yetis 2007), Pollution Routing Problem (Bektas & Laporte 2011) for the minimization of CO_2 emission,

and fuel emission optimization in the VRP (Qian & Eglese 2016).

Despite the numerous versions of IRPs and the emerging Green VRPs, the integration of energy aspect in the IRP is still a new area of study. Since 2014, there are a growing number of papers dealing with energy efficiency in the IRP. (Treitl, Nolz & Jammernegg 2014) and (Alkawaleet, Hsieh & Wang 2014) are among the first to integrate the environmental aspect to the IRP. Early papers concern the integration of environmental consideration into the IRP and its impacts. More recently, (Cheng, Qi, Wang & Zhang 2016) studies the multi-period IRP with carbon emission and proposed a genetic algorithm to solve it. In (Cheng, Yang, Qi & Rousseau 2017), the authors study the modeling of green IRP with heterogeneous fleet. Several objectives are proposed and a branch-and-cut solution algorithm is applied.

3 Multi-graph Representation of Real Road Networks

Inspired by the work in (Garaix, Artigues, Feillet & Josselin 2010), the real road network representation is constructed as follows. Let us consider a *simple* graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where the set $\mathcal{V} = V \cup V'$ of vertices contains a set V of nodes including the locations of the set \mathcal{C} of customers and a depot 0, and a set V' of intermediate points representing, e.g., traffic lights, junctions. Each arc $(i, j) \in \mathcal{A}$ represents a link between two points $i, j \in \mathcal{V}, i \neq j$ on the road network and is characterized by two attributes, namely the travel time and the energy consumption. The graph \mathcal{G} can be incomplete. By solving a multi-objective shortest path problem between each pair $(i \in V, j \in V \setminus \{i\})$ of nodes on \mathcal{G} , one can obtain a set of Pareto-optimal paths $\mathcal{P}_{i,j}$, which yields the multi-graph $G = (V, A)$. This graph contains only the nodes in $V = \mathcal{C} \cup \{0\}$, where each pair of nodes in V is linked by an arc $(i, j)^l \in A$ with $l \in \mathcal{P}_{i,j}$. Each arc is characterized by a fixed travel time $\tau_{i,j}^l$ and an energy consumption per unit of mass denoted $e_{i,j}^l$ as estimated in (He, Briand & Jozefowicz 2016) based vehicle dynamics.

4 Problem Presentation

Under the Vendor Managed Inventory (VMI), the inventory levels of the set of customers \mathcal{C} is monitored during a time horizon with totally H hours. Each customer $i \in \mathcal{C}$ consumes a *single* product gradually. Each customer $i \in \mathcal{C}$ has a series of consumption events, denoted by D_i . Each consumption $d \in D_i$ is defined by a time T_i^d and an amount of product P_i^d . Each customer also has an initial inventory level denoted by I_i^0 , a minimum inventory level of zero, and a maximum inventory level denoted by \bar{I}_i respectively. There is a *homogeneous* fleet K of vehicles at disposal

for the deliveries. Each vehicle $k \in K$ has the same capacity B and a curb weight W which is the empty vehicle weight (both in kg). At the beginning of the time horizon, the vehicles are all at the depot 0 with an initial quantity not higher than its capacity B .

The problem is to decide on the time and quantity of each delivery to each customer $i \in \mathcal{C}$ by each vehicle $k \in K$, and the route of each vehicle to perform the deliveries. The objective minimizes the total energy consumed by the vehicles required for the routing while respecting the following constraints: (1) *inventory balance*: the inventory level should be consistent across the different time periods; (2) *inventory capacity*: the inventory level of each customer should never go below the minimum and beyond the maximum levels; (3) *mass-flow balance*: the delivery quantity should be consistent with the mass transported on each arc; (4) *vehicle capacity*: the quantity of product transported by a vehicle should be non-negative and never larger than the vehicle capacity; (5) *vehicle flow*: the vehicles should depart from the depot, perform a tour among the customers to be visited, and return back to the depot; (6) *precedence*: the difference between the arrival time of two consecutive visits in the same route should not be smaller than the travel time between the two locations.

In the following, we make the assumption that the number of vehicles in the fleet is enough to cover the total quantities demanded in the customer consumption events in the whole time horizon. The delivery time by vehicle $k \in K$ to customer $i \in \mathcal{C}$ is assumed to be the same as the arrival time of k to i and the delivery is supposed to be instantaneous. By convention, each time a customer is visited before a certain consumption event, the sequence of the operations is the delivery first, followed by the consumption and finally the computation of the inventory level before the next consumption. The Hour 0 is an artificial start of the horizon without any consumption or delivery. Moreover, the capacity of the depot is considered to be infinite and it contains as many products as needed.

5 Mathematical Model

The following variables are to be decided in the EMIRP model. For each arc $l \in \mathcal{P}_{i,j}$ linking each pair of locations $(i, j) \in V \times V$ and for each vehicle $k \in K$, continuous variable $m_{i,j}^{k,l} \in [0, B]$ represents the mass transported by the vehicle k on the path l from the node i to the node j . For each consumption $d \in D_i$ of customer $i \in \mathcal{C}$, continuous variable $r_i^d \in [0, B]$ represents the total quantity delivered to i before the consumption event d . For each consumption $d \in D_i$ of customer $i \in \mathcal{C}$ and for each vehicle $k \in K$, variable $q_i^{d,k} \in [0, \bar{I}_i]$ represents the amount of product given to customer i by the vehicle k before the con-

sumption event d . For each vehicle $k \in K$ and each pair of nodes $i, j \in V$, variable $t_{i,j}^k \in [0, H]$ indicates the arrival time of vehicle k to node j if the previous node visited is node i . For each arc $l \in \mathcal{P}_{i,j}$ linking each pair of nodes $i, j \in V$ and for each vehicle $k \in K$, variable $x_{i,j}^{k,l} \in \{0, 1\}$ is a binary variable equal to 1 if the vehicle k travels from i to j using arc l , 0 otherwise. For each consumption $d \in D_i$ of customer $i \in C$ and each vehicle $k \in K$, variable $y_i^{d,k} \in \{0, 1\}$ is a binary variable equal to 1 if the vehicle k makes a delivery to customer i before his consumption d , 0 otherwise.

5.1 Complete Model

The objective (1) is to minimize the total energy consumption E defined by Eq. 1. The first term is proportional to the flow transported on each arc. The second is related to the empty weight W of the vehicles, which can be viewed as a fixed cost when a path linking a pair of stations is traversed by a vehicle.

$$E = \sum_{k \in K} \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} e_{i,j}^l m_{i,j}^{k,l} + W \sum_{k \in K} \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} e_{i,j}^l x_{i,j}^{k,l} \quad (1)$$

EMIRP: minimize E such that:

Transportation flow balance

$$\sum_{j \in V} \sum_{l \in \mathcal{P}_{i,j}} m_{j,i}^{k,l} - \sum_{j \in V} \sum_{l \in \mathcal{P}_{i,j}} m_{i,j}^{k,l} = r_i^k \quad \forall i \in C, \forall k \in K \quad (2)$$

$$r_i^k = \sum_{r \in D_i} q_i^{d,k} \quad \forall i \in C, \forall k \in K \quad (3)$$

Link routing and flow

$$m_{i,j}^{k,l} \leq B x_{i,j}^{k,l} \quad \forall k \in K, \forall (i,j) \in V \times V, \forall l \in \mathcal{P}_{i,j} \quad (4)$$

Link visit and routing

$$\sum_{j \in V} \sum_{l \in \mathcal{P}_{j,i}} x_{j,i}^{k,l} = \sum_{d \in D_i} y_i^{d,k} \quad \forall i \in C, \forall k \in K \quad (5)$$

Routing

$$\sum_{j \in V} \sum_{l \in \mathcal{P}_{0,j}} x_{0,j}^{k,l} \leq 1 \quad \forall k \in K \quad (6)$$

$$\sum_{j \in V} \sum_{l \in \mathcal{P}_{j,i}} x_{j,i}^{k,l} - \sum_{j \in V} \sum_{l \in \mathcal{P}_{i,j}} x_{i,j}^{k,l} = 0 \quad \forall k \in K, \forall i \in V \quad (7)$$

Arrival time

$$\sum_{j \in V \setminus \{i\}} t_{i,j}^k - \sum_{j \in V \setminus \{i\}} t_{j,i}^k \geq \sum_{j \in V \setminus \{i\}} \sum_{l \in \mathcal{P}_{i,j}} \tau_{i,j}^l x_{i,j}^{k,l} \quad \forall i \in V, \forall k \in K \quad (8)$$

$$\sum_{j \in V \setminus \{i\}} t_{j,i}^k \geq \sum_{d \in D_i \setminus \{1\}} (T_i^{d-1} + 1) y_i^{d,k} \quad \forall i \in C, \forall k \in K \quad (9)$$

$$\sum_{d \in D_i} T_i^d y_i^{d,k} \geq \sum_{j \in V \setminus \{i\}} t_{j,i}^k \quad \forall i \in C, \forall k \in K \quad (10)$$

Link visit and amount of delivery

$$q_i^{d,k} \leq \bar{I}_i y_i^{d,k} \quad \forall i \in C, \forall k \in K, \forall d \in D_i \quad (11)$$

Inventory monitoring

$$I_i^0 + \sum_{k \in K} \sum_{u=1}^d q_i^{u,k} - \sum_{u=1}^d P_i^u \geq 0 \quad \forall i \in C, \forall d \in D_i \quad (12)$$

$$I_i^0 + \sum_{k \in K} \sum_{u=1}^r q_i^{u,k} \leq \bar{I}_i \quad \forall i \in C, \forall d \in D_i \quad (13)$$

Additional constraints

$$\sum_{d \in D_i} y_i^{d,k} \leq 1 \quad \forall i \in C, \forall k \in K \quad (14)$$

$$\sum_{l \in \mathcal{P}_{i,j}} x_{i,j}^{k,l} \leq 1 \quad \forall (i,j) \in V \times V, \forall k \in K \quad (15)$$

$$\sum_{i \in C} \sum_{d \in D_i} q_i^{d,k} \leq B \quad \forall k \in K \quad (16)$$

Variable domains

$$x_{i,j}^{k,l} \in \{0, 1\} \quad \forall (i,j) \in V \times V, \forall k \in K, \forall l \in \mathcal{P}_{i,j} \quad (17)$$

$$m_{i,j}^{k,l} \in [0, B] \quad \forall (i,j) \in V \times V, \forall k \in K, \forall l \in \mathcal{P}_{i,j} \quad (18)$$

$$y_i^{d,k} \in \{0, 1\} \quad \forall i \in C, \forall k \in K, \forall d \in D_i \quad (19)$$

$$q_i^{d,k} \in [0, \bar{I}_i] \quad \forall i \in C, \forall k \in K, \forall d \in D_i \quad (20)$$

$$r_i^k \in [0, B] \quad \forall i \in C, \forall k \in K \quad (21)$$

$$t_{i,j}^k \in [0, H] \quad \forall (i,j) \in V \times V, \forall k \in K \quad (22)$$

Constraints (2) are for the flow balance of the mass transported inside the network. Constraints (3) links variable r and q . Constraints (4) link routing variables x and mass flow variables m . Constraints (6) ensure that one vehicle performs at most one route. Constraints (7) ensure that the number of incoming arcs to a customer is the same with the number of outgoing arcs from this customer for each vehicle. Constraints (8)–(10) are the flow formulation for the timing of each delivery (Bektas & Lysgaard 2015). Constraints (8) are for the precedence of visits, in which the term $\sum_{j \in V \setminus \{i\}} t_{i,j}^k$ sets the arrival time to the next location visited after the location i and the term $\sum_{j \in V \setminus \{i\}} t_{j,i}^k$ gives the arrival time to i . Constraints (9) and (10) bound the arrival time to each customer i between two consecutive consumption events of this customer. Constraints (5) say that if the vehicle visits a customer, then there is at least one delivery before one of the customer's consumptions. This set of constraints links the routing variables x with the delivery decision variables y . Constraints (11) link the delivery decision variables y and the delivery quantity variables q . They impose that if there is no delivery, then the delivered quantity is 0. Otherwise, the delivered quantity is no

larger than the customer capacity. Constraints (12) and (13) bound the inventory level in the acceptable range *before and after* each consumption or delivery. Additionally, Constraints (14) ensure that there is at most one delivery to each customer per vehicle. Constraints (15) make sure that at most one path in the multi-graph is used in each route. Constraints (16) is for the capacity of the vehicles. Constraints (17)–(22) define the variable domains.

6 Solution Method

A decomposition method based on Lagrangian relaxation is presented in this section. The idea is to relax the linking constraints (3), (5), (9) and (10) between inventory management and routing, while keeping the transportation flow balance and the exact timing of visits in the routing part.

6.1 Lagrangian Relaxation

Inspired by the decomposition proposed in (Chien, Balakrishnan & Wong 1989), the Lagrangian relaxation of **(EMIRP)** that dualizes constraints (3), (5), (9) and (10) respectively with the multipliers $\alpha_i^k \in \mathbb{R}$, $\beta_i^k \in \mathbb{R}$, $\gamma_i^k \in \mathbb{R}^+$, $\delta_i^k \in \mathbb{R}^+$ for each $i \in C$ and $k \in K$. The relaxed problem **(L($\alpha, \beta, \gamma, \delta$))** is to minimize a new objective $\mathcal{E}_L(\alpha, \beta, \gamma, \delta)$ with the Lagrangian multipliers defined in Eq. (23), under the constraints (2), (4), (6)–(8), (11)–(22) where the value $\tilde{e}_{i,j}^{k,l}$ is a modified energy cost defined in Eq. (24).

Since the sets of variables $\{q, y\}$ and $\{x, r, m, t\}$ are now independent of each other, the relaxed problem **(L($\alpha, \beta, \gamma, \delta$))** can be decomposed into (1) an inventory allocation problem (IA) with variables q and y for the inventory monitoring; (2) for each vehicle k , a capacitated cycle problem (CC) with fixed charges and precedence constraints on a multi-graph.

$$\begin{aligned} \mathcal{E}_L(\alpha, \beta, \gamma, \delta) = & \sum_{k \in K} \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} e_{i,j}^{k,l} m_{i,j}^{k,l} \\ & + \sum_{k \in K} \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} \tilde{e}_{i,j}^{k,l} x_{i,j}^{k,l} \\ & + \sum_{k \in K} \sum_{i \in C} \alpha_i^k r_i^k \\ & + \sum_{k \in K} \sum_{i \in C} (\delta_i^k - \gamma_i^k) \sum_{j \in V \setminus \{i\}} t_{j,i}^k \\ & - \sum_{k \in K} \sum_{i \in C} \alpha_i^k \sum_{d \in D_i} q_i^{d,k} \\ & - \sum_{k \in K} \sum_{i \in V} \beta_i^k \sum_{d \in D_i} y_i^{d,k} \\ & + \sum_{k \in K} \sum_{i \in V} \gamma_i^k \sum_{d \in D_i \setminus \{1\}} (T_i^{d-1} + 1) y_i^{d,k} \\ & - \sum_{k \in K} \sum_{i \in V} \delta_i^k \sum_{d \in D_i} T_i^d y_i^{d,k} \end{aligned} \quad (23)$$

$$\begin{aligned} \tilde{e}_{i,j}^{k,l} = & W e_{i,j}^{k,l} + \beta_j^k \\ & \forall k \in K, (i, j) \in V \times V, l \in \mathcal{P}_{i,j} \end{aligned} \quad (24)$$

6.2 Inventory Allocation Sub-problem

The problem of Inventory Allocation (IA) consists in allocating vehicle deliveries to customer consumption events under the constraints of inventory balance, inventory capacity and vehicle capacity. For visiting each customer $i \in C$ before each consumption $d \in D_i$ by each vehicle $k \in K$, there is a fixed cost $\nu_i^{d,k}$ is induced by the multipliers β_i^k , γ_i^k and δ_i^k as defined in Eq. (25) under the additional assumption that $T_i^0 = -1$.

$$\begin{aligned} \nu_i^{d,k} = & -\beta_i^k + \gamma_i^k (T_i^{d-1} + 1) - \delta_i^k T_i^d \\ & \forall i \in C, \forall d \in D_i, \forall k \in K \end{aligned} \quad (25)$$

The values of α_i^k define a cost proportional to the quantity delivered to each customer $i \in C$ by each vehicle $k \in K$. The objective is to minimize the total costs induced by the multipliers as defined in Eq. (26), while making sure that the inventory level of each customer always stays inside the limits and that the vehicle never carries more than its capacity. The problem is to find the best delivery quantity and time to each customer by each vehicle. Since the exact time of each delivery makes no difference in this problem, we only need to decide the consumption event before which each visit is to make.

$$- \sum_{k \in K} \sum_{i \in C} \alpha_i^k \sum_{d \in D_i} q_i^{d,k} + \sum_{k \in K} \sum_{i \in V} \sum_{d \in D_i} \nu_i^{d,k} y_i^{d,k} \quad (26)$$

The decision variables are: $q_i^{d,k} \in [0, \bar{I}_i]$ for the amount of product delivered to each customer $i \in C$

after each consumption $r \in D_i$ by each vehicle $k \in K$; $y_i^{d,k} \in \{0, 1\}$ for the decision of whether vehicle $k \in K$ visit customer $i \in C$ before his consumption $d \in D_i$ (equal 1) or not (0). The model is composed by constraints (11)–(14), (16), and variables domains (19) and (20) derived from the relaxed problem.

In the following, the problem *with heterogeneous fleet* is demonstrated to be NP-hard even with one single consumption per customer and with non-negative fixed visiting costs ν . The fixed visiting costs can be restricted to be non-negative because if there is a negative fixed visiting cost $\nu_i^{d,k}$ before a certain consumption r of a customer i by a vehicle k , it will always be profitable to deliver to i before the consumption event r by vehicle k . By redefining the fixed costs for these visits to zero, we obtain a problem instance with non-negative fixed visiting costs.

Proposition 1. *Inventory Allocation (IA) problem with heterogeneous fleet is NP-hard.*

Proof. This proof shows that IA can be transformed from the MINCOSTFIXEDFLOW problem.

First of all, an instance of the MINCOSTFIXEDFLOW is defined on a directed graph G . The vertex set V contains two special vertex, namely, a source and a sink. Each arc $(i, j) \in A$ is associated with a non-negative cost $c_{i,j}$ and a non-negative capacity $u_{i,j}$. The problem is to find a minimum cost subset $A' \subset A$ such that in (V, A') , the flow from the source s to the sink t is at least F .

It is demonstrated to be strongly NP-hard even on bipartite graphs in (Krumke, Noltemeier, Schwarz, Wirth & Ravi 1999).

From an instance of the MINCOSTFIXEDFLOW problem on a bipartite graph, we construct an instance of IA problem and demonstrate that the solution of the MINCOSTFIXEDFLOW problem exists if and only if IA gets an optimal solution.

We apply the reduction from the problem with one source and multiple sinks, which is obviously reducible to the basic version of the problem with single source and single sink.

Consider now an instance of the MINCOSTFIXEDFLOW problem with one source and multiple sinks on a bipartite graph G . The set of vertices $V = \{s\} \cup T \cup K \cup C$ contains a special source vertex s , a set T of sink vertices and two disjoint sets K and C . There is an arc from s to each vertex $k \in K$, and from each vertex $i \in C$ to a sink $t_i \in T$. Each vertex $i \in C$ is linked to each vertex $k \in K$. The problem is to find a minimum cost subset A' of A such that in (V, A') the flow from the source s to each sink $t_i \in T$ is at least F_i .

Note that the bipartite graph $(K \cup C, A)$ can be incomplete, but this does not influence the analysis of NP-hardness, since the graph \mathcal{G} can be transformed to a complete one by associating a very large cost with the arcs in the complete graph but not in A . In the following, without loss of generality, \mathcal{G} is considered to be complete.

We now construct an instance of IA: K defines the set of vehicles; C defines the set of customers; $\{D_1, \dots, D_{|C|}\}$ defines the set of consumption events of customers; the capacity of each vehicle is equal to the capacity of each arc (s, k) with $k \in K$; the capacity of each customer $i \in C$ is equal to F_i ; each customer has one single consumption with an amount F_i ; the initial level of each customer is 0; the fixed cost $\nu_i^{1,k}$ of visiting a customer $i \in C$ by vehicle $k \in K$ before his only consumption 1 is defined by the sum of fixed costs of the arcs (s, k) , (k, i) , (i, t) : $c_{s,k} + c_{k,i} + c_{i,t}$. If one obtains a minimum cost subset $A' \subset A$ such that the flow in (V, A') from source s to each sink $t_i \in T$ is at least F_i , then one can define a solution of IA by setting $y_i^{1,k}$ to 1 if arc $(k, i) \in A'$, to 0, otherwise, with $q_{i,k}^r$ set to the corresponding flow amount on arc (k, i) . If one obtains a solution of the above instance of IA, one can identify a minimum cost subset A' with arcs $(k, i) \in A'$ if vehicle k is set to visit customer i before his consumption. This transformation is obviously in polynomial time. \square

For the homogeneous fleet case, the problem can be proved to be NP-hard by a reduction from Knapsack problem. Take an instance of the Knapsack problem with n objects, each object with a profit of p_i , a weight w_i and a container with capacity C . A solution of the Knapsack problem is a subset of objects $S \subset \{1, \dots, n\}$ such that the total profit is the maximum. An instance of IA can be transformed in the following way. Suppose there are one single vehicle with capacity C , n customers each with $I_i^0 = 0$ and only one consumption event with quantity w_i . Set the fixed cost $\nu_i^{1,k}$ to $-p_i$ and the variable cost $\alpha_i^k = 0$. The IA problem finds a subset of customers such that the total costs of visit ($\sum_{i \in C} -p_i y_i$) is the minimum.

6.3 Capacitated Cycle (CC) Sub-problem

The exact sequencing of each visit to each customer by each vehicle is decided in a set of sub-problems called “capacitated cycle” (CC). Since the vehicles are independent from each other, there is one CC per vehicle. For each vehicle k , a cycle problem with fixed costs on a multi-graph has to be solved under precedence and capacity constraints. In this problem, each vehicle has to make a tour that starts and ends at the depot and travels among profitable customers to make deliveries.

Each time the vehicle k travels from site i to site j using the arc $l \in \mathcal{P}_{i,j}$, there is a fixed energy cost $\tilde{e}_{i,j}^{k,l}$ related to the Lagrangian multiplier β_j^k as given in Eq. (24). A cost proportional to the amount of commodity flow on each arc $l \in \mathcal{P}_{i,j}$ from site i to site j is defined by the cost $e_{i,j}^l$ of the original (EMIRP). Each time vehicle k delivers any product to customer i , there is a cost proportional to the amount of delivery defined by α_i^k . Additional costs proportional to the arrival time of vehicle k to customer i are given by $\delta_i^k - \gamma_i^k$. The problem is to minimize the total cost defined in Eq. 27, with constraints (2), (4), (6)–(8), (15), (17)–(18) and (21)–(22).

$$\begin{aligned} & \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} e_{i,j}^l m_{i,j}^{k,l} + \sum_{(i,j) \in V \times V} \sum_{l \in \mathcal{P}_{i,j}} \tilde{e}_{i,j}^{k,l} x_{i,j}^{k,l} \\ & + \sum_{i \in C} \alpha_i^k r_i^k + \sum_{i \in C} (\delta_i^k - \gamma_i^k) \sum_{j \in V \setminus \{i\}} t_{j,i}^k \end{aligned} \quad (27)$$

The problem is NP-hard as it contains as a special case the problem of finding a Hamiltonian tour. Indeed, given simple graph $G = (V, A')$, by setting $e_{i,j}^l = 0$, $\tilde{e}_{i,j}^{k,l} = 0$ for all $(i, j) \in A'$, $e_{i,j}^l = 1$, $\tilde{e}_{i,j}^{k,l} = 1$ for all $(i, j) \in A \setminus A'$ and $\alpha_i^k = -1$, $\delta_i^k - \gamma_i^k = 0$ for all $i \in V$, the Hamiltonian tour on the graph G is also the min-cost cycle with cost equal to $-n$ with n the number of vertices in V .

6.4 Solution Algorithm Based on Lagrangian Relaxation

The Lagrangian relaxation is known to give a lower bound to the original problem (Geoffrion 1974). Algorithm 1 proposes the Lagrangian relaxation solution scheme. The algorithm goes on iteratively, keeping and updating an upper bound UB and a lower bound LB to the original problem.

At the beginning, a heuristic method based on the urgency of customers is applied to solve the original problem. If the heuristic can find a feasible solution, the UB is set to the objective of this initial solution. Otherwise, the heuristic gives a set of tours and a list of lost customers. Fictive vehicles are added to make direct deliveries to these lost customers by taking the path with the maximum energy. The value of the UB is set to the sum of total energy costs of the tours given by the heuristic and all the fictive direct shippings.

Initial values of Lagrangian multipliers can have a strong influence on the quality of the lower bound obtained with the Lagrangian relaxation. In our algorithm, all the multipliers are set to zero initially except for the values of α , which are computed by Eq. (28). In this way, the visits to customers are

Algorithm 1 Lagrange relaxation based decomposition method

- 1: Solve the initial problem by a heuristic method
 - 2: Initialisation of the multipliers
 - 3: $LB \leftarrow -\infty$, $UB \leftarrow +\infty$
 - 4: Number of iteration $n \leftarrow 0$
 - 5: **while** stop condition not satisfied **do**
 - 6: Solve IA
 - 7: **for all** vehicle $k \in K$ **do**
 - 8: Solve CC
 - 9: Get the current solution $S^{(n)}$ with the combined objectives of IA and CC $\mathcal{E}_L^{(n)}(S^{(n)})$
 - 10: Update the multipliers with the current solution $S^{(n)}$
 - 11: **if** $\mathcal{E}_L^{(n)} > LB$ **then**
 - 12: $LB \leftarrow \mathcal{E}_L^{(n)}$
 - 13: Apply a repairing algorithm to get a feasible solution from the solution of IA and CC and get its objective $\bar{\mathcal{E}}^{(n)}$
 - 14: **if** $\bar{\mathcal{E}}^{(n)} < UB$ **then**
 - 15: $UB \leftarrow \bar{\mathcal{E}}^{(n)}$
 - 16: $n \leftarrow n + 1$
-

always profitable in the relaxed problem at the beginning of the algorithms.

$$\alpha^{(0)} = - \sum_{j \in V \setminus \{i\}} \sum_{l \in \mathcal{P}_{i,j}} e_{i,j}^l \quad \forall i \in C, \forall k \in K \quad (28)$$

In each iteration, problem $(\mathbf{L}(\alpha, \beta, \gamma, \delta))$ is decomposed into sub-problems IA and CC. By solving IA, a delivery assignment of each customer to each vehicle is obtained. By solving CC for each vehicle, a routing plan for deliveries to customers is obtained. Then, a sub-gradient method is used to update the multipliers. These values of multipliers will define new coefficients in objective functions for IA and CC and the algorithm passes to new iteration.

The sum $\mathcal{E}_L^{(n)}$ of objective values of IA and each CC in each iteration n gives a lower bound to the initial problem. The LB is updated if the current lower bound is better. The solutions obtained by the relaxation can be repaired to get a feasible solution to the original EMIRP. To repair a solution, the values of the variables x are set to 1 if their values in the current solution of the relaxed problem is 1, otherwise, they are undetermined. A Mixed Integer Linear Programming (MILP) is then solved to get a complete solution to the original problem. If a better feasible solution can be found within a certain time limit, the upper bound UB is updated. The algorithm continues in this way until a certain stop condition. In our case, the condition is when $UB = LB$ (which is impossible in practice) or when a total time limit is reached.

According to the Lagrangian duality, inside each iteration, the objective $\mathcal{E}_L(\alpha, \beta, \gamma, \delta)$ as defined by (23)

is to be minimized in relation to continuous variables $\alpha_i^k \in \mathbb{R}$, $\beta_i^k \in \mathbb{R}$, $\gamma_i^k \in \mathbb{R}^+$ and $\delta_i^k \in \mathbb{R}^+$. At the end of each iteration, a classic sub-gradient method (Pajouh & Balasundaram 2010) is applied for the update of the multipliers.

7 Experimentation and Preliminary Results

Some preliminary experimentations are performed to test the solution method.

7.1 Instance Preparation

The instances are generated based on the IRP benchmarks with single vehicle on a horizon of 3 hours (Archetti, Bertazzi, Laporte & Speranza 2007). The energy consumption per unit of mass on each arc is estimated with the same rule as in (He et al. 2016) with two types of roads (national and highway). A randomized waiting time per stop is set on each arc according to the road type of this arc. If the arc is national, the waiting time is set to 1 second per stop. If the arc is a highway, the waiting time is 1 hour on the whole arc. The travel duration on each arc is then computed with the vehicle speed, the distance of the arc and the waiting time, assuming a fixed acceleration. A matrix of travel duration as well as a matrix of energy costs per unit of mass are obtained. In addition, one period in a benchmark instance is considered to be one day with 24 hours. To generate the series of customer consumption, the customer demands per period in the initial instances are broken down into a random number of consumption events. Each consumption event is associated with an arrival time between 0 and 24 and a randomized quantity. The sum of the quantity of all the consumption events in one period is equal to the demand rate per period in the initial instances. The fleet in the generated instances contains maximum 3 vehicles, which corresponds to one single vehicle for each period in the initial instances.

The multi-graph is obtained from the initial network of customers. As explained in the end of Section 3, a Shortest Path Problem (SPP) is solved twice with different objectives. One is to minimize total travel time and the other is to minimize the energy per unit of mass. The solutions with these two objectives give the two extremities of the Pareto front. The problem is finally solved on this multi-graph.

7.2 Preliminary Results

The complete model is coded in C++ and solved using IBM ILOG CPLEX 12.6.1. with 2 hours time limit. Table 1 is a comparison of the results obtained with decomposition method and the commercial solver. The first column is the instance num-

ber under the format $(a)n(b)$, with (a) the instance set and (b) the number of customers in the instance. The column “LB” indicates the best lower bound obtained after 2 hours of solution. The column “UB” shows the best upper bound after 2 hours of solution with the number given in parentheses for the number of vehicles actually used in the solution. If no feasible solution is found from the relaxation at the end of the time limit, the UB value corresponds to the upper bound of the initial solution and a question mark is given in parentheses. The column “GAP” presents the optimality gap defined by Eq. (29). For the Lagrangian decomposition, the number of iterations after 2 hours of solution is also reported in the column “# iter”. In addition, if the initial heuristic can not find a feasible solution, the upper bound is a fictive one with a question mark in parentheses. For instances with 5 customers, both CPLEX and the decomposition method can find the optimal solution. However, the Branch & Cut by CPLEX can prove the optimality but the gap in the decomposition method is still about 30%. The solver is only able to find optimal solutions to instances with 5 customers within 2 hours time limit. Compared to the results given by CPLEX in Table 1, the lower and upper bounds are still to be improved in the decomposition method. For instances with 15 customers, CPLEX cannot find feasible solution for any of the instances but the decomposition method can find feasible solution for 1 out of 5 instances.

$$\text{GAP} = \frac{\text{UB} - \text{LB}}{\text{UB}} \quad (29)$$

Table 1: Comparison of results

Inst.	Lagrangian decomp.				CPLEX		
	LB	UB ($ K $)	GAP	# iter	LB	UB ($ K $)	GAP
1n5	170.291	218.874 (2)	22.20%	344	218.874	218.874 (2)	0
2n5	119.663	164.015 (1)	27.04%	300	164.015	164.015 (1)	0
3n5	199.783	281.374 (2)	29.00%	205	281.374	281.374 (2)	0
4n5	199.978	226.184 (2)	11.59%	186	226.184	226.184 (2)	0
5n5	110.820	158.914 (1)	30.26%	288	158.914	158.914 (1)	0
1n10	396.379	651.748 (2)	39.18%	120	530.4486	676.0565 (3)	21.54%
2n10	305.136	583.867 (2)	47.74%	105	425.9035	533.8335 (2)	20.22%
3n10	203.765	405.838 (2)	49.79%	146	322.3254	374.9160 (3)	14.03%
4n10	362.325	847.365 (3)	57.24%	101	419.3095	551.1373 (3)	23.92%
5n10	397.507	667.904 (3)	40.48%	112	508.9303	627.4120 (3)	18.88%
1n15	508.565	583.55 (?)	91.28%	84	651.7868	–	–
2n15	543.413	1085.60 (3)	49.94%	92	683.1724	–	–
3n15	491.502	5998.66 (?)	91.81%	96	716.6038	–	–
4n15	484.708	8091.55 (?)	94.01%	79	608.6757	–	–
5n15	568.785	6007.45 (?)	90.53%	85	694.8486	–	–

8 Conclusion and Future Works

This paper presents the EMIRP. It is closer to the reality with a continuous-time monitoring of inventory levels and continuous-time travel durations on the multi-graph. The link between inventory management and routing is tightened by the fact that the

consumption events of the customers happens at the same time as the vehicles travel in the road network and that the selection of routes is dependent both on the travel time and the delivery quantity.

The work presented in this chapter is still on-going and it can be improved in many ways. First, realist instances can be generated upon a real road network with more intermediate points. A multi-objective shortest path problem can be solved to compute Pareto-optimal paths between each pair of nodes. For the solution method, the application of Lagrangian relaxation can be improved. The IA sub-problem might be solvable by dynamic programming. The information given by the solution of IA can also help the solution of the CC sub-problems. Some symmetry breaking cuts need to be added to distinguish different vehicle routes. In addition, this decomposition method can be integrated with a fast heuristic to derive feasible solutions from the solutions of the relaxed problem.

References

- Air Liquide, EURO & ROADEF (2016). ROADEF/EURO Challenge 2016: Inventory Routing Problem, <http://challenge.roadef.org/2016/en/index.php>.
- Alkawaleet, N., Hsieh, Y.-F. & Wang, Y. (2014). Inventory Routing Problem with CO₂ Emissions Consideration, *Logistics Operations, Supply Chain Management and Sustainability*, Springer International Publishing, Cham, pp. 611–619.
- Andersson, H., Hoff, A., Christiansen, M., Hasle, G. & Løkketangen, A. (2010). Industrial aspects and literature survey: Combined inventory management and routing, *Computers and Operations Research* **37**(9): 1515–1536.
- Archetti, C., Bertazzi, L., Laporte, G. & Speranza, M. G. (2007). A Branch-and-Cut Algorithm for a Vendor-Managed Inventory-Routing Problem, *Transportation Science* **41**(3): 382–391.
- Bektaş, T. & Laporte, G. (2011). The Pollution-Routing Problem, *Transportation Research Part B: Methodological* **45**(8): 1232–1250.
- Bektaş, T. & Lysgaard, J. (2015). Optimal vehicle routing with lower and upper bounds on route durations, *Networks* **65**(2): 166–179.
- Cheng, C., Qi, M., Wang, X. & Zhang, Y. (2016). Multi-period inventory routing problem under carbon emission regulations, *International Journal of Production Economics* **182**: 263–275.
- Cheng, C., Yang, P., Qi, M. & Rousseau, L. M. (2017). Modeling a green inventory routing problem with a heterogeneous fleet, *Transportation Research Part E: Logistics and Transportation Review* **97**: 97–112.
- Chien, T. W., Balakrishnan, A. & Wong, R. T. (1989). An Integrated Inventory Allocation and Vehicle Routing Problem, *Transportation Science* **23**(December 2014): 67–76.
- Coelho, L. C., Cordeau, J.-F. & Laporte, G. (2013). Thirty Years of Inventory Routing, *Transportation Science* **48**(1): 1–19.
- Garaix, T., Artigues, C., Feillet, D. & Josselin, D. (2010). Vehicle routing problems with alternative paths: An application to on-demand transportation, *European Journal of Operational Research* **204**(1): 62–75.
- Geoffrion, A. M. (1974). Lagrangean relaxation for integer programming, *Approaches to integer programming*, Springer, pp. 82–114.
- He, Y., Briand, C. & Jozefowicz, N. (2016). A mass-flow based milp formulation for the inventory routing with explicit energy consumption, *Proceedings of 5th the International Conference on Operations Research and Enterprise Systems*, pp. 242–251.
- Kara, I., Kara, B. Y. & Yetis, M. K. (2007). Energy Minimizing Vehicle Routing Problem, *Verlag Berlin Heidelberg* **1**: 62–71.
- Krumke, S. O., Noltemeier, H., Schwarz, S., Wirth, H.-C. & Ravi, R. (1999). Flow Improvement and Network Flows with Fixed Costs, in P. Kall & H.-J. Lüthi (eds), *Operations Research Proceedings 1998: Selected Papers of the International Conference on Operations Research Zurich, August 31 – September 3, 1998*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 158–167.
- Lin, C., Choy, K., Ho, G., Chung, S. & Lam, H. (2014). Survey of Green Vehicle Routing Problem: Past and future trends, *Expert Systems with Applications* **41**(4): 1118–1138.
- Pajouh, F. M. & Balasundaram, B. (2010). Gradient-Type Methods, *Wiley Encyclopedia of Operations Research and Management Science*, John Wiley & Sons, Inc.
- Qian, J. & Eglese, R. (2016). Fuel emissions optimization in vehicle routing problems with time-varying speeds, *European Journal of Operational Research* **248**(3): 840–848.
- Treitl, S., Nolz, P. C. & Jammernegg, W. (2014). Incorporating environmental aspects in an inventory routing problem. A case study from the petrochemical industry, *Flexible Services and Manufacturing Journal* **26**(1-2): 143–169.

A FRAMEWORK FOR SUPPLY CHAIN RISK MODELING

S.-E. BEN JBARA, P. DAVID, G. ALPAN

Univ. Grenoble Alpes, CNRS, G-SCOP, 38 000 Grenoble, France

Saleh-Eddine.Ben-Jbara@grenoble-inp.fr, pierre.david@grenoble-inp.fr, gulgün.alpan-gaujaj@grenoble-inp.fr

ABSTRACT: *Controlling supply chain risks is an important issue for companies. Far from being only the prerogative of natural disasters, the disruptions of today's supply chains can sometimes be caused by minor events amplified by the increasing complexity of industrial organizations, causing severe economic losses. Risk management in supply chains is a recent theme and the proposed methods and tools to analyse risks are not yet fully mature to meet practitioners' expectations. One of the powerful solutions to analyse risks is using simulation. Despite its effectiveness to cover the complexity of the supply chains, the use of simulation still suffers from its difficulty of implementation. The aim of our work is to facilitate the use of simulation for risk analysis of supply chains. Thus, we have developed a modelling framework, which enables an easier construction of models of supply chain structure, behaviour and associated risks. This is done through the proposition of a set of meta-models and libraries, defined on the basis of the SCOR reference model. In addition, we propose a translation guide for the translation of the conceptual model of supply chains into a simulation model enabling testing risk scenarios. A library of simulation modules accompanies this translation.*

KEYWORDS: *Supply chain risks, Framework for simulation, Metamodeling, SCOR*

1 INTRODUCTION

In the last years, several supply chains (SCs) were subject to disruptions and witnessed negative impacts on their performance. According to a study made by Simchi Levi et al. (2013) from PwC consulting firm, over a sample of 209 companies with a global footprint, disruptions incurred negative impact on the financial performance of many companies: 54% of the companies said that their sales revenue has dropped and 64% of them suffered a decline in their customer service levels. Chopra and ManMohan (2014) state that surveys have shown that managers do little to prevent incidents since the solutions to reduce risks are not weighed against SC cost efficiency. Also, according to a recent study by Marchese and Paramasivam (2013) from Deloitte consulting firm on 600 supply chains, the top executives revealed that many companies do not master SC risk management (SCRM). In fact, only 33% used risk management approaches to proactively and strategically manage SC risks based on their operating environment conditions and 45% felt that their risk management was only "somewhat effective" or "not effective" at all. Many reasons can explain this. The major reason is the low implementation of the instruments of supply chain risk management as suggested by Jüttner (2005). Even if they are implemented many companies have an immature risk management process. Indeed, the study of Simchi Levi et al. (2013) shows that 59 % of the investigated companies have immature processes in place to effectively address incidents; their SCRM processes are neither proactive nor flexible. Another reason is the fact that companies misunderstand SC risk management. SCRM is still understood in many industries primarily as

a company-specific task, while it is necessary to focus not only on the company's own risks but also the risks of the partners (Jüttner 2005).

The awareness about having effective SCRM processes increases everyday within the industry. SC practitioners need to be assisted for improving their SCRM. Even if many risk management approaches exist and are part of the toolbox of managers they still lack user-friendliness and adapted processes to deploy them.

Some of the SCRM techniques are based on simulation. The literature works highlight numerous advantages of simulation for SC risk studies. For instance Chopra and ManMohan (2014) suggest conducting stress testing to evaluate "what if " scenarios and hence to assess risks. Simulation is also stated to have many advantages for SC problems. For instance, Longo and Mirabelli (2008) state that simulation is better for capturing the stochastic behaviour of SCs. Nevertheless, many works also highlight a problem of adoption of simulation despite its capacity to capture the dynamics of complex SCs. Wu et al. (2006) explain the problem by the lack of usability of the existing tools, while Cigolini et al. (2011) explain it by the lack of user-friendly commercial solutions, the lack of internal skills and/or the lack of time to develop a simulation model from scratch.

The use of simulation for SC risk assessment is quite dynamic in the past years. Most of the works develop system-oriented models. Some of them give an interest in proposing a general model for some specific risks, such as disruptions due to forecasting errors, stock-outs, bull-whip effect, etc. (e.g. (Guertler & Spinler 2015), (Ghadge et al. 2013), (Li et al. 2016)). However, this approach is limited by the kind of risks studied. Hence, we think that proposing general modelling frameworks for generic risks presenting similar features is an inter-

esting research direction. Very few works give methodological perspectives and frameworks for assisting SC practitioners in analysing their risks.

Therefore, the main target in this work is assisting the SC practitioners in creating the simulation models of their SC risks through promoting a quicker and easier construction of simulation models and through enabling risk scenarios experimentation.

The rest of the paper is organised as follows. In section 2, we provide an overview of our framework. In section 3, we present the meta-model to specify the structural and operational aspects of the supply chain as well as its potential risks. Section 4 gives insights on how to translate this meta-model into a simulation model. And finally, section 5 gives the concluding remarks.

2 FRAMEWORK OVERVIEW

Our modelling framework for simulation aims to fill in some of the lacks encountered in the literature. To cite some of them, first of all, most of the works are case specific. Second, most of the frameworks proposed in literature (Persson et al. 2012), (Cope et al. 2007) and (Sprock and McGinnis 2014) fail in covering all the aspects that enable an easy use (including the coverage of the SC domain knowledge, the communication of well-structured building blocks, integration of risks).

We identified a set of best practices for the development of an effective modelling framework for simulation. This work addresses the recommended practices cited below:

- Providing a graphical description of a meta-model is very beneficial. Choosing a good definition meta-modelling language increases the expressiveness of the meta-model and the perceived ease of use.
- Capturing the SC domain knowledge using reference model increases the fidelity of the modelling constructs to the modelled reality.
- Including concepts for risks will enhance the analysis capability of the framework outputs and enlarge the scope of its use.

- Providing a translation guideline and a simulation library to help SC practitioners rapidly and easily build their own simulation models based on conceptual SC models.

Figure 1 illustrates the support provided by the framework for each step of SC risk analysis: To analyse the risks threatening the SC, the first step is to build the conceptual model of the SC. The conceptual model needs to capture the elements that form the SC, namely, the SC structure, the SC behaviour and the risks threatening the SC. The support provided by the framework for this step is composed of meta-models to describe the structure and behaviour of the SC together with libraries that specify a set of building blocks and how they can be connected together to model a given SC. In this step, we also provide a meta-model of risks that covers numerous risks cited in the literature.

The second step that has to be conducted by the SC practitioner is to translate the conceptual model into a simulation model. The support provided by the framework for this step is a translation guideline that enables translating each element of the conceptual model into an element of the simulation model. This translation is also simplified by using a library of simulation modules.

Finally, the SC practitioner needs to experiment his/her model to test risk scenarios. The framework assists the SC practitioners through defining methods for the definition and the analysis of the scenario.

In this article, we describe the methodology adopted to develop the framework and the associated tools. We first explain the choice made for meta-modelling tool, and the principles followed to define meta-model for the structure, the behaviour and the risks, respectively. Secondly, we explain the choices made for translating the modelling constructs and the conceptual model into an executable simulation model. We also explain the adoption of the discrete event simulation (DES) and the choice of ARENA as an example of simulation platform.

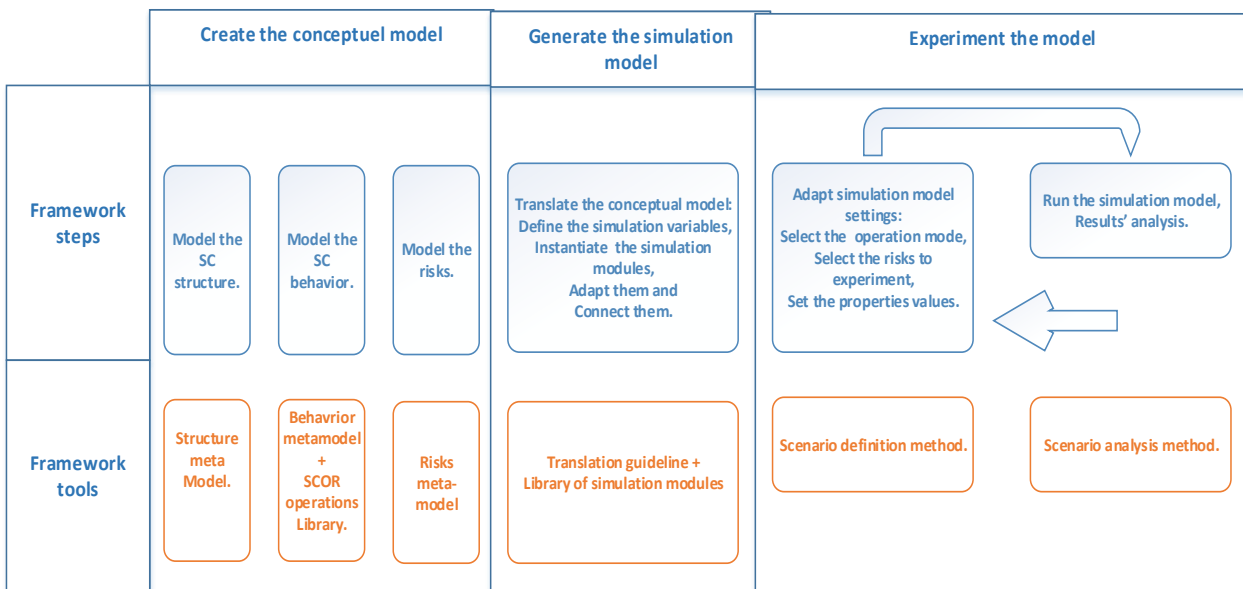


Figure 1: A framework for SC risk analysis

3 DEVELOPMENT OF THE SC MODELLING APPROACH

The first mission of the framework is to support the practitioner in describing the SC that is a complex system by nature. The idea is to use a systematic approach guiding the user in modelling the relevant aspects for risk analysis. For that purpose our approach is to propose a set of predefined building blocks expressing the concepts commonly manipulated by the SC owner (e.g. factories, supplier, stock policies, sourcing policy...). Technically, we define a meta-model to capture the SC domain and the SC risk domain. In the next paragraphs, we provide an overview of the theoretical and technical foundations of our meta-model.

Meta-modelling refers to the definition of a modelling language that permits the creation of models using the vocabulary of a given domain knowledge. The meta-model describes a set of building blocks and their relations. The users need only to instantiate the building blocks (value their features) to model his/her system.

Most of the works in the literature do not provide a complete description of a meta-model (such as the works of Labarthe et al. (2007) and Cope et al. (2007)). Furthermore, many papers fail in communicating about the semantics of their meta-models when they provide a graphical description of them (such as the works of Kitagawa et al. (2000) and Long and Zhang (2014)).

In the next section, we describe the strategy adopted to design the meta-model of SCs and risks that aim at capturing the domain knowledge and expressing it.

3.1 SysML as a meta-modelling language

The features of the SC domain and the various relationships within this domain need to be well described by the meta-models so that they can be adopted by the users.

To assure that this criterion is respected we use a widely adopted meta-modelling language that provides a normalized syntax and a semantic in a way that enhances the communication capability and the understanding of the developed meta-models. Our choice goes to SysML, which is an object-oriented (OO) modelling language that extends UML. SysML takes advantage of the capability of the OO constructs (such as packages, classes, and methods) to provide a natural way to capture complex systems. In fact, as stated by Coad et al. (1991), the OO concepts are aligned with the natural interpretation of the SCs expert domain to view the system as collections of related objects. The SysML specification already defines how the basic modelling constructs behave (inheritance mechanisms, association, inclusion of behaviour, features behaviour to quote some). Therefore our efforts are concentrated on providing an adapted semantic that express the SC domain.

3.2 Designing the meta-model constructs

To propose constructs that are easy to use and that enable capturing the SC domain, we choose to mix two

strategies. We propose general constructs that provide high flexibility and freedom to the modeller in describing different scenarios (e.g. resources, stock, operations, ...) and we reduce the customization effort required to specify the details of the general constructs through providing libraries of domain specific constructs (e.g. list of customized operations, for instance, for making or delivering products). The meta-model constructs are expressed as SysML blocks.

3.2.1 SC structure meta-model

The structure of the SC is understood as the static parts of the supply chain: the actors, the infrastructure (namely, the Facilities, the Buffers and the Resources), the products, and the transportation within the SC (namely, the Routes, the Paths and the related Transfer or Transportation Resources). A global view of the structure meta-model is given in Figure 2. Actors are linked together through the Contract that defines the terms of the exchange between the Actors of the SC. An Actor holds a set of Facilities. Each Facility may hold Resources and Buffers. TransportationResource and TransferResource are specific types of Resources. The Buffers are linked together through Paths. A Path is associated with TransferResources. Facilities are linked together through Routes. A Route is associated with TransportationResources.

The SC practitioner can model his/her SC structure using the proposed meta-model. He/She needs to pick the meta-model elements that fit with his/her supply chain and then specify the property values according to his/her supply chain.

The meta-model is built through various views. Each view regroups a set of elements that either defines a structural level of the SC network (e.g. Actor's network level) or supports a given activity (e.g. transportation). We define four views of the SC structure meta-model: the actor's network view, the product view, the infrastructure view and the transportation network view.

3.2.2 The SCOR reference model for SC behaviour

The SC behaviour meta-model is developed following the proposal made in (Ben Jbara et al. 2015). We defined views in the meta-model showing the flows (e.g. products, orders) and sets of functions realized by SC actors. We choose to use the SCOR reference model as a basis for the development of the library of functions. The Supply Chain Operations Reference-model (SCOR) has been developed and endorsed by the Supply-Chain Council (SCC) in 1996 as a cross-industry standard diagnostic tool for supply chain management. It provides a unified terminology and standard descriptions of processes that can be used to describe SCs. The SC functions are captured by SCOR model as a set of processes in three hierarchical levels. The fourth level, which is the implementation level that decomposes process elements, is out of the boundary of the SCOR model. SCC states that it is up to the company to decompose their own specific process elements.

Level 1 (as named in the SCOR model) is the top level that defines the process types: it defines the scope and the contents of the SCOR model. It consists of six global process types: Plan, Source, Make, Deliver, Return and Enable

Level 2 is the configuration level that defines the process categories: These categories enable a company to im-

plement their operations' strategy through the configuration they choose for their supply chain. Three policies for managing the supply chain are defined for the processes Source, Make and Deliver: policies linked to stock, linked to order and linked to engineering.

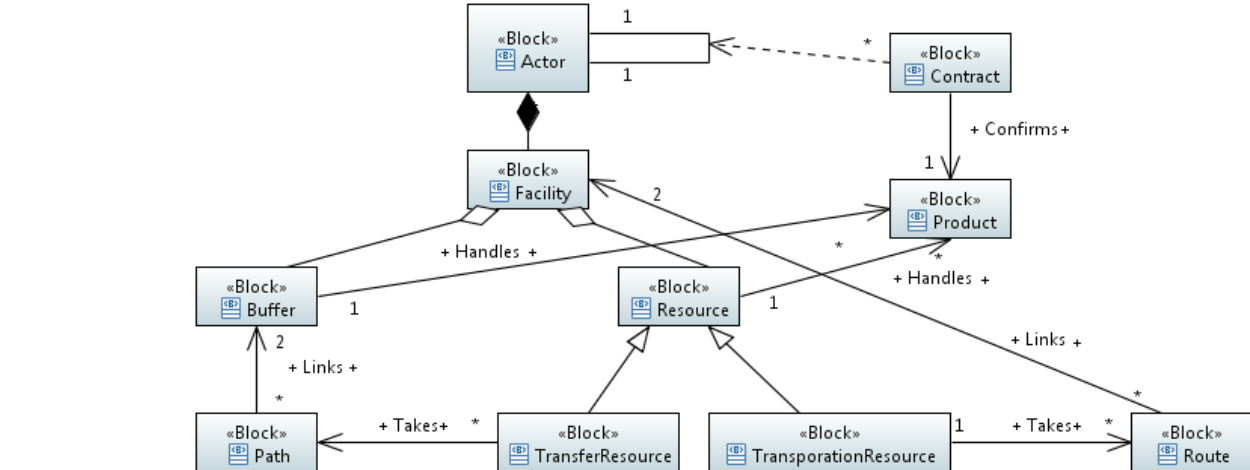


Figure 2: SC Structure meta-model blocks definition diagram

plement their operations' strategy through the configuration they choose for their supply chain. Three policies for managing the supply chain are defined for the processes Source, Make and Deliver: policies linked to stock, linked to order and linked to engineering.

The level 3 is the process element level that defines, for every process category, the different elements that compose it (e.g. Receive Product, Verify product). We use the process elements proposed by level 3 of SCOR to define domain-specific Operation constructs. For instance, following the SCOR notation, we use the "sM1.2/sM2.2 issueMaterial" to define the Operation "(sMi.2) ISSUEMATERIAL". Furthermore, we use the inputs and the outputs that are specified by SCOR for each process element to define a block for flows and the associated set of properties.

SCOR is adopted by numerous works in the past: Pundoor and Herrmann (2006) propose a framework to build DES models using the SCOR textual syntax. Persson et al. (2012) and Long (2014) use level two and three of SCOR in an aggregated way. Unlike our work, they neither specify the features of the exchanged variables defined by SCOR and nor consider the different ways of executing an operation. Since 2008, Gensym proposes the e-SCOR simulation tool where the building blocks are designed using SCOR processes. Dong et al. (2006) propose the IBM SmartSCOR, which is a simulation and optimization tool that uses the SCOR model to design some modelling constructs. Sprock and McGinnis (2014) propose a SCOR-compliant supply chain reference architecture that permits simulation model generation.

In difference to the above-mentioned works, we propose to go further and to define advanced constructs to capture SC features from SCOR. We note that, most of the

SCOR based frameworks proposed in the literature do not provide a description of how they capture the functioning of the SCOR behaviour elements such as the works of Sprock and McGinnis (2014) and Long and Zhang (2014). Hence, we propose to define detailed algorithms for the operations captured from SCOR.

Furthermore, the previous works do not describe how

they capture the inputs/outputs of the SCOR process elements. In our proposition, the input/output flows transferred between SCOR functions are modelled through well-defined modelling constructs.

As highlighted by many authors the relationships between the SC actors are not clearly described when presenting the SCOR processes. Hence, we propose specific constructs to extend SCOR for capturing the relations and the interactions that exist between the SC partners.

We propose to model the behavioural aspects of the supply chain using a set of Operation constructs. Each Operation is described by:

- An input/output table, summarizing the model element on which the operation acts (e.g. flows and elements of the structure). Table 1 illustrates the PRODUCE operation, where the type of each input and output corresponds to a class of the meta-model.
- A set of assumptions on the execution of the operation since SCOR does not provide a detailed information on how the workflows are to be executed.
- An algorithm that links the input variables to output variables, subject to a set of assumptions.

A core of 9 operations has been developed to model standard SCs. They are listed in Table 2.

SCOR Inputs	Retained Inputs and Outputs	Designations
	Inputs	
Workflow	pO [1..*]: ProductionOrder [1..*]	The received production order that informs about what to produce and the required quantity.
	rP: Product	The Product to be produced.
	iB: Buffer [1..*]	The input buffers where components are taken.
	oB: Buffer	The buffer where manufactured products are put.
	uR: Resource	The resource used for manufacturing.
	Outputs	
Information feedback, Waste produced, Workflow.	iFd: Informationfeedback [1..*]	The notification about the execution state.
	pO [1..*]: ProductionOrder [1..*]	The production order with a modified status.

Table 1: I/O for the PRODUCE Operation

A meta-model is also built to define the set of flows manipulated by a SC (e.g. orders, product flows, information feedback). The concept of process is also developed that permits to show interrelations between the Operations. In our framework, it takes the shape of an Activity diagram that shows the sequence of Operations and the exchanged flows. In this kind of views the Operation owners (e.g. SC actors) may be specified.

The Operations library
PRODUCE (sMi.3.1)
TEST (sMi.3.2)
ISSUE MATERIAL (sMi.3.3)
PICKANDPACK (C.sDi.9-sDi.10)
LOADVEHICLE (sDi.11)
SHIPPRODUCT (sDi.12+A+sSRi.5)
RECEIVE (sSi.2+A+sDi.13+A+sDRi.3)
VERIFY (sSi.3+A+sDi.13+A+sDRi.3)
TRANSFER (sSi.4+A+sDRi.4+A+ sD1.8)

Table 2 : The library of operations

3.2.3 Defining the risks modelling constructs

Finally, we focus on specifying Risk constructs to be integrated in the SC meta-model. The SC risks are numerous and heterogeneous. Ho et al. (2015) reported 20 research papers providing a SC risks categorization. We observe, four ways of classifying risks: by the likelihood of realization, the origin of the risk, the controllability (i.e. the capability of controlling the triggers) and impacts. Most of the categorizations are based on the risk “impact”. In fact, some of the proposed categorization refers to “impacted elements” (see for example (Tang & Nurmaya Musa 2011)), others refer to the “impacted functions” (e.g. (Bogataj & Bogataj 2007)), to the “nature of impacts” (e.g. (Talluri et al. 2013)) or to “level of impacted activities” (e.g. (Min & Zhou 2002)). An interesting categorization in this group is the one proposed by Saleh Ebrahimi et al. (2012), which focuses on the manner by which the model elements are impacted. This categorization is model oriented and concentrates on modelling risks and their impacts. Indeed, it is more efficient for risk assessment to define a modelling way for every risk category rather than defining a modelling way for each risk apart. This classification is impact oriented, so

it allows generating many types of risks whatever the cause is. For example, a machine breakdown or a strike in a plant may be modelled in the same way by an increasing production cycle time. We thus adopt the categorization proposed by Saleh Ebrahimi et al. (2012) and provide modelling constructs for each risk group. We provide a meta-model that specifies risk attributes and their interaction with other SC elements. Hence, the SC practitioners only need to identify the group in which the risk they want to test belongs to and to model it using the corresponding meta-model construct. Following (Saleh Ebrahimi et al. 2012), three risk classes are defined in the meta-model:

- Model’s variables modification (e.g. decreasing the capacity of a resource after a risk event)
- Operation modification (e.g. errors in a shipment operation)
- Model element destruction (e.g. complete breakdown of a factory or a road after an earthquake)

4 DEVELOPMENT OF TRANSLATION GUIDELINES AND SIMULATION MODULES

One of the current difficulties in using simulation for risks analysis is the construction of the simulation models. The SC practitioner needs to understand and to capture the various features of its SC and to express it as a simulation model using the syntax provided by the simulation software. Even if the current simulation software are useful and effective, their simulation modules are not specific for the SC domain and are of a low level of abstraction. Hence, numerous simulation bricks need to be combined and customized to simulate a small part of a SC.

The solution we develop in this work is to provide a translation guideline that enables the SC practitioner to directly translate its conceptual model into a simulation model expressed using a DES language. The translation is accompanied by a set of SC domain specific simulation modules.

We develop a correspondence between the SC meta-model and the constructs used in a DES.

We illustrate the translation approach by developing a set of SC domain specific simulation modules corre-

sponding to the library of Operations. The development of simulation modules includes the development of simulation algorithms, the declaration of the simulation variables, and the development of a user interface to set the parameter values.

To develop these modules we use the ARENA simulation software that has an integrated tool for generating new simulation modules. Many commercial simulation software exists in the market. Cimino et al. (2010) give a survey on the commonly used DES software. In this survey ARENA seems to be well perceived by the one hundred interviewed simulation experts. It has the best scores for user ability, modular construction, domain coverage (logistics and manufacturing) and the user community.

The translation is done in two steps: first the SC structure is defined and then the selected SC behaviour is generated.

4.1 Translation of SC structure

There are two methods to translate a structure construct.

- Using the simulation variables

Each property of the SC structure construct is translated into an array variable. The first column of the array is used to identify the related construct. The other columns are used for referencing other constructs in relation. For instance, the *cycleTime* property of the Resource construct is translated into an array variable of two dimensions where the first column contains the identifier of the related resource and the second contains the identifier of the products handled by this resource. An example is given for a buffer in Table 3.

- Using the predefined simulation modules

Some of the simulation software provides specific simulation modules to present physical constructs. When predefined simulation modules exist it is better to use them to get advantage of the predefined variables and functions. For example, ARENA provides a simulation pattern for defining the Resource construct. In this case a direct mapping can be made from our SC meta-model to the targeted module of the simulation tool.

For all structure constructs we defined the equivalent module in ARENA.

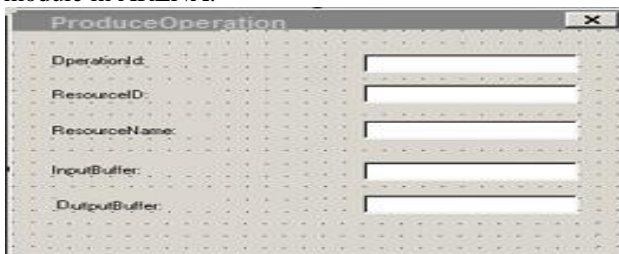


Figure 3: Dialog box for the Produce Operation simulation module

Model objects
<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">Buffer1: Buffer</p> <p style="text-align: center;">slots</p> <ul style="list-style-type: none"> Identifier: 1 Designation: ReceiveOutputBuffer Capacity: 100,100 InventoryLevel: 0,0 HandledProduct: Bd,Lv </div>
ARENA model variables
<p>The capacity variable which is an integer array of 2 dimensions is initialized with the following values:</p> <ul style="list-style-type: none"> • Capacity: <ul style="list-style-type: none"> (1,2)= 100 (1,3)=100 <p>The Inventory level variable which is an integer array of 2 dimensions is initialized with the following values:</p> <ul style="list-style-type: none"> • InventoryLevel: <ul style="list-style-type: none"> (1,2)= 100 (1,3)=100 <p>Where 1 refers to the <i>identifier</i> of Buffer1, 2 refers to the <i>identifier</i> of the Product Lv and 3 refers to the <i>identifier</i> of the Product Bd.</p>

Table 3: Example of a Buffer translation

4.2 Translation of the SC Behavior

The generation of the simulation model is made to emulate the behaviour of the SC. The translation of the SC behaviour declared in the conceptual model allows animating the listed flows and operations. To guide this creation we developed simulation modules corresponding to the Operation constructs available in the library given in Table 2. We also developed the correspondence between the flows declared in the conceptual model (see for example the I/O of Operations) and a set of simulation entities manipulated by ARENA model. Finally, the process modelling the interaction between the Operations are shown through the simulation chart of ARENA.

An advantage of using software such as ARENA is the possibility to develop simulation module libraries in relation to our Operations library. Hence we defined simulation blocks easing simulation parameterization. As an example, we show in Figure 3 the dialog box created for the parameterization of the PRODUCE Operation. This dialog box enables specifying the used resource, input and output buffers as mentioned in the conceptual model for the PRODUCE Operation illustrated in Table 1.

Associated to each dialog box, a sub-model is defined for each operation. This sub-model gives the operation logic based on the algorithm defined for each operation in the Operations library. The other I/O flows not already taken into consideration (for instance, production order for the PRODUCE Operation given in Table 1) are simulation entities to be read from a database (for example, from the information systems of the company)

The translation of the process of the conceptual model is made through interconnecting the Operation sub-models

following the connections expressed in the activity diagram showing the SC process organization.

Simulation modules are generated for Risk constructs in a similar manner. Hence, we also developed a set of risks simulation modules. A risk module is parameterized through a dialog box. As for the operations, behind a risk dialog box, there is a sub-model connectable to the rest of the simulation model.

An Arena model is thus created for a process or a group of processes through the interconnection of modules showing operations or risks. An example of a created model is given in Figure 4. It illustrates an excerpt of a trading goods process between two actors with a destruction of flow risk.

The obtained model enables conducting various “what if” scenarios of risks from one SC configuration. It is easy to change the tested risks by including/removing risks modules. The user may use classical techniques to define the scenarios to test and to set the simulation parameters such as the simulation length or the number of replications.

5 CONCLUSION

In this paper, we explain the methodology adopted for the development of a framework to support SC risk analysis by simulation. The provided tools enable modelling the SC and the associated risks, translating the conceptual model into a simulation model.

The method is tested on a fictitious example in (Ben Jbara 2018). A real case is of course necessary to validate its

applicability. Nevertheless, we believe that the developed concepts support the creation of effective simulation models by structuring the conceptual model that is necessary to build a simulation model. It also clarifies the elements to describe the SC structure, the way to connect them and provide mechanisms to embed a diversity of behaviour existing in SCs. To the best of our knowledge the closest work is by Sprock and McGinnis (2014). Our work differs from theirs by including SCOR in structuring the definition of the SC behaviour and by explicitly including the risks.

In its current form the model does not have an expression of the SC control issues (ex: inventory management strategies, production planning rules). This is the next perspective to include in further development of the framework. Despite the ease of use brought by a detailed data structure, constructing the models is still time consuming since the SC behaviour is complicated by nature (huge amount of data, many alternatives of interactions between actors or of policies).

ACKNOWLEDGMENTS

We would like to thank the Region Auvergne Rhône Alpes that supported this research.

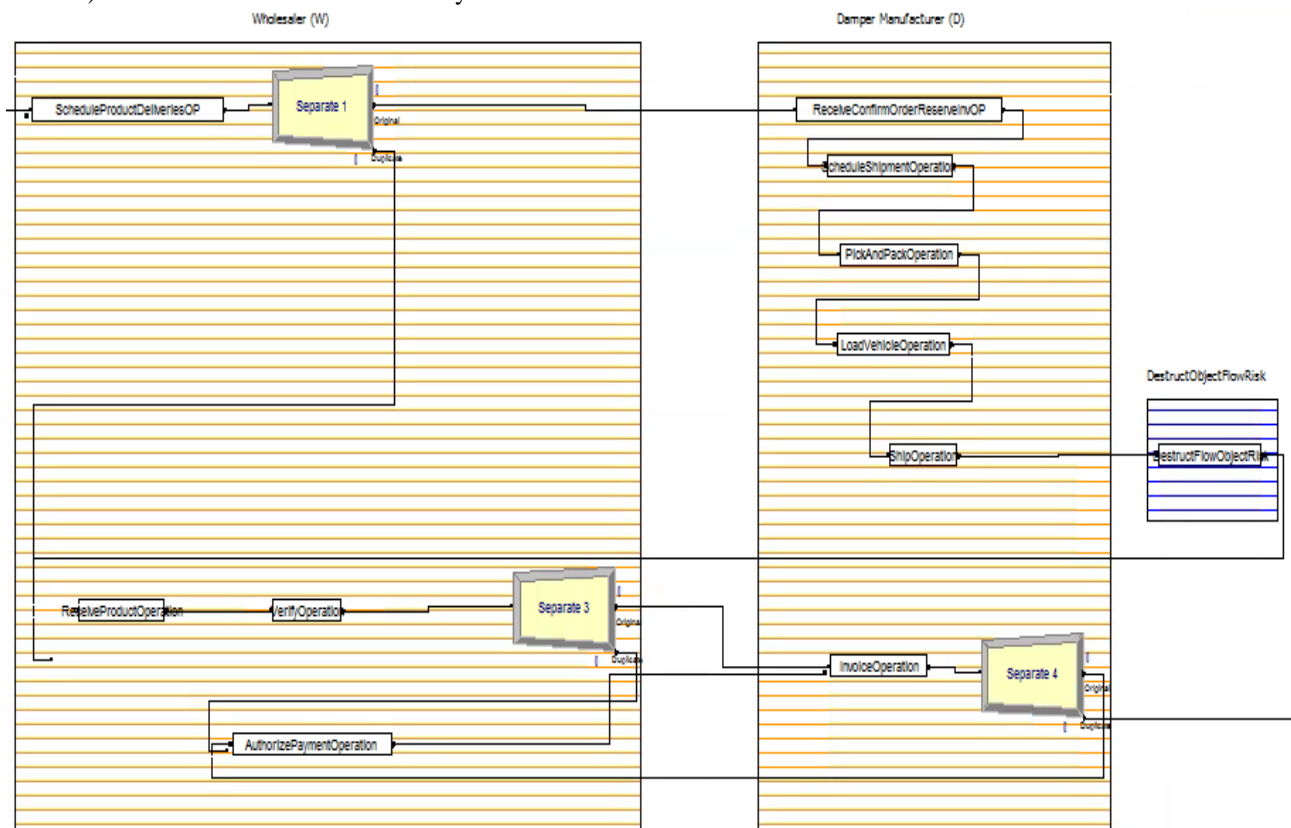


Figure 4: ARENA Model of a trading goods process subject to a risk of losing shipments

REFERENCES

- Ben Jbara, S. E. 2018. Risk management in supply chains: a simulation and model-based approach. *PhD. Thesis of Grenoble INP*, defended the 31st of January 2018.
- Ben Jbara, S. E., David, P., Alpan, G., 2015. A Modeling Framework Based on SCOR Model Towards Supply Chain Risk Management. In *45th International Conference on Computers & Industrial Engineering CIE45*.
- Bogataj, D. and Bogataj, M., 2007. Measuring the supply chain risk and vulnerability in frequency space. *International Journal of Production Economics*, 108(1–2), pp.291–301.
- Cigolini, R., Pero, M. and Rossi, T., 2011. An object-oriented simulation meta-model to analyse supply chain performance. *International Journal of Production Research*, 49(19), pp.5917–5941.
- Cimino, A., Longo, F. and Mirabelli, G., 2010. A General Simulation Framework for Supply Chain Modeling: State of the Art and Case Study. *International Journal of Computer Science Issues*, 7(2), pp.1–9.
- Clark, T., Sammut, P. and Willans, J., 2015. *Applied metamodelling: a foundation for language driven development* arXiv:1505., arXiv preprint.
- Coad, P., Yourdon, E. and Coad, P., 1991. *Object-oriented design* NJ: Yourdon press., ed., Englewood Cliffs.
- Chopra, S. and ManMohan, S.S., 2014. Managing risk to avoid supply-chain breakdown. *MIT Sloan management review*, 34(5), pp.360–387.
- Cope, D. et al., 2007. Supply chain simulation modeling made easy: An innovative approach. *2007 Winter Simulation Conference*, pp.1887–1896.
- Dong, J. et al., 2006. IBM SmartScor - A SCOR Based Supply Chain Transformation Platform Through Simulation And Optimization Techniques. In *Proceedings of the 2006 Winter Simulation Conference*.
- Ghadge, A. et al., 2013. A systems approach for modelling supply chain risks. *Supply chain management: an international journal*, 18(5), pp.523–538.
- Guertler, B. and Spinler, S., 2015. When does operational risk cause supply chain enterprises to tip? A simulation of intra-organizational dynamics. *Omega (United Kingdom)*, 57, pp.54–69.
- Ho, W. et al., 2015. Supply chain risk management: a literature review. *International Journal of Production Research*, 53(16), pp.5031–5069.
- Jüttner, U., 2005. Supply chain risk management Understanding the business requirements from a practitioner perspective. *The International Journal of Logistics Management*, 16(1), pp.120–141.
- Kitagawa, T. et al., 2000. A Description Language based on Multi Functional Modeling and a Supply Chain Simulation Tool. In *2000 IEEE*.
- Labarthe, O. et al., 2007. Toward a methodological framework for agent-based modelling and simulation of supply chains in a mass customization context. *Simulation Modelling Practice and Theory*, 15(2), pp.113–136.
- Li, C., 2013. Controlling the bullwhip effect in a supply chain system with constrained information flows. *Applied Mathematical Modelling*, 37(4), pp.1897–1909.
- Long, Q. and Zhang, W., 2014. An integrated framework for agent based inventory-production-transportation modeling and distributed simulation of supply chains. *Information Sciences*, 277(March), pp.567–581.
- Longo, F. and Mirabelli, G., 2008. An advanced supply chain management tool based on modeling and simulation. *Computers and Industrial Engineering*, 54(3), pp.570–588.
- Marchese, K. and Paramasivam, S., 2013. *The Ripple Effect How manufacturing and retail executives view the growing challenge of supply chain risk*, 2013 Deloitte Development LLC.
- Min, H. and Zhou, G., 2002. Supply chain modeling: past, present and future. *Computers & Industrial Engineering*, 43(1–2), pp.231–249.
- Persson, F. et al., 2012. Supply chain dynamics in the SCOR model- A simulation modeling approach. In *Proceedings of the 2012 Winter Simulation Conference*. pp. 3821–3832.
- Pundoor, G. and Herrmann, J., 2006. A hierarchical approach to supply chain simulation modelling using the Supply Chain Operations Reference model. *International Journal of Simulation and Process Modelling*, 2(3–4), pp.124–132.
- Saleh Ebrahimi, D., David, P. and Alpan, G., 2012. A Model Based Specification For A decision Support Tool For Supply Chain Risk Management. In *CIE42 Proceedings*. Cape Town, South Africa.
- Simchi Levi, D., Vassiladis, C. and Kyriatoglou, I., 2013. Making the right risk decisions to strengthen operations performance. In *PwC and the MIT Forum for Supply Chain Innovation*.
- Sprock, T. and McGinnis, L.F., 2014. Towards automated access to supply chain analyses. In *Proceedings of the 2014 POMS Annual Conference*.
- Talluri, S. (Sri) et al., 2013. Assessing the Efficiency of Risk Mitigation Strategies in Supply Chains. *Journal of Business Logistics*, 34(4), pp.253–269.
- Tang, O. and Nurmaya Musa, S., 2011. Identifying risk issues and research advancements in supply chain risk management. *International Journal of Production Economics*, 133(1), pp.25–34.
- Wu, T., Blackhurst, J. and Chidambaram, V., 2006. A model for inbound supply risk analysis. *Computers in Industry*, 57(4), pp.350–365

A new class of spectral risk measures

Mohammed Berkouch ; Ghizlane Lakhnati

Marcelo Brutti Righi

LGII, ENSA
Ibn Zohr University,
Agadir - Morocco

Federal University of
Rio Grande do Sul,
Porto Alegre - Brazil

mohammed.berkouch@edu.uiz.ac.ma ; g.lakhnati@uiz.ac.ma marcelo.righi@ufrgs.br

ABSTRACT: *The aim of this paper is to introduce a class of spectral risk measures that extends the Gini-type measure of risk and variability, by taking risk aversion into consideration. Our class of risk measures is coherent and catches variability, an important concept for risk management. The analysis is made under the Choquet integral representation framework. A practical application is provided.*

KEYWORDS: *risk measures, variability measures, risk aversion, signed Choquet integral, Extended Gini Shortfall.*

1 INTRODUCTION

In modern risk management, a large number of risk measures have been proposed in the literature. These measures are mappings from a set of random variables (financial losses) to real numbers. Among the most interesting and potentially most promising propositions has been the theory of spectral risk measures (SRMs), proposed by (Acerbi, 2002). SRMs are characterized by a weighting function ϕ which reflects the decision-maker's attitude towards risk, however there has been little guidance on the choice of ϕ in the literature.

"... no natural choice is provided by purely financial arguments and the function ϕ appears as the instrument by which an investor can express her subjective attitude toward risk". (Acerbi, 2002)

Recently, (Furman et al., 2017) introduced Gini Shortfall (GS) risk measure which is a composition between Expected Shortfall (ES) and tail based Gini coefficient. Despite being a SRM, GS supposes that all individuals have the same attitude towards risk, while agents differ in the way they take personal decisions that involve risk because of discrepancies in their risk aversion. Thus, since the weighting function ϕ in SRMs is supposed to express the subjective attitude towards risk, it is more legitimate for ϕ to be dependent on a risk aversion parameter.

To incorporate such psychological behavior in tail risk analysis, we introduce a generalized version of GS. That means that GS is actually only one possible choice of a large class of spectral risk measure risk measures that we called Extended Gini Shortfall (EGS). This approach allow us to consider both like-

lihood of loss and variability which are two pillars of the notion of risk.

2 MAIN OBJECTIVES

- Introduce the concept of EGS risk measures.
- Explore properties about EGS.
- Provide an interpretation.
- Illustrate the results by means of a practical application.

3 METHODOLOGY

Our strategy will be to employ the concept of Extended Gini coefficient introduced by Yitzhaki and (Schechtman, 2005) with Theorem 2.1 in (Furman et al. 2017) as a guideline to construct our risk measures class, and explore its properties by exploiting the Choquet integral representation.

In this paper, $X \in L^0$ is a rv modeling financial profit and loss, F_X denotes the cdf of X , p is a confidence level, and r is a risk aversion degree.

We begin by exposing definitions of both risk and variability measures. We assume throughout the paper that all functionals respect the following property, which is essential in order to obtain a functional directly from its distribution function.

Definition 3.1. *A functional f is said to be law invariant if it fulfills the following property:*

(A) *Law Invariance: If $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ have the same distributions under \mathbb{P} , succinctly $X \stackrel{d}{=} Y$, then $f(X) = f(Y)$.*

Definition 3.2. A risk measure is a functional $\rho : \mathcal{X} \rightarrow (-\infty, \infty]$, which may fulfill the following properties:

(B1) *Monotonicity:* $\rho(X) \leq \rho(Y)$ when $X, Y \in \mathcal{X}$ are such that $X \leq Y$ \mathbb{P} -almost surely.

(B2) *Translation invariance:* $\rho(X + m) = \rho(X) + m$ for all $m \in \mathbb{R}$ and $X \in \mathcal{X}$.

(A1) *Positive homogeneity:* $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda > 0$ and $X \in \mathcal{X}$.

(A2) *Subadditivity:* $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$.

(A3) *Co-monotonic additivity:* $\rho(X + Y) = \rho(X) + \rho(Y)$ for every co-monotonic pair $X, Y \in \mathcal{X}$.

A risk measure is monetary if it satisfies properties (B1) and (B2), and it is coherent if it satisfies furthermore (A1) and (A2).

Remark 3.3. For interpretations of these properties, we refer the reader to (Föllmer and Schied, 2011, Chap 4), (Delbaen, 2012), and (McNeil et al., 2015). For example, both functionals VaR and ES are monetary and co-monotonically additive, whereas ES is even coherent.

Definition 3.4. Let $p \in (0, 1)$. We have that the Value at Risk and the Expected Shortfall are functionals $\text{VaR}_p : L^0 \rightarrow (-\infty, \infty]$ and $\text{ES}_p : L^1 \rightarrow (-\infty, \infty]$ defined conform:

$$\text{VaR}_p(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, \quad (1)$$

$$\text{ES}_p(X) = \frac{1}{1-p} \int_p^1 \text{VaR}_q(X) dq. \quad (2)$$

Definition 3.5. A functional $\nu : \mathcal{X} \rightarrow [0, \infty]$ is a measure of variability, which may fulfill the following properties¹:

(C1) *Standardization:* $\nu(c) = 0$ for all $c \in \mathbb{R}$.

(C2) *Location invariance:* $\nu(X + c) = \nu(X)$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

¹Inspired from the deviation measures notion of (Rockafellar et al., 2006).

(A1) *Positive homogeneity:* $\nu(\lambda X) = \lambda \nu(X)$ for all $\lambda > 0$ and $X \in \mathcal{X}$.

(A2) *Subadditivity:* $\nu(X + Y) \leq \nu(X) + \nu(Y)$ for all $X, Y \in \mathcal{X}$

(A3) *Co-monotonic additivity:* $\nu(X + Y) = \nu(X) + \nu(Y)$ for every co-monotonic pair $X, Y \in \mathcal{X}$.

A measure of variability is coherent if it further satisfies (C1), (C2), (A1) and (A2).

Remark 3.6. For instance, the most classical measures of variability are the Variance and the Standard Deviation. The variance functional satisfies properties (A), (C1), (C2) but not (A1) or (A2), hence it is not coherent. On the other side, the standard deviation functional, since satisfying all aforementioned properties, is coherent. Neither the variance nor the standard deviation is co-monotonically additive.

3.1 Extended Gini Shortfall

The notion of signed Choquet integral plays a pivotal role thereafter. It originates from (Choquet, 1954), in the framework of capacities, and is further characterized and studied in decision theory by (Schmeidler, 1986, 1989).

Definition 3.7. A function $h : [0, 1] \rightarrow \mathbb{R}$ is called a distortion function when it is non-decreasing and satisfies the boundary conditions $h(0) = 0$ and $h(1) = 1$. Let h be a distortion function, the functional defined by the equation:

$$I(X) = \int_0^\infty (1-h(F_X(x)))dx - \int_{-\infty}^0 h(F_X(x))dx \quad (3)$$

for all $X \in \mathcal{X}$ is called the (increasing) Choquet integral. Whenever $h : [0, 1] \rightarrow \mathbb{R}$ is of finite variation, I is called the signed Choquet integral.

When h is right-continuous, then equation(3) can be rewritten as (Wang et al. 2017):

$$I(X) = \int_0^1 F_X^{-1}(t)dh(t). \quad (4)$$

Furthermore, when h is absolutely continuous, with ϕ a function such that $dh(t) = \phi(t)dt$, then equation(4) becomes:

$$I(X) = \int_0^1 F_X^{-1}(t)\phi(t)dt. \quad (5)$$

In this case, ϕ is called the *weighting functional* of the signed Choquet integral I .

The following theorem is enunciated with a complete proof in (Furman et al., 2017), it gives the characterization for co-monotonically additive and coherent measures of variability.

Theorem 3.8. *Let $\nu : L^q \rightarrow \mathbb{R}$ be any L^q -continuous functional. The following three statements are equivalent:*

(i) ν is a co-monotonically additive and coherent measure of variability.

(ii) There is a convex function $h : [0, 1] \rightarrow \mathbb{R}$, $h(0) = h(1) = 0$, such that

$$\nu(X) = \int_0^1 F_X^{-1}(u)dh(u), \quad X \in L^q. \quad (6)$$

(iii) There is a non-decreasing function $g : [0, 1] \rightarrow \mathbb{R}$ such that

$$\nu(X) = Cov[X, g(U_X)], \quad X \in L^q. \quad (7)$$

We recall Extended Gini coefficient (EG) is a measure of variability, widely used in Economics and Finance, that depends on a risk aversion parameter r . EG is formally expressed in the covariance term by:

$$EGini_r(X) = -2rCov[X, (1-F_X(X))^{r-1}], r > 1 \quad (8)$$

There are special cases of interest for the Extended Gini:

- For $r = 2$: the Extended Gini coefficient becomes the simple Gini.
- For $r \rightarrow \infty$: the Extended Gini reflects the attitude of a max-min decision-maker who expresses risk only in terms of the worst outcome.
- For $r \rightarrow 1$: the Extended Gini tends to zero and represents the attitude of a risk-neutral individual who does not care about variability.

Now, we turn the focus to an adaptation to the tail of the distribution function. In this sense, we introduce the Tail Extended Gini (TEG):

Definition 3.9. *The Tail Extended Gini is a functional $TEGini_{r,p} : L^1 \rightarrow [0, \infty]$ defined conform:*

$$TEGini_{r,p}(X) = \frac{-2r}{1-p} Cov[X, (1 - F_X(X))^{r-1} | X > x_p], \quad (9)$$

where, $r > 1$ and $0 < p < 1$.

Thus, TEG represents a version of EG conditioned on VaR being violated.

Proposition 3.10. *The Tail Extended Gini is standardized, location invariant, positively homogeneous and co-monotonic additive. Moreover, it is a signed Choquet integral conform:*

$$TEGini_{r,p}(X) = \frac{2}{(1-p)^2} \int_p^1 F_X^{-1}(u)[g_r(u)+(1-p)^{r-1}]du. \quad (10)$$

with, $g_r(u) = -r(1-u)^{r-1}$.

Then, we define our main contribution, Extended Gini Shortfall, which is a linear combination of the Expected Shortfall risk measure with the Tail Extended Gini measure of variability:

Definition 3.11. *The Extended Gini Shortfall is a functional $EGS_{r,p}^\lambda : L^1 \rightarrow (-\infty, \infty]$ defined conform:*

$$EGS_{r,p}^\lambda(X) = ES_p(X) + \lambda TEGini_{r,p}(X), \lambda \geq 0 \quad (11)$$

where, λ is a loading parameter.

3.2 Properties of EGS

In order to be a reasonable tool for risk management, the properties of coherent risk measures are desired. But, TEG is not sub-additive (as it can be shown with a counter-example), and as a measure of variability is not monotone. However, when λ is zero, then EGS obviously inherits all the properties of the ES which is coherent, but when λ is sufficiently large, then the TEG starts to dominate ES, and thus coherence of EGS cannot be expected. Intuitively, as suggested by (Furman et al., 2017), there might be a threshold that delineates the value of λ for which EGS is coherent.

Despite the fact that TEG is not coherent, we show that penalizing the ES by TEG gives rise to a coherent measure of risk that quantifies both the magnitude and the variability of tail risks.

Proposition 3.12. *The Extended Gini shortfall is translation invariant, positively homogeneous, and co-monotonically additive, which can be represented as a signed Choquet integral conform:*

$$EGS_{r,p}^\lambda(X) = \int_0^1 F_X^{-1}(u) \phi_{r,p}^\lambda(u) du \quad (12)$$

where,

$$\phi_{r,p}^\lambda(u) = \frac{1}{(1-p)^2} [1-p+2\lambda(g_r(u)+(1-p)^{r-1})] \mathbb{1}_{[p,1]}(u) \quad (13)$$

and $u \in [0, 1]$.

Moreover, the Extended Gini Shortfall is a coherent risk measure for $\lambda \in [0, 1/(2(r-1)(1-p)^{r-2})]$.

3.3 Interpretation

From the previous Proposition, we get the Extended Gini Shortfall is part of the spectral risk measures class, introduced in (Acerbi, 2002), characterized by the weighting function $\phi_{r,p}^\lambda$, which enables to reflect the individual's subjective attitude toward risk. In (Furman et al., 2017) a specific case is introduced ($r = 2$), which assigns the same weighting function to all decision-makers.

It is valid to point out that by the spectral representation with a partial derivatives analysis we can provide an interpretation about how this weighting function $\phi_{r,p}^\lambda$ behaves in relation to changes of each parameter among u, r, p , and λ .

Proposition 3.13. *Consider the weighting function*

$$\begin{aligned}\phi_{r,p}^\lambda(u) &= \phi(u, r, p, \lambda) \\ &= \frac{1 - p + 2\lambda[(1 - p)^{r-1} + g_r(u)]}{(1 - p)^2} \mathbb{1}_{[p,1]}(u),\end{aligned}$$

where $u \in [0, 1]$, $p \in (0, 1)$, $r > 1$, and $\lambda \in [0, 1/(2(r-1)(1-p)^{r-2})]$. We have that:

(i) *The interval for values of λ that make EGS subadditive has superior limit non-decreasing in p , if $r \geq 2$ and non-increasing in p otherwise. Moreover, the interval is non-decreasing in r if, and only if, $r \geq 1 - \frac{1}{\ln(1-p)} = r_0$;*

(ii) *ϕ is non-decreasing in u ;*

(iii) *ϕ is non-decreasing in λ ;*

(iv) *ϕ is non-decreasing in p if, and only if*

$$u \geq 1 - \left(\frac{1 - p - 2\lambda(r-3)(1-p)^{r-1}}{4\lambda r} \right)^{\left(\frac{1}{r-1}\right)};$$

(v) *ϕ is non-decreasing in r if, and only if*

$$(1-p)^{(1-p)} \geq (\exp\{(1-u)^{r-1}[r \ln(1-u) + 1]\})^{\left(\frac{1}{r-1}\right)}.$$

The non-linear behavior of the partial derivatives found by the previous proposition is related to the function $g_r(u) = -r(1-u)^{r-1}$, which is central to the theory we develop, which is not trivial for $r \neq 2$, the case considered in (Furman et al., 2017). As one can easily note, it recurrently appears in the expression for partial derivatives. Hence, our contribution is by extending the canonical case to a situation where more complexity regarding r can be addressed.

Since p reflects the prudence level, which is usually close to 1 in practice, then r_0 in item (i) is in general close to 1. Thus, for practical matters, the superior limit for λ is non-decreasing in most relevant values of r . The most complex parameter sensitivities are regarding the prudence level p and the generalization term r , respectively in items (iv) and (v), because both $\frac{\partial \phi}{\partial p}$ and $\frac{\partial \phi}{\partial r}$ are expressions that can assume both positive and negative values since some terms in numerator possess distinct signs without a domination. This can be explained due to the fact that ϕ is a weighting function and changes in p and r alter how much 'mass' is put to any probability level u . Because $\int_0^1 \phi_{r,p}^\lambda(u) du = 1$, it is necessary that the increase on $\phi(u)$ for some values of u be compensated by some decrease in others.

Regarding prudence level, item (iv) in the previous proposition emphasizes that ϕ non-decreases in p for larger values of u . This is in consonance with practical intuition because more weight is put to extreme probabilities. Moreover, this is corroborated when we verify the variations of ϕ regarding u and p , where we obtain the non-negative expression

$$\frac{\partial^2 \phi}{\partial u \partial p} = \frac{4\lambda r(r-1)(1-u)^{r-2}}{(1-p)^3} \mathbb{1}_{[p,1]}(u).$$

When we consider the sensibility of variations regarding u and r , in a case we get

$$\frac{\partial^2 \phi}{\partial u \partial r} = \frac{2\lambda(1-u)^{r-2}[(2r-1) + (r^2-r)\ln(1-u)]}{(1-p)^2} \mathbb{1}_{[p,1]}(u),$$

there is divergence about sign – first term inside brackets is non-negative while the second is non-positive, corroborating with the previous argument. Nonetheless, ϕ is non-decreasing in r , a situation where r behaves more like a risk aversion coefficient, when a non-decreasing function of p , $(1-p)^{(1-p)}$ is greater than a threshold that is an expression depending on both u and r . Repeating the argument of practical choices for the prudence level, p will be close to 1 and ϕ will be increasing in r for more values of u . In this sense, r must be understood as a generalization parameter for a family of EGS risk measures rather than a single linear risk aversion coefficient.

4 APPLICATION

In practice, the assessment of $EGS_{r,p}^\lambda$ can be reduced to evaluating its discrete version proposed by (Acerbi, 2002) as a consistent estimator for spectral risk measures:

$$\widehat{EGS}_{r,p}^\lambda(X) = \sum_{i=1}^N X_{(i)} \phi_i, \quad (14)$$

where, $\{X_{(i)}; i = 1, \dots, N\}$ are the ordered statistics given by the N-tuple $\{X_1, \dots, X_N\}$ of observations,

and ϕ_i is the natural choice for a suitable weighting function given by:

$$\phi_i = \frac{\phi_{r,p}^\lambda(i/N)}{\sum_{k=1}^N \phi_{r,p}^\lambda(k/N)} \quad i = 1, \dots, N \quad (15)$$

and satisfying $\sum_i \phi_i = 1$.

Remark 4.1. When r is sufficiently high, $(1 - p)^{r-1} - r(1 - u)^{r-1} \rightarrow 0$ thus $\phi_{r,p}^\lambda \rightarrow \frac{1}{1 - p}$.

That is to say, under this condition, $EGS_{r,p}^\lambda$ and ES_p give similar outcomes.

This remark ensures that, for a highly risk averse investor, we can simply use ES_p . In other words, ES_p can be considered as a limit of $EGS_{r,p}^\lambda$ for a quite high risk aversion degree. Furthermore, since more averse the decision-maker is less risk he takes then we can claim that $EGS_{r,p}^\lambda$ is at least equal to ES_p i.e., $EGS_{r,p}^\lambda \geq ES_p$.

An illustration of the practical usage of EGS is provided with a numerical example. Our dataset consists of daily returns from the MASI index covering the period of 15th November 2016 to the 15th November 2017 (Figure 1), which includes a total of $N = 250$ observations. The MASI (Moroccan All Shares Index) is a stock index that tracks the performance of all securities listed in the Casablanca Stock Exchange located at Casablanca in Morocco.²

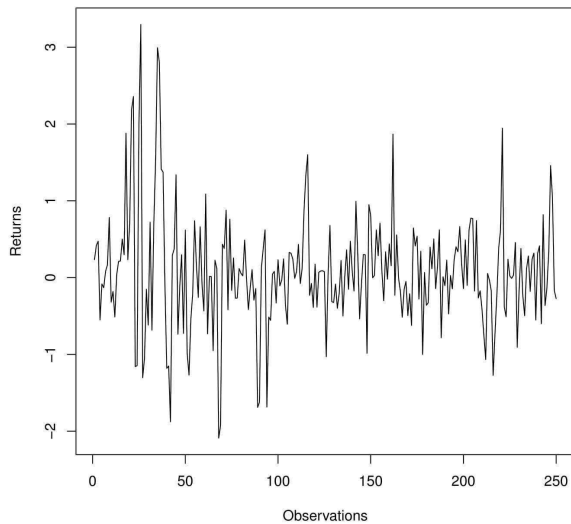


Figure 1: Graph of the daily observed MASI return.

By using the sorted returns series we calculate the VaR_p value to identify the concerned losses. Then, we affect to each value its own weight in accordance

²<http://www.casablanca-bourse.com>

Sorted Returns(%)	Weights
⋮	⋮
-1.146	0.04
-1.151	0.05
-1.160	0.05
-1.181	0.06
-1.270	0.06
-1.273	0.07
-1.304	0.08
-1.621	0.08
-1.685	0.09
-1.690	0.10
-1.880	0.10
-1.924	0.11
-2.090	0.11

Table 1: Weighted losses beyond VaR for $p=95\%$ and $r=2$.

with the parameters p, r , and λ as shown below in Table 1.

To fulfill the condition $\lambda \in [0, 1/(2(r - 1)(1 - p)^{r-2})]$ we take λ arbitrarily as the midpoint of the interval, thus $\lambda = \frac{1}{4(r - 1)(1 - p)^{r-2}}$.

In Table 2 below, we report the calculation results of $\widehat{EGS}_{r,p}^\lambda$ for different values of p and r .

$\widehat{EGS}_{r,p}^\lambda$	$r = 2 (GS_p^\lambda)$	$r = 3$	$r = 6$	$r = 20$	$r = 30$
$p = 90\%$					
$VaR = 0.73\%$	1.32%	1.28%	1.25%	1.22%	1.22%
$ES = 1.21\%$					
$p = 95\%$					
$VaR = 1.11\%$	1.58%	1.55%	1.52%	1.50%	1.49%
$ES = 1.49\%$					
$p = 99\%$					
$VaR = 1.79\%$	1.99%	1.98%	1.97%	1.96%	1.96%
$ES = 1.96\%$					

Table 2: Outcomes of the EGS estimator according to p and r .

This empirical exercise using the daily returns for the MASI index between 15th November 2016 and 15th November 2017 is a historical approach that illustrates the practical use of $EGS_{r,p}^\lambda$ in the real world by considering the psychological attitude of the decision-

maker. The obtained results confirm earlier remarks: first, more the decision-maker is risk averse less risks he takes and then smaller is the amount of capital required to hedge his position; furthermore, we have $EGS_{r,p}^\lambda \geq ES_p$ and in a highly risk averse context ($r \geq 20$ for this survey) values of both risk measures are similar.

FORTHCOMING RESEARCH

The main contribution in this paper is on theoretical results, but this approach gives rise to further forthcoming investigations: capital requirement, portfolio management,... In this sense, parametric risk-forecasting of EGS is a subject that deserves a further and separate survey which will be made in a forthcoming work.

REFERENCES

- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**, 1505-1518.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**, 203-228.
- Ceriani, L. and Verme, P. (2012). The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *Journal of Economic Inequality*, **10**, 421-443.
- Dana, R.-A. (2005). A representation result for concave Schur-concave functions. *Mathematical Finance*, **15**, 613-634.
- Delbaen, F. (2012): *Monetary Utility Functions*. Osaka University Press, Osaka.
- Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, **5**, 131-295.
- Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*. (Third Edition.) Walter de Gruyter, Berlin.
- Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin: Journal of the International Actuarial Association*, **36**, 433-462.
- Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocation and heavy-tailed risks. *Journal of Banking and Finance*. In Press, Accepted Manuscript.
- Giorgi, G.M. (1990) Bibliographic portrait of the Gini concentration ratio. *Metron*, **48**, 183-221.
- Mao, T. and Wang, R. (2016). Risk aversion in regulatory capital principles. Working paper.
- McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. (Revised Edition.) Princeton University Press, Princeton, NJ.
- Righi, M. and Ceretta, P. (2016). Shortfall Deviation Risk: an alternative to risk measurement.
- Righi, M. (2017). A composition between risk and deviation measures. Working Paper.
- Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviation in risk analysis. *Finance and Stochastics*, **10**, 51-74.
- Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**, 255-261.
- Schmeidler, D. (1989). Subject probability and expected utility without additivity. *Econometrica*, **57**(3), 571-587.
- Shalit, H., and Yitzhaki, S. (1984). Mean-Gini, portfolio theory and the pricing of risky assets. *Journal of Finance*, **39**(5 (December)), 1449-1468.
- Wang, R., Wei, Y. and Willmot., G.E. (2017). Characterization, Robustness and Aggregation of Signed Choquet Integrals. Working paper.
- Yaari, M.E. (1987). The dual theory of choice under risk. *Econometrica*, **55**, 95-115.
- Yitzhaki, S. (1983) On an extension of Gini inequality index. *International Economic Review*, **24**, no. **3**, 617-628.
- Yitzhaki, S. (1998) More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, **8**, 13-30.
- Yitzhaki, S. and Schechtman, E. (2005). The properties of the extended Gini measures of variability and inequality. *Metron*, LXIII(3), 401-443.
- Yitzhaki, S. and Schechtman, E. (2013). *The Gini Methodology: A Primer on a Statistical Methodology*. Springer, New York, NY.

Algorithme d'approximation pour un problème d'affectation d'une commande multiproduit multifournisseur

Azeddine CHEREF, Christian ARTIGUES,
Emmanuel HÉBRARD, Pierre LOPEZ
Sandra U. NGUEVEU

Vincent CHARVILLAT

Yves Régent

CNRS, LAAS-CNRS

Université de Toulouse, INSA, INP, France
cheref@laas.fr, artigues@laas.fr, hebrard@laas.fr
lopez@laas.fr, ngueveu@laas.fr

Université de Toulouse

IRIT, INP, France
Vincent.Charvillat@enseeiht.fr

Devatics

Toulouse, France
yregent@devatics.com

RÉSUMÉ : Dans le contexte de la vente en ligne, étant donné une commande constituée d'un ensemble de produits et un ensemble de magasins, le problème que nous présentons dans ce papier consiste à répartir la commande sur un ensemble de magasins de sorte à minimiser le coût de la commande. Ce problème se retrouve dans plusieurs contextes et comporte beaucoup de similitudes avec le problème de localisation d'entrepôts. Afin d'introduire l'algorithme de résolution proposé, nous présentons une heuristique d'approximation pour cas où les produits sont disponibles dans les magasins en quantité illimitée. Le problème étant une généralisation du problème couverture par ensembles, il est par conséquent NP-difficile. Ainsi, une heuristique avec garantie de performance basée sur l'approche d'arrondi en programmation linéaire est proposée pour le problème traité.

MOTS-CLÉS : Approximation, Problème de localisation, Arrondi en programmation linéaire.

1 INTRODUCTION

Dans le contexte de la vente en ligne, une commande d'un client peut être constituée de plusieurs produits en quantités différentes. De plus, les produits commandés sont vendus par différents magasins et chacun d'eux dispose d'une certaine quantité pour chaque produit. L'objectif pour le client est d'acheter tous les produits de la commande en minimisant le prix global. Ce prix dépend du coût des produits ainsi que du coût d'acheminement depuis les magasins. En effet, les coûts des produits peuvent être différents selon les magasins sélectionnés et les coûts de transports dépendent également des magasins.

Le problème considéré dans ce papier consiste alors à sélectionner un ensemble de magasins de sorte à satisfaire la commande du client au meilleur coût. Ce problème comporte beaucoup de similitudes avec le problème de localisation d'entrepôts (*Facility Location*) et peut être considéré comme une extension de celui-ci. En effet, le problème de localisation d'entrepôts consiste à déterminer un ensemble d'emplacements afin d'implanter des installations (voir (Melo, Nickel & da Gama 2009) et (Daskin, Snyder & Berger 2005) pour plus de détails). Le problème général de localisation d'entrepôts standard ne prend pas en compte les capacités ainsi que les caractéristiques des installations potentielles (capacités, types

de produits, etc). Ce problème a été étendu et des capacités sont prises en compte dans le choix des implantations ainsi que des caractéristiques spécifiques à chaque emplacement. Ainsi, dans le problème étudié dans ce papier, les commandes du client et les magasins correspondent respectivement aux installations et aux magasins. Cependant, ce problème diffère du problème de localisation d'entrepôts car les magasins ainsi que les produits disponibles dans chacun d'eux sont connus.

Le problème que nous traitons étant NP-difficile, nous nous intéressons dans ces travaux aux algorithmes d'approximations, i.e., des heuristiques avec une garantie sur la qualité de la solution et le temps d'exécution. Une attention particulière est portée au problème de localisation d'entrepôts dans la littérature sur les algorithmes d'approximation. Shmoys et al dans (Shmoys, Tardos & Aardal 1997) proposent le premier algorithme pour ce problème avec un facteur d'approximation constant. Ces travaux sont étendus au cas multiproduit par (Ravi & Sinha 2010). Ainsi, ces auteurs étudient le problème de localisation d'entrepôts multiproduits et où le coût d'ouverture d'un entrepôt comprend un coût fixe et un coût variable qui dépend des produits disponibles dans cet entrepôt. Nos travaux s'inspirent de ces deux papiers et nous utilisons l'approche d'arrondi en programmation linéaire (*LP-rounding*) introduite par (Lin &

Vitter 1992) afin d'approcher la solution. Néanmoins, nous invitons les lecteurs à se référer au livre de Williamson et Shmoys (Williamson & Shmoys 2011) ainsi qu'aux travaux suivants (Chudak 1998), (Guha & Khuller 1999) et (An, Singh & Svensson 2017) pour plus de détails sur les techniques d'approximation pour le problème de localisation d'entrepôts.

2 DESCRIPTION DU PROBLEME

Afin de poser la formulation du problème, nous introduisons les notations suivantes. Soit $D = \{1, \dots, n\}$ l'ensemble des produits demandés et pour chaque produit $j \in D$, on note par d_j la quantité demandée. L'ensemble de magasins est noté $M = \{1, \dots, t\}$ et pour chaque magasin $i \in M$, on note par $P_i \subseteq D$ l'ensemble des produits disponibles dans le magasin i . Les stocks des magasins étant connus, on note par u_{ij} la quantité de produits j disponible dans le magasin i . Nous supposons sans perte de généralité que la quantité demandée par le client est disponible dans les magasins (i.e., $\sum_{i \in M} u_{ij} \geq d_j, \forall j \in D$). En effet, si un produit $j \in D$ n'est pas disponible en quantité suffisante, on pose $d_j = \sum_{i \in M} u_{ij}$. L'objectif considéré consiste à satisfaire la demande du client à coût minimum. Les coûts pris en compte dans notre modèle représentent les coûts des produits ainsi que les coûts de transport. Le coût d'un produit j dans le magasin i est noté c_{ij} et le coût de transport à partir d'un magasin i est noté f_i . Notez que ce coût ne dépend pas de la quantité transportée.

3 FORMULATION DU PROBLÈME

Nous introduisons un modèle de programmation linéaire en nombres entiers pour le problème. On commence par présenter les variables utilisées.

- $g_{ij} \in \{0, \dots, d_j\}$: indique le nombre d'unités du produit $j \in D$ acheminés depuis le magasin i ,
- $z_i \in \{0, 1\}$: indique si le magasin i est sélectionné.

$$\min \sum_{i \in M} f_i z_i + \sum_{i \in M} \sum_{j \in P_i} c_{ij} g_{ij} \quad (1)$$

$$\sum_{i \in M: j \in P_i} g_{ij} = d_j, \quad \forall j \in D \quad (2)$$

$$g_{ij} \leq u_{ij} z_i, \quad \forall i \in M, \forall j \in P_i \quad (3)$$

$$g_{ij} \in \{0, \dots, d_j\}, \quad \forall i \in M, \forall j \in P_i \quad (4)$$

$$z_i \in \{0, 1\}, \quad \forall i \in M \quad (5)$$

Les contraintes (2) font en sorte que chaque commande soit satisfaite. Les contraintes (3) garantissent la faisabilité de la solution en tenant compte

des quantités disponibles dans chaque magasin. Cette contrainte nous assure également que lorsque un produit $j \in D$ est transféré du magasin $i \in M$ alors le magasin i doit être sélectionné. La fonction objectif (1) minimise le coût total composé de coûts fixes qui correspondent aux magasins sélectionnés et des coûts de transport.

4 CAS DE LA CAPACITÉ ILLIMITÉE (PROBLÈME DE LOCALISATION D'ENTREPÔTS)

Dans des applications réelles, un magasin peut être spécialisé dans une catégorie de produits et ainsi disposer d'une grande quantité de produits de cette catégorie. Nous présentons dans cette section une heuristique avec garantie de performance pour le problème dans lequel les produits sont supposés disponibles en quantité illimitée dans les magasins. Ainsi la quantité u_{ij} de produits $j \in D$ disponibles dans le magasin P_i vérifie $u_{ij} = +\infty$ si $j \in P_i$, 0 sinon.

Noter que le problème présenté dans cette section est similaire au problème de localisation d'entrepôts avec des distances non-métriques (voir (Hochbaum 1982)). Par conséquent, nous présentons cette méthode afin d'introduire l'algorithme dédié au cas où les capacités sont considérées limitées et qui constitue notre contribution.

Dans ce cas particulier et afin de formuler le problème, nous utilisons les variables suivantes:

- $x_{ij} \in \{0, 1\}$: indique si commande j est satisfaite par le magasin i ,
- $y_i \in \{0, 1\}$: indique si le magasin i est sélectionné.

Il faut noter que la quantité transportée pour chaque produit n'est pas prise en compte. En effet, lorsqu'un produit est disponible dans un magasin, celui-ci est disponible en quantité illimitée. Par conséquent, dans une solution optimale, une commande $j \in D$ est satisfaite par un seul magasin $i \in M$ tel que $j \in P_i$.

$$\min \sum_{i \in M} f_i y_i + \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x_{ij} \quad (6)$$

$$\sum_{i \in M: j \in P_i} x_{ij} = 1, \quad \forall j \in D \quad (7)$$

$$x_{ij} \leq y_i, \quad \forall i \in M, \forall j \in D \quad (8)$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in M, \forall j \in D \quad (9)$$

$$y_i \in \{0, 1\}, \quad \forall i \in M \quad (10)$$

4.1 Méthode de résolution

La résolution du problème étant NP-difficile, nous utilisons la programmation linéaire afin d'approcher la solution optimale du problème à variables entières présenté ci-dessus. L'algorithme suivant retourne une solution réalisable au problème et garantit un ratio de performance.

1. Résoudre la relaxation linéaire de la formulation présentée ci-dessus (contraintes (6) - (10)) . Soit (x', y') la solution (fractionnaire) obtenue.
2. En utilisant les techniques de filtrage ("filtering"), trouver une nouvelle solution fractionnaire. Soit (x'', y'') la solution obtenue (voir section 4.1.1).
3. Arrondir la solution fractionnaire (x'', y'') afin d'obtenir une solution entière notée x''', y''' . A cet effet, nous avons utilisé le problème de multicouverture par multiensembles (*Multiset Multi-cover*) (voir section 4.1.2).

A noter que dans (Ravi & Sinha 2010), les auteurs étudient une extension du problème de localisation d'entrepôts et utilisent le problème de couverture par ensembles (*Set Cover*) dans le même but. Enfin, nous obtenons un ratio d'approximation égal à $\log k + 2$. où k est le nombre maximum de produit dans tout les magasins.

4.1.1 Filtrage

Etant donné la solution optimale de la relaxation du problème noté (x', y') , cette étape consiste à maintenir un sous ensemble de variables non nulles x' dans une valeur positive et de mettre les autres variables à zero. Nous reprenons dans cette partie le raisonnement utilisé par (Ravi & Sinha 2010, Shmoys et al. 1997, Lin & Vitter 1992). En effet, le problème que nous traitons constitue un cas particulier du problème traité dans (Ravi & Sinha 2010). Ainsi, la solution (x'', y'') est atteinte en suivant les étapes suivantes.

- Soit X_j l'ensemble des magasins $i \in M$ tels que $x'_{ij} > 0$ soient triés par ordre croissant des coûts c_{ij} .
- Soit $0 < \alpha < 1$
- Etant donné α , trouver pour tout $j \in D$ le plus petit coût $c_j(\alpha)$ tel que $\sum_{i \in X_j: c_{ij} \leq c_j(\alpha)} x'_{ij} \geq \alpha$
- Soit pour tout $j \in D$, $\alpha_j = \sum_{i \in X_j: c_{ij} \leq c_j(\alpha)} x'_{ij}$
- Créer une nouvelle solution fractionnaire (x'', y'') où x'' et y'' prennent les valeurs suivantes :

- $x''_{ij} = x'_{ij}/\alpha_j$ si $c_{ij} \leq c_j(\alpha)$, 0 sinon.
- $y''_i = \min\{1, y'_i/\alpha\}$

Lemme 1. La solution (x'', y'') est réalisable pour la relaxation du problème défini par les contraintes (6 - 10) et pour tout $j \in D$, $c_j(\alpha) \leq \frac{1}{1-\alpha} \sum_{i \in M} c_{ij} x'_{ij}$.

Preuve. Pour tout $j \in D$, on a.

$$\begin{aligned} \sum_{i \in M} c_{ij} x'_{ij} &\geq \sum_{i \in M: c_{ij} \geq c_j(\alpha)} c_{ij} x'_{ij} \\ &\geq c_j(\alpha) \sum_{i \in M: c_{ij} \geq c_j(\alpha)} x'_{ij} \end{aligned} \quad (11)$$

D'après la définition de $c_j(\alpha)$,

$$\sum_{i \in M: c_{ij} < c_j(\alpha)} x'_{ij} \leq \alpha$$

Alors,

$$\sum_{i \in M} x'_{ij} = 1 \Rightarrow \sum_{i \in M: c_{ij} \geq c_j(\alpha)} x'_{ij} \geq 1 - \alpha$$

En remplaçant dans (11), nous obtenons $c_j(\alpha) \leq \frac{1}{1-\alpha} \sum_{i \in M} c_{ij} x'_{ij}$. \square

Noter qu'un résultat similaire au Lemme exposé est proposé dans (Shmoys et al. 1997).

4.1.2 Arrondi

Afin d'obtenir une solution entière réalisable au problème défini par les contraintes 6-10, Nous arrondissons dans cette section la solution (x'', y'') . De manière similaire à (Ravi & Sinha 2010), nous utilisons le problème de couverture par ensembles qui consiste à trouver le sous-ensemble de magasins permettant de couvrir la demande et dont le coût total d'ouverture est minimal. Ainsi, la solution entière obtenue notée (x''', y''') est obtenue en suivant les étapes suivantes.

Nous commençons par construire une instance du problème de couverture de l'ensemble des demandes $D = 1, \dots, n$ par les ensembles $\{s_i \mid 1 \leq i \leq t\}$ de produits que peuvent fournir chaque magasin i .

On utilise les notations suivantes :

- Un ensemble de magasins $S = \{i \in M : y''_i > 0\}$
- L'ensemble des produits disponibles dans chaque magasin : $s_i = \{j \in P_i : x''_{ij} > 0\}$
- Un coût f_i pour chaque magasin $i \in S$
- Soit k une borne supérieure du nombre de produits disponibles dans chaque magasin, i.e., $k = \max_{i \in S} |s_i|$. Clairement, k est borné par n .

L'instance étant construite, soit h_i une variable telle que $h_i = 1$ si le magasin $i \in S$ est sélectionné, égale

à 0 sinon. La formulation du problème de couverture par ensembles est ainsi définie comme suit.

$$\min \sum_{i \in S} f_i h_i \quad (12)$$

$$\sum_{i \in S: j \in s_i} h_i \geq 1, \quad \forall j \in D \quad (13)$$

$$h_i \in \{0, 1\}, \quad \forall i \in S, \forall j \in s_i \quad (14)$$

Les contraintes (13) garantissent que chaque commande $j \in D$ est affectée à au moins un magasin et l'objectif est de minimiser coût total des magasins sélectionnés. Ainsi ce problème de couverture par ensembles peut être résolu avec un ratio d'au plus $\log(k+1)$ (voir (Vazirani 2001)). A noter que cette valeur représente le ratio entre la solution optimale et la solution relaxée du problème de couverture par ensembles. Etant donné une couverture par ensembles (solution des contraintes (12–14)), nous construisons la solution (x''', y''') de la façon suivante :

- La commande $j \in D$ est affectée au magasin $i' \in S$ tel que $i' = \operatorname{argmin}_{i \in S} \{c_{ij} : z_i = 1, j \in s_i\}$.
- $x''_{ij} = 1$ si $i = i'$ et 0 sinon.
- $y''_i = 1$ si $z_i = 1$ et 0 sinon.

Lemme 2. La solution entière (x''', y''') est réalisable pour le problème défini par les contraintes (6 - 10).

Preuve. Nous pouvons voir que pour chaque produit $j \in D$, il existe un magasin $i \in M$ pour lequel la variable $x''_{ij} > 0$ car $\sum_{i \in M: j \in P_i} x''_{ij} = 1$. Ainsi, chaque produit est disponible dans au moins un magasin s_i . Par conséquent, la solution (x''', y''') est réalisable car chaque commande est affectée à exactement un magasin. \square

4.1.3 Facteur d'approximation

Dans cette section, un facteur d'approximation est calculé. Celui-ci représente le quotient entre les valeurs de l'objectif correspondant aux solutions (x''', y''') et (x', y') . Afin de calculer ce facteur d'approximation, nous calculons séparément le facteur d'approximation sur la partie de l'objectif qui correspond aux coûts de transport et celui correspondant ou coûts des produits. Ainsi nous obtenons.

- Facteur d'approximation sur la partie de l'objectif qui correspond aux coûts de transport dans la fonction objectif :

$$\sum_{i \in M} f_i y'_i \leq \frac{\log k + 1}{\alpha} \sum_{i \in M} f_i y''_i$$

Preuve. Etant donnée l'instance utilisée pour le problème de couverture par ensembles et le fait

que le facteur d'approximation pour le problème de couverture par ensembles est au plus égal à $\log(k+1)$, on a.

$$\sum_{i \in M} f_i y''_i \leq (\log k + 1) \sum_{i \in M} f_i y''_i$$

Du fait que $y''_i = \min\{1, y'_i/\alpha\}$, on a.

$$\sum_{i \in M} f_i y''_i \leq \frac{1}{\alpha} \sum_{i \in M} f_i y'_i$$

Alors.

$$\sum_{i \in M} f_i y''_i \leq \frac{\log k + 1}{\alpha} \sum_{i \in M} f_i y'_i$$

\square

- Facteur d'approximation sur la partie de l'objectif qui correspond aux coûts des produits dans la fonction objectif :

$$\sum_{i \in M} \sum_{j \in D} d_j c_{ij} x''_{ij} \leq \frac{1}{1-\alpha} \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x'_{ij}$$

Preuve. Etant donné que pour tout $j \in D$, $\sum_{i \in M} c_{ij} x''_{ij} \leq c_j(\alpha)$, on a.

$$\begin{aligned} \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x''_{ij} &\leq \sum_{j \in D} d_j c_j(\alpha) \\ &\leq \frac{1}{1-\alpha} \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x'_{ij} \end{aligned}$$

\square

- Alors pour la fonction objectif complète nous obtenons le facteur d'approximation suivant.

$$\begin{aligned} \sum_{i \in M} f_i y''_i + \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x''_{ij} &\leq \frac{\log k + 1}{\alpha} \sum_{i \in M} f_i y'_i + \frac{1}{1-\alpha} \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x'_{ij} \\ &\leq \max\left\{\frac{\log k + 1}{\alpha}, \frac{1}{1-\alpha}\right\} \\ &\quad \left(\sum_{i \in M} f_i y'_i + \sum_{i \in M} \sum_{j \in D} d_j c_{ij} x'_{ij}\right) \end{aligned}$$

Ainsi, en posant $\alpha = \frac{\log k + 1}{\log k + 2}$ nous obtenons un facteur d'approximation égal à $\log k + 2$.

5 CAS DE LA CAPACITÉ LIMITÉE

Nous proposons dans cette section une heuristique avec garantie de performance pour le cas où les produits sont disponibles dans les magasins en quantités limités. La programmation linéaire est également utilisée afin d'approcher la solution optimale du problème à variables entières. Notons que du fait des capacités sur les magasins, il est impossible de trouver un ratio de performance basé sur la relaxation

linéaire du problème (voir (Shmoys et al. 1997) pour plus de détails). Par conséquent, nous proposons une (ρ, ρ') -approximation qui trouve en temps polynomial une solution avec un facteur ρ du coût de l'optimum mais où les capacités sont étendues d'un facteur ρ' . Un aperçu de l'algorithme utilisé est présenté dans ce qui suit.

5.1 Méthode de résolution

L'algorithme suivant est utilisé afin de résoudre le problème pour le cas où les quantités disponibles dans chaque magasin sont limitées. Cet algorithme est similaire à celui utilisé pour le cas où les capacités sont considérées illimitées.

1. Résoudre la relaxation linéaire du problème défini par les contraintes (1) - (5)). Soit (g', z') la solution (fractionnaire) obtenue (voir section 5.1.1)..
2. Trouver une nouvelle solution fractionnaire. Soit (g'', z'') la solution obtenue (voir section 5.1.2).
3. Trouver une solution réalisable (entière). Soit (g''', z''') la solution obtenue (voir section 5.1.3).

Finalement, nous obtenons un facteur d'approximation égal à $(\log k' + 2, \frac{\log k' + 2}{\log k' + 1})$ où k' est le nombre maximum de produits dans tout les magasins.

5.1.1 Formulation du problème relaxé

Nous présentons un modèle de programmation linéaire pour le problème a variables relaxées. Les variables suivantes utilisées.

- $g'_{ij} \in [0, 1]$: indique la proportion de la demande d_j satisfaite par le magasin i ,
- $z'_i \in [0, 1]$: indique si le magasin i est sélectionné.

$$\min \sum_{i \in M} f_i z'_i + \sum_{i \in M} \sum_{j \in D} d_j c_{ij} g'_{ij} \quad (15)$$

$$\sum_{i \in M} g'_{ij} = 1, \quad \forall j \in D \quad (16)$$

$$d_j g'_{ij} \leq u_{ij} z'_i, \quad \forall i \in M, \forall j \in D \quad (17)$$

$$0 \leq g'_{ij} \leq 1, \quad \forall i \in M, \forall j \in D \quad (18)$$

$$0 \leq z'_i \leq 1, \quad \forall i \in M \quad (19)$$

Cette formulation représente une relaxation du la formulation définie par les contraintes (1)–(5)). En effet, un remplacement de g_{ij} par $d_j g'_{ij}, \forall i \in M, \forall j \in D$ rend les deux formulations identiques.

On peut remarquer qu'en se basant sur la programmation linéaire, il est impossible de trouver un algorithme d'approximation pour le problème considéré à capacité limitée. nous invitons les lecteurs à se référer au papier de (Shmoys et al. 1997) pour plus de détails.

5.1.2 Filtrage

Cette étape est semblable à celle présentée en section 4.1.1 pour le cas où les quantités dans les magasins sont considérés limitées. Ainsi, les différentes étapes utilisées afin d'obtenir la nouvelle solution fractionnaire (g'', z'') ne sont pas exposées. Nous rappelons cependant certaines valeurs.

- $g''_{ij} = g'_{ij}/\alpha_j$ if $c_{ij} \leq c_j(\alpha)$ et 0 sinon.
- $z''_i = \min\{1, y'_i/\alpha\}$
- $\forall j \in D, c_j(\alpha) \leq \frac{1}{1-\alpha} \sum_{i \in M} c_{ij} x'_{ij}$.

La solution ainsi obtenue notée (g'', z'') n'est pas forcément réalisable pour les contraintes (1–4). En effet, soit une solution dans laquelle il existe $i \in M$ et $j \in D$ tel que $d_j g'_{ij} = u_{ij}$. Dans ce cas, $z'_i = 1$. En remplaçant (g', z') par (g'', z'') , cette contrainte n'est pas respectée (cette contrainte correspond à la contrainte (17) dans la formulation relaxée). Nous étendons par conséquent la quantité de produits disponibles dans chaque magasin afin d'avoir une solution (g'', z'') réalisable. Soit u'_{ij} la nouvelle quantité de produit j disponible dans le magasin i telle que $u'_{ij} = \frac{1}{\alpha} u_{ij}$.

5.1.3 Arrondi

Dans cette section, nous arrondissons la solution (g'', z'') afin d'obtenir une solution entière notée (g''', z''') qui soit réalisable sur les contraintes (1) - (5)). A cet effet, nous utilisons le problème de multicouverture par multiensembles (*Multiset Multicover*) comme sous problème du cas que nous traitons.

Problème de multicouverture par multiensembles : Ce problème est une généralisation du problème de couverture par ensembles dans lequel chaque élément e est associé à un nombre de couvertures r_e indiquant le nombre de fois ou celui-ci doit être couvert par toute couverture réalisable. En plus du nombre de couvertures, chaque élément apparaît dans un ensemble S avec une multiplicité de $m(S, e)$.

La solution entière (g''', z''') est alors obtenue par les étapes suivantes.

Nous commençons par construire une instance du problème de multicouverture par multiensembles de la manière suivante :

- Un ensemble de n membres $D = 1, \dots, n$ représentant les demandes
- Une demande de couverture $r_j = d_j$ pour chaque produit $j \in D$
- Un ensemble de magasins $S = \{i : i \in M, z_i'' > 0\}$
- L'ensemble des produits disponibles dans chaque magasin : $s_i = \{j \in P_i : x_{ij}'' > 0\}$
- La multiplicité $m(s_i, j) = u_{ij}'$ pour chaque produit $j \in s_i$ qui correspond au quantités disponibles
- Un coût f_i pour chaque magasin $i \in S$
- Soit k' une borne supérieure du nombre de produits disponibles dans chaque magasin, i.e., $k' = \max_{i \in S} \{\sum_{j \in s_i} u_{ij}'\}$.

Etant donné l'instance, on note par h_i' une variable qui vaut 1 si le magasin $i \in S$ est sélectionné et 0 sinon. Le problème de multicouverture par multi-ensembles peut être formulé comme suit.

$$\begin{aligned} \min \sum_{i \in S} f_i h_i' \\ \sum_{i \in S: j \in s_i} m(s_i, j) h_i' \geq r_j, \quad \forall j \in D \quad (20) \\ h_i' \in \{0, 1\}, \quad \forall i \in S, \forall j \in s_i \end{aligned}$$

Les contraintes (20) garantissent la satisfaction des demandes et l'objectif est de minimiser le coût total induit par les magasins sélectionnés. Ce problème peut être résolu en temps polynomial avec un ratio d'au plus $\log(k+1)$ par rapport à la solution relaxée (voir (Rajagopalan & Vazirani 1998) pour plus de détails). A noter que ce ratio d'approximation est logarithmique sur la taille du plus grand ensemble et non sur n . Dans notre cas, la taille du plus grand ensemble correspond au nombre maximum de produits disponibles dans un magasin.

Etant donné une solution au problème de multicouverture par multiensembles présenté ci-dessus, la solution entière (g''', z''') est construite à l'aide de l'algorithme suivant.

- $z_i''' = 1$ si $h_i' = 1$ et 0 sinon.
- Pour chaque commande $j \in D$,
 - Soit $R_j = \{i : h_i' = 1, j \in s_i\}$.
 - Soit $d_j' = d_j$,
 - Pour $l = 1$ à $|R_j|$, faire
 - * Si $d_j' \leq 0$,
 - $x_{ij}''' = \min\{d_j', u_{ij}'\}/d_j$,

$$\begin{aligned} \cdot d_j' = d_j - u_{lj}' \\ * \text{ Sinon } x_{lj}''' = 0 \end{aligned}$$

Il faut noter qu'un autre algorithme que celui présenté ci-dessus ne peut améliorer le facteur d'approximation théorique. Enfin, la solution (g''', z''') est réalisable pour le problème défini par les contraintes (15-19). Soit $\forall i \in M, \forall j \in D, g_{ij} = djg_{ij}'''$ et $\forall i \in M, z_i = z_i'''$. La solution (g, z) est réalisable pour le problème défini par les contraintes (1-5).

5.1.4 Facteur d'approximation

Le facteur d'approximation est obtenu de manière similaire à celle utilisée pour le cas où les quantités de produits disponibles dans les magasins sont illimitées. Ainsi, pour $\alpha = \frac{\log k+1}{\log k+2}$, nous obtenons un facteur d'approximation égal à $(\log k + 2, \frac{\log k+2}{\log k+1})$.

6 Conclusion

Nous avons proposé des algorithmes d'approximation avec garantie de performance pour un problème d'affectation d'une commande multiproduit multi-fournisseur dans le cadre de la vente en ligne. Par la suite nous proposerons une méthode exacte et des heuristiques basées sur la programmation linéaire en nombres entiers pour la résolution pratique du cas général. Ces travaux sont réalisés dans le cadre du projet "One Stock Performance", financé par le programme "Easynov" (FEDER / REGION Midi-Pyrénées Languedoc Roussillon) en partenariat avec la société DEVATICS.

References

- An, H.-C., Singh, M. & Svensson, O. (2017). Lp-based algorithms for capacitated facility location, *SIAM Journal on Computing* **46**(1): 272–306.
- Chudak, F. A. (1998). *Improved Approximation Algorithms for Uncapacitated Facility Location*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 180–194.
- Daskin, M. S., Snyder, L. V. & Berger, R. T. (2005). *Facility Location in Supply Chain Design*, Springer US, Boston, MA, pp. 39–65.
- Guha, S. & Khuller, S. (1999). Greedy strikes back: Improved facility location algorithms, *Journal of Algorithms* **31**(1): 228 – 248.
- Hochbaum, D. S. (1982). Heuristics for the fixed cost median problem, *Mathematical Programming* **22**(1): 148–162.
URL: <https://doi.org/10.1007/BF01581035>

- Lin, J.-H. & Vitter, J. S. (1992). ϵ -approximations with minimum packing constraint violation (extended abstract), *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, STOC '92, ACM, New York, NY, USA, pp. 771–782.
- Melo, M., Nickel, S. & da Gama, F. S. (2009). Facility location and supply chain management a review, *European Journal of Operational Research* **196**(2): 401 – 412.
- Rajagopalan, S. & Vazirani, V. V. (1998). Primal-dual rnc approximation algorithms for set cover and covering integer programs, *SIAM Journal on Computing* **28**(2): 525–540.
- Ravi, R. & Sinha, A. (2010). Approximation algorithms for multicommodity facility location problems, *SIAM Journal on Discrete Mathematics* **24**(2): 538–551.
- Shmoys, D. B., Tardos, E. & Aardal, K. (1997). Approximation algorithms for facility location problems (extended abstract), *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, ACM, New York, NY, USA, pp. 265–274.
- Vazirani, V. V. (2001). *Approximation Algorithms*, Springer-Verlag New York, Inc., New York, NY, USA.
- Williamson, D. P. & Shmoys, D. B. (2011). *The design of approximation algorithms*, Cambridge university press.

AMELIORATION DE L'EFFICACITE ENERGETIQUE DANS UN RESEAU ELECTRIQUE EN APPLIQUANT LES METHODES D'APPRENTISSAGE ET DE CLUSTERING

M. HASCURI, M. DERRHI

Département Génie Electrique et Industriel
Ecole Nationale des Sciences Appliquées de Tanger
ENSAT
Ancienne Route de l'Aéroport, Km 10, Ziaten. BP :
416. Tanger - Maroc
mohamedhascuri@gmail.com, mderrhi@yahoo.fr

M. AIT RAMI

Ecole des Nouvelles Sciences et Ingénierie de Tanger
ENSIT,
Bd Taki Eddine Al Hilali, villa n°131, Q. Souriyennes -
TANGER
aitrami@autom.uva.es

RESUME :

La réduction de la facture énergétique est aujourd'hui l'une des principales priorités des professionnels et des clients. Cependant, l'instabilité des prix de l'énergie nous invite à réfléchir à de nouvelles solutions autres que l'amélioration des énergies renouvelables, afin de faire face aux pertes d'énergie élevées dans un réseau électrique. Récemment, de nombreux travaux ont porté sur des réseaux électriques intelligents appelés «Smart Grids». Ces réseaux particuliers sont basés sur la mesure en temps réel des paramètres électriques, permettant une gestion efficace de l'énergie et réduisant les pertes. D'une part, ce travail vise à découvrir les différentes techniques utilisées pour établir les mesures et les profils de charge, et à trier les consommateurs qui ont des consommations élevées. D'un autre côté, il présente de possibles améliorations futures permettant une gestion avancée de l'énergie.

MOTS-CLES : Technologies de la communication ; Réseau électrique ; Réseau intelligent ; Smart Metering ; Surveillance de charge non intrusive; Signatures.

1 INTRODUCTION

À mesure que la sensibilisation aux économies d'énergie et à l'énergie verte augmente, des changements importants ont été introduits dans nos réseaux. En conséquence, nous disposons aujourd'hui d'un large éventail d'options pour produire de l'énergie, mais comme la consommation n'est toujours pas gérée, ces options ne seront pas aussi efficaces qu'elles le devraient.

Afin d'optimiser la consommation d'énergie, et d'éviter les surcharges importantes pendant les heures de pointe, un consommateur peut répartir sa consommation sur la journée pour ne pas utiliser d'appareils gourmands pendant les heures de pointe. Cependant, des informations en temps réel sur l'énergie consommée sont nécessaires pour réaliser un planning de consommation permettant d'effectuer cette amélioration.

Pour obtenir ces informations, des compteurs intelligents sont utilisés pour fournir des données sur les différents paramètres électriques (puissance active et réactive, courant, tension, déphasage, taux d'harmoniques,...). Malheureusement, ces appareils sont coûteux et l'installation d'un compteur intelligent à chaque point du circuit

électrique ne serait pas abordable, surtout dans les situations où de nombreux appareils sont utilisés. Cela nous conduit à choisir une stratégie de surveillance de charge non intrusive (Non Intrusive Load Monitoring)

Dans une installation domestique, la démarche serait de construire une base de données contenant les signatures des différents appareils électriques, et, tout en utilisant des algorithmes de traitement de signal et d'estimation, pouvoir identifier les appareils en marche à partir du profile de charge global.

Nous pourrions par la suite généraliser cela sur une installation urbaine alimentée à partir d'un poste électrique HTA/BT. Il conviendra dans un premier lieu, de détecter les différents groupes de clients connectés à ce poste, afin de mieux comprendre l'évolution de la consommation de chaque groupe. Ensuite, En partant des catalogues des signatures des appareils, examiner pour chaque groupe de clients, les types de dispositifs électriques susceptibles de faire partie des appareils présents chez ce groupe. Enfin, proposer à chaque groupe de clients, les économies à faire ou la planification optimale de sa consommation.

2 IDENTIFICATION DES CHARGES À L'AIDE DES SIGNATURES ET DE LA SURVEILLANCE NON INTRUSIVE DE LA CHARGE.

2.1 Etat d'art

2.1.1 Identification des appareils électriques

L'identification des charges est un processus basé sur la collecte de données sur les consommateurs, ensuite l'usage des algorithmes de clustering pour identifier le dispositif.

Dans la plupart des travaux de recherche, les consommateurs à identifier représentent des appareils individuels dans le cadre d'une installation domestique, le concept de «signatures» correspond alors au profil de consommation en énergie active et réactive, à la forme des harmoniques générés par cet appareil, à l'évolution du courant traversant l'appareil ou encore à la tension entre ses bornes, ainsi, on établit pour chaque appareil une signatures formée des éléments déjà mentionnés.

En suivant cette démarche dans l'article présenté par (W K Lee, G S K Fung, H Y Lam, F H Y Chan, 2004), les auteurs proposent 2 approches pour identifier un dispositif:

1. Mesurer et représenter les caractéristiques de charge.
2. Utiliser des algorithmes de traitement du signal et d'estimation pour désagréger le signal et identifier la charge.

Un exemple de signature obtenue pour des appareils domestiques est présenté ci-après :

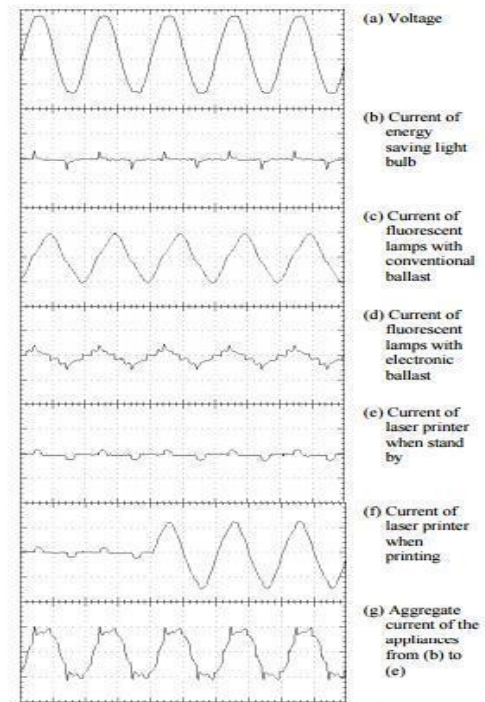


Figure 2 : Signatures de quelques appareils domestiques

Dans le but d'identifier un consommateur électrique, plusieurs algorithmes ont été testés. On distingue 2 familles d'algorithmes dits d'apprentissage :

Apprentissage supervisé :

Dans ce cas, tous les paramètres électriques qui peuvent être utilisés dans le processus d'identification (puissance active, puissance réactive, courant ...) sont définis au début de l'algorithme, pour éviter toute difficulté au cours du processus.

Apprentissage non supervisé

La seule entrée de l'algorithme est un ensemble de données d'apprentissage et le système définit les données qui seront utilisées par lui-même.

Le choix entre ces 2 familles d'algorithmes tient en compte 2 paramètres principaux : la précision et le temps d'exécution. Ces calculs ont été réalisés par (V. Abeykoon, N. Kankanamdu, A. Senevirathna, P. Ranaweera, R. Udawalpola, 2016).

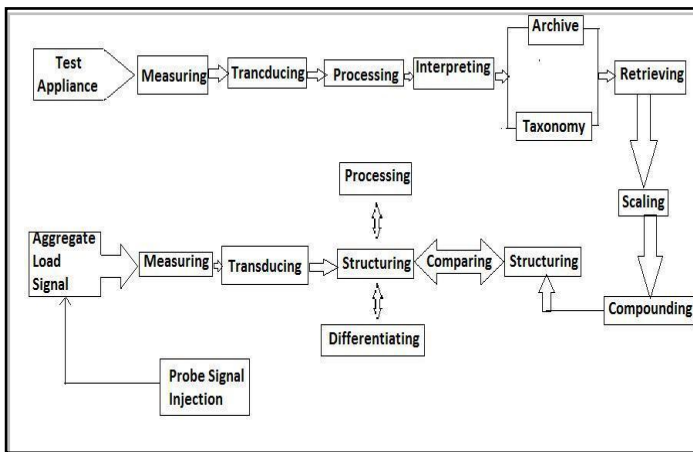


Figure 1 : Principe de l'identification des charges

Type	Classifiant	Précision (%)	Exécution (s)
Supervisé	SVM	97	0.010
Supervisé	ANN	96	13.000
Non supervisé	Mean Shift	94	0.013
Non supervisé	Silhouette	98	0.012
Non supervisé	K-Means	98	0.15

Tableau 1: Principe de l'identification des charges

Dans cet article, on note que les auteurs ont essayé d'obtenir une compréhension complète des comportements des appareils étudiés pendant chaque mode de fonctionnement. Par conséquent, on peut conclure qu'un appareil n'a pas forcément 2 modes de fonctionnement (ON/OFF) comme dans le cas d'une charge purement résistive, ou purement inductive, mais qu'il dispose parfois de plusieurs mode de fonctionnement, i.e. plusieurs valeurs de consommation. Cette donnée pourrait nous être utile dans la suite de notre démarche en vue de réduire la consommation d'un appareil au lieu de l'arrêter carrément.

La démarche suivie pour analyser le comportement de ce genre d'appareils est une démarche probabiliste faisant appel à des méthodes stochastiques.

Ainsi, (Jacob A. Mueller, Anusha Sankara, Jonathan W. Kimball, Bruce McMillin, 2014), ont recours aux modèles cachés de Markov afin de renseigner l'état d'un appareil à un moment donné en utilisant l'apprentissage supervisé reposant sur un historique de données et l'archive des usages antérieurs de l'appareil. Ils présentent également les différents travaux de recherche qui s'intéresse à l'identification des charge dans le cadre d'une installation domestique et les différents défis que présente cette tâche : Activation simultanée des appareils, présence d'appareils de même caractéristiques électriques,...

2.1.2 Identification des profils de charges

Si au niveau d'un appareil, le passage d'un niveau de consommation à un autre est expliqué par l'activation ou la désactivation d'un élément interne de l'appareil, on peut généraliser ce concept à un client ayant désactivé ou activé un appareil au cours de la journée.

Dans ce même cadre, des travaux de recherche se sont intéressés à la détermination des profils de charge typiques pour des clients appartenant à des secteurs d'activité bien déterminés. (Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi, 2017) présentent une démarche pour relever des profils de charge typique correspondant à des classes de clients différentes. Ensuite, les profils de charges journaliers sont classés en

utilisant des algorithmes de classification (ANN, SVM, les arbres de décisions,...) suivant ces classes.

(Dongsik Jang, Jiyong Eom, Min Jae Park, Jae Jeung Rho, 2016) ont eu recours à un processus dit « Dirichlet Process Mixture Model » (Le modèle de mélange de processus de Dirichlet) utilisé pour créer des classes de clients suivant la similarité entre leurs profils de consommation respectifs. Ainsi, en réalisant des mesures sur 2 classes de consommateurs : Commercial et résidentiel, on a pu ressortir de chaque classe des sous-classes gouvernées par des paramètres tels que les nombres des occupants, le pays, le type du logement,...

Or, la démarche NILM (Non Intrusive Load Monitoring) repose sur le principe de l'utilisation d'un seul compteur intelligent placé en amont de l'installation, de ce fait, on n'aura pas un accès direct à ces profils pour pouvoir les classer, d'où la nécessité de l'usage des algorithmes de désagrégation.

(Yongli Zhu, Songtao Lu, 2014) présentent une démarche d'analyse en composantes indépendantes assistée par l'analyse en ondelette. Cette démarche, par le biais d'un traitement préliminaire visant à réduire la complexité du problème, et par une analyse en ondelette menant à la mise en valeur des traits correspondant à chaque type de consommateurs (Résidentiel ou commercial), a permis la désagrégation d'un profil de charge enregistré à partir d'une sous-station, en 2 profils de charge : résidentiel et commercial

Dans le même cadre, (Young-II Kim, Jong-Min Ko, Seung-Hwan Choi, 2011) proposent de générer des profils de charge typiques en utilisant les algorithmes de classification (classification hiérarchique, K-means, Fuzzy C-means). Ce travail vise essentiellement l'élaboration du profil de charge pour des clients qui ne dispose pas de compteurs à lecture automatique « Automatic Meter Reading ».

2.2 Application des méthodes de clustering et d'apprentissage pour l'estimation des consommations

Notre point de départ serait un poste de transformation alimentant n groupes de clients (G_1, G_2, \dots, G_n). Ces groupes ont été choisis en fonction de leurs types de charges (leurs secteurs d'activité), le volume de leurs consommations, leurs heures de pointes, et l'ancienneté de leurs installations... La figure 3 montre le modèle étudié :

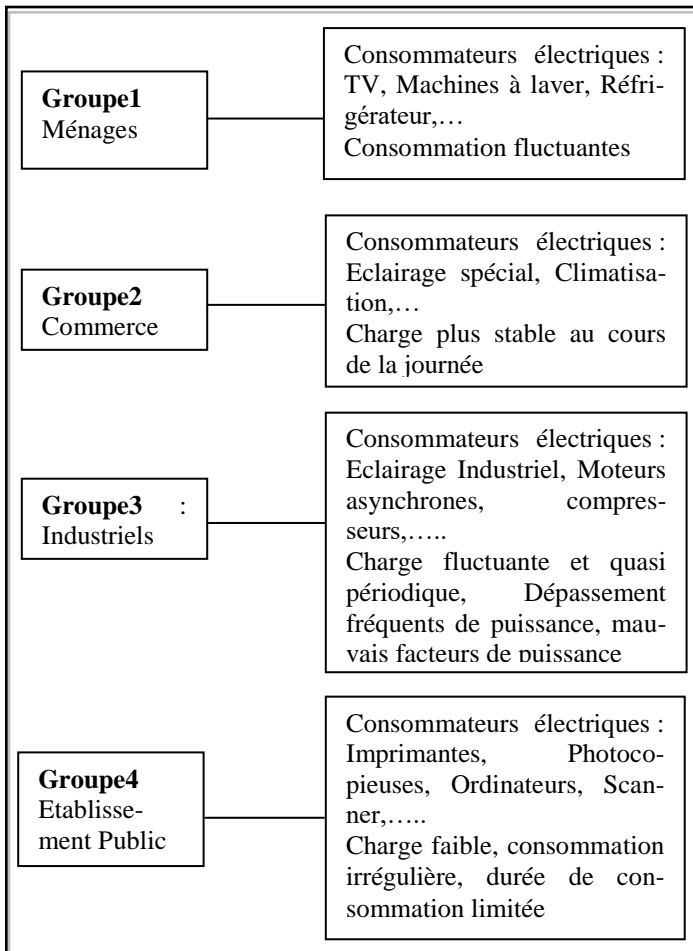


Figure 3: Exemple de détermination de clusters pour n=4

Notons M_i le vecteur des données électriques correspondant au groupe i tel que :

$$M_i(t) = \begin{bmatrix} P_i(t) \\ Q_i(t) \\ \cos\phi(t) \end{bmatrix} \tag{1}$$

Avec :

$P_i(t)$: Puissance active appelée par le groupe i à l'instant t

Q_i : Puissance réactive appelée par le groupe i à l'instant t

$\cos\phi_i$: Facteur de puissance correspondant au groupe i à l'instant t

Or, au moment de l'établissement du contrat d'abonnement, une puissance dite souscrite est déterminée sur la base d'un bilan des puissances des dispositifs électriques présents chez le client. On pourrait donc écrire la vecteur précédent sous la forme :

$$M_i(t) = \begin{bmatrix} \mu_i(t)P_{si} \\ \eta_i(t)Q_{si} \\ \cos\phi(t) \end{bmatrix} \tag{2}$$

Avec :

$\mu_i(t)$: le rapport de la puissance active instantanée sur la puissance souscrite du groupe i

$\eta_i(t)$: le rapport de la puissance active instantanée sur la puissance souscrite du groupe i

P_{si} : La puissance active souscrite du groupe i

Q_{si} : La valeur de la puissance réactive appelée par le groupe i permettant de maintenir le facteur de puissance à une valeur supérieure à 0,8 pour une valeur de puissance active égale à P_{si} .

Déterminer la puissance appelée par le groupe i à l'instant t revient à trouver les valeurs de $\mu_i(t)$ et $\eta_i(t)$.

Ces 2 paramètres peuvent être considérés comme des variables aléatoires, prenant des valeurs discrètes (correspondant au nombre des appareils fonctionnant à l'instant t chez les consommateurs du groupe i , ainsi qu'aux valeurs de la puissance consommée par chacun de ces appareils), avec des probabilités calculées sur la base des historiques de consommation, des profils de charge typiques, des conditions économiques (surtout dans le cadre d'un groupe de clients industriels), des conditions climatiques,...

On se propose donc d'adopter dans un premier lieu, une analyse en composantes indépendantes afin de dégager les différents profils de consommation (groupes de consommateurs). En fait, cette analyse repose sur le principe suivant :

Supposons que n sources S envoient une information en même temps et qu'on reçoit le mélange de cette information. En principe, on devra être capable de séparer ce mélange de n informations en réalisant au moins n observations. Ceci se présente sous la forme suivante :

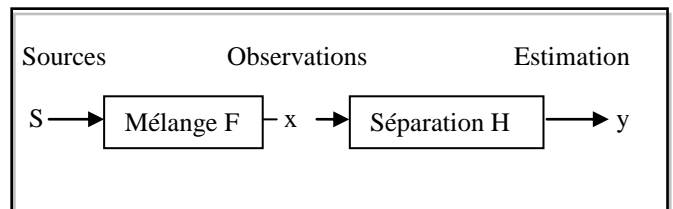


Figure 4: Modèle de séparation des sources

Ce modèle nous servira à trier au début les différents groupes alimentés par le poste électrique. Une période d'entraînement est nécessaire pour que le système devienne capable de distinguer les différents groupes de consommation présents.

Ensuite, en se basant sur les données collectées sur chaque groupe, jugées ayant une influence sur la consommation, prédire l'état de la consommation à un instant donné à travers un modèle de Markov caché, comme dans la figure 5 :

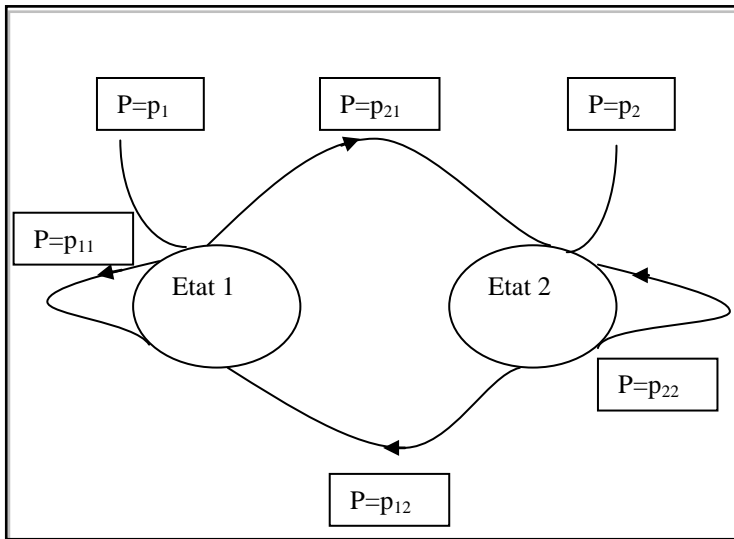


Figure 5 : Modèle caché de Markov

Avec :

- p_i : La probabilité pour que la consommation soit à l'état i
- p_{ii} : La probabilité pour que la consommation reste à l'état i
- p_{ij} : La probabilité pour que la consommation passe de l'état i à l'état j .

3 AMELIORATION DES RAPPORTS DE CONSOMMATION EN UTILISANT LA PROGRAMMATION LINEAIRE

Soit le bilan de puissance instantané suivant, correspondant à un poste de transformation HTA/BT :

$$P_i(t) = \sum_{i=1}^n \mu_i(t) P_{si} \tag{3}$$

$$Q_i(t) = \sum_{i=1}^n \eta_i(t) Q_{si} \tag{4}$$

Il s'agit de minimiser P_i et Q_i à chaque instant ou plus précisément à chaque changement d'état de la consommation. Ceci doit tenir en compte que certains appareils fonctionnant chez le groupe de clients i ne peuvent être arrêtés à un instant donné, d'autres ne peuvent marcher en mode économie d'énergie, d'où la présence des contraintes donnant à ce problème la forme d'un problème de programmation linéaire.

Les données correspondant aux puissances souscrites sont déjà connues, il reste à déterminer les combinaisons des facteurs μ_i et η_i permettant d'avoir un minimum de consommation.

4 ETUDE DE CAS :

Considérons l'exemple des 3 groupes suivants :

- Cinéma
- Restaurants
- Administrations publics

Ci-après l'évolution de la charge au cours de la journée. On remarque que la charge totale prend l'allure de la charge correspondant au groupe le plus gourmand au niveau de la consommation (les restaurants dans notre cas). Les données correspondent à la synthèse des mesures effectuées au cours de la même période de l'année. Ce modèle reste valable pour d'autres groupes ayant les mêmes habitudes et installés dans des endroits ayant les mêmes conditions climatiques et économiques.

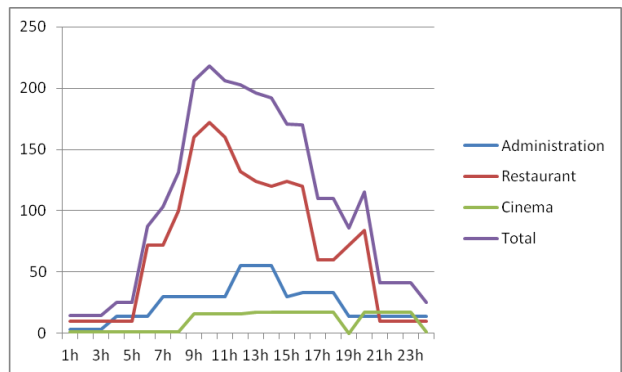


Figure 6 : Profils de charge journalier 3 groupes de consommateurs différents (Puissance active)

Supposons qu'on effectue une étude sur des administrations dans une autre zone de la ville B où le nombre d'administrations est le double. En tenant compte des conditions économiques différentes de la région, ainsi que la différence de l'attitude des employés des 2 groupes d'administrations, on peut estimer la probabilité p_1 pour que la consommation des administrations dans la zone B atteigne la valeur 110 kW approximativement à 13h et qu'elle passe à la valeur 60kW à 15h avec une probabilité p_2 . La consommation des administrations a 5 états discrets.

Vu que les profils des administrations correspondent à la période d'été, il est très probable que l'augmentation de la consommation pendant la période de midi est justifiée par l'activation des climatisations.

Ainsi, après avoir déterminé les conditions de température optimales pour les fonctionnaires, on peut définir les contraintes du problème de minimisation de la consommation comme étant les températures maximales à ne pas dépasser afin de réduire la consommation des climatiseurs.

5 CONCLUSION :

Après avoir classé les clients suivant des groupes de consommateurs homogènes, fait l'analogie avec des cas de groupes de consommateurs déjà étudiés tout en prenant en considération les différences aux niveaux économiques, climatiques ou culturels, et en détectant les moments de la journée où la consommation risque de s'élever, on procède à l'amélioration de l'efficacité énergétique et de la gestion d'énergie en calculant les taux optimaux d'utilisation de la puissance. On communique par la suite à chaque groupe les modes d'utilisation de la puissance installée à chaque moment de la journée en vue de baisser la consommation et adapter la consommation à la production. Dans la suite de notre étude, on procédera au déploiement de cette stratégie sur une zone de la ville de Tanger contenant des groupes de consommateurs de natures très différentes et on essaiera d'appliquer la démarche de 3 étapes pour réduire la consommation : Création des groupes de consommateurs, Prévision des états de consommation de chaque groupe, et amélioration des taux d'utilisation de la puissance installée.

REMERCIEMENTS

Nous remercions la direction du CED des Sciences et Techniques de l'ingénieur de Tanger ainsi que celle du LTI (Laboratoire des Technologies Innovantes) et

l'ENSIT (Ecole des Nouvelles Sciences et Ingénierie de Tanger) pour le soutien et l'accompagnement en vue de fournir tous les éléments nécessaires pour la réalisation de nos travaux de recherche

REFERENCES

- Alfonso Capozzoli, Marco Savino Piscitelli, Silvio Brandi, 2017. Mining Typical Load Profiles in buildings to support energy management in the smart city context. *9th International Conference on Sustainability I Energy and Buildings (SEB-17)* Crete, Greece 2017, pp 865-874
- Dongsik Jang, Jiyong Eom, Min Jae Park, Jae Jeung Rho, 2016. Clustering disaggregated load profiles using a Dirichlet process mixture model. *Energy* 92, pp 507-516
- Jacob A. Mueler, Anusha Sankara, Jonathan W. Kimball, Bruce McMillin, 2014. Hidden Markov Models for Nonintrusive Appliance Load Monitoring. *North American Power Symposium (NAPS)*, Pullman, USA 2014
- V. Abeykoon, N. Kankanamdurage, A. Senevirathna, P. Ranaweera, R. Udawalpola, 2016. Real Time Identification of Electrical Devices through Power Consumption Pattern Detection. *International Conference on Micro and Nano Technologies, Modelling and Simulation*, Malaysia 2016, pp. 11-16
- W K Lee, G S K Fung, H Y Lam, F H Y Chan, 2004. Exploration on Load Signatures. *International Conference on Electrical Engineering (ICEE) Korea*, Vol. 53, Issue. 2
- Yongli Zhu, Songtao Lu, 2014. Load Profile disaggregation by Blind Source Separation : a Wavelet-assisted Independent Component Analysis Approach. *2014 IEEE PES General Meeting*, pp 1-5
- Young-II Kim, Jong-Min Ko, Seung-Hwan Choi, 2011. Methods for Generating TLPs (Typical Load Profiles) for Smart Grid-Based Energy Programs. *Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium*

AN APPROACH TO SOLVE INVENTIVE PROBLEMS BASED ON THE SUBSTANCE-FIELD ANALYSIS AND SYSTEMS DYNAMICS

Delgado-Maciel J. AUTHOR, Cortés-Robles G. CO-AUTHOR

Tecnológico Nacional de México - Instituto
Tecnológico de Orizaba Avenida Oriente 9 No. 852 Col.
Emiliano Zapata C.P. 94320 Orizaba, Veracruz
jdelgadom@ito-depi.edu.mx, gcortes@ito-depi.edu.mx

Negny S. CO-AUTHOR

Institut National Polytechnique de Toulouse
CNRS UMR 5503, PSI/Génie Industriel – INPT –
ENSIACET, 118 Route de Narbonne, 31077 Toulouse
Cedex 04 – France
Stephane.Negny@ensiacet.fr

Maldonado-Macias A. CO-AUTHOR

Department of Industrial Engineering and
Manufacturing – Institute of Engineering and
Technology – Autonomous University of Ciudad Juarez,
Avenida del Charro 450 Norte, Col. Partido Romero,
Ciudad Juárez, Chihuahua, Mexico
amaldona@uacj.mx

ABSTRACT: *The Theory of Inventive Problem Solving (TRIZ) has their foundations on the knowledge extracted from the evolution of technical systems, scientific knowledge, and an effort to capitalize knowledge from several domains. Despite the TRIZ capacity to solve inventive problems, it lacks from three competencies: (1) TRIZ does not have a tool to observe the causality among the most important variables within an inventive problem; (2) It cannot follow the transformation of a system through time; (3) TRIZ cannot unveil the hidden relationships among different conflicts in a system to determine the right problem to solve and the most promising solving path. Nevertheless, there is a technical approach to explore the behavior of a system that can deal with the TRIZ drawbacks: The System Dynamics (SD) Modeling and Simulation Framework. This article demonstrates the feasibility to combine both approaches to produce a different problem-solving tool that can deal with inventive problems, particularly with problems modeled through a set of functions. The TRIZ tool that best deals with such kind of problems is the Substance-Field Modeling (SFM) tool. Once the user proposes a model, the next tool to use is a set of solving strategies (known as 76 Standard Solutions). These strategies suffer an adaptation process to propose at least one solving path. The combination of SFM and the SD approach is illustrated through a case study to discuss the advantage and limitation of a different solving framework.*

KEYWORDS: *Inventive Problems Modeling, System Dynamics Simulation, TRIZ, Substance-Field Analysis*

1 INTRODUCTION

The Theory of Inventive Problem Solving (TRIZ) is a useful tool for solving problems in three main classes: (1) situations model as a physical or technical contradiction; (2) when it is necessary to define the trends of evolution (ToE) of a system; and (3) when a problem involves the model of at least one function through the Substance-Field Analysis (SFA). The SFA is a TRIZ analytical tool for modeling problems related to existing technological systems. According to TRIZ, every system is created to perform at least one useful function (Altshuller 1999). The term useful does not refer to the concept of something beneficial or harmful, just to the fact that the system accomplishes their design purpose. The term 'substance' is used to describe an object of any level of complexity. A substance is used indistinctly to represent physical objects, systems or materials (Altshuller 1984).

The action or means of accomplishing it is called 'field.' According (Bulley et al. 2015) the term 'field' is not defined by its intention (set of formal attributes) but by its extension (set of formal objects: mechanical, chemical, thermal, electrical, gravitational and magnetic field). Further, the 'field' concept change from source to source (Altshuller 1984), (Salamatov 1999), and (Savransky 1998). A Substance-Field Model (SFM) is a graph used to describe a system or to formulate a problem as a function or as a set of functions. If the function depicts a problem where available knowledge cannot lead to a satisfactory solution, then the graph is linked to a set of generic rules called 76 standard solutions. A standard solution describes a solving path that can deal efficiently with the problem. To make a complete SFM are necessary two substances and a field or two fields and one substance. The construction of a SFM encompasses four steps: (1) Identify the components, (2) Construct the model, (3) Correlate the model with one of the 76 Standard Solutions, and (4) To develop at least one solution

concept and validate their feasibility (Terninko 1998) and (Helfman et al. 2015). The SFA has some advantages over other inventive tools: (1) It creates simple models through ideas extracted from a knowledge base; (2) It represents a complex system through a synergy of small and simple models; (3) It allows the construction of modeling functions, and (4) It shows the relationship between the field and the substances.

Despite the usefulness of the SFA modeling, it also has some important limitations: (1) The TRIZ Toolbox is not able to model a conflict and the relation of their variables taking into account the temporal dimension (Delgado-Maciel et al. 2017). Neither can solve problems simultaneously, which significantly limits their ability to deal with complex problems (Terninko 1998). (2) TRIZ solves inventive problems sequentially. Available methodologies do not propose any guide to assist the solver in the decision making to discern what problem is the most significant. (3) The TRIZ solving process does not propose any approach or technique to evaluate what problem has the greater impact (positive or negative) on the system. Thus, it is not possible to analyze how the system containing the conflict will change or what will be the effect of removing one conflict in the system, neither to distinguish between two or more potential solutions (Delgado-Maciel et al. 2017). Nevertheless, there is an approach that can deal with the lack of dynamism of TRIZ: The System Dynamics Modeling and Simulation. The System Dynamics (SD) is used to analyze and model temporal behavior in complex environments. SD identifies feedback loops between the system components, as well as its behavior through time allowing the construction of simulation models (Forrester 1961), (Sterman 2000). To evaluate the feasibility to combine both solving approaches, this article has five sections. Section two briefly describes the theory of SFA and SD that generates a research opportunity. This section explores the potential for combining the SD approach with the SFA through a proposed methodology. Section three presents the combined approach through a case study. The fourth section discusses the benefits of SFA and SD by creating a synergy between them. The last section presents the conclusion and future work of this paper.

2 BACKGROUND

The SD was described by Jay Forrester in the 60's (Forrester 1961) and initially was called Industrial Dynamics. SD has a methodology for analyzing and modeling complex systems, which combines diagrams, models, and analytic tools to explore the complex relations within a system and understand how the system changes over time (Forrester 1988), (Sterman 2000). The actual SD methodology is the product of (Forrester 1988) and (Sterman 2000) work, which synthesized the steps for modeling dynamic simulation models. The methodology has four phases: (1) Conceptualization, (2) Formulation, (3) Evaluation, and (4) Implementation. Several research works in the innovation management domain apply SD. For instance, (Xue and Xu 2017)

analyze the factors and the internal logic between them that influence the capacity of innovation in a company of information technologies. The authors use Vensim PLE software to simulate the model. (Dash et al. 2010) Study the problem of decision-making based on the risk of technological innovation from the entrepreneurial team. In this work, the authors identify the differences between an individual and a team's decision-making where decisions are mainly affected by risk perceptions and preferences. The perspective of agent-based modeling is behind the SD model. The proposed approach is validated through a case study in the decision-making about the risk of technological innovation in a Chinese automotive company. (Walrave and Raven 2016) Develop an SD model that integrates the concept of the engines of innovation, following the literature on emerging systems of technological innovation, with the notion of "transition pathways" that developed as part of multilevel thinking. (King and Burgess 2017) Present a new ERP model to understand the relationships between critical success factors better and to encourage the exploration of the most appropriate implementation strategies using SD. From a different perspective, the TRIZ theory which has an extended field of application in the innovation management field suffers from several drawbacks. About this subject, (Rantanen and Domb 2008) explains the importance of combining TRIZ with other approaches and tools to improve the efficiency of a problem-solving process. The mix of other techniques or methodologies is a successful improvement strategy that produces a synergy where both techniques increase their advantages (Cortes et al. 2009). Similarly, the SD methodology can be applied to model systems and analyze each component and their variables over time, but it cannot guide the user to solve a problem, just to explore the potential effects of different scenarios. Hence, this article proposes the integration of the modeling processes of SFA with the SD approach. This integration can model an inventive problem and also propose at least one solving path or strategy. A case study depicts the application of the SFA+SD approach. The methodology described in this article is a synthesis from (Terninko 1998), (Salamatov 1999), (Forrester 1988) and (Sterman 2000). Table 1 shows the four phases and the most significant activities.

<p>Phase 1 Description: Identify elements</p>	<ol style="list-style-type: none"> 1) Define the purpose of the model. 2) Identify critical variables. 3) Establish relationships among variables. 4) Develop the causal loop diagram (CLD). 5) Identify the substances and field, interacting inside the model.
<p>Phase 2 Formulation: Making diagrams</p>	<ol style="list-style-type: none"> 1) Develop the Forrester diagram. 2) Simulate the model only to observe the conflicts. 3) Build the SF Model. 4) Comparison and verification the SFM and Forrester diagram to secure the compatibility of both models.

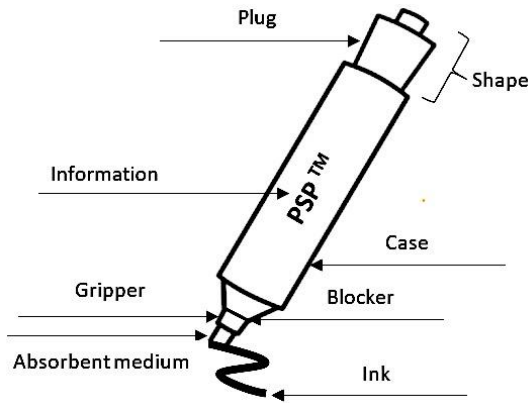


Figure 2: Physical parts of a dry-erase marker

3.2 Formulation: Making diagrams

The second part consists in to build diagrams corresponding to SFA and SD methodologies. Both will provide information about the conflicts in the system.

1) Develop the Forrester diagram: The CLD is transformed into a Forrester Diagram, making use of a simulation software: Stella ©. At this moment the endogenous and exogenous variables are transformed into flows, auxiliary, and level variables. Forrester Diagram allows the introduction of additional variables that do not appear in the CLD, such as quantity, text necessary, capacity medium, use, limit, demand, and the quantity of ink. These variables make it possible to complete the missing spaces of information within the programming of the model. Figure 3 shows Forrester diagram.

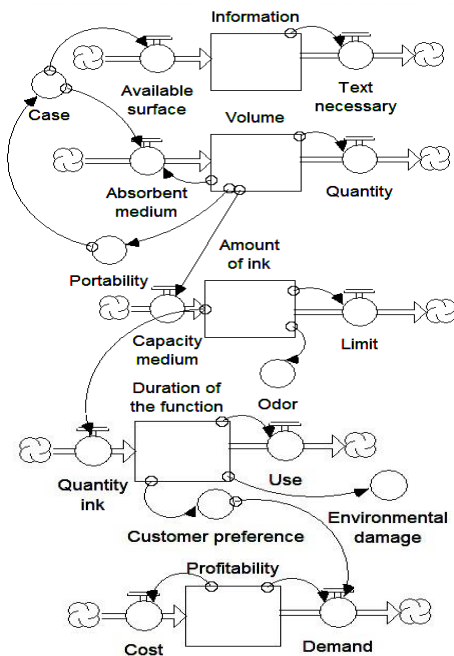


Figure 3: Forrester Diagram

Level variables such as information, volume, ink quantity, function duration, and profitability generate a differential equation measured over time. In this paper, the most significant equation is the volume, because it in-

volves the portability, length, and the absorbent medium of the object.

$$V(t) = V(t_0) + \int_0^t (A - Q)dt \quad \dots \quad (1)$$

Where $V(t)$ = Volume of the object in any instant of time, $V(t_0)$ = Initial volume, A = The size of the absorbent medium and Q = Quantity of ink. The volume of the object is closely related to its portability.

2) Simulate the model only to observe the conflicts: A simulation model is useful to observe conflicts in the system. Stella © allows creating graphs inside its interface. Some conflicts found are:

A) If the marker case increases, the object volume increase then portability decreases.

B) If the ink level increases, also do the duration of the main function, nevertheless, manufacturing cost also increases affecting the cost-effectiveness.

C) If the duration of the main function increases, similarly does the customer preference generating high-cost effectiveness (higher demand), and a decrease in environmental damage.

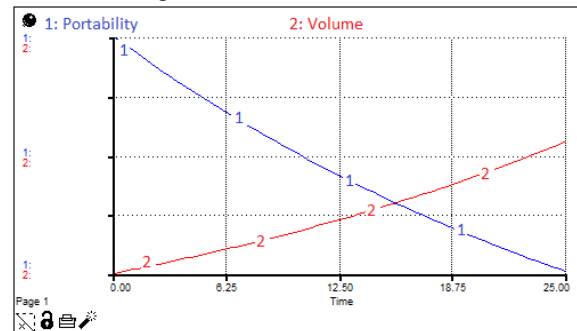


Figure 4: Graph 1. Portability – Volume

Figure 4 represents the first conflict: If the object volume increases, then it will de-crease the portability of the object.

3) Build the SF Model: Once the Forrester diagram was built, the SFM is modeled. Table 2 shows the nomenclature used in SFM. Each arrow has a different meaning in the model.

Analysis	Nomenclature
1) Application	→
2) Desired effect	→
3) Insufficient desired effect	- - - - - →
4) Excessive	→
5) Harmful effect	~ ~ ~ ~ ~ →
6) Inexistent effect	- - - - -
7) Transformation of model	→
8) Uncontrolled effect	- - - - - →

Table 2: Nomenclature used in SFM (Delgado-Maciél et al. 2017)

Figure 5 shows the development of the SFM. The model is composed of five substances and two fields that are related to each other to form a unique system. This one is formed of four interlaced levels. The first level is composed of S_1 , S_2 , and a mechanical field, referring to the level of ink acting on the blackboard. The second shows

the relationship of the absorbent medium to the ink level through a gravitational field. The third level denotes the application of the case on the absorbent medium through a mechanical field. The last level uses the previous field to form the final relation between the case and plug.

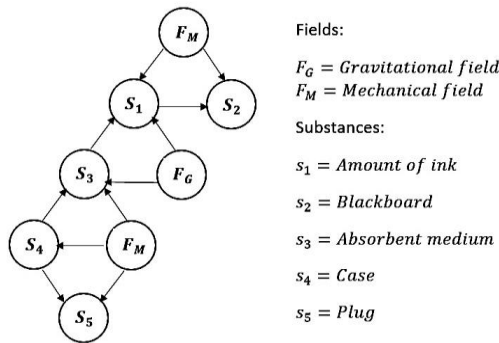


Figure 5: SFM and variables

Figure 5 is useful to model the inventive problems identified in step 4 with the help of the CLD. There are some inventive problems in the system: the ink volume, the length, the absorbent medium, and the ink spill.

<p>Conflict 1. Length – Volume: If S_1 increases its volume, S_3 must increase its length. S_1 should maintain its volume at S_3 without accelerating the wear rate.</p>
<p>⇒ Object volume</p>
<p>Conflict 2. Ink volume – Volume: If S_1 is excessive its volume is stored in S_5, generating a harmful effect.</p>
<p>⇒ Ink spill</p>
<p>Conflict 3. Absorbent medium – ink spill: If S_4 is insufficient to contain S_3, then leakage will occur.</p>
<p>⇒ Leakage</p>

Table 3: Conflicts and SFM

4) Comparison and verification of the SFM and Forrester diagram to secure the compatibility of both models: Both models have high compatibility in their variables that allow representing the system through causal relations (Forrester diagram) and using functions (SFM). The endogenous and exogenous variables must be similar

between both models to ensure compatibility between them.

3.3 Evaluation: Simulation

Phase 3 consists of using the model to simulate the variables programmed. In this section, the sensitivity analysis uses the 76 Standard Solutions.

1) Simulate the model and test the relationships of the CLD and the SFM. In this step, the user can modify any variable using the Stella © interface. The modification of the model variables produces information about the relevance of the CLD and the SD simulation model.

2) Test the model under different assumptions: Once the CLD and SFM assumptions have been verified, the next phase demands a simulation of the model to test different scenarios. The importance lies in establishing policies to solve the conflicts described in Table 3. The 76 Standard Solutions are a set of strategies to transform the system and thus, useful to create different scenarios.

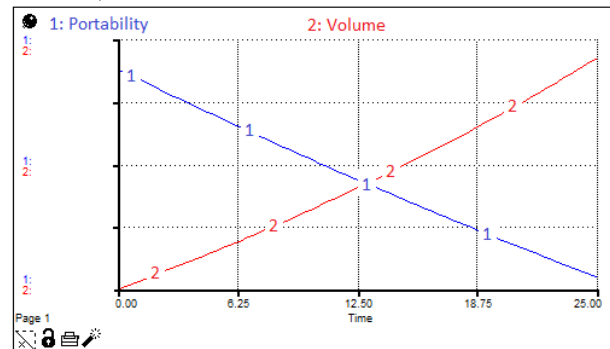


Figure 6: Graph 2. Portability – Volume

Comparing figure 4 and figure 6, the application of policies give rise to modifications in the graphs of the variables analyzed. The time horizon of the simulation concerns to the last 25 years. In this period the product had little variation in its original design. However, the trend shows that design modifications impact the volume of the object. As the object becomes smaller, portability increases; generating an inventive problem. The user can manipulate the model using sensitivity analysis to test different alternatives that meet their requirements and provide a solution to the conflicts listed in Table 3.

3) Document the model response with sensitivity analysis: Sometimes it is difficult to analyze the variations. Sensitivity analysis allows the evaluation of different scenarios simultaneously to compare different alternatives.

4) Consider solutions from the 76 Standard Solutions: Finally, the proposed archetypes are used from the 76 standard solutions. Salamatov (1999) offers a description of the general structure of these solving strategies and a useful explanation of how to apply each solution standard. Example: The SFM shown in figure 6 can be solved using class 5 (helpers) of a standard solution, through group 5-1 (introduction of substances under restricted conditions) and group 5-2 (introductions of fields under restricted conditions).

3.4 Application: Find solutions

The last phase is to collect all conclusions and observations from the set of scenarios and evaluation of the SFM and SD.

1) Observation and analysis of the behavior of the model under different scenarios: The graphical interface of Stella © allows the visualization of different policies and evaluate the best scenario. Figure 6 shows some examples. In figure 7, the top graph represents an increase in the portability (line 1) and a reduction in the volume (line 2). The bottom graph shows the opposite case: an increase in the volume (line 2) generates a degradation in the portability (line 1).

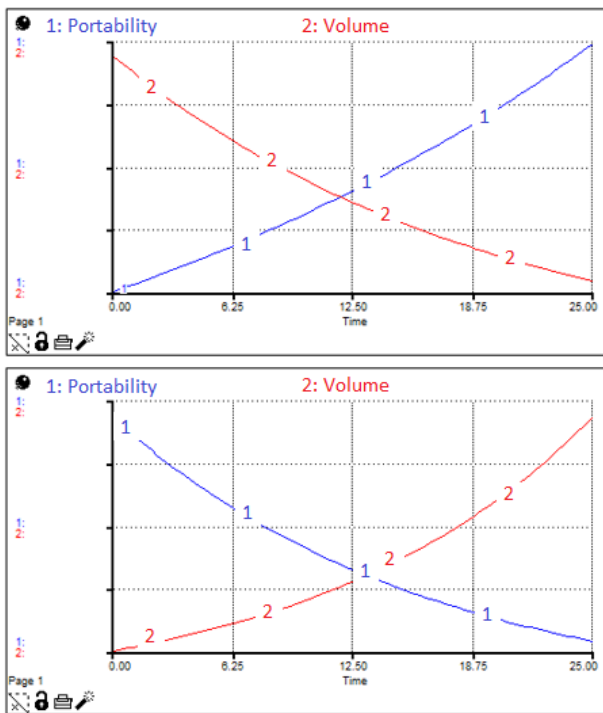


Figure 7: Proposed policies

2) Develop at least one solution: After analyzing SFM and CLD, a solution is proposed using the 76 standard solutions. In section 3.3 it was described that the groups 5-1 and 5-2 are similar to the model presented. A possible solution is the standard 5-1-1-2: “If it is necessary to introduce a substance into the system, and it is not allowed, a field can be introduced instead of the substance.” The transformation of SFM shown in Table 3 in a new SFM is shown in figure 8.

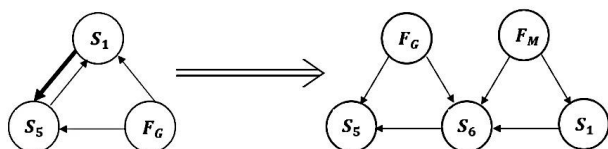


Figure 8: Transformation SFM using 76 standard solution

According to the standard solution, a new variable (substance S_6) was added to find a solution: S_6 represents a device that replaces ink by using the capillarity of the

absorbent medium while the object is not in use. Hence, when the object is in repose, the absorbent medium has physical contact with a certain amount of ink stored in a small object. The object transfers the ink to the absorbent medium by contact.

3) Validate the solution: A quick and effective way to validate the solution is to introduce the new data into Stella © to simulate a new scenario and verify if the solution proposed by the SFM will have the expected impact. Once the simulation is done, the graphical interface allows the user to visually analyze the proposal. Figure 9 shows a graph after applying the proposed solution.

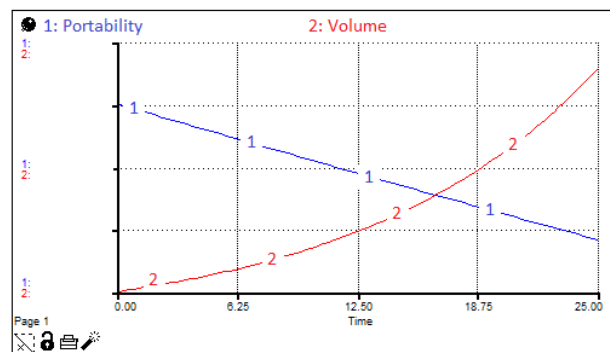


Figure 9: Graph after applying standard solutions

According to Figure 9, after applying the new solution, the portability (line 1) decreased more smoothly. The image shows a less pronounced decrease in portability and a normal increase in the volume variable (line 2).

4) Recommend the best alternative: In this case, the solution performs as expected. However, it is possible to use another solution from the 76 standard solutions (groups 5-1 and 5-2 for the case described)

4 DISCUSSION AND RESULTS

One of the great challenges of 21st-century science is the analysis of complex systems (Bar-Yam 2003). Complex systems consist of agents or elements that interact with each other in their environment in ways that can often be described regarding relatively simple rules (Jacobson 2017). Some interactions at a micro level within a system lead to the creation of new patterns at a higher level (macro system). This is carried out through positive or negative feedbacks and sensitivity of the initial conditions. Feedbacks influence each component of the system, causing the smallest variation in the system to produce significant changes. Frequently, the solving process of an inventive problem uncovers the lack of knowledge. The solver cannot propose an effective solution or simply does not know how to materialize a function, even if the expected result is clear. Hence, the solver or user only has his experience and past events to deal with it. According to (Savransky 2000) the complexity of the systems during the modeling process of inventive problems is not related to the number of variables in the model, but to the number of conflicts, their nature and

the solution found. The case study shown above reflects the feasibility of creating a synergy between SFA and SD. It is important to notice that both tools have different approaches to problem-solving. The proposed methodology provides an opportunity for research because of its ability to encompass two different areas of knowledge through a series of steps. This methodology represents a new approach for problem-solving of certain types of complex systems: Those that include the analysis of functions and require continuous simulation. Each of the phases of Table 1 shows a close relationship between both techniques, generating a single tool.

SFA is a tool capable of modeling the functions of certain systems. SD has the tool for modeling and simulation, but does not offer any guideline or tool for problem-solving, neither for proposing inventive solutions. The proposed synergy between SFA and SD assists the user in the decision making process through an analysis of the system over time. Although there are techniques that involve simulation, innovation, and inventiveness, the proposed SFM + SD approach produces some unique advantages that other tools do not possess. Table 4 shows a brief comparison between Discrete Simulation (DS), Substance-Field Analysis (SFA), System Dynamics (SD) and the combinations of the Substance-Field Modeling and System Dynamics (SFM + SD).

Advantage	DS	SFA	SD	SFA+SD
Modeling simultaneous problems	-	-	X	X
Modeling complex systems	X	X	X	X
Performance analysis of functions	-	X	-	X
Solution of inventive problems	-	X	-	X
Use of tools to simulation	X	-	X	X
Based on mathematical models	X	-	X	X

Table 4: Comparison between different approaches

The results of this integration generate a valuable contribution to the user since it allows the user to model and simulate an inventive problem. The first stage of the SFM+SD methodology gives the user the ability to analyze the relationships between the system variables. This step produces valuable information for modeling inventive problems. This information satisfies the requirements of SFA and SD. Probably, the most important part of phase one is the design of the CLD because it is the support for the Forrester diagram and the SFM. The second part consists of performing the SFM and the Forrester diagram in parallel. Both approaches produce complementary information and a wide perspective about the problem and their conditions. During the third stage, the simulation model created in the previous

phase is executed. The sensitivity analysis plays an important role because it allows testing different scenarios and guiding the user to the solution process. In the fourth step, it is necessary to ensure the compatibility between the two diagrams. Finally, the last phase involves a decision-making process where the user evaluates the best solution based on previous simulations. The main advantage of the Stella © interface is reflected in the use of visual tools allowing a better analysis, including sensitivity analysis, which is a way to measure the impact on the result of a calculation produced by any change in the variables. Stella proposes in its interface the tools to conceive a sensitivity analysis, and hence, it is possible to do a simultaneous analysis of different variables in a single graphic. Figure 10 shows an example of sensitivity analysis in the volume variable.

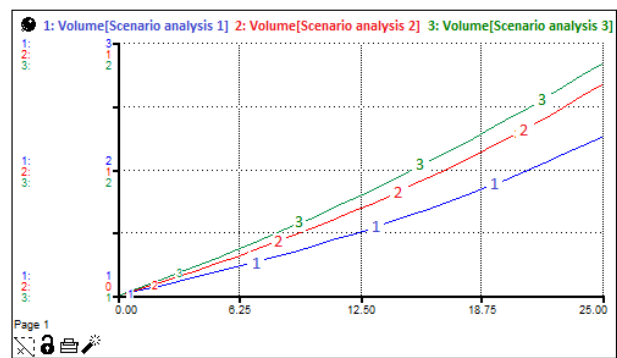


Figure 10: Sensitivity analysis

The SD compatibility is defined whenever it is necessary to carry out simulations over time. Perhaps the main contribution of SD to SFA is the ability to model a system and be able to quantify it through equations, giving mathematical support for the modeling process in TRIZ. The synergy between SFA and SD is possible because both techniques have some common steps within their methodologies, particularly during modeling that involves the representation of the variables that form the system, giving rise to the SFM and the CLD.

The validation process involves moving the conceptual design to a physical plane. Hence, the next steps are considered to carry out this crucial step:

- 1) Design of scenarios obtained from the simulation model.
- 2) Selection of alternatives based on the sensitivity analysis of the simulation.
- 3) The physical construction of prototypes.
- 4) Feedback and return of experiences to improve the design process, the conceptual design, and to identify new opportunities.

The final result of the previous steps results in a physical prototype that allows validating what was analyzed during the simulation.

5 CONCLUSION AND FUTURE WORK

The synergy between SFA and SD generates a research opportunity for proposing new methods and unveil new

resources for problem-solving. In this combination, TRIZ acquires a tool capable of modeling conflicts mathematically and analyzing their respective behaviors over time. In turn, SD extends its field of application via the assimilation of some tools for inventive problem-solving.

The proposed methodology aims to demonstrate the feasibility of combining SFA with SD. The main contribution of this work is the demonstration of a mutual benefit between both techniques: SFA is still a tool with significant limitations, perhaps the most important is its inability for analyzing systems with continuous variables. This work points out the opportunity to combine both techniques to get a richer and more flexible modeling process. The intention of the article is not the proposal of a universal method of modeling, but to propose a function-oriented approach that is verified through simulation. This creates the need to strengthen SFA with tools capable of overcoming their weaknesses.

Future work suggests using the SFA + SD methodology to solve problems with a greater degree of complexity. The proposed case has a didactic purpose, and it includes only one object of daily use. The next evaluation of the methodology is to apply it to solve conflicts generated in the design of new products or the improvement of processes. These applications have a greater number of variables, and therefore their complexity is higher. While the development of the methodology only shown synergy between SFA and SD, however, both techniques are capable of including new approaches from other knowledge tools.

ACKNOWLEDGMENTS

The National Council on Science and Technology (CONACYT) for the grant number 637564, FC-2433 and the Secretariat of Public Education (SEP) through the Tecnológico Nacional de México and PRODEP sponsored this work. Additionally, the ROPRIN working group (Industrial Process Optimization Network) supported this work.

REFERENCES

- Altshuller G., 1984. *Creativity as an Exact Science*. CRC Press.
- Bar-Yam Y., 2003. *Dynamics of complex systems*. Reading, MA: Addison-Wesley.
- Bultey A., et al., 2015. A Proposal of a Systematic and Consistent Substance-field Analysis. *Procedia Engineering*, 131, p. 701-710.
- Carver J., 2000. Dry Marker and Eraser System. US Patent 6,048,121, 11 April 2000
- Cortes G., et al., 2009. Case-based reasoning and TRIZ: A coupling for innovative conception in Chemical Engineering. *Chemical Engineering and Processing: Process Intensification*, 48(1), p. 239-249.
- Dash D., 2009. Modeling technological innovation risks of an entrepreneurial team using system dynamics: An agent-based perspective. *Technological Forecasting and Social Change*, 77(6), p. 857-869.
- Delgado-Maciel J., et al., 2017. Application of the Systems Dynamics Approach to Model Inventive Problems. *Conference LNCS on Hybrid Artificial Intelligent Systems (HAIS'2017)*, Logroño, Spain, vol. 10334, p. 494-506
- Forrester J., 1961. *Industrial Dynamics*. The MIT Press.
- Forrester J., 1988. *Principles of Systems*. Productivity Press. Productivity Press
- Helfman Y., et al., 2015. Substance Field Analysis and Biological Functions. *Procedia Engineering*, 131, p. 372-376.
- Jacobson M., et al., 2017. Designs for learning about climate change as a complex system. *Learning and Instruction*, p. 1-14
- King F. and Burguess T., 2006. Beyond critical success factors: A dynamic model of enterprise system innovation. *International Journal of Information Management*, 26(1), p. 59-69.
- Rantanen K. and Domb E., 2008. *New problem solving applications for engineers and manufacturing professionals*. CRC Press
- Salamatov Y., 1999. *TRIZ: the Right Solution at the Right Time: A Guide to Innovative Problem Solving*. Institute of Innovative Design.
- Savransky S., 2000. *Engineering of Creativity: Introduction to TRIZ Methodology of Inventive Problem Solving*. CRC Press.
- Sterman J., 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Education.
- Terninko J., et al., 1998. *Systematic Innovation: An Introduction to TRIZ (Theory of Inventive Problem Solving)*: CRC Press LLC
- Xue C. and Xu Y., 2017. Influence Factor Analysis of Enterprise IT Innovation Capacity Based on System Dynamics. *Procedia Engineering*, 174, p. 232-239.

AN EFFECTIVE COLUMN GENERATION ALGORITHM FOR THE INVENTORY-ROUTING PROBLEM WITH TRANSSHIPMENT

Wouter Lefever, El-Houssaine Aghezzaf

Khaled Hadj-Hamou

Department Industrial Systems Engineering
and Product Design, Ghent University

DISP, INSA Lyon

Technologiepark 903

21 avenue Jean Capelle

Zwijnaarde 9052, Belgium

69621 Villeurbanne, France

wouter.lefever@ugent.be, elhoussaine.aghezzaf@ugent.be khaled.hadj-hamou@insa-lyon.fr

ABSTRACT: *The Inventory-Routing Problem with Transshipment is a combined inventory management and routing problem in which the possibility of outsourcing deliveries to a subcontractor is allowed. In this paper, we reformulate the problem using the Dantzig-Wolfe decomposition framework. The Dantzig-Wolfe linear relaxation is solved by column generation. The columns are generated as needed by solving a selective traveling salesman problem. In order to improve the linear relaxation lower bound two sets of valid inequalities are added to the master problem. To further accelerate the optimization procedure a greedy heuristic is developed to create a good initial set of columns and an appropriate branching strategy is chosen. The experimental results demonstrate the strength of the linear relaxation lower bound. The presented column generation algorithm manages to solve instances with up to 30 retailers to optimality.*

KEYWORDS: *Inventory-Routing, Dantzig-Wolfe decomposition, Column Generation, Valid Inequalities, Branch-and-Bound.*

1 INTRODUCTION

In the late 1980's Vendor-Managed Inventory (VMI) emerged as a very promising strategy to improve supply chain performance. Applications by Wal-mart, Procter & Gamble and other successful VMI initiatives contributed to the popularization of VMI. Today it is one of the most investigated supply chain strategies.

In a VMI system the supplier manages the retailer's inventory. The freedom to choose the timing and the size of the deliveries allows him to combine orders of multiple retailers more efficiently. The only restriction the supplier has to respect is to make sure no stock-outs can occur at each retailer. The suppliers that are engaged in a VMI system can integrate inventory management and transportation planning and take decisions to optimize the costs of both subsystems simultaneously. The mathematical problem that supports these decisions is known as the Inventory Routing Problem (IRP).

In this paper, we deal with a specific variant of the IRP: the Inventory-Routing Problem with Transshipment (IRPT). The concept of transshipments within inventory-routing was first formally introduced by

Coelho et al. (2012). In this problem not only the supplier can deliver goods to the retailers, but also transshipments between retailers or between the supplier and a retailer are possible. These transshipments are carried out by a subcontractor. The objective is to determine the quantities delivered to all retailers by both the supplier and the subcontractor in the different periods of the planning horizon while minimizing the total cost. This total cost consists of inventory costs, transportation costs of the supplier and costs of the outsourced deliveries.

From an academic point of view research on the IRPT is scarce. Since its introduction by Coelho et al. (2012) and a follow-up study by Coelho and Laporte (2013), there have been no further studies on modeling and solving the IRPT. Further investigation into the structure of this problem is thus definitely worthwhile.

The remainder of this paper is organized as follows. In Section 2, we present a literature survey on models, solution approaches and applications of the IRPT. In Section 3, we present the current mathematical model of the IRPT. In Section 4, we reformulate the problem using the Dantzig-Wolfe decomposition framework. We discuss the linear programming master

problem and the resulting pricing subproblem. We also present two sets of valid inequalities, a greedy heuristic to create an initial set of columns and an appropriate branching strategy. In Section 5, we present the experimental results. Finally in Section 6, we conclude the paper with some remarks about further research.

2 LITERATURE REVIEW

The IRPT is first formally introduced in Coelho et al. (2012). The authors present a Mixed Integer Program for the IRPT. The inclusion of transshipments makes the problem very hard to solve and the experimental results illustrate the weakness of the linear relaxation. In order to obtain a good quality solution the authors develop a heuristic based on the adaptive large neighborhood search (ALNS) framework presented by Pisinger and Ropke (2007).

In Coelho and Laporte (2013) the formulation of the IRPT is improved. Weaker subtour elimination constraints are replaced by the Dantzig-Fulkerson-Johnson constraints. The authors present a branch-and-cut algorithm which includes a solution improvement procedure that approximates the cost of a new solution resulting from vertex removals and/or insertions. This branch-and-cut algorithm is applied on 160 instances of the IRPT. Results show that 158 instances can be solved to optimality.

The flexibility of the IRPT is illustrated in Coelho et al. (2014). In this paper the Dynamic and Stochastic IRP (DSIRP) is investigated. In this problem the retailer demands are gradually revealed over time. The authors propose four different policies to solve the problem. Results show that a proactive policy which uses demand forecasts and transshipments greatly improves the solutions quality.

The interest of a transshipment option within the context of inventory-routing is also investigated by Chrysochoou and Ziliaskopoulos (2015). They study the IRPT under stochastic demand. The resulting stochastic problem is solved using the L-shaped method. The experimental results demonstrate the potential of the presented solution approach.

In this article, we will further investigate the IRPT. To the best of our knowledge we are the first to reformulate the IRPT using the Dantzig-Wolfe decomposition framework and solve it using column generation. Earlier papers (Desaulniers et al., 2015; Taş et al., 2014; Errico et al., 2016) have already demonstrated the interest of applying column generation within the context of vehicle-routing. By reformulating the problem we aim to strengthen the mathematical formulation of the IRPT and improve the weak linear relaxation lower bound. Our main con-

tributions are:

- The reformulation of the IRPT using the Dantzig-Wolfe decomposition framework.
- The development of an effective column generation algorithm.
- The development of a greedy heuristic to create a good initial set of columns.
- The design of an intelligent branching strategy.

3 PROBLEM DESCRIPTION

In this section we will present the general characteristics of the IRPT, but for a detailed description of the problem we refer to Coelho et al. (2012). The problem is defined on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{0, \dots, n\}$ is the vertex set and \mathcal{E} is the edge set. Vertex 0 represents the supplier and the vertices of $\mathcal{V}' = \mathcal{V} \setminus \{0\}$ represent retailers. All the actors, supplier and retailers, have inventory holding costs of h_i per unit and per period. The retailers have a minimum inventory level, which is fixed at 0, and a maximum inventory level, which is fixed at its inventory holding capacity C_i . The supplier has a minimum inventory level of 0, but has no maximum inventory level.

In every period $p \in \mathcal{P} = \{1, \dots, \rho\}$ of the planning horizon a quantity r^p of the product is made available at the supplier. The supplier will use this quantity and its initial inventory to ship quantities q_i^p to each retailer i so that they can fulfill their demands d_i^p . For this purpose a single vehicle of capacity Q is available. This vehicle can perform one route, beginning and ending at the suppliers depot, in every period p . The routing costs c_{ij} are associated with the length of the edge $(i, j) \in \mathcal{E}$.

In the IRPT transshipments w_{ij}^p can be made by a subcontractor later in the time period. A transshipment can start from the depot or from any retailer in a subset $\mathcal{R} \subseteq \mathcal{V}'$. The transshipments are only made by direct shipping and the unit cost associated with transshipping products from i to j is b_{ij} .

For the inventory management the maximum level policy is applied. That means that the delivered quantities have to respect the minimum and maximum inventory level. The retailer's inventory I_i^p is calculated at the end of each period after deliveries, transshipments and satisfying the demand.

The objective of the problem is to minimize the total cost consisting of the inventory holding costs, the transportation costs of the supplier and the costs of the outsourced deliveries. Coelho et al. (2012)

present a model for the IRPT using the following binary variables: x_{ij}^p is equal to 1 if and only if retailer j immediately follows retailer i on the route of the supplier's vehicle in period p and 0 otherwise, and y_i^p is equal to 1 if and only if either the depot ($i = 0$) or retailer i is visited in period p and 0 otherwise. Let the quantity of product delivered from the supplier to each retailer i in each time period p be q_i^p , and let the amount of product delivered directly from $i \in \mathcal{R} \cup \{0\}$ to retailer $j \in \mathcal{V}'$ at period p using the outsourced carrier be w_{ij}^p .

$$\min \sum_{i \in \mathcal{V}} \sum_{p \in \mathcal{P}} h_i I_i^p + \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}, i < j} \sum_{p \in \mathcal{P}} c_{ij} x_{ij}^p + \sum_{i \in \mathcal{R} \cup \{0\}} \sum_{j \in \mathcal{V}'} \sum_{p \in \mathcal{P}} b_{ij} w_{ij}^p \quad (1)$$

$$\text{s.t. } I_0^p = I_0^{p-1} + r^p - \sum_{i \in \mathcal{V}'} q_i^p - \sum_{i \in \mathcal{V}'} w_{0i}^p \quad p \in \mathcal{P} \quad (2)$$

$$I_i^p = I_i^{p-1} + q_i^p + \sum_{j \in \mathcal{R} \cup \{0\}} w_{ji}^p - \sum_{j \in \mathcal{V}'} w_{ij}^p - d_i^p \quad p \in \mathcal{P}, i \in \mathcal{V}' \quad (3)$$

$$I_i^p \leq C_i \quad p \in \mathcal{P}, i \in \mathcal{V}' \quad (4)$$

$$q_i^p \leq C_i - I_i^{p-1} \quad p \in \mathcal{P}, i \in \mathcal{V}' \quad (5)$$

$$q_i^p \leq C_i y_i^p \quad p \in \mathcal{P}, i \in \mathcal{V}' \quad (6)$$

$$\sum_{i \in \mathcal{V}'} q_i^p \leq Q y_0^p \quad p \in \mathcal{P} \quad (7)$$

$$\sum_{j \in \mathcal{V}', i < j} x_{ij}^p + \sum_{j \in \mathcal{V}', i > j} x_{ji}^p = 2y_i^p \quad p \in \mathcal{P}, i \in \mathcal{V} \quad (8)$$

$$\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}, i < j} x_{ij}^p \leq \sum_{i \in \mathcal{S} \setminus \{m\}} y_i^p \quad \mathcal{S} \subseteq \mathcal{V}', p \in \mathcal{P}, m \in \mathcal{S} \quad (9)$$

$$x_{0j}^p \in \{0, 1, 2\} \quad j \in \mathcal{V}', p \in \mathcal{P} \quad (10)$$

$$x_{ij}^p \in \{0, 1\} \quad i, j \in \mathcal{V}', p \in \mathcal{P} \quad (11)$$

$$y_i^p \in \{0, 1\} \quad i \in \mathcal{V}, p \in \mathcal{P} \quad (12)$$

$$q_i^p \geq 0 \quad i \in \mathcal{V}', p \in \mathcal{P} \quad (13)$$

$$I_i^p \geq 0 \quad i \in \mathcal{V}, p \in \mathcal{P} \quad (14)$$

$$w_{ij}^p \geq 0 \quad i \in \mathcal{R} \cup \{0\}, j \in \mathcal{V}', p \in \mathcal{P} \quad (15)$$

In the IRPT, the total cost to be minimized is the sum of inventory holding costs at the supplier and at the retailers, the routing costs for the supplier's vehicle and the transshipment costs (1). Constraints (2) and (3) are inventory balance constraints for the supplier and the retailers. Constraints (4) limit the inventories of the retailer by their maximum capacity. Constraints (5)-(6) define the quantities deliv-

ered. If retailers i is not visited in period p , then constraints (6) enforce that the quantity delivered to it will be zero. Otherwise, if retailer i is visited in period p , then constraints (6) limit the quantity delivered to the retailer's inventory holding capacity, and this bound is tightened by constraints (5), making it impossible to deliver more than what would fill this capacity. Constraints (7) state that the vehicle capacity is not exceeded. Constraints (8)-(9) guarantee that a feasible route is determined to visit all retailers served in period p . Finally, constraints (10)-(15) enforce integrality and non-negativity conditions on the variables. Note that the routing variable x_{0j}^p can take the value 2. In this case the supplier visits only one retailer in the period p , a so-called direct shipment.

4 DANTZIG-WOLFE DECOMPOSITION

In the model for the IRPT constraints (2)-(7) manage the inventory of the supplier and retailers. They also define the quantities that are delivered and guarantee that the vehicle capacity is respected. These constraints are only dependent on the locations (supplier or retailers) that are visited and not on the order in which they are visited. On the other hand constraints (8)-(9) determine the route that is followed by the supplier's vehicle and are independent of the inventory variables I_i^p and delivered quantities q_i^p . We will exploit this structure to formulate a Dantzig-Wolfe decomposition of this problem. In this decomposition we will define the tours that are made by the supplier's vehicle as columns that are introduced in the master problem which manages the inventory of the supplier and retailers and the delivered quantities.

4.1 Master Problem

In the linear programming master problem (LPM) we introduce a variable λ_p^k associated with every column. Every variable λ_p^k represents a certain tour $k \in \mathcal{K}$ made by the supplier in period p . \mathcal{K} represents the set of all possible tours made by the supplier starting and ending at the depot and visiting a subset of retailers of \mathcal{V}' . The corresponding LPM is:

$$\min \sum_{i \in \mathcal{V}} \sum_{p \in \mathcal{P}} h_i I_i^p + \sum_{i \in \mathcal{R} \cup \{0\}} \sum_{j \in \mathcal{V}'} \sum_{p \in \mathcal{P}} b_{ij} w_{ij}^p + \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \left(\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}, i < j} c_{ij} x_{ij}^p \right)^{(k)} \lambda_p^k \quad (16)$$

s.t. (2) – (5)

$$q_i^p - C_i \sum_{k \in \mathcal{K}} b_{ip}^k \lambda_p^k \leq 0 \quad i \in \mathcal{V}, p \in \mathcal{P} \quad (17)$$

$$\sum_{i \in \mathcal{V}'} q_i^p - Q \sum_{k \in \mathcal{K}} a_p^k \lambda_p^k \leq 0 \quad p \in \mathcal{P} \quad (18)$$

$$\sum_{k \in \mathcal{K}} \lambda_p^k \leq 1 \quad p \in \mathcal{P} \quad (19)$$

$$\lambda_p^k \geq 0 \quad k \in \mathcal{K}, p \in \mathcal{P} \quad (20)$$

(13) – (15)

where the coefficients of matrix A_p^k and matrix B_{ip}^k are given by

$$a_p^k = \begin{cases} 1 & \text{if the depot is included in the tour} \\ & \text{associated with variable } \lambda_p^k, \\ 0 & \text{otherwise.} \end{cases}$$

$$b_{ip}^k = \begin{cases} 1 & \text{if retailer } i \text{ is included in the tour} \\ & \text{associated with variable } \lambda_p^k, \\ 0 & \text{otherwise.} \end{cases}$$

Compared to the MIP formulation of section 3 the LPM is different in several ways. The only variables that are included in the LPM are I_i^p , q_i^p , w_{ij}^p and λ_p^k . The variables I_i^p , q_i^p and w_{ij}^p only act on inventory component of the IRPT. The variables λ_p^k represent the tours made by the supplier. In the objective function (16) the summation over the routing variables x_{ij}^p with their associated costs c_{ij}^p has been replaced by a summation over the λ_p^k with the corresponding cost of the represented tour. The inventory management (2)-(5) of the MIP formulation has remained the same. In the new equations (17)-(18) the binary variables y_0^p and y_i^p have been replaced by their appropriate summation over the variables λ_p^k , $\sum_{k \in \mathcal{K}} a_p^k \lambda_p^k$ and $\sum_{k \in \mathcal{K}} b_{ip}^k \lambda_p^k$.

Constraint (19) guarantees that there is at most one tour per period p . Constraint (20) limits the domain of the λ_p^k variables and constraints (13)-(15) limit the domain of the variables I_i^p , q_i^p and w_{ij}^p .

4.2 Pricing Subproblem

To find the column that yields the lowest reduced cost, we must solve the pricing subproblem. Associating μ_i^p , π^p and σ^p to the dual prices of constraints (17), (18) and (19), we formulate the following sub-

problem for every $p \in \mathcal{P}$:

$$\min \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}, i < j} c_{ij} x_{ij} - Q \pi^p y_0 - \sum_{i \in \mathcal{V}'} C_i \mu_i^p y_i - \sigma^p \quad (21)$$

$$\text{s.t. } \sum_{j \in \mathcal{V}', i < j} x_{ij} + \sum_{j \in \mathcal{V}', i > j} x_{ji} = 2y_i \quad i \in \mathcal{V} \quad (22)$$

$$\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}, i < j} x_{ij} \leq \sum_{i \in \mathcal{S} \setminus \{m\}} y_i \quad \mathcal{S} \subseteq \mathcal{V}', m \in \mathcal{S} \quad (23)$$

$$x_{0j} \in \{0, 1, 2\} \quad j \in \mathcal{V}' \quad (24)$$

$$x_{ij} \in \{0, 1\} \quad i, j \in \mathcal{V}' \quad (25)$$

$$y_i \in \{0, 1\} \quad i \in \mathcal{V} \quad (26)$$

The subproblem will only act on the routing component of the IRPT and contains only the routing variables x_{ij} and the binary variables y_i . Since the tours of the different periods in the planning horizon are independent of each other, we can formulate a subproblem for every $p \in \mathcal{P}$ rather than one subproblem for the whole planning horizon. In the subproblem the objective function (21) depends on the transportation costs of the tour and the dual values μ_i^p , π^p and σ^p of the master problem. Constraints (22) and (23) are degree and subtour elimination constraints like constraints (8) and (9) in the original problem. Constraints (24)-(26) enforce the integrality of the variables x_{ij} and y_i .

The pricing subproblem is similar to a selective Travelling Salesman Problem (STSP). Finding the optimal solution is an NP-complete problem as it envelops solving the TSP.

4.3 Valid Inequalities

In Lefever et al. (2017) the mathematical formulation of the IRPT was examined. The authors proposed two sets of valid inequalities for the IRPT. In order to strengthen the mathematical formulation of the IRPT and improve the Dantzig-Wolfe linear relaxation lower bound, we will add these valid inequalities to the master problem:

- The minimum number of visits.
 $\forall i \in \mathcal{V}', p_1, p_2 \in \mathcal{P}, p_2 \geq p_1:$

$$\sum_{p=p_1}^{p_2} y_i^p + \frac{\sum_{p=p_1}^{p_2} \sum_{j \in \mathcal{R} \cup \{0\}} w_{ji}^p}{\sum_{p=p_1}^{p_2} d_i^p} \geq \frac{\sum_{p=p_1}^{p_2} d_i^p - I_i^{p_1-1}}{\sum_{p=p_1}^{p_2} d_i^p} \quad (27)$$

- The minimum number of subdeliveries.

$$\forall i \in \mathcal{V}', s \in \mathcal{P} \mid \mathcal{P}^-_{is} \neq \emptyset:$$

$$\sum_{p \in \mathcal{P}^-_{is}} y_i^p + \frac{\sum_{p \in \mathcal{P}^-_{is} \cup \{s - |\mathcal{P}^-_{is}|\}} \sum_{j \in \mathcal{R} \cup \{0\}} w_{ji}^p}{d_i^s} \geq 1 \quad (28)$$

Valid inequality (27) determines the minimum number of times the supplier has to visit a retailer i over a period interval $[p_1, p_2]$. If the retailer is not visited enough, then one or multiple transshipments of sufficient size should be made. Valid inequality (28) ensures that at least 1 delivery, either by the retailer or the subcontractor, is made during the periods of the set \mathcal{P}^-_{is} . The set \mathcal{P}^-_{is} contains all the periods up to period s , in which a delivery can be made so that the delivered units could still be in stock when a first in, first out (FIFO) inventory system is applied.

By replacing the binary variables y_0^p and y_i^p by their corresponding summations $\sum_{k \in K} b_{ip}^k \lambda_p^k$ and $\sum_{k \in K} a_p^k \lambda_p^k$ the valid inequalities (27) and (28) can be added to the formulation. In the pricing subproblem the objective coefficients of the binary variables y_i^p appearing in the valid inequalities (27) and (28) should be changed correspondingly.

4.4 Generation Of Columns

Because the pricing subproblem is an NP-complete problem, the majority of the solution time is spent on solving this problem. To accelerate this procedure we will add the column obtained after solving the subproblem for period p to all the periods of the planning horizon.

Furthermore, we develop a greedy heuristic in order to obtain an initial set of columns. In our heuristic columns are iteratively built using a randomized greedy criterion. Based on the optimal solutions of smaller instances, we remark that solutions often consist of periods in which either no tour is made or a tour is made to a large subset of the retailers. If a tour is made, the number of retailers in the tour is typically between 60% and 80% of the retailers. In our heuristic we exclude all tours with too few or too many retailers.

To build the columns we will use the inventory holding cost and the inventory capacity of a retailer as greedy criterion, because these retailers have more potential to serve as satellite facilities. Denote \mathcal{H}^+ the set of retailers with higher inventory holding costs than the supplier, and \mathcal{C}^+ the set of retailers with a larger than average inventory holding capacity. Then, each execution of the heuristic consists of the following steps: We select a random number of retailers visited in our column between $[0.6|\mathcal{V}'|, 0.8|\mathcal{V}'|]$. We

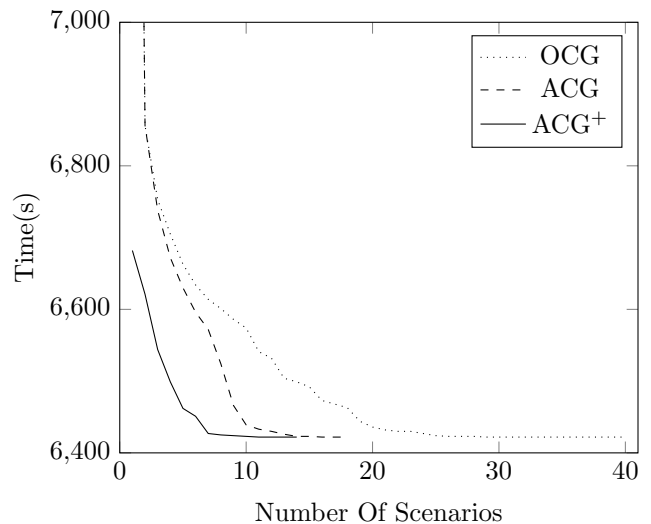


Figure 1: Sequence of events during a period.

randomly select one retailer from the set \mathcal{H}^+ and one retailer from the set \mathcal{C}^+ until 75% of the retailers in our column is assigned. The other retailers in the column are selected from the set $\mathcal{V}' \setminus \{\mathcal{H}^+ \cup \mathcal{C}^+\}$. We execute the heuristic $|\mathcal{P}| |\mathcal{V}'|$ times. In this way, we obtain a large enough number of columns to avoid the ineffective first iterations. By randomly choosing retailers from the sets \mathcal{H}^+ , \mathcal{C}^+ and the $\mathcal{V}' \setminus \{\mathcal{H}^+ \cup \mathcal{C}^+\}$ there is enough diversity in the initial column set.

In addition to the columns generated by the greedy heuristic, we will also generate all the columns in which only one retailer is visited, so-called direct shipments, a priori. The advantage of this strategy is that the value 2 must no longer be included in the domain of the x_{0j} variables in the subproblem. This strengthens the formulation of the pricing subproblem and accelerates its solution time.

In figure 1 the effect of these improvements on the column generation procedure are shown for an instance with 30 retailers and a planning horizon of six periods. The dotted line shows the original column generation (OCG), when the obtained column is added only to the period for which the pricing subproblem is solved. By adding this column to every period of the pricing subproblem, the column generation procedure is accelerated (ACG), the dashed line. Remark that the tail at the end of the procedure is much shorter than in the original column generation. By starting with a larger initial set of columns (ACG⁺), the first iterations are more effective and the Dantzig-Wolfe linear relaxation can be found in 14 iterations.

4.5 Branching Strategy

When the subproblem does not generate any more columns, we have obtained the optimal solution of the Dantzig-Wolfe linear relaxation. If the solution

is integer, $y_0^p = \sum_{k \in K} a_p^k \lambda_p^k$ and $y_i^p = \sum_{k \in K} b_{ip}^k \lambda_p^k$ are integer, we have found the optimal solution to the IRPT. Otherwise, a branch-and-bound tree must be explored, and additional columns might be generated at each branch.

In order to minimize the number of nodes that should be explored in our branch-and-bound tree, we will apply the following branching scheme: First, we will branch on the fractional y_0^p values. If all these values are integer, we branch on the y_i^p value whose value is closest to 0.5. In this way, we aim to create branches whose linear relaxation is much higher than the linear relaxation of their parent.

An example of our branching strategy is shown in figure 2. In this example an instance with five retailers and a planning horizon of three periods is considered. The pruned nodes are shown in grey.

In figure 2(a) the algorithm branches first on the variable y_0^3 . In the upper branch this variable is set to zero. In this branch the supplier can not deliver goods in the third period. In the lower branch the variable is set to 1 and the supplier must deliver goods to at least one retailer (enforced by the degree constraints). This has a large impact on the linear relaxation. In figure 2(b) the algorithm branches on a random y_i^p variable. It is clear that this branching strategy has less effect on the linear relaxation which probably will lead to more nodes in the branch-and-bound tree.

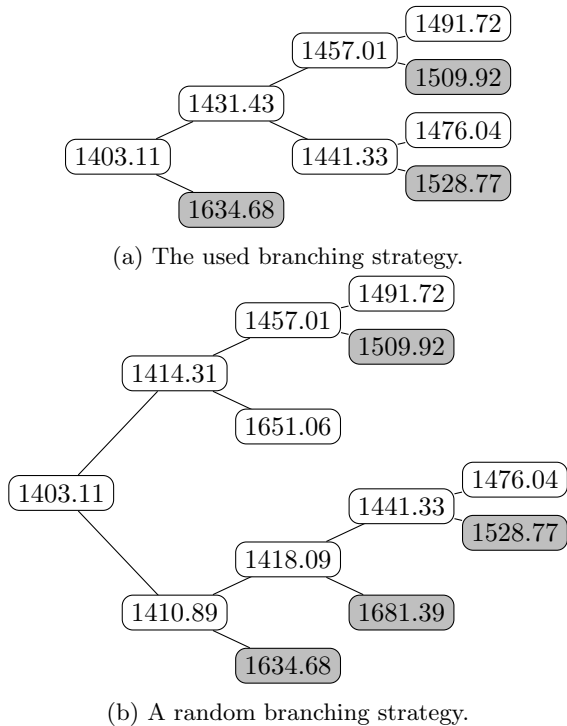


Figure 2: Difference between different branching strategies.

In order to rapidly obtain a good estimate of the

value of the optimal solution, we will use a depth-first branch-and-bound strategy. This allows us to prune nodes earlier in the branch-and-bound tree. If a breadth-first approach would be used, then lower branch of figure 2(a) would need to be explored before the solution values 1491.72 and 1476.04 are found.

5 COMPUTATIONAL STUDY

To evaluate the performance of the column generation algorithm we conducted experiments on a benchmark set of 80 instances. The experiments were executed on a computer with an i7-3770 3.4 GHz processor and 32 GB of memory. The algorithm was coded in Java and the optimization problems were solved with Gurobi 7.0.

The 80 benchmark instances, created by Archetti et al. (2007), are divided in 16 sets. The instances differ in three characteristics: high or low inventory holding costs, short or long planning horizon (3 or 6 periods) and number of retailers (ranging between 5 and 30 retailers). The results are always aggregated for 5 instances with the same characteristics.

In table 1 and table 2 the values of the linear relaxation of the original model (LP), the linear relaxation of the Dantzig-Wolfe decomposition (DWLP) and the integer problem solution (MIP) are reported. There is a significant increase in linear relaxation value when the problem is formulated using the Dantzig-Wolfe decomposition framework. The average gap between the linear relaxation of the original formulation and the integer problem solution value is 18.31% for the low inventory cost instances and 7.19% for the high inventory cost instances. When the Dantzig-Wolfe decomposition framework is applied, the MIP-gap decreases respectively to 5.43% and 2.29%.

Name	LP	DWLP	MIP
Short planning horizon instances			
abn05	699.86	725.14	744.89
abn10	1295.21	1526.74	1577.31
abn15	1430.71	1750.69	1840.07
abn20	1762.93	2121.47	2278.04
abn25	2052.60	2484.78	2578.75
abn30	2401.09	2838.51	2964.08
Long planning horizon instances			
abn05	2232.11	2366.59	2554.45
abn10	2957.50	3556.11	3978.71

Table 1: Comparison of the linear relaxations of the low inventory cost instances.

In table 3 and table 4 the performance of our column generation algorithm is evaluated. For small instances, short planning horizon and up to 15 retailers, the problem can be solved very quickly. For larger instances the computation time increases very

Name	LP	DWLP	MIP
Short planning horizon instances			
abn05	1610.55	1637.25	1660.27
abn10	3707.46	3942.17	3999.03
abn15	4643.30	4965.86	5054.50
abn20	6287.41	6652.37	6818.76
abn25	8007.44	8453.51	8557.89
abn30	9951.79	10385.19	10533.45
Long planning horizon instances			
abn05	4427.93	4566.68	4742.19
abn10	6926.71	7546.26	7940.05

Table 2: Comparison of the linear relaxations of the high inventory cost instances.

rapidly. There are two factors that explain this steep increase: the complexity of the pricing subproblem and the number of nodes in the branch-and-bound tree. For instances with more retailers the complexity of the pricing subproblem has a high influence on the computation time. This is also reflected by the performance of the column generation algorithm on the instances of the sets abs20 and abs25. Both for the low and high inventory cost instances the branch-and-bound tree explores on average less nodes for the instances of the set abs25 than for the instances of the set abs20, 84 and 184 versus 165 and 209. However, because the subproblem is more complex for the instances of the set abs25, the algorithm spends more time per node which results in a longer overall solution time.

For the long planning horizon instances the longer solution time can be explained by the large number of nodes that need to be explored. For these instances the gap between the linear relaxation and the integer solution value is larger than for the short planning horizon instances. Hence, more branching needs to happen before nodes can be pruned or before an integer solution is found. Another explanation for the large number of nodes in these instances is the fact that branching in these instances has less effect than for short planning horizon instances. If in a short planning horizon instance only one tour is made, then branching on a y_i^p variable decides whether the retailer is replenished via the supplier or via the subcontractor. In a long planning horizon instances, multiple tours are made so that this retailer can always be included in another tour.

6 CONCLUSION

In this paper, we investigated the mathematical formulation of the Inventory-Routing Problem with Transshipment. The problem was reformulated using the Dantzig-Wolfe decomposition framework and solved using column generation. In order to strengthen the mathematical model, valid inequalities

Name	Column Generation Algorithm		
	Time (s)	#Columns	#nodes
Short planning horizon instances			
abn05	0.05	67	5
abn10	0.34	181	13
abn15	3.44	397	51
abn20	27.41	1367	165
abn25	53.41	1212	84
abn30	438.55	4471	365
Long planning horizon instances			
abn05	4.18	152	391
abn10	693.08	1728	5666

Table 3: Performance of the Column Generation Algorithm for the low inventory cost instances.

Name	Column Generation Algorithm		
	Time (s)	#Columns	#nodes
Short planning horizon instances			
abn05	0.04	75	5
abn10	0.34	141	13
abn15	3.49	389	63
abn20	33.51	1620	209
abn25	100.84	2333	184
abn30	755.60	6899	661
Long planning horizon instances			
abn05	2.96	132	288
abn10	202.86	1438	2429

Table 4: Performance of the Column Generation Algorithm for the high inventory cost instances.

ities were added to the master problem. Furthermore, a greedy heuristic was developed to generate an initial set of columns. This initial set of columns was enriched by all the direct shipment columns. This strengthened the formulation of the pricing subproblem. Finally, an intelligent depth-first branching strategy was designed so that uninteresting branches can be pruned earlier in the optimization process.

We believe that this work opens up to a number of extensions. The experimental results showed that the column generation algorithms stops being competitive when many retailers are involved. Solving the pricing subproblem to optimality is a computational barrier when applying column generation to these instances. A effective heuristic which replaces the pricing subproblem would accelerate the solution procedure a lot and might lead to very high quality solutions.

Another important research direction is the investigation of the Dantzig-Wolfe linear relaxation. The results demonstrated that by reformulating the problem the linear relaxation lower bound increased considerably. The strength of this bound plays a major role in column generation. Further investigation can significantly reduce the number of nodes that need

to be explored in the branch-and-bound tree and improve the competitiveness of the column generation algorithm.

REFERENCES

- Archetti, C., Bertazzi, L., Laporte, G., and M. G. Speranza, 2007. A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation Science*, 41(3), p. 382-391.
- Chrysochoou, E., and A. Ziliaskopoulos, 2015. Stochastic Inventory Routing Problem with Transshipment Recourse Action. 10th Conference on Stochastic Models of Manufacturing and Service Operations (SMMSO'2015), Volos, Greece, p. 17-24.
- Coelho L. C., Cordeau, J. F., and G. Laporte, 2012. The inventory-routing problem with transshipment. *Computers and Operations Research*, 39(11), p. 2537-2548.
- Coelho, L. C., and G. Laporte, 2013. The exact solution of several classes of inventory-routing problems. *Computers and Operations Research*, 40(2), p. 558-565.
- Coelho, L. C., Cordeau, J. F., and G. Laporte, 2014. Heuristics for dynamic and stochastic inventory-routing. *Computers and Operations Research*, 52, p. 55-67.
- Desaulniers, G., Rakke, J. G., and L. C. Coelho, 2015. A branch-price-and-cut algorithm for the inventory-routing problem. *Transportation Science*, 50(3), p. 1060-1076.
- Errico, F., Desaulniers, G., Gendreau, M., Rei, W., and L. M. Rousseau, 2016. A priori optimization with recourse for the vehicle routing problem with hard time windows and stochastic service times. *European Journal of Operational Research*, 249(1), p. 55-66.
- Lefever W., Aghezzaf E. H., and K. Hadj-Hamou, 2017. An Improved Formulation for the Inventory-Routing Problem with Transshipment. Submitted to *Computers and Operations Research*
- Pisinger, D., and S. Ropke, 2007. A general heuristic for vehicle routing problems. *Computers and Operations Research*, 34(8), p. 2403-2435.
- Taş, D., Gendreau, M., Dellaert, N., Van Woensel, T., and A. G. De Kok, 2014. Vehicle routing with soft time windows and stochastic travel times: A column generation and branch-and-price solution approach. *European Journal of Operational Research*, 236(3), p. 789-799.

ANALYSING ECO-SYSTEM OF ‘-ILITIES’ FOR RESILIENCE EVALUATION IN SYSTEM OF SYSTEMS

A. Behrang MORADI
B. Nicolas DACLIN
C. Vincent CHAPURLAT

LGI2P - IMT Mines Ales, BP 30100
7, rue Jules Renard 30100 ALES, France

A. behrang.moradi@mines-ales.fr

B. nicolas.daclin@mines-ales.fr

C. vincent.chapurlat@mines-ales.fr

ABSTRACT: This paper addresses the analysis of the ecosystem of ‘-ilities’ to formalize and evaluate the resilience in the context of the System of Systems Engineering (SoSE). It presents the very first concepts to formalize and evaluate the resilience. On the one hand, the analysis of several metrics and indicators of resilience as well as the relationship between resilience and other non-functional requirements are examined. On the other hand, the relation between other ‘-ilities’ ecosystem and resilience are highlighted and mapped with the dynamic of the resilience allowing an accurate evaluation. A set of indicators is to be chosen to evaluate the resilience and particularly indicators that are associated with these ecosystem of ‘-ilities’ to each zone of the resilience life-cycle. The expected benefit of such method is to allow SoS resilience characterization by analyzing the ecosystem, in order to enable its evaluation.

KEYWORDS: System of systems engineering, Resilience, Non-functional property, ‘-ilities’ ecosystem.

1 INTRODUCTION

“A system of systems (SoS) is an integration of a finite number of constituent systems which are independent and operable, and which are networked together for a period of time to achieve a certain higher goal” (Gorod et al. 2008). According to Maier (Maier 1996), a SoS is composed of separate constituent systems and have five characteristics:

Operational Independence: If the SoS is disassembled into its component systems the component systems must be able to usefully operate independently.

Managerial Independence: The component systems are separately acquired and integrated but maintain a continuing operational existence independent of the SoS.

Evolutionary independence: SoS is not created once and for all, but evolves as constituent systems, or functions thereof, are added, removed, or modified.

Geographic distribution: The parties collaborating in a SoS are distributed over a large geographic extent.

Emergent behaviors: They are only exhibited at the SoS level and cannot be achieved by any of the constituent systems operating independently of the other constituent systems.

According to this definition, the formation of a SoS is not necessarily a permanent phenomenon, but rather a necessity for the integration and the networking in a centralized way for a specific objective. Many systems can be seen as SoS, e.g. air transport networks, Franco-German high speed train networks, smart grid system, electric power distribution networks. For instance, in a

smart grid (Figure 1), the power distribution system is independent of the provider unit system and is able to carry out its mission and it can distribute, share necessary information with other systems, as well as modify power plant in order to evolve and aggregate the desired global mission.



Figure 1: Smart grid system of systems¹

In that context, the SoS engineering (SoSE) requires having means to model, to simulate and to evaluate global behaviour of a SoS as well as a set of functional and non-functional requirements that a SoS must meet to fulfil its mission adequately. The problem addressed concerns the evaluation of non-functional requirements of SoS. It is about specifying and developing method for the evaluation of resilience, and applying this method

¹http://solutions.3m.com/wps/portal/3M/en_EU/SmartGrid/EU-Smart-Grid/

within the SoS. The proposed method aims to define mathematical model to verify if the resilience expectation is satisfied when a SoS is under design or existing. The paper is structured as follow. After this brief introduction, the problematic of this research work is pointed out. The concept of resilience is presented as well as different methods to evaluate this requirement. The next section presents the as the founding principles for the development of our approach to analyze and evaluate resilience. The final section presents the conclusion and the future perspective for this research.

2 PROBLEMATIC AND EXPECTED OUTCOMES

Among the non-functional requirements, a set is called ‘-ility’ (ISO 15288, 2011) and represents “*the desired properties of systems, such as flexibility or maintainability (usually but not always ending in ‘-ility’), that often manifest themselves after a system has been put to its initial use. These properties are not the primary functional requirements of a system’s performance, but typically concern wider system impacts with respect to time and stakeholders than are embodied in those primary functional requirements.*” (De Weck, 2012). For instance, we can mention the *flexibility, robustness, safety, interoperability* or else, *survivability*. Currently 74 to 83 ‘-ilities’ are identified in the literature. The network of ‘-ilities’ correlations schematized Figure 2 presents a set of identified ‘-ilities’ as well as their dependencies (De Weck, 2012). These dependencies show causal relationships but cannot detail the semantic nature of the link, which could include complement, substitution or other compromise between pairs of ‘-ilities’.

The proposed approach to evaluate the resilience intends to relies on the correlation network by studying and analyzing the relationships that can exists between ‘-ilities’. For instance the variation of given ‘-ilities’ that can result from the variation of another ‘-ilities’ from which it is linked. In that way, the criteria that define a given link must be defined. Thus, the here presented research work proposes to study and analyze the ‘-ilities’ ecosystem as a whole rather than to study a single ‘-ility’ in isolation and focuses on the resilience.

Resilience in the context of SoS is an important requirement because it must be mastered and maximized (Chin et al. 2013) in order to effectively cope with disruptive events and maintain acceptable levels of services and performance. It exist several definitions of resilience and usually, the resilience is defined as “*the ability of the system to resist, absorb, recover or adapt to disturbances and diminish the consequences as well as to recover quickly and effectively*” (Haines, 2009), (Hollnagel, 2013), (Royce, 2014). Lastly, resilience is practiced in various application area *e.g.* critical infrastructure monitoring, or security (Francis & Bekera 2014), (Pflanz, 2012), (Berkeley, 2010), (Andrews, 2014).

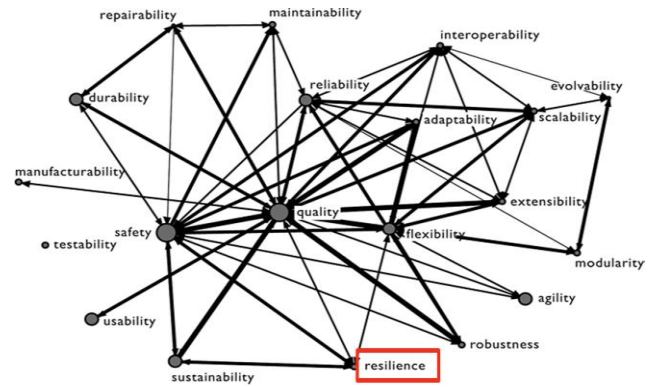


Figure 2: Correlation network of ‘-ilities’ (De Weck, 2012)

As a consequence this research work aims to define, specify and evaluate the resilience within a SoSE context. The prevention, preparation and management of negative perturbations and their effects within the framework of the security and protection of systems have become a major concern due to the increasing complexity of the SoS. The disturbance force may lead to heavy loss of performance and affect a SoS so that the measure and evaluation, of non-functional requirement such as resilience, is a part of the improvement of SoS’s performance and the mitigation of negative aftermaths.

3 RESILIENCE DEFINITION AND EVALUATION: A SURVEY

3.1 Definition of resilience

Several definitions describe resilience in the literature. (Hollnagel et al. 2012) Hollnagel define the resilience as “*the ability of a system or organization to react to and recover from disturbances at an early stage with minimal effect on its dynamic stability*”. (Scott, 2010) defines resilience as a combination of avoidance, survival, and recovery and considers brittleness to be the opposite of resilient. He defines resilience as the ability of a system to absorb a disruption without loss of capability. Both suggest that resilient systems are able to manage disastrous situation due to several capacities requested throughout the classical phases of the disaster management lifecycle: **Anticipation (Preparation, Prevention)**, **Response (Absorption, Adaptation)** and **Recovery** (Figure 3).

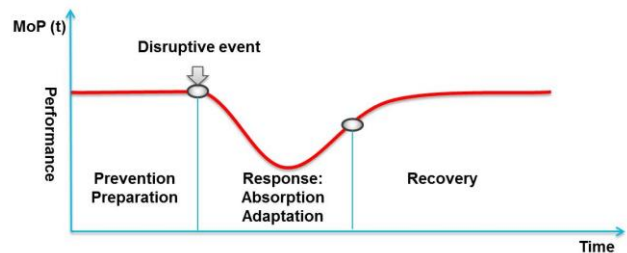


Figure 3: The classical resilience dynamic

Zone1-Anticipation (Prevention and preparedness) aims at identifying and minimizing the risks and prepara-

tion to face, its equipment and fittings, and the natural hazards of the environment (UNESCO 2017).

Zone2-Response (Absorption and Adaptation). The absorptive capability refers to an endogenous ability of a system to reduce the negative impacts caused by disruptive events and minimize consequences with less effort (Nan, 2017), (Royce, 2014). The capacity of a system to absorb negative effects can be improved during the prevention phase by increasing system’s **robustness**. The adaptive capability (**adaptability**) refers to an endogenous ability of the system to adapt to disruptive events through self-organization (**flexibility, interoperability**) to minimize consequences and it can be enhanced by using emergency systems.

Zone3-Recovery is defined as a return to a qualified acceptable condition. The *restorative* capability refers to an ability of the system to rapidly be repaired (**maintainability, repairability**) and returned to an as much as possible normal and reliable functioning mode that meets the requirements for an acceptable and desirable level of quality of service and expected control (Nan, 2017). The study and synthesis of different definitions shows that authors suppose that resilience can be characterized by zone such as *Anticipation, Recovery and Response*, each ones considering more or less zone in their study (e.g. Nogal considers only the resilience as a problem of response and recovery) (Figure 4). As part of our research work to assess resilience within SoS, we consider the three areas of the dynamic of resilience, as defined by (Cox, 2011).

Anticipation	Response	Recovery	Reference
✓	✓	✓	Cox, 2011
✗	✓	✓	Nogal, 2016
✗	✓	✓	Nan, 2017
✗	✓	✓	Royce, 2014
✗	✓	✓	Garguilo, 2016
✗	✗	✓	Devanandham, 2012

Figure 4: Resilience dynamic

3.2 Evaluation of resilience

Resilience can be assessed in different ways: by measuring the performance or loss of performance of a system before and after the disruptive event, the potential loss of functionality, the loss of quality of service, the effectiveness of the security barriers, or else, the activities recovery. Whatever the way used to measure resilience, the main goal is to have information about the strengths or weaknesses of a system to prioritize actions to improve ability of resilience. Numerous works to evaluate the resilience are available in the literature e.g. (Royce, 2014) and (Zobel, 2011) each with its own points of view.

(Royce, 2014) states the measurement of resilience is a function of the 3 capacities (absorption, adaptation and recovery) as well as recovery time, through the meas-

urement of the performance and its evolution. His method is based on a resilience analysis framework and a metric for measuring resilience. His analysis framework consists of system identification, resilience objective setting, vulnerability analysis, and stakeholder engagement. (Zobel, 2010) provides a quantitative measure of resilience in the face of multiple disaster-related events. He introduces the concept of the resilience triangle, which incorporates measures of both the robustness against initial loss due to a disaster and the rapidity of the recovery process. He extends the concepts of the resilience triangle and predictive resilience in disaster by considering the trade-offs between several criteria for each event (multi events), as well as for any multi-events situation. (Chan & Fekri 2008) developed the Resiliency-Connectivity metric that measures the resilience of communication networks in the face of adversarial influences. Their measure takes into account the probability of link compromise and the probability of overall network connectivity. Their work is based on sudden disasters and the initial impact of each event as well as the recovery time of the system before the next event occurs. A mathematical model is developed for the measure of resilience. Robustness is used to increase the level of SoS resilience so that it recovers its performance and returns to an acceptable state (Figure 5), thus, the link between resilience and robustness is clearly exhibited in this measure.

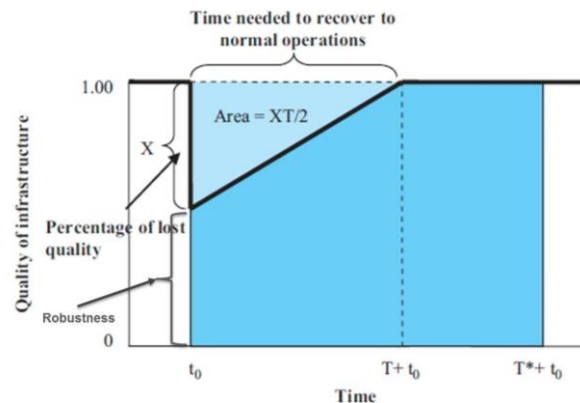


Figure 5: The predicted resilience triangle

The works presented above shows different approaches to assess resilience (Figure 6). Among these works, it is to note that Zobel highlights the link between resilience and another ‘-ility’ (robustness in the Zone2-response and rapidity in zone3- recovery). Its measurement shows the importance of the consideration of the link between ‘-ilities’ since the more the robustness is high, the more the resilience can be efficient in term of response. Thus, the ecosystem of ‘-ilities’ and their relationship can be used and study to evaluate a given ‘-ility’.

Reference	Ecosystem of '-ilities'	Method/approach	Zone
Royce, 2014	✘	Resilience factor	Response, Recovery
Devanandham, 2011	✘	Potential losses	Recovery
Zobel, 2011	✔ (Robustness)	Loss of features	Response, Recovery

Figure 6: Synthesis of different resilience assessment methods

4 APPROACH PROPOSAL

4.1 Means of analysis of resilience and '-ilities' ecosystem

We consider the representation of the assessment of resilience by defining, analyzing and formalizing the '-ilities' ecosystem (Figure 7).

In this proposition, we focus on the analysis of the resilience and its environment and highlight the various components to evaluate the resilience. Using the ecosystem of '-ilities' in terms of evaluation and measuring the level of resilience, allows to treat resilience in a non-isolated way. The four components considered in our work are: *influence* (1), *orientation* (2), *dependence* (3) and *propagation* (4).

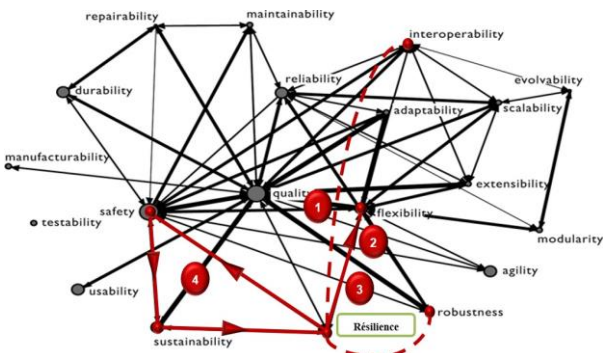


Figure 7: Correlation network of '-ilities' based on (De Weck, 2012)

4.1.1 The influence

This component aims to identify the influence between resilience and other '-ilities'. The influence is defined as *an action, exercised by one '-ility' onto another*. In other words the influence characterizes a change of a given -ility and coming from another '-ility'.

Influence is about the identification of the link between '-ilities' belonging to the network. Thus the knowledge of the influence is based on the construction of a network that consider a given '-ility' under study and links with other '-ilities' that can modify it. Some influences are identified and defined (De Weck, 2012) (Figure 7). However, as claimed (De Weck, 2012) some might exist but are not yet identified. The reason is that some '-ilities' are quite recent and other ones are not enough studied to identify possible or new links in the network. For instance, interoperability influences the resilience

but it is not highlighted in the graph. Indeed, Interoperability is *the ability of entities to communicate and interact effectively* (ISO/TC 184/SC 5 2011) so, it influences resilience since, interactions between systems to share information can support the system to adapt or mitigate to a disruptive event. In order to assess resilience based on the study resilience, it is necessary to identify all influences between resilience and other '-ilities'. The knowledge of influence will be used, further, to precise if this one is negative or positive. This aspect is important since if an influence is not identified the future defined metric and further, the evaluation will be incorrect. Currently, others influences (with resilience) than those presented in the primary graph are identified (e.g. interoperability, robustness). This identification relies mainly on the analysis of the literature.

4.1.2 The orientation

The orientation defines whether the influences is unidirectional (ex. quality *influences* resilience) or bidirectional (ex. quality *influences* resilience and vice versa).

Precisely, the orientation allows to know the direction of the changing, that means by which '-ilities' a given one is influenced and which ones it influences itself. This orientation must be considered "from" resilience "to" another '-ilities' as well as "from" another '-ilities' to resilience. The objective is to identify the ecosystems of '-ilities' that increase (but also decrease) the resilience of a SoS. Defining and identifying first, all the influences with resilience and their orientation, allows to get a network of '-ilities' acting as a basis to assess resilience further. Figure 8 shows the identified ecosystems of '-ilities' that influence resilience (e.g. interoperability, sustainability, reparability...) as well as the orientation of these influences. For instance, there is a dual direction between resilience and safety (safety ↔ resilience), meaning that by decreasing or increasing the level of safety, the resilience is impacted and vice versa.

Let's note that the two first components influence and orientation are quite generic to be applied to each '-ilities' of the graph that could be studied deeper.



Figure 8: Graph of influences and orientations for resilience assessment

4.1.3 The dependence

The dependence characterizes is the level of a studied '-ility' regarding to the level of other '-ility' by which it is influenced. The dependence expresses the operability of a given '-ility' (namely targeted ility and noted *IOp*, for *Ilities Operability*), depending on criteria characterizing

another ‘-ility’ (namely source ‘-ility’) from which it is influenced. The operability is the ability to achieve an ‘-ility’. **Operability** is a state where an ‘-ility’ is functioning at some level (quantitative or qualitative).

The dependence relies on three aspects namely the **importance**, the **incidence** and the **state**.

The importance (noted α_s), is defined as a value that expresses the participation of the operability of the source ‘-ility’ to the operability of the target ‘-ility’. The importance scale goes from *low* to *high* and via *null*. For instance, the robustness has an importance ranked to high since the operability of the robustness influences highly the operability of the resilience.

The incidence (noted β_s) is defined as the direction of operability of the source ‘-ility’. The incidence is characterized by two directions such as *positive* and *negative*. For instance, the incidence of robustness is positive that means in the current state it is considered as good.

The state (noted γ_s) is defined as the current level of the source ‘-ility’. This level can be obtained by classical metrics or any means used usually to evaluate an ‘-ility’. For instance, the current state of robustness is high (and positive as well). In another state it could be considered as low (but always positive).

The analysis of the dependence of ecosystem of ‘-ilities’ finds its founding principles, in this work, in the methodology of the analysis of the functional dependency network (FDNA) (Paul R. Garvey & C Ariel Piento 2010). The proposed approach integrates the methodology (FDNA) on the graph of influence in order to formalize and formulate the dependence of the ‘-ilities’ network. The proposed influence graph allows a visual representation of complex dependencies between ‘-ilities’ (source, target) on the one hand and, the design of analytic formalisms that measure and trace dependencies between ‘-ilities’ on the other hand. As a consequence, we assert the ‘-ilities’ ecosystem can be represented as a directed graph whose entities are ‘-ilities’ that depict the **importance**, the **incidence** and the **state** of source – target relationships.

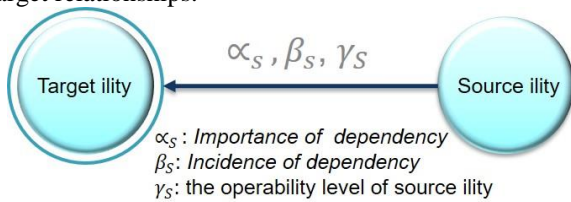


Figure 9: Graph of Dependency between a source and target ility

Thus, the evaluation of a given ility (Iop_i) is a function depending on the parameters defined above (γ_s , α_{st} and β_{st}) and can be noted:

$$Iop_T = f(\alpha_S, \beta_S, \gamma_S) \quad (1)$$

Where :

Iop_T : the operability level of target ‘-ility’

α_S : the importance of dependency

β_S : the incidence of dependency

γ_S : the operability level of source ‘-ility’

Lastly since, a target ‘-ility’ can depend on several source ‘-ility’, the operability level can be expressed by the general function such as:

$$Iop_T = f(Iop_{s1}, \dots, Iop_{sn}) \quad (2)$$

$$Iop_T = F(f(\alpha_{s1T}, \beta_{s1T}, \gamma_{s1T}) \dots f(\alpha_{snT}, \beta_{snT}, \gamma_{snT})) \quad (3)$$

Iop_T : the operability level of target ‘-ility’

α_{snT} : the importance of dependence

β_{snT} : the incidences of dependency

γ_{sn} : the incidence of operability level of sources ‘-ility’

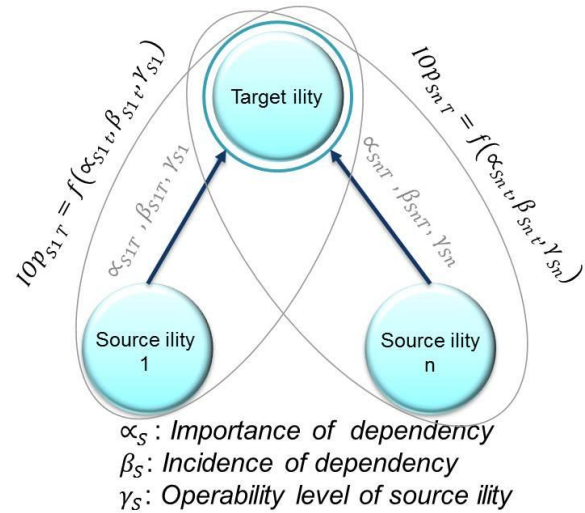


Figure 10: Model dependency relationships ‘-ilities’ ecosystem

Iop_T is obtained by, first, calculating the resulting operability of the target depending on each source Iop_{st} . Once each intermediate operability levels are obtained, they are aggregated to get the final operability level of the target Iop_T . At this stage, we focus on calculating the resulting operabilities from each ‘-ility’ source on the ‘-ility’ target and, the further step will be to define the function allowing to aggregate each operability to get the final operability of the target.

4.1.4 The propagation

The propagation represents a relation “starting from” an ‘-ility’ and returning to this ‘-ility’ *via* another ‘-ilities’ (e.g. resilience → safety → sustainability → resilience). At this stage, we consider and limit the propagation study to a chain of 3 ‘-ilities’. In this section, we present the ‘-ilities’ that have links with the resilience and thus we analyze their relations by practicing the components defined above (influence, orientation, dependence). For instance, figure 11 show the *sustainability* influences the *resilience*, the *resilience* influences the *safety* and finally, the *safety* influences the *sustainability*. That means that a variation in sustainability leads and propagates to a variation in resilience and a variation in resilience leads to a variation in safety and finally leads to variation in sustainability.

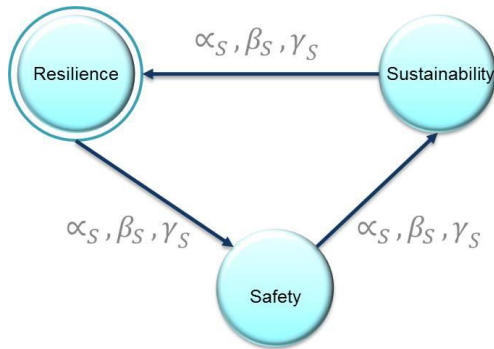


Figure 11: ‘-ilities’ ecosystem Graph Models Dependence Relationships

The problematic inherent to this chain (and more widely to the propagation) is to find and start from an initial -ility that allows to analyse the chain. Then the analysis can start relying on the principles of dependence as defined previously. From this stage, the modifications are propagated through the chain until to back on the initial ility. In that case, the obtained results for each -ility belonging to the chain must be analyzed. In this current state, our research attempts to develop mechanisms to analyse the chain of -ility.

Figure 12 summarizes the relation between resilience and sustainability in agreements with the four components for the relation analysis and evaluation. In this end, each relationship between resilience and other ‘-ilities’ (which impact resilience) have to be identified and formalized. In the same way the components such as dependence and propagation must formally established in order to allow the evaluation of the resilience (e.g. quantification or qualification of “high increase” or else definition of effect in the propagation chain).

Influence	Orientation	Dependance	Propagation
Resilience-Sustainability	S → R	Strength Criticality (α, β, γ)	

Figure 12: Characterization of the relation between resilience and sustainability

4.2 Resilience evaluation according to its dynamic

The resilience life-cycle (Figure 3) (*anticipation, response and recovery*) must be considered. Indeed, the aim is to know the level of resilience for each considered zone in order to implement adapted solution to improve resilience if a specific zone is failing. To this purpose, ‘-ilities’ ecosystem and their influences are mapped in each zone of the dynamic of resilience (e.g. flexibility can be expected during the prevention, Figure 13). This positioning is relative to the analysis of resilience and ‘-ilities’ ecosystem that have an identified links in each area. The objective is to get a resilience indicator for each zone in order to identify precisely a problem of resilience. In the end, the aggregation of each indicator will provide an evaluation of the resilience of the SoS.

Figure 13 shows the mapping of all the ‘-ilities’ that is identified and positioned in different phases of resilience life-cycle, sometimes in various phases (e.g. flexibility is positioned in zone 2, zone 3 and robustness in zone 1, zone 2, zone 3). Mapping makes it possible to establish the resilience indicators associated with these ‘-ilities’ and for each zone of belonging. These indicators ultimately enable the level of resilience to be accurately identified and assessed, so aggregation of all these indicators will provide an assessment of the overall resilience of the SoS. The novelty of this method is the analysis of resilience in the network of ‘-ilities’, so as to analyse globally the ‘-ilities’ that influence resilience without having to study them in isolation. All ‘-ilities’ that influence resilience have been identified. In order to assess resilience in the ‘-ilities’ ecosystem without having to consider all the ‘-ilities’ (it may not be all relevant depending on the system), the user will choose the ‘-ilities’ that have the most impact and important on the system. The choice of ‘-ilities’ to consider as a priority must be based on the life cycle of ‘-ility’.

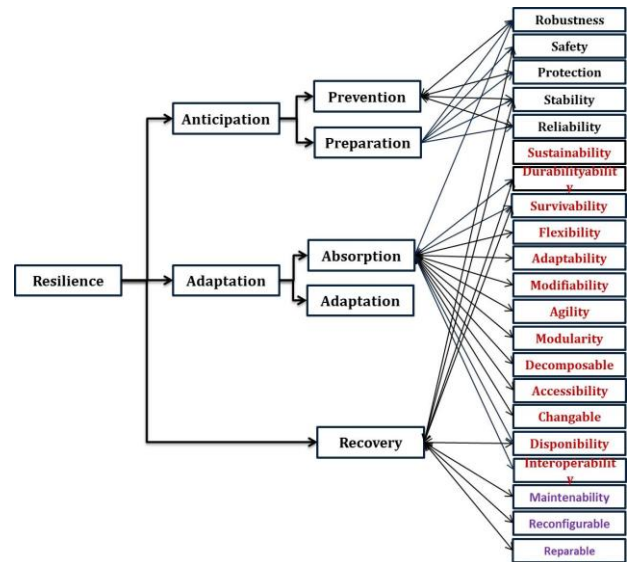


Figure 13: Resilience dynamic and ‘-Iilities’ Mapping

5 CONCLUSION

The very first development of the here presented work aims, in the end, to evaluate the resilience in the context of SoSE relying on the dynamic of the resilience and on the study of the resilience as an ‘-ilities’ belonging to a wider set of ‘-ilities’. On the one hand, several resilience metrics are studied and analyzed as well as the relationship between resilience and the other ‘-ilities’ ecosystem. On the other hand the ‘-ilities’ ecosystem and their dependencies (influences) are mapped by positioning them according to the dynamic of resilience. In order to assess resilience, we will assign an indicator associated to each zone of resilience. We presented a graph model of ‘-ilities’ ecosystem that influence resilience and the main components are identified and highlighted and will be developed in the next steps in order to evaluate resilience. Future work is related to the formalization and

development of indicators to evaluate resilience as well as mechanisms to study the propagation in the network.

REFERENCES

- Andrews, Z. et al., Fault Modelling in System-of-Systems Contracts. Available at: <https://arxiv.org/pdf/1404.7775.pdf> [Accessed November 8, 2017].
- Berkeley Iii, A.R. & Wallace, M., 2010. National Infrastructure Advisory Council A Framework for Establishing Critical Infrastructure Resilience Goals Final Report and Recommendations by the Council. Available at: <https://www.dhs.gov/xlibrary/assets/niac/niac-a-framework-for-establishing-critical-infrastructure-resilience-goals-2010-10-19.pdf> [Accessed November 8, 2017].
- Chan, K. & Fekri, F., 2008. A Resiliency-Connectivity metric in wireless sensor networks with key predistribution schemes and node compromise attacks. *Physical Communication*, 1(2), pp.134–145. Available at: www.elsevier.com/locate/phycom [Accessed January 9, 2018].
- Chin, K.S. et al., 2013. FRAMEWORK FOR MANAGING SYSTEM-OF-SYSTEMS ILITIES. , (14), pp.56–65. Available at: https://www.dsta.gov.sg/docs/default-source/dsta-about/framework_for_managing_system_of_systems_ilities.pdf?sfvrsn=2 [Accessed June 19, 2017].
- Francis, R. & Bekera, B., 2014. A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering and System Safety*.
- Gorod, A., Sauser, B. & Boardman, J., 2008. System-of-Systems Engineering Management: A Review of Modern History and a Path Forward-of-system engineering (SoSE) man- agement conceptual areas, system-of-systems engineering management framework, system-of-system engineering management matrix (SoSEMM). *IEEE SYSTEMS JOURNAL*, 2(4). Available at: http://boardmansauser.com/Systemics/Publications_files/IEEESJGorodSauserBoardman.pdf [Accessed September 14, 2017].
- Haimes, Y.Y., 2009. On the Definition of Resilience in Systems. *Risk Analysis*, 29(4), pp.498–501. Available at: <http://doi.wiley.com/10.1111/j.1539-6924.2009.01216.x> [Accessed November 8, 2017].
- Hollnagel, E., 2013. Resilience engineering in practice: A guidebook. Available at: <https://books.google.fr/books?hl=fr&lr=&id=1YHXCAAQBAJ&oi=fnd&pg=PR13&dq=resilience+engineering+in+practice+a+guide+book+eric+hollnagel&ots=NAUjk8E4m&sig=w9VqUwZRzRSEpd5htmsejRLeQ> [Accessed May 11, 2017].
- Hollnagel, E., Woods, D.D. & Leveson, N., 2012. *Resilience engineering: Concepts and precepts*, ISO/TC 184/SC 5, 2011. *Advanced automation technologies and their applications — Part 1: Framework for enterprise interoperability*, Maier, M.W., 1996. Architecting Principles for Systems-of-Systems. *INCOSE International Symposium*, 6(1), pp.565–573. Available at: <http://doi.wiley.com/10.1002/j.2334-5837.1996.tb02054.x> [Accessed May 9, 2017].
- Nan, C. & Sansavini, G., 2017. A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering and System Safety*.
- Pflanz, M. & Levis, A., 2012. An approach to evaluating resilience in command and control architectures. In *Procedia Computer Science*.
- Scott, J., 2010. The Principles of Infrastructure Resilience. *Domestic Preparedness*, pp.1–4. Available at: https://www.researchgate.net/profile/Scott_Jackson7/publication/255993485_Principles_of_Infrastructure_Resilience/links/5564779708ae9963a12020e7.pdf [Accessed May 3, 2017].
- UNESCO, 2017. Disaster Planning: prevention, preparedness, response, recovery. , p.8.
- De Weck, O.L., Ross, A.M. & Rhodes, D.H., 2012. Investigating Relationships and Semantic Sets amongst System Lifecycle Properties (Ilities). *Third International Engineering Systems Symposium CESUN 2012, Delft University of Technology, 18-20 June 2012*, (June), pp.18–20.
- Zobel, C.W., 2011. Representing perceived tradeoffs in defining disaster resilience. *Decision Support Systems*, 50(2), pp.394–403. Available at: http://ac.els-cdn.com/S0167923610001764/1-s2.0-S0167923610001764-main.pdf?_tid=c0148934-664b-11e7-8356-00000aacb361&acdnt=1499786299_520c3614ace1805f1175e96a94442c3a [Accessed July 11, 2017].

IRONAPPLICATION OF SIMULATION AND LEAN TECHNIQUES FOR IMPROVEMENT OF GLULAM PRODUCTION: A CASE STUDY

M. Ghiyasinab, N. Lehoux, C. Cloutier

Department of mechanical engineering, Laval University
 Québec (QC)
 Marzieh.Ghiyasinab.1@ulaval.ca,
 Nadia.Lehoux@gmc.ulaval.ca,
 Caroline.Cloutier@gmc.ulaval.ca

S. Ménard

Department of Applied Sciences, Civil Engineering,
 University of Québec at Chicoutimi, Saguenay (Québec)
 Sylvain_Menard@uqac.ca

ABSTRACT:

Modular construction manufacturing helps to improve standardization and increase productivity in the construction industry. Glued laminated timber (glulam) is a type of engineered wood which can be used in modular construction as an environment-friendly product. This article considers improvements in the production of curved glulam by investigating the sources of waste and non-value adding activities in the production processes. A simulation model is built for glulam production and validated by a case-study for a gridshell project. Sources of waste are identified, and lean methods are suggested for improvement. The lean solutions are tested in the simulation model and results demonstrate 27% improvement in cycle time and 77% improvement in wait time. Additionally, the impact of applying only 50% elimination of non-value adding activities is compared with 100% elimination.

KEYWORDS: *Lean approach, simulation, cycle time, construction industry, glulam*

1 INTRODUCTION

Several manufacturing sectors have profited from the application of lean concepts to increase productivity. Following manufacturers, the construction industry also implemented and adopted lean concepts to reduce wastes and increase productivity. However, productivity in the construction industry has improved more slowly than in the manufacturing sectors. According to the McKinsey Global Institute report, global labour-productivity in construction has grown by only 1 percent per year over the past two decades, contrasted with the growth of 2.8 percent in the world economy and 3.6 percent in manufacturing (McKinsey Global Institute, 2017).

One of the reasons for lower productivity in the construction industry comes from the nature of the projects themselves and lack of standardization since every project is different. According to the McKinsey report, a drastic increase in productivity would be possible if construction was to tend towards a manufacturing-like system with a much higher degree of standardization and modularization and a greater share of off-site production instead of on-site construction. Modular construction manufacturing (MCM) has a transition role from on-site construction projects to off-site production. MCM consists of three main stages: panel prefabrication, production line in the factory, and on-site module installation (Moghadam *et al.* 2012a).

Engineered wood products seem to be an appropriate solution for MCM. Prefabricated engineered wood products, such as glued laminated timber (glulam) can be

used in residential and nonresidential construction projects. Some companies manufacture these products in high volume, while there are companies who manufacture engineered wood products for specific construction projects. They consider mass customization instead of mass production, as the sizes and characteristics of the product will vary for every construction project. However, making improvements in this area helps to increase the productivity while having a greater control over the quality produced in the construction industry.

Current environmental awareness increases the motivation to use wood as a construction material (Kuzman and Sandberg, 2016) since it is an environment-friendly material which reduces CO₂ emissions. Engineered wood products can be used for many types of structure because of their physical properties and aesthetic appearance (Structural timber association, 2014). Structural glulam is one of the oldest engineered wood products and is still very competitive in modern construction. Glulam is fabricated from wood boards which are glued together so that they form a beam cross-section of the desired shape (Malo and Angst, 2008).

In this research, manufacturing of glulam is analysed and simulated. The simulation model is validated with a case-study of curved glulam production for a gridshell project conducted in a small and medium-sized enterprise (SME). For this firm, the seven sources of waste according to lean principles are analysed, and the elimination of three of them is targeted (non-value adding (NVA) activities, transportation, and defects). Lean tools such as single minute exchange of dies (SMED), U-shaped layout, total

manufacturing and assembly (DFMA) are the suggested improvements tested with the simulation. The impact of completely eliminating NVA activities (100%) versus partial reduction (50%) is also verified. The results show that partial elimination of NVA activities are worthwhile.

This paper considers standardization in the construction area by focusing on glulam production. Previous studies which performed a simulation methodology has been applied to a variety of production systems, however, to our knowledge, production of engineering wood such as glulam has not been considered in this type of study. The result of this research provides a simulation model for glulam production in SMEs which can be adapted to different cases. Simulation results may also be seen as a decision-making tool for the managers who can preview the impacts of the proposed changes. Another contribution concerns the identification and the prioritization of sources of waste in production of prefabricated parts for construction, considering the specific characteristics of this system. Developing a step by step approach that starts from feasible and less expensive improvements could motivate the managers to be aware of the potential results and to implement lean approach.

In the next section of this article, a literature review of lean concepts and the simulation approach are presented. Then a simulation model of glulam production is introduced. Subsequently, the case study and adopted model are described. Then the results and discussion are presented, followed by conclusions.

2 LITERATURE REVIEW

Lean manufacturing techniques gained great attention when the International Motor Vehicle Program (IMVP) at the Massachusetts Institute of Technology (MIT) published its findings in 1988. Lean manufacturing practices aim to reduce all forms of waste and inefficiency from production flow towards achieving efficient and flexible systems (Womack and Jones, 1996). Since then, the lean theory has been adopted widely in manufacturing industries. Implementation of lean techniques in the construction industry, which is called lean construction, did not grow as quickly as lean manufacturing. The reason being that construction is done on-site and that each project is constructed in a different location (Solomon, 2004). According to (Howell, 1999), lean construction results from the application of a new form of production management to construction. Koskela (1992) states that lean construction shares the goals of lean production: elimination of waste, cycle time reduction, and variability reduction. One of the top priorities in the lean approach is the elimination of wastes and the activities which consume the time of workers and other resources, without generating value for the final product. These kinds of activities are called non-value-adding activities in terms of the lean thinking theory (Nikakhtar *et al.*, 2015).

Computer simulation is defined as the process of making a mathematically and logically explained model of a real-world system (Farrar *et al.*, 2004). AbouRizk (2010) mentioned four phases for building a simulation model: product abstraction phase, process abstraction and modelling phase, experimentation phase, and decision-making phase. Simulation can be used for analysing dynamic systems even where there is uncertainty, and the initial model can be used to test different scenarios for improvement. Ferrin *et al.* (2005) denoted the advantages of simulation such as identifying the problems and documentation. Computer simulation provides an excellent environment to implement the principles of lean production (AbouRizk, 2010; Abedi Saidabad and Taghizadeh, 2015).

The literature contains research on using simulation for implementing lean in a variety of sectors and industries. Although most of the cases are in mass production and large-sized companies, some research show how the lean approach can be applied in SMEs (Mahfouz *et al.* 2011; Faisal, 2016), job-shops (Li, 2003) or the process industry (Abdulmalek and Rajgopal, 2007). Mahfouz *et al.* (2011) developed a simulation-based optimization model to improve lean parameters in a packaging manufacturing SME. They measured three key performance factors, cycle time, WIP (work in process), and workforce utilization, after applying lean tools such as facility layout and preventive maintenance. Faisal (2016) examined the implementation of lean manufacturing using value stream mapping (VSM) for SMEs in the leather industry. Implementing the pull system showed improvements in the average throughput and decrease in WIP. Li (2003) carried out a simulation experiment to compare the effects of applying the just-in-time concept to the performances of push and pull systems. Abdulmalek and Rajgopal (2007) adapted lean principles for the process sector using VSM and prepared a simulation model to illustrate potential benefits of reducing lead-time and WIP. They followed lean methods such as the pull system, setup reduction and total productive maintenance (TPM).

Dotoli *et al.* (2015) presented a lean warehousing approach for the strategic design and re-engineering of production warehouses. Their proposed approach integrates the Unified Modeling Language (UML), Value Stream Mapping, and Genba Shikumi methodologies. Genba is a philosophy that recommends to spend more time in the production line while helping in identifying sources of waste. Alvarez *et al.* (2009) studied the use VMS as a tool in lean manufacturing implementation, in particular for an efficient introduction of kanban and milkrun techniques. Dotoli *et al.* (2014) presented a novel hierarchical and iterative technique for the analysis and management of the internal logistics of manufacturing systems. The method effectively integrates the value stream mapping (VSM) tool, the analytic hierarchy process (AHP) approach, and discrete event simulation.

Parthanadee and Buddhakulsomsiri (2014) used VSM and simulation to improve the efficiency of the batch production system commonly found in SMEs. VSM was used to identify the problems and finding solutions, and simulation was used to analyse the result of applying improvement suggestions. Abedi Saidabad and Taghizadeh (2015) evaluated the function of production line through computer simulation and made improvements using the principles of production line such as concurrency of the operations and the least distance. The results showed improvements in waiting time, cycle time, and efficiency in a case study of an iron foundry. Tokola *et al.* (2015) studied lean manufacturing methods and how simulation is used to consider them. They analysed 24 articles which addressed lean and manufacturing together. 46% of the papers used VSM to analyse the system and extract NVA activities. Kanban (38%), layout (29%), pull (25%), and WIP (25%) are the other tools that were mostly applied. The effects of takt time (21%) and SMED (17%) are easy to model as well. Reductions in WIP, lead time, labour, and floor space were frequent results in the simulations. Nagi *et al.* (2017) used a pull simulation model in a multiproduct assembly line. They applied line balancing and controlling the WIP and developed a design of experiments (DOE) method to be tested by the simulation model. The result showed 14% improvement in throughput rate. Zarrin and Azadeh (2017) simulated a manufacturing organization with maintenance strategy to evaluate the impacts of resilience engineering principles on lean practices. The results showed that redundancy among the resilience engineering principles has the most impact on system performance.

As for lean construction, Farrar *et al.* (2004) presented a systematic approach for the application of lean theory in computer simulation models and implemented it for a case of road construction. The improvements in hourly production rate, resource utilization, and project duration resulted through the elimination of NVA activities and from the implementation of a methodology of pulling material through the process. Nikakhtar *et al.* (2015) demonstrated improvements in cycle time and process efficiency by using multi-skilled teams and sensitivity analysis with simulation for the concrete pouring process. Abbasian-Hosseini *et al.* (2014) applied lean and simulation to determine the best combination of resources in a case study for the bricklaying process. Identifying NVA activities and following lean methods such as JIT and pull system lead to a 40% improvement in productivity.

Yu *et al.* (2013) and Moghadam (2012b) worked on applying simulation for lean based improvements in manufacturing for modular construction. Ritter *et al.* (2017) mentioned that considerable effort is needed for production line flow balancing because of a large amount of variation and customization in the home construction industry. They used simulation to evaluate the state of production in a case study and analysed the result of

improvements such as labour reallocation which showed an increase in production rate.

Identification of seven sources of waste and finding solutions to eliminate them are mainly considered in the methodologies of implementing lean manufacturing. However, the priority and importance of the sources of waste, and how to eliminate them, are different based on the nature of each industry. This research describes the seven sources of waste in the production of prefabricated parts for construction. As a result, this paper investigates, through simulation, the implementation of lean concepts for production of prefabricated parts for construction that shares the characteristics of both manufacturing and construction industries, which does not seem to have been considered in previous studies.

3 METHODOLOGY AND SIMULATION MODEL

With the aim of implementing lean approach using simulation for a case study, this research follows a comparable methodology to the other studies that use a combination of simulation and lean approach. In this study, cycle time and wait time were used as key performance factors. Cycle time is considered in the work of Mahfouz *et al.* (2011) in packaging manufacturing, in Abedi Saidabad and Taghizadeh (2015) for the iron factory, in Farrar *et al.* (2004) for a construction project, and in Nikakhtar *et al.* (2015) for concrete pouring. Some researchers applied VSM to analyse NVA activities (see for example Abbasian-Hosseini *et al.* (2014)). This study follows the same method to identify NVA activities. Pull system and JIT are lean tools regularly used in the literature. In the system of glulam production for a construction project, the production is based on the customer demand. Therefore, pull system was the one considered. Abdulmalek and Rajgopal (2007) considered set-up reduction and TPM which are also used in this study as they are appropriate tools in the studied system. A general simulation model for glulam production was created. It was developed gradually by adding details and verification was tested in each step. The model was adapted to a case-study of production of curved glulam for a gridshell project. The validation of the model was tested by the confirmation of experts in the production line and hypothesis testing. Based on the production line and the simulation model analysis, the sources of waste were identified and prioritized for this production system and solutions suggested using lean tools. Lean tools were considered based on the needs of the factory as a solution to eliminate the sources of waste and NVA activities. The characteristics of the production line and feasibility of the suggested solutions were also considered. Comparing to other studies, the approach used in this study analysed the result of each improvement separately as well the result of implementing all improvements simultaneously. In this way, it was possible to have an overview of the benefits of applying lean tools in each step. Moreover, since completely elimination of waste may be costly and difficult for an SME at the beginning of implementing

lean techniques, the impact of applying only 50% elimination of NVA activities was compared with 100% elimination.

3.1 Creation of the model

For the first step, a simulation model of glulam production was developed in Arena 15 based on the general production of glulam and data gathering from an SME in the province of Quebec, Canada. A typical glulam manufacturing process consists of lumber drying (when purchasing green lumber), grading, trimming, finger jointing or end jointing, planing, face bonding, finishing, and fabrication (Puettmann and Wilson, 2005). Since the suppliers provide dried wood, usually there is no need for drying in SMEs. 100% of the raw material is inspected to check for the humidity level, elasticity and visual defects and knots. According to the level of quality, wood layers are graded in different degrees such as B, C, D where B is the best quality level. One of the advantages of glulam is the possibility of combining the lower qualities of lumber with the higher quality ones. The method consists of placing high-quality laminations in the outer parts of the cross-section where stresses normally are highest, and lower quality laminations in the inner zones to make combined glulam (Malo and Angst, 2008).

The simulation model was built, using the basic and advanced process modules for production flow and advanced transfer modules for transportation between stations. Based on the available data, the simulated process times were defined as distributions, listed in appendix 1. In particular, the triangular distribution was used for processes, as a minimum, maximum, and most likely values were available. Process time of pressing was assumed as a constant distribution based on the need for spending the accurate time of pressing. The arrival of raw material was simulated based on the schedule of everyday work in the factory. Arriving wood is the main entity and the dimensions and degrees are assigned to it using an assign module. The entity picture changes for each degree of quality to facilitate visualisation. Figure 1 illustrates the conceptual model of the simulation and the main process modules.

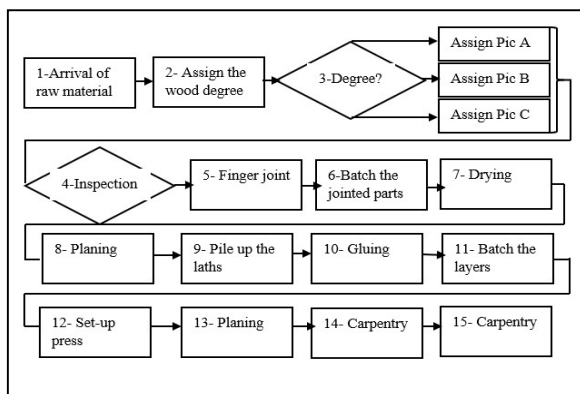


Figure 1: Conceptual model for simulation of glulam production

3.2 Verification of the model

Model verification can be briefly described as “building the model correctly” (Chung, 2004). In order to avoid errors and problems, it is important to start with a simpler model and improve it gradually. This means breaking up the complete model into a simpler and smaller model and then adding to it, which is easier to debug. In programming, this method is called the divide-and-conquer approach. In each step, the model is run, and errors and syntaxes are checked, and more details will be added to the next steps (Chung, 2004). This method was applied to develop a valid model. After building the model, some tools and modules were used to visualize and check how the model works. These tools include: using different entity pictures for different types of entity, following entities, changing entity pictures, changing resource pictures, displaying values and plots and writing to an output file.

4 CASE STUDY

4.1 Adapting the simulation model to the case-study

The case study concerns an SME which is active in the architectural design of buildings, production of glulam for the projects, and construction of those projects. Three different shapes of glulam are manufactured in this factory which are straight beams, round beams, and curved beams (arcs). The production of glulam for a particular construction project was observed and the simulation model was applied to this project. The project is a gridshell structure and all of the parts are curved beams. Timber gridshells are free-formed structures which are defined as structures "with the shape and strength of a double curvature shell but made of a grid instead of a solid surface" (Douthe *et al.* 2006). Glulam seems to be an appropriate material for these modern and free-formed structures because of its aesthetic appearance, lightness, and its flexible size and shape. The production of curved glulam for this project was observed and the simulation model was updated based on the data gathering and analysis. Figure 2 shows a screenshot picture of the simulation model for this gridshell project.

4.2 Validation of the model

To analyse the results of a simulation model, a single run of the model is not sufficient (Hassan and Gruber, 2008). Hassan and Gruber (2008) validated their model based on ten replications and Abbasian-Hosseini *et al.* (2014) validated their model using 20 replications. In this research, the model is based on one real project and was replicated 10 times. The results are observed, and the average cycle time is compared with the cycle time for the real project. The comparison is shown in table 1 (times are in hours).

production processes. In this case, unorganized workplaces and lack of standard work charts are an important source of motion. 5S, visual control, and optimization of space are lean methods that could reduce the motion.

5.1.4 Waiting

In this case, waiting does not add value to the product. Therefore, wait time is considered as a performance factor which should be reduced.

5.1.5 Overprocessing (Non-value adding activities)

Overprocessing refers to the activities which do not add value to the product. In this case, inspection and set-up time for the press are sources of waiting. Pile-up is another process that does not add value to the product. These three processes and the suggestion to eliminate them are discussed in the following.

Inspection

In this case study, one operator is assigned to the inspection process and does most of the work by visual inspection using a Mechanical Timber Grader (MTG). It is necessary to reduce the time of inspection as it stops the production line when arriving wood is waiting for inspection. There are three possibilities to reduce the inspection time. (1). Purchasing an automatic wood scanner which costs between \$400,000 and \$800,000 (X-Ray model). (2). Recruitment of another operator. (3). Getting help from the current production staff. The price of the inspection machine is not currently affordable for the company, therefore the feasible solution would be to increase the human resources.

Press set-up

Another activity that needs to be reduced as much as possible is press set-up. The set-up process takes 4 hours each time the machine is used to produce curved beams. For the gridshell project, using only curved beams, set-up time is significant. According to lean production, the set-up time should be less than 10 minutes (Delago *et al.* 2016), using the Single Minute Exchange of Die (SMED) method. The following steps are suggested to reduce the set-up time from 4 hours to 2 hours. (1). Maximise the set-up activities that can be done while the press machine is working (external elements). (2). Simplify and streamline all elements. (3). Standardize procedures for the workers to follow and limit wasted time. (4). Create auxiliary tools and equipment to make the set-up quicker.

Pile-up

In the production process, there is a station before gluing where the laths are piled up and an operator decides which combination of wood layers should be glued together. For example, to produce a 5 layer glulam the two top layers must be of grade B, for the central layer grade C is acceptable and the two bottom layers should be grade B. This process is required because the laths are produced by batches of quality degree instead of based on the glulam's needs. This process can be omitted and the operator will

be freed if laths are produced using a daily plan based on glulam production needs.

The results of eliminating NVA activities

The lean approach aims to have a system with zero NVA activities and waste. Simulating the proposed methods to eliminate inspection, set-up, and pile up, showed that the cycle time could be reduced from 178 hours to 135 hours which is an improvement of 24 %. Wait time would be reduced from 202 hours to 162 hours, a 20 % improvement. The result shows noticeable improvements for this one project which suggests more improvements in the long term. However, it is important to find feasible solutions that respect the real factory's situation. In most cases, inspection is assumed to be a NVA activity (Nikakhtar *et al.*, 2015). In the production of engineered wood, inspection is an important process to ensure the quality restrictions. Moreover, checking for knots and grading is part of the production process and cannot be eliminated. Hence, instead of eliminating NVA activities completely, the factory in this case study thought it more realistic to aim for reducing inspection and set-up times by 50% instead of 100 %. The simulated results show that cycle time reduces from 178 hours to 146 hours (18% improvement instead of 24%) while wait time is reduced from 202 hours to 186 hours (13% improvement instead of 20%). Figure 3 and figure 4 illustrate the results and comparison.

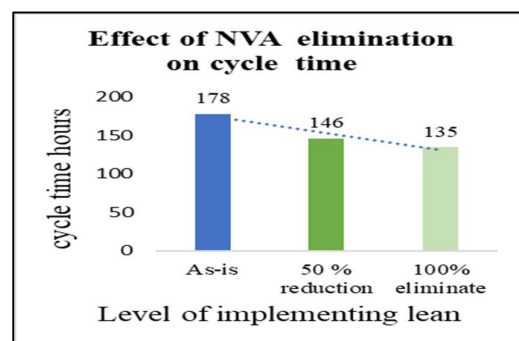


Figure 3: Comparing the effect of NVA activity elimination on cycle time

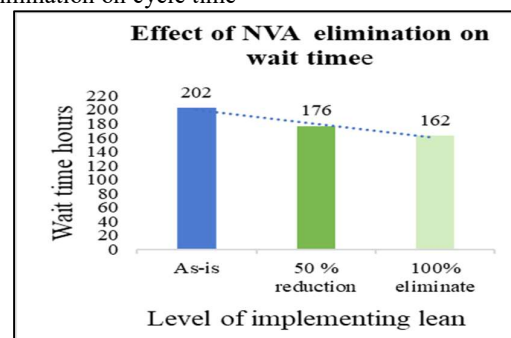


Figure 4: Comparing the effect of NVA activity elimination on wait time

5.1.6 Transportation

It can be recognized from the factory layout in figure 5, that the production flow is not efficient. There are significant unnecessary transfers from the pile-up station to gluing, from the cold press to the planing 2 station, and from the finishing station to the exit door. Eliminating the pile up process and changing the layout to make a U-shaped production line improves the production flow. Figure 6 illustrates the suggested layout.

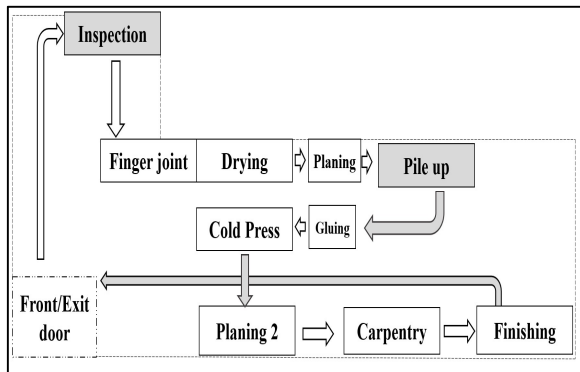


Figure 5: Current factory layout with unnecessary transport

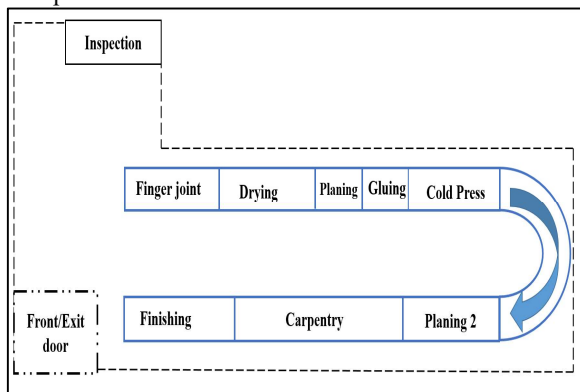


Figure 6: Suggested U-shaped layout

This U-shaped layout was simulated and table 2 shows that the cycle time was reduced from 178 hours to 176 hours which is not a noticeable improvement (1.12%). On the contrary, the wait time has been reduced from 202 hours to 62 hours (69.53%) which is significant. Reduction in wait time results in a smoother flow of material in the production line. This change will lead to improvements in cycle time in the future.

Transfer	As-is	Lean	Improvement
Average cycle time	178	176	1.12%
Wait time	202	62	69.53%

Table 2: Results of changing the layout

5.1.7 Defects (failure and rework)

The last category of waste which is considered in this study relates to defects. Two sources of defect are machinery failure and rework.

Failure

Based on the observations and operators' information, the simulation model represents the failure rate of machinery with a triangular distribution. In the factory, the sudden breakdown of machinery occurs, however, the rate of failure has not been documented. The first step of improvement is to document every time that failure occurs and hold regular meetings to analyse the reasons. The second step is to implement the TPM method. According to Abdulmalek and Rajgopal (2007), Total Productive Maintenance (TPM) involves: (1). Constantly monitoring machinery breakdown and finding the failure rate. (2). Discussions to find the source of the problems. (3). Autonomous maintenance Safety and environment management. (4). Planned maintenance instead of condition-based maintenance.

According to the factory's work schedule, almost every morning the press machine is loaded with the layers of laths which have been produced the day before. However, since the factory is closed on weekends and planed laths cannot be kept more than 24 hours, the press is idle on Monday morning while laths are being produced. Therefore, Monday morning is an appropriate time for preventive maintenance of the press. For the other equipment, lunchtime is suggested. The proposed times for preventive maintenance is presented in table 3.

Machinery	Maintenance		Weekday	Time
	Uptime (day)	Downtime (minutes)		
Finger joint	7	120	Tuesday	12h-13h
Drying	7	120	Wednesday	12h-13h
Planing 1	7	120	Thursday	12h-13h
Planing 2	7	120	Friday	12h-13h
Press	7	120	Monday	8h-10h

Table 3: Proposed preventive maintenance times

Rework

The main amount of rework by the factory for this project came from the inconsistencies in design which were detected at the carpentry station (second-last station). For the gridshell project, large curved glulam beams were built and then cut into smaller parts in the carpentry station. This design created waste in both material and processes. It also was the source of mistakes and rework. Design for manufacturing and assembly (DFMA) methods focus on taking time in the design phase to prevent the errors in the final products. According to DFMA, the following steps must be considered: (1). Use standard, off-the-shelf parts rather than custom components. (2). Design for ease of fabrication. (3). Aim

for mistake-proof design. (4). Design with predetermined assembly technique in mind.

Elimination of rework and failure in the simulation model shows that cycle time reduces from 178 hours to 175,7 hours which is not significant (1.23%) while the wait time is noticeably reduced from 202 hours to 89 hours (56% improvement). Table 4 lists these results.

Failure/rework	As-is	Lean	Improvement
Average cycle time	178	175.7	1.23%
Wait time	202	89	56%

Table 4: Results of eliminating failure and rework

6 DISCUSSION

Applying all changes simultaneously

The result of improvements in the three targeted sources of waste shows improvements in the cycle time and wait time. What would be the result of applying all the improvements at the same time? The result of modifying the simulation model with complete elimination of NVA, transfer and failure, demonstrates 27% reduction in cycle time (from 178h to 129h) and 77% reduction in wait time (from 202h to 47h). Additionally, the result of more realistic improvements which means reducing the time of inspection and set-up by 50% was investigated. The results show that with partial improvements of NVA together with the elimination of transfer and defects, the cycle time still has 26% improvement (from 178h to 132h) and wait time indicates 75% improvement (from 202h to 50h). Charts in figure 7 and figure 8 illustrate the comparison between the partial and complete implementation of the lean approach. The result is interesting as there is a drastic reduction from the actual production system and partial implementation of a lean approach, while the difference between the partial and complete implementation is very small. This result helps managers understand that using lean approaches to even partially reduce wastes can be profitable. These first steps could lead to continuous improvement and reach higher levels of productivity in the future.

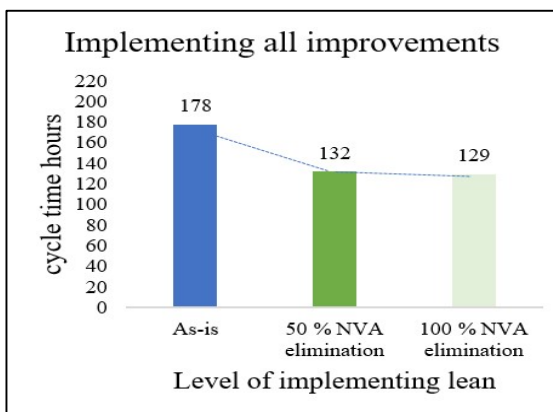


Figure 7: Comparing the effect of implementing all improvements on cycle time

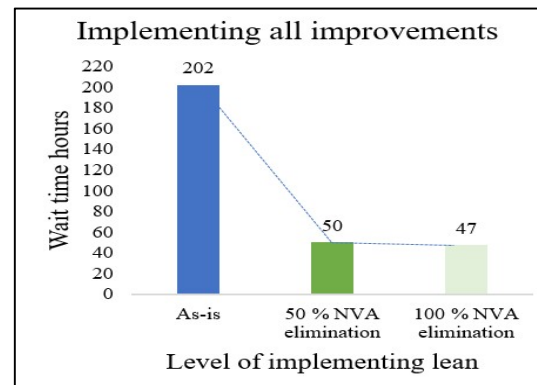


Figure 8: Comparing the effect of implementing all improvements on wait time

7 CONCLUSION

This research considers applying the lean approach in a glulam manufacturing company. Firstly the production of glulam is simulated using Arena software. Then the basic simulation model is modified for a case-study in an SME which manufactures glulam to be used in its construction project. The seven sources of waste were identified, and improvements were suggested based on the lean approach. Three main sources of waste were highlighted as NVA activities, transfer, and defects. NVA activities include inspection, set-up and pile up. Simulating the elimination of these three activities leads to a 24% improvement in cycle time and 20% improvement in wait time. Feasible solutions such as adding an operator to inspection and SMED result in a 50% decrease in inspection and set up time. The pile up process can be eliminated by providing a standard action plan. Improving transport by changing the factory layout to a U-shape layout gives 1.12 % improvement in cycle time and 69.53 % improvement in wait time. The machinery failure rate can be improved by using a TPM system, and reworks can be eliminated by DFMA. These improvements lead to a 1.23% reduction in cycle time and 56% reduction in wait time. It is noticeable that improvement in NVA activities indicates a higher impact on cycle time, while improvement in transfer and defects shows a higher impact on wait time.

The impact of all improvements together is also considered. The result of eliminating three sources of waste would be 27% reduction in cycle time and 77% reduction in wait time. In the case of partial elimination of NVA activities, the results would be 26% reduction in cycle time and 75% reduction in wait time, which is very close to the result of the complete elimination of NVA activities. This research attempts to apply the lean approach in an SME that manufactures glulam. The area of manufacturing for construction needs to be studied and improved in the purpose of increasing the productivity in construction projects.

Appendix 1: Process distribution Parameters of glulam production

Process	Distribution	parameters (minutes)
Inspection	Triangular	a= 22, m= 25, b= 28
Finger joint	Triangular	a= 1.1, m=1.5, b= 2
Drying	Triangular	a= 12,m=13, b= 14
Planing 1	Triangular	a= 5, m=6, b= 7
Pile up	Triangular	a= 12, m= 15, b= 18
Gluing	Triangular	a= 30, m=35, b= 40
Cold Press	Constant	Value: 150
Planing 2	Triangular	a= 21, m=24, b= 27
Carpentry 1	Triangular	a= 120, m=126, b= 132
Carpentry 2	Triangular	a= 120, m=126, b= 132
Carpentry 3	Triangular	a= 120, m=126, b= 132
Finishing1	Triangular	a= 36, m=41, b= 46
Finishing2	Triangular	a= 36, m=41, b= 46

REFERENCES

- Abbasian-Hosseini, S. A., A. Nikakhtar, and P. Ghoddousi, 2014. Verification of lean construction benefits through simulation modelling: A case study of bricklaying process. *KSCE J. Civ. Eng.*, 18(5), P.1248–1260.
- Álvarez, R., R. Calvo, M. M. Peña, and R. Domingo, 2009. Redesigning an assembly line through lean manufacturing tools. *Int J Adv Manuf Technol*, 43, P. 949–958
- Abdulmalek, F.A., and J. Rajgopal, 2007. Analyzing the benefits of lean manufacturing and value stream mapping via simulation: A process sector case study. *Int. J. Production Economics*, 107, P. 223–236.
- Abedi Saidabad, A., and H.Taghizadeh, 2015. Performance and Improvement of Production Line Function Using Computer Simulation (Case Study: An Iron Foundry). *American Journal of Computational Mathematics*, 5(2015), P. 431-446
- AbouRizk, S, 2010. Role of Simulation in Construction Engineering and Management, *Journal of Construction Engineering and Management-ASCE*, 136, P. 1140-1153.
- Chung, C. A., 2004. Simulation Modeling Handbook: A Practical Approach, *CRC Press, US*.
- Delago, L., M. Machado, F. Brito, G. Landgraf, M. Schroeder, and C. Torezzan, 2016. Learning lean philosophy through 3d game-based simulation, *Winter Simulation Conference, Washington D.C.*
- Dotoli, M., P. Fanti, G. Iacobellis, G. Rotunno, 2014. An Integrated Technique for the Internal Logistics Analysis and Management in Discrete Manufacturing Systems, *the International Journal of Computer Integrated Manufacturing*, Vol. 27, No. 2, pp. 165-180.
- Dotoli, M., N. Epicoco, M. Falagario, N. Costantino, and B. Turchiano, 2015. An integrated approach for warehouse analysis and optimization: A case study, *Journal of Computers in Industry*, 70, p. 56-69.
- Douthe, C., O. Baverel, and J. Caron, 2006. Form-finding of a grid shell in composite materials, *J. IASS* 47(150), P. 53-62.
- Faisal A. M, 2016. Simulation modelling and analysis of value stream mapping for the implementation of lean manufacturing in labour-intensive small and medium-sized enterprises. *ICEEOT*, p. 3567-3569.
- Farrar, J., S. AbouRizk, and X. Mao, 2004. Generic implementation of lean concepts in simulation models, *Lean Constr. J.* 1, P.1–23.
- Ferrin, D. M., M. J. Miller, and D. Muthler, 2005. Lean sigma and simulation, so what's the correlation? V2, *Winter Simulation Conference*, Piscataway, N. J.
- Hassan, M. M., and S. Gruber, 2008. Simulation of concrete paving operations on Interstate-74, *J Constr Eng Manag.* 134(1), p. 2-9.
- Howell, G.A., 1999. What is lean construction, *International Group of Lean Construction 7th Annual Conference*, Berkeley, USA
- Koskela, L.,1992. Application of the new production philosophy to construction, *CIFE technical report#72*, Stanford University
- Kuzman, M. K., and D. Sandberg, 2016. A new era for multi-storey timber buildings in Europe, *70th Forest Products Society International Convention*, Forest Products Society, Peachtree Corners, GA, USA.
- Li, J.W., 2003. Simulation-based comparison of push and pull systems in a job-shop environment considering the context of JIT implementation, *International Journal of Production Research*, 41(3), P. 427-447.
- Mahfouz, A., J. Shea, A. Arisha, 2011. Simulation-Based Optimization Model for the Lean Assessment in SME: A Case Study, *Winter Simulation Conference*, Phoenix, 11-14 December 2011, 2403-2413.
- Malo, K. A., and V. Angst, 2008. Glued laminated timber. *Handbook 1 Timber Structures*, CZ/06/B/F/PP/
- McKinsey Global Institute, 2017. Reinventing construction: a route to higher productivity. *McKinsey & Company*, February 2017
- Moghadam, M., A. Alwisy, and M. Al-Hussein, 2012a. Integrated BIM/Lean Base Production Line Schedule Model for Modular Construction Manufacturing. *Construction Research Congress (ASCE 2012)*, West Lafayette, USA, pp. 1271 – 1280.
- Moghadam. M., M. Al-Hussein, S. Al-Jibouri, and A. Telyas, 2012b. Post-simulation visualization model for effective scheduling of modular building construction, *Can. J. Civ. Eng.* 39, P. 1053–1061.
- Nagi, M., F. F Chen, and H. Wan, 2017. Throughput Rate Improvement in a Multiproduct Assembly Line Using Lean and Simulation Modeling and Analysis, *27th International Conference on Flexible Automation and Intelligent Manufacturing*, Modena, Italy

- Nikakhtar, A., S.A., Abbasian-Hosseini, K.Y., Wong, and A. Zavichi, 2015. Application of lean construction principles to reduce construction process waste using computer simulation: a case study, *Int. J. Services and Operations Management*, 20(4), P.461-479.
- Parthanadee, P., and J. Buddhakulsomsiri, 2014. Production efficiency improvement in batch production system using value stream mapping and simulation: a case study of the roasted and ground coffee industry, *Production Planning and Control*, 25(5), P. 425-446.
- Puettmann, M. E., and J. B. Wilson, 2005. Gate-to-gate life-cycle inventory of glued-laminated timbers production, *Wood and Fiber Science*, pp. 99 – 113.
- Ritter, C., Y. Zhang, R. Dupuis, and M. Al-Hussein, 2017. Simulation of Production Line Improvement in Modular Home Manufacturing, *25th Annual Conference of the International Group for Lean Construction*. Heraklion, Greece, pp 737-744.
- Sargent, R.G., 2011. Verification and validation of simulation models, *Winter Simulation Conference*, Phoenix, AZ, USA.
- Solomon, J., 2004. Application of the Principles of Lean Production to Construction, *Master's thesis*, University of Cincinnati, Cincinnati, 9 July.
- Structural timber association, 2014. Glued laminated timber structures, *Rev 0 - 11.11.14/ EB008*.
- Tokola, H., E. Niemi and V. Väistö, 2015. Lean Manufacturing Methods in Simulation Literature: Review and Association Analysis, *Winter Simulation Conference*, 2239–2248. Piscataway, New Jersey.
- Womack, J. P., and D.T., Jones, 1996. Lean thinking: Banish waste and create wealth in your corporation, *Simon & Schuster*, N.Y.
- Yu H., M. Al-Hussein, S. Al-Jibouri, A.Telyas, 2013. Lean transformation in a modular building company: A case for implementation. *Journal of management in engineering*, 29 (1), P. 103-111
- Zarrin.M., and A. Azadeh, 2017. Simulation optimization of lean production strategy by considering resilience engineering in a production system with maintenance policies, *Simulation: SAGE Journals*, 93(1), P. 49–68.

APPORT D'AGILITE AUX BIORAFFINERIES : VERS UNE DEMARCHE COLLABORATIVE ET UNE SERVICISATION DES PROCÉDES

Michelle HOUNGBÉ, Anne-Marie BARTHE-DELANOË, Stéphane NEGNY

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS,
Toulouse, France

michelle.houngbe@ensiacet.fr, annemarie.barthe@ensiacet.fr, stephane.negny@ensiacet.fr

RESUME : *Enjeu industriel d'actualité, le développement de bioraffineries se trouve freiné par les irrégularités de production, la dispersion et la variété des sources d'approvisionnement. Ces impératifs de flexibilité imposent à la bioraffinerie de devenir agile et de dépasser la vision silo et disciplinaire de son écosystème pour développer une vision plus globale permettant d'identifier les synergies possibles, du récoltant jusqu'au consommateur.*

Le projet ARBRE propose de virtualiser la bioraffinerie en un réseau collaboratif, où les opérations unitaires sont servicisées. Ceci permet de configurer rapidement le procédé, de réutiliser des unités et du matériel existants et de proposer de nouvelles solutions adaptées aux variabilités actuelles de l'écosystème.

MOTS-CLES : *Agilité, bioraffinerie, collaboration, servicisation, industrie du futur.*

1 INTRODUCTION

L'accroissement de la production et de l'usage des bioénergies, et des produits biosourcés, est un défi d'avenir s'inscrivant dans la transition énergétique et environnementale. En France, le Grenelle de l'Environnement a fixé entre autres objectifs d'ici 2020 une part de 23% d'énergies renouvelables dans la consommation finale (Assemblée Nationale et Sénat 2010) qui repose pour moitié sur l'usage de la biomasse, mais aussi de doubler le volume de matières premières d'origine végétale dans l'industrie chimique, avec pour effet la création de 5000 emplois directs (Le Déaut et al. 2016).

Le développement des bioraffineries et de la chimie du végétal s'est accéléré ces dernières années. Toutefois, la valorisation de la biomasse, matière vivante provenant des plantes et des animaux, est soumise à de fortes contraintes liées à la nature même de la biomasse : irrégularités de production en qualité et quantité, variété et dispersion de l'approvisionnement, saisonnalité et disponibilité. De plus, d'autres facteurs freinent le développement des bioraffineries tels que la difficulté de l'ancrage dans le territoire local, et la nervosité de la demande finale. En effet, la bioraffinerie, qu'elle soit de première, de deuxième ou de troisième génération, est généralement conçue comme une seule usine physique qui intègre l'ensemble des processus de transformation de la biomasse : pré-traitement, fermentation, séparation et purification.

Du fait de cette structure physique figée héritée de la chimie du pétrole, la bioraffinerie reste un système extrêmement spécialisé, de par le procédé mis en œuvre, les matières premières valorisées, et les produits résul-

tant de la transformation. En outre, le passage du pilote de laboratoire à l'échelle industrielle représente également un obstacle à la mise en œuvre de procédés innovants de traitement de la biomasse, tant du point de vue des opérations physico-chimiques mises en œuvre que du point de vue économique et financier. Enfin, le traitement de la biomasse, tel qu'il est réalisé aujourd'hui, nécessite de lourds investissements financiers initiaux et engendre des coûts opératoires importants, alors que la difficulté d'approvisionnement menace le bon fonctionnement de la bioraffinerie. Les bioraffineries telles que conçues actuellement ne sont donc pas en mesure de s'adapter d'une part, à la forte variabilité de l'offre et de la demande et, d'autre part, à la difficulté d'implantation des usines.

La problématique sous-jacente à ces constats relève d'un besoin d'apport d'agilité aux bioraffineries, au niveau des systèmes physique et d'information. En effet, le mode de production des bioraffineries doit être repensé d'un point de vue structurel et organisationnel, en y intégrant certains concepts de l'usine du futur. Ainsi, l'agilité apportée aux bioraffineries permettra de stimuler le développement industriel, en prenant en compte les facteurs économiques, sociaux et environnementaux de leurs écosystèmes. Nous pouvons nous poser la question suivante :

Comment favoriser le développement des bioraffineries et leur donner l'agilité nécessaire pour évoluer dans un environnement instable ?

Pour répondre à cette question, il est possible d'adopter une vision plus large du traitement de la biomasse en ne le limitant pas au procédé chimique, mais en prenant aussi en compte son écosystème (agriculteurs, indus-

triels, clients). Dans cette optique, la transformation de la biomasse peut être considérée comme une chaîne logistique.

Cette dernière est définie comme un réseau d'organisations –qui supporte des flux physiques, informationnels et financiers–, dans différents processus et activités, qui fournissent un produit ou un service, dans le but de satisfaire le client (Christopher 1992). Cependant, la multiplication des opérations de transport entre les organisations de la chaîne peut apparaître comme un frein du point de vue économique. En effet, dans le cas du traitement de la biomasse, les matières premières et les produits intermédiaires sont à faible valeur ajoutée du fait de leur haute teneur en eau, rendant ainsi leur transport peu rentable.

Par contre, si l'on s'intéresse à la conception de la chaîne logistique du point de vue des produits à valeur ajoutée, alors la notion de décentralisation devient intéressante pour apporter l'agilité nécessaire aux bioraffineries. En effet, cela permet l'accès à des ressources humaines (compétences) et matérielles (équipements pour la réalisation d'opérations unitaires) plus variées et donc plus à même de répondre au besoin d'un procédé donné pour traiter un type de biomasse identifié à un instant t , en fonction de l'offre, de la demande, etc. sans avoir à investir dans l'achat d'équipements ou la construction d'un site dédié. De plus, la réactivité, la flexibilité, la robustesse, la stabilité ou encore la résilience sont devenues des facteurs clés dans la définition de la performance d'une entreprise et a fortiori, d'une chaîne logistique. Par conséquent, la notion de performance ne se limite plus au traditionnel triptyque « Coût, Qualité, Délai » mais à une notion beaucoup plus globale (Humez 2008).

Pour faire face à ces obstacles, le projet ARBRE (Agilité pour les BioRaffineriEs) propose de fournir cette agilité nécessaire aux bioraffineries en poursuivant deux objectifs majeurs: (i) l'adoption d'une démarche systémique (penser la structure de la bioraffinerie comme un système de systèmes); (ii) l'identification des synergies possibles entre acteurs et territoires.

Dans la suite de cet article, nous nous intéresserons d'abord à l'état de l'art concernant l'agilité des bioraffineries et plus particulièrement la décentralisation du traitement de la biomasse. Puis nous nous intéresserons à la notion de servicisation et d'économie de fonctionnalité, à l'Internet Physique avant d'aborder les différents apports des technologies de l'information et de la communication dans ce contexte. Enfin, la proposition du projet ARBRE sera détaillée, avant de conclure.

2 AGILITE DES BIORAFINERIES : ETAT DE L'ART

Dans le cadre du projet ARBRE, l'agilité est vue comme « la capacité d'un système à réaliser, dans le laps de temps le plus court possible (réactivité), la détection de

son inadéquation à l'environnement dans lequel il évolue et la mise en œuvre de l'adaptation nécessaire » (Barthe-Delanoë et al. 2014). Dans ce cadre, nous allons détailler les leviers existants (ou à envisager) pour réaliser l'agilité des bioraffineries, au moment de leur conception mais aussi pendant leur exploitation.

2.1 La décentralisation du traitement de la biomasse

Les besoins actuels de réduction des coûts et de flexibilité maximale de la chaîne logistique peuvent être satisfaits en ayant recours à des systèmes de production décentralisés et à la mise en place de stratégie de collaboration inter-organisationnelles.

Ces collaborations peuvent se traduire par l'établissement d'Entreprise Virtuelle voire d'Organisation Virtuelle (Camarinha-Matos et al. 1998) formée par plusieurs entreprises afin d'atteindre un (ou des) objectif(s) commun(s). Si les travaux menés dans ce sens sont appliqués à l'industrie des biens et des services, voire aux institutions et administrations (Montarnal 2015), en ce qui concerne l'industrie chimique, ce type de comportement décentralisé et collaboratif n'a jamais été établi (Yue et al. 2014).

En dépit d'une littérature fournie sur les sujets de la modélisation et de l'optimisation des chaînes logistiques des bioraffineries et qui s'appuient sur l'analyse des différents intrants, produits et procédés (Yue et al. 2014), très peu d'articles s'intéressent à la décentralisation du procédé de traitement de la biomasse (tout type confondu). Ces rares travaux étudient des réseaux collaboratifs qui ne mettent en jeu que deux acteurs (Kim et al. 2011) et/ou ne prennent en compte que le critère du coût financier du transport des produits intermédiaires pour évaluer la faisabilité de la décentralisation (Bowling et al. 2011). Les contraintes propres au champ d'application du traitement de la biomasse ne sont pour le moment pas abordées, comme le caractère périssable de certains produits agricoles, les phénomènes de dégradation naturels, etc. en vue de proposer un processus collaboratif pertinent et réaliste.

2.2 La servicisation et l'économie de fonctionnalité

La servicisation des entreprises manufacturières qualifie la transition de la vente de biens matériels vers la vente d'offres intégrées de produits et services immatériels (Vandermerwe and Rada 1988). Il est à noter que dans cette logique de servicisation, les relations entre entreprises (B2B) ou entre entreprises et consommateurs (B2C) s'inscrivent dans une dimension collaborative plutôt que transactionnelle (Peillon 2016). Il est à noter que, pour le moment, le secteur de la transformation de la biomasse ne se positionne pas dans une logique de servicisation.

Par contre, une passerelle existe entre le degré de servicisation et l'économie de fonctionnalité. L'économie de fonctionnalité est définie par Stahel (Stahel 2006) comme l'économie « [...] qui optimise l'usage ou la fonction des biens et services, se concentre sur la gestion des richesses existantes, sous la forme de biens, de connaissances et de capital naturel. L'objectif économique en est de créer une valeur d'usage la plus élevée possible pendant le plus longtemps possible, tout en consommant le moins de ressources matérielles et d'énergie possible ». Dans cette logique, la pérennité du matériel existant, sa réutilisation, sa modularité et sa facilité de maintenance deviennent des enjeux majeurs (Fondation Concorde 2010), enjeux soulignés par le rapport n°380 présenté en 2016 au Sénat concernant la stratégie française vis-à-vis de la bioéconomie.

Actuellement mise en œuvre au niveau de quelques entreprises et objet de recherche (Chalal 2014; Derrouiche and Dubruc 2014), la servicisation peut être rapprochée de l'économie de coopération (que l'on pourrait considérer comme étant le premier niveau de l'économie de fonctionnalité) que l'on retrouve traditionnellement dans le monde agricole.

A travers les Coopératives d'Utilisation de Matériel Agricole (CUMA), les agriculteurs se regroupent pour investir ensemble dans du matériel, embaucher le personnel nécessaire à l'usage du matériel et partager l'usage de l'équipement. Aujourd'hui, près d'un agriculteur sur deux est en CUMA (FNCUMA 2017). Certaines CUMA ont des activités spécifiques liées à la valorisation de la biomasse : bois, bûches et plaquettes, compostage (FNCUMA 2017).

Partant de ce constat, la servicisation du traitement de la biomasse s'inscrit dans un comportement existant au niveau des exploitants agricoles. Par contre, cette vision est en rupture avec la logique de fonctionnement des autres acteurs de la chaîne de traitement de la biomasse, telles que les entreprises de transformation.

2.3 L'Internet Physique

L'Internet Physique est un concept innovant qui s'inspire du réseau Internet (au sens du système d'information) pour concevoir et organiser les flux physiques. Il est défini par ses concepteurs comme étant « un système logistique global tirant profit de l'interconnexion des réseaux d'approvisionnement par un ensemble standardisé de protocoles de collaboration, de conteneurs modulaires et d'interfaces intelligentes pour une efficacité et une durabilité accrues. » (Ballot and Montreuil 2014).

Si parmi les solutions à l'étude on retrouve la standardisation (des conteneurs, des remorques de camions), la définition de réseaux souterrains de transport de marchandises (comme en Suisse), de grands groupes tels que Procter&Gamble ou Bayer s'intéressent à la modularisation des procédés via la « conteneurisation » de leurs outils de production.

Cette modularisation à l'échelle d'un conteneur standard réalisée par Bayer (dans le cadre du projet européen FP7 F3Factory) permet non seulement de déployer rapidement des unités et de créer des usines modulaires et flexibles (usines mobiles), mais aussi d'acheminer l'outil de production au plus près de la source d'approvisionnement en matière première ou du consommateur, réduisant in fine les besoins en transports (et les coûts économiques, énergétiques et environnementaux associés).

Si pour l'heure cette modularisation en est encore au stade expérimental (Lier et al. 2017), sa généralisation ouvrirait de nouvelles perspectives dans l'industrie de process. En effet, elle faciliterait la capacité d'adaptation des acteurs de la chaîne logistique face à la variabilité en dimensionnant l'outil de production en fonction de la quantité, de la qualité, du type des ressources, de leur localisation mais aussi de la demande du marché. Elle permettrait également la mutualisation d'équipements, abaissant ainsi les coûts d'investissement initiaux dans l'outil de production pour la réalisation d'opérations unitaires.

2.4 Les technologies de l'information et de la communication

Comme défini précédemment, l'agilité se réalise en deux étapes symbiotiques : la détection et l'adaptation. L'évolution des nouvelles technologies numériques permet d'assurer la capacité à collecter et analyser données, informations et connaissances. Celles-ci peuvent être exploitées pour caractériser le système de transformation de la matière et son contexte, fournissant ainsi une connaissance de la situation (au sens de « situational awareness » (Endsley 1995)), facilitant la tâche de détection et permettant de guider la décision d'adaptation.

Compte tenu du degré de connectivité du monde agricole (eFarming), de la disponibilité et du volume de sources de données exogènes concernant les conditions météorologiques, les territoires (OpenData par exemple), la problématique de l'exploitation des données, de l'information et de la connaissance constitue une autre piste de réponse à la problématique de l'agilité du traitement de la biomasse.

A cette fin, les technologies relevant du Complex Event Processing (CEP) qui permettent d'évaluer automatiquement une confluence d'événements et une prise de mesures, sur des périodes temporelles (ou fenêtres) de durées variables (Etzion and Niblett 2010) s'inscrivent pleinement dans le cadre de notre étude. Cette corrélation d'événements peut être causale, temporelle et/ou spatiale. Elle est définie à l'aide de règles. Le CEP est une technologie de réseau qui va déclencher des actions et créer de la connaissance à propos de la situation à partir de systèmes distribués basés sur les messages, provenant de bases de données et d'applications, ceci en temps réel.

Une telle technologie permettrait de supporter le niveau informationnel de la bioraffinerie agile et, dans une optique de décentralisation, servicisation et modularisation, d'ainsi faciliter la caractérisation des acteurs du réseau, de leurs services, et de l'écosystème. Une autre fonctionnalité intéressante est la possibilité de détecter certaines variabilités (comme des phénomènes météorologiques, des prévisions de vente en baisse, des cours sur les marchés financiers, des rumeurs sur les réseaux sociaux) afin d'anticiper leurs conséquences (diminution du rendement de production de certaines parcelles, nécessité de se positionner sur un nouveau produit).

3 ARBRE : UNE PLATEFORME DE CONCEPTION ET DE SUIVI DES BIORAFFINERIES AGILES

Initié en 2017 dans le cadre du programme Jeunes Chercheuses et Jeunes Chercheurs (JCJC) pour une durée de 4 ans, le projet ANR ARBRE (Agilité pour les BioRaffineriEs) est mené au sein du Laboratoire de Génie Chimique à Toulouse et compte dans son comité de pilotage le partenaire industriel SOLAGRO. Ce projet vise à proposer un cadre méthodologique outillé permettant de concevoir et de mettre en oeuvre des bioraffineries agiles, suivant une structure collaborative et modulaire.

3.1 Les objectifs et la philosophie de ARBRE

Pour faire face aux obstacles mentionnés précédemment, deux objectifs majeurs sont poursuivis dans le projet ARBRE pour fournir cette agilité nécessaire aux bioraffineries.

Dans un premier temps, il s'agit d'adopter une démarche systémique, et de penser la structure de la bioraffinerie comme un système de systèmes (Figure 1), où le procédé de transformation n'est plus réalisé sur un seul et unique site mais est réparti sur différentes usines.

Ainsi, chacune des opérations du procédé de transformation est servicisée, permettant (i) la réutilisation d'unités et de matériels existants, (ii) la coexistence sur un même site du traitement de produits conventionnels et de produits biosourcés, (iii) la reconfiguration facilitée du procédé en fonction des perturbations de l'écosystème (qualité, quantité, coût, etc.).

Ceci nous amène au concept innovant de « bioraffinerie virtuelle » (Figure 2) : c'est la collaboration du récoltant, de différents sites de production, et du consommateur — au niveau de leurs systèmes physiques et d'information — qui construit la bioraffinerie en fonction du procédé retenu et du bioproduit ciblé. Le rôle et l'implication de chaque acteur peuvent changer pour s'adapter aux objectifs et aux contraintes du moment.

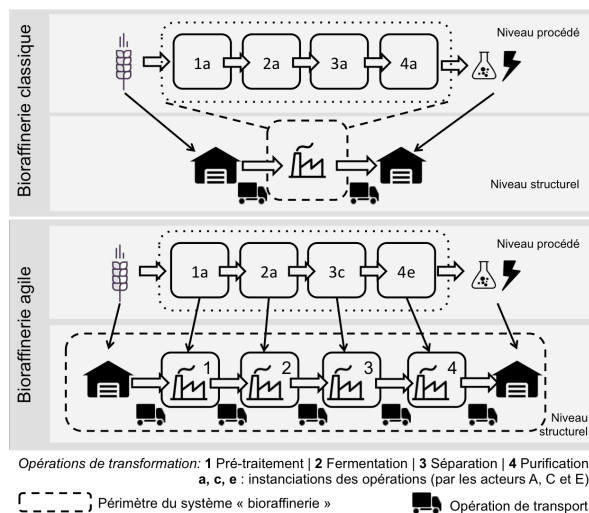


Figure 1. Bioraffinerie classique vs. Bioraffinerie agile (telle que proposée par le projet ARBRE)

La connaissance sur la collaboration, en termes d'objectifs, d'acteurs, de services, mais aussi d'opportunités, de menaces, d'environnement, de performance est représentée sous la forme d'un modèle, instance d'un méta-modèle. Pour obtenir ce méta-modèle, nous nous baserons sur le méta-modèle « core » réalisé par (Bénaben et al. 2016), qui définit un méta-modèle de collaboration, générique vis-à-vis du domaine métier. Il a été utilisé avec succès pour des applications liées à la collaboration industrielle (B2B), à la gestion de crise (crise civile, crise humanitaire, crise routière). Ce méta-modèle core est ensuite spécifié en couches, c'est-à-dire en méta-modèles étendus à des domaines métier différents. Afin de réaliser le méta-modèle de la collaboration relative au traitement de la biomasse, un appel à des experts académiques et industriels du domaine a été effectué (en complément d'une revue de littérature exhaustive) afin d'en valider le contenu et la cohérence au niveau métier.

Dans un second temps, au niveau du procédé lui-même, il est nécessaire de faciliter la recherche de nouvelles solutions adaptées aux variabilités. Compte tenu du caractère « virtuel » de la bioraffinerie et de ce nouveau mode de fonctionnement, il est envisageable de proposer des approches basées sur une méthodologie d'innovation couplée à la gestion des connaissances (type couplage TRIZ et Raisonnement à partir de cas) (Negny and Le Lann 2008; Robles et al. 2011) pour créer différents procédés alternatifs en fonction de la biomasse utilisée, des produits biosourcés recherchés ou de toute autre contrainte exprimée.

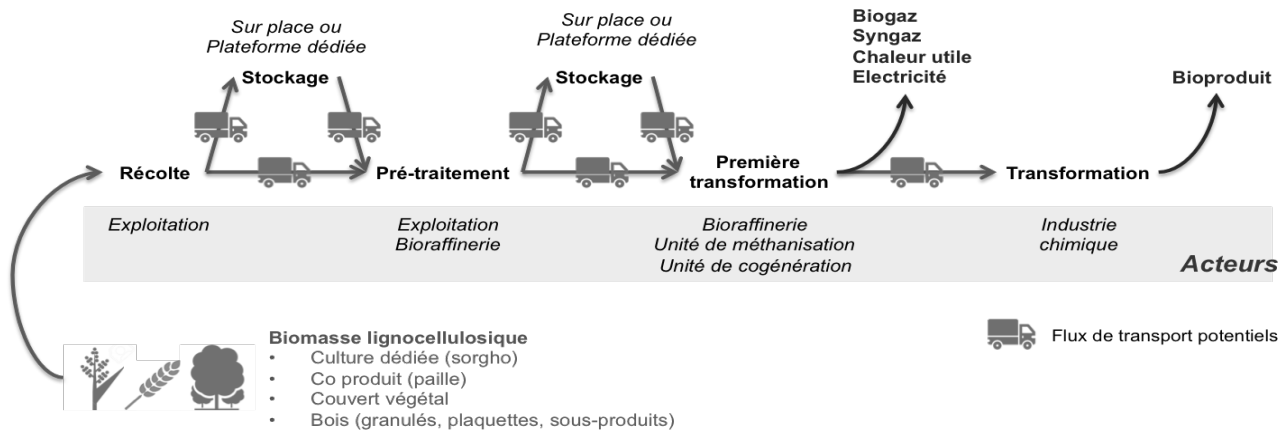


Figure 2. La bioraffinerie virtuelle, un réseau collaboratif

L'implémentation du procédé retenu se fait ensuite en recherchant les récoltants, les acteurs industriels, les transporteurs, ainsi que les consommateurs respectant les contraintes susnommées, en utilisant par exemple des métaheuristiques type colonies de fourmis (Montarnal 2015) sur une base de connaissance. Cette base de connaissance contiendra les services proposés par les acteurs potentiels du réseau collaboratifs, ainsi que les différentes caractéristiques (fonctionnelles et non fonctionnelles) liées à ces services.

D'autre part, à ce stade de définition du réseau collaboratif, le choix du stockage (lieu dédié ou « sur le champ »), l'éloignement géographique (notion de chaîne logistique courte), le coût, le délai de traitement, l'impact sur l'environnement neutre ou positif (Colonna et al. 2015; Vanholme et al. 2013), etc. sont intégrés dans le processus de sélection des partenaires.

3.2 Le cadre méthodologique et logiciel de ARBRE

Cette vision en rupture d'une bioraffinerie agile s'inscrit donc sur deux échelles spatiales (le procédé et l'ensemble du réseau) et en deux phases:

1. Au niveau de la phase de conception : la conception d'un procédé innovant et de la bioraffinerie virtuelle, en adéquation avec leur écosystème à l'instant t_0 . En d'autres termes, il s'agit de concevoir le réseau collaboratif initial.
2. Au niveau de la phase d'exploitation : au cours du temps, les variabilités doivent être détectées et modélisées afin que le procédé et/ou l'ensemble du réseau soient adaptés en conséquence. Dans le cadre du projet ARBRE, il s'agit de tester la robustesse et la dynamique du réseau en comparant le réseau collaboratif initial déduit avec ceux obtenus en modifiant la valuation des variabilités. L'écart calculé entre les différents réseaux modélisés (en utilisant la méthodologie décrite dans Barthe-Delanoë et al. (2014) permettra d'aider les décideurs dans le choix de la solution la plus adaptée.

Cet apport d'agilité peut être supporté par l'évolution des nouvelles technologies dans le domaine de la connexion de données et de l'exploitation de données en temps réel. L'acceptation de cette évolution sera facilitée car certains acteurs ont déjà amorcé cette transition vers le numérique. Un récent sondage IPSOS montre que les grandes exploitations françaises utilisent en moyenne 4 outils ou équipements de eFarming et que les exploitants sont très largement favorables (65%) à l'usage des nouvelles technologies du numérique (IPSOS 2016). Deux tiers des agriculteurs interrogés dans le cadre de l'étude menée envisagent entre autres de s'équiper de capteurs (pour les animaux, les bâtiments, les machines agricoles) d'ici les prochaines années.

Cette dimension liée à l'Internet des Objets (Atzori et al. 2010; Kopetz 2011) relève de la collecte, création, exploitation et visualisation de données, d'informations et de connaissances. Elle est exploitée dans la méthodologie proposée afin de participer à la caractérisation de la structure de la bioraffinerie et de son contexte, permettant ainsi l'adéquation de la prise de décision avec une représentation juste et actualisée du terrain.

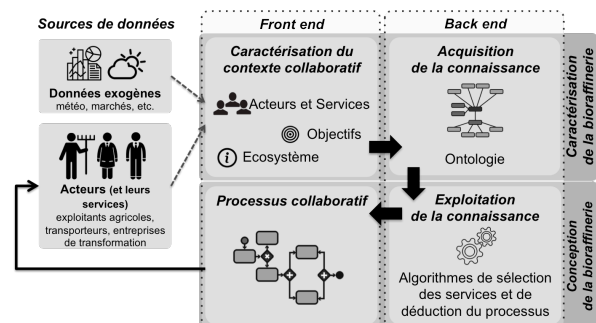


Figure 3. La plateforme logicielle ARBRE

La méthodologie développée dans le projet ARBRE sera implémentée dans un outil logiciel de type application Web. En termes d'architecture (voir Figure 3), l'utilisateur aura accès à une interface de caractérisation du contexte collaboratif (front end), pour saisir manuellement certaines données. D'autres données relevant de l'IoT (prévisions météorologiques, santé économique, etc.) pourront être collectées et traitées de façon automatique et en temps réel en utilisant le Complex Event Pro-

cessing. Le second type d'interface consistera en la restitution du processus collaboratif déduit (« quels acteurs réalise quel service à quel moment ? »). Les mécanismes d'acquisition de la connaissance et de déduction du processus collaboratifs seront exécutés côté serveur (back end) et demeureront invisibles pour l'utilisateur.

4 CONCLUSION

La proposition portée par le projet ARBRE vise à dépasser la vision en silo et disciplinaire de l'écosystème des procédés de transformation de la matière et de l'énergie. Cette vision en rupture a pour objectif est de développer une vision plus globale permettant d'identifier les synergies possibles à l'échelle d'un territoire (département, région). L'approche développée dans ce projet consiste à penser le procédé monolithique du traitement de la biomasse comme une chaîne logistique décentralisée. Cette vision permet d'apporter l'agilité nécessaire en conception mais aussi durant l'exploitation de la bioraffinerie.

En la modélisant comme un réseau collaboratif, où les acteurs proposent des services, il devient alors possible de construire une bioraffinerie adaptée aux objectifs du moment, tout en minimisant les coûts d'implémentation et d'investissement, et en favorisant la réutilisation des matériels et compétences existants. Il s'agit également d'inclure l'usage du numérique pour améliorer l'efficacité de la dynamique et de l'ensemble des fonctions du réseau collaboratif et de ses interactions avec son écosystème.

Le projet ARBRE en étant à ses débuts, le développement du cadre méthodologique présenté ici est en cours de réalisation. Outre la réflexion portée sur la conception du méta-modèle de la collaboration dédiée au traitement de la biomasse, un effort particulier est mené pour serviciser le procédé en lui-même. Il s'agit en effet de trouver une façon de représenter, à la fois au niveau du système physique et au niveau du système d'information, les différentes composantes du procédé comme des services proposés par l'entreprise. Afin de valider la pertinence des concepts exposés dans le méta-modèle, ce dernier a d'abord été soumis à des experts académiques et industriels. La prochaine étape consiste maintenant à modéliser des systèmes existants de bioraffinerie (cas industriels) afin de vérifier la couverture du méta-modèle.

REMERCIEMENTS

Les auteurs remercient les membres du projet ANR ARBRE (2017-2021) pour leurs conseils et leurs remarques. Ces travaux de recherche sont financés par l'Agence Nationale de la Recherche dans le cadre du projet ARBRE (numéro : ANR-17-CE10-0006).

REFERENCES

- Assemblée Nationale, and Sénat. (2010). *LOI n° 2010-788 du 12 juillet 2010 portant engagement national pour l'environnement. 2010-788*.
- Atzori, L., Iera, A., and Morabito, G. (2010). "The Internet of Things: A survey." *Computer Networks*, 54(15), 2787–2805.
- Ballot, E., and Montreuil, B. (2014). *L'Internet Physique : Le réseau des réseaux des prestations logistiques*. Transports Recherche Innovation, La Documentation Française.
- Barthe-Delanoë, A.-M., Truptil, S., Bénaben, F., and Pingaud, H. (2014). "Event-driven agility of interoperability during the Run-time of collaborative processes." *Decision Support Systems*, 59, 171–179.
- Bénaben, F., Lauras, M., Truptil, S., and Salatgé, N. (2016). "A Metamodel for Knowledge Management in Crisis Management." *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 126–135.
- Bowling, I. M., Ponce-Ortega, J. M., and El-Halwagi, M. M. (2011). "Facility Location and Supply Chain Optimization for a Biorefinery." *Industrial & Engineering Chemistry Research*, 50(10), 6276–6286.
- Camarinha-Matos, L. M., Afsarmanesh, H., Garita, C., and Lima, C. (1998). "Towards an architecture for virtual enterprises." *Journal of Intelligent Manufacturing*, 9(2), 189–199.
- Chalal, M.-E. (2014). "Aide à la décision pour la servicisation des PME industrielles: une approche par modélisation et simulation." Saint-Etienne, EMSE.
- Christopher, M. (1992). *Logistics Strategy and Supply Chain Management*. Pitman.
- Colonna, P., Tayeb, J., and Valceschini, E. (2015). "Nouveaux usages des biomasses." *Le Déméter*, 275–305.
- Derrouiche, R., and Dubruc, N. (2014). "Approche d'analyse et de simulation des différents scénarii système produit-service (pss) : cas d'Envie Loire."
- Endsley, M. R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors*, 37(1), 32–64.
- Etzion, O., and Niblett, P. (2010). *Event Processing in Action*. Manning Publications Co., Greenwich, CT, USA.
- FNCUMA. (2017). *CUMA Chiffres clés, édition 2017*. Paris, 23.
- Fondation Concorde. (2010). *L'économie de fonctionnalité: vers un nouveau modèle économique durable*. Fondation Concorde, Paris.

- Humez, V. (2008). "Proposition d'un outil d'aide à la décision pour la gestion des commandes en cas de pénurie : une approche par la performance." PhD Thesis, <<http://ethesis.inp-toulouse.fr/archive/00000839/>> (Mar. 20, 2017).
- IPSOS. (2016). "L'innovation est déjà dans le pré !" *IPSOS FRANCE*, <<http://www.ipsos.fr/comprendre-et-maitriser-son-marche/2016-09-08-l-innovation-est-deja-dans-pre>> (Mar. 16, 2017).
- Kim, J., Realf, M. J., Lee, J. H., Whittaker, C., and Furtner, L. (2011). "Design of biomass processing network for biofuel production using an MILP model." *Biomass and Bioenergy*, 35(2), 853–871.
- Kopetz, H. (2011). "Internet of Things." *Real-Time Systems*, Real-Time Systems Series, Springer US, 307–323.
- Le Déaut, J.-Y., Courteau, R., and Sido, B. (2016). *De la biomasse à la bioéconomie : une stratégie pour la France*. Session Ordinaire 2015-2016, SENAT.
- Lier, S., Paul, S., Ferdinand, D., and Grünwald, M. (2017). "Modular Process Engineering: Development of Apparatuses for Transformable Production Systems." *ChemBioEng Reviews*, 4(1), 60–70.
- Montarnal, A. (2015). "Deduction of inter-organizational collaborative business processes within an enterprise social network." Ecole des Mines d'Albi-Carmaux.
- Negny, S., and Le Lann, J.-M. (2008). "Case-based reasoning for chemical engineering design." *Chemical engineering research and design*, 86(6), 648–658.
- Peillon, S. (2016). "La servicisation des entreprises industrielles." *La Revue des Sciences de Gestion*, (278–279), 131–140.
- Robles, G. C., Rodriguez, A. M., Negny, S., and LeLann, J. M. (2011). "The TRIZ-CBR synergy: a knowledge based innovation process." *International Journal of Manufacturing Technology and Management*, 22(3), 233.
- Stahel, W. R. (2006). *The Performance Economy*. Palgrave Macmillan.
- Vandermerwe, S., and Rada, J. (1988). "Servitization of business: Adding value by adding services." *European Management Journal*, 6(4), 314–324.
- Vanholme, B., Desmet, T., Ronsse, F., Rabaey, K., Van Breusegem, F., De Mey, M., Soetaert, W., and Boerjan, W. (2013). "Towards a carbon-negative sustainable bio-based economy." *Frontiers in Plant Science*, 4, 174.
- Yue, D., You, F., and Snyder, S. W. (2014). "Biomass-to-bioenergy and biofuel supply chain optimization: Overview, key issues and challenges." *Computers & Chemical Engineering*, Selected papers from ESCAPE-23 (European Symposium on Computer Aided Process Engineering - 23), 9-12 June 2013, Lappeenranta, Finland, 66, 36–56.

APPORTS DES PARADIGMES SOCIAUX DANS L'INTERNET DES OBJETS INDUSTRIEL: VERS DES OBJETS COMMUNICANTS INDUSTRIELS SOCIAUX

Eddy BAJIC, Oussama HAJLAOUI

Université de Lorraine, CRAN, UMR 7039, CNRS,
Boulevard des Aiguillettes B.P.70239, 54506 Vandœuvre-lès-Nancy
eddy.bajic@univ-lorraine.fr, oussama.hajlaoui7@etu.univ-lorraine.fr

RESUME : *L'interaction entre objets communicants hétérogènes, dans un contexte Internet of Things (IoT) est contrainte par des problèmes d'interopérabilité, de sécurité et de confidentialité. Ces aspects sont prépondérants dans les classes d'applications ouvertes, et peuvent en limiter l'évolution face à l'augmentation remarquable du nombre d'objets communicants, et de leur caractère de plus en plus nomade dans un contexte Industrie 4.0. Une nouvelle approche émerge au travers du paradigme Social Internet of Things (SIoT) qui consiste à procurer à des objets communicants semi-autonomes, des capacités d'interaction inspirées des relations sociales humaines. Dans cet article, nous présentons l'évolution du concept IoT vers SIoT, ainsi que les résultats des travaux de recherche majeurs réalisés dans ce mouvement. Nous analysons les apports des paradigmes sociaux dans les interactions entre objets communicants industriels. Nous étendons ces travaux en proposant une structuration d'un Objet Communicant Industriel Social (OCIS), et la définition d'un protocole d'interaction sociale entre OCIS, en spécifiant les notions d'équilibre et de monétisation de services. Un cas d'application dans un contexte industriel est présenté.*

MOTS-CLES : *IoT, SIoT, objet communicant, modélisation, interaction, service, monétisation de service.*

1 INTRODUCTION

Le terme *Internet of Things* (IoT), désigne la mise en réseau d'objets, connectés à Internet, et possédant la capacité d'interagir avec leurs environnements, et d'échanger des données avec d'autres objets. *IERC* (IoT european research cluster) le définit comme une «infrastructure de réseau dynamique avec des capacités de configuration automatique basées sur des protocoles de communication standard et interopérables où les « choses » physiques et virtuelles ont des identités, des attributs physiques et des personnalités virtuelles. Les « choses » utilisent des interfaces intelligentes et sont intégrées de manière transparente dans le réseau d'information (Minerva et al., 2015). Les industries manufacturières ont bénéficié amplement des progrès de IoT (Acatech, 2013) où les objets communicants sont déployés de manière intensive permettant d'améliorer l'automatisation de la production et d'assurer le suivi en temps réel. Le domaine d'application Industriel de l'IoT a donné naissance à *Industrial Internet of Things* (IIoT) (Breivold, 2015).

L'augmentation du nombre d'objets communicants qui peuplent l'environnement IoT, estimé à 50 milliards à l'horizon 2020 (Cisco, 2011), introduit des problèmes d'interopérabilité et limite l'aspect évolutif des réseaux d'objets communicants (Nitti et al., 2012). Ces problématiques s'amplifient avec le nomadisme croissant des objets communicants que l'on observe dans les infrastructures industrielles et grand public. Ceci a favorisé l'émergence d'une nouvelle approche appelée *Social Internet of Things* (SIoT). Cette approche initiée

par (Atzori et al., 2012), permet aux objets communicants d'acquiescer des comportements similaires à ceux des êtres humains et d'établir des interactions avec d'autres objets d'une manière autonome.

Dans cet article, nous présentons, à travers un état de l'art, les travaux scientifiques majeurs réalisés sur les interactions sociales SIoT. Nous étendons les théories développées dans les travaux analysés et nous proposons une structuration d'un Objet Communicant Industriel Social (OCIS). Nous proposons ensuite un protocole d'interaction sociale entre OCIS dont nous analysons les apports dans les interactions entre objets communicants industriels (OCI), et son impact sur les services fournis. Nous présentons finalement, un cas d'application industriel selon l'approche SIoT.

2 ETAT DE L'ART

Afin de déterminer les modèles d'interaction et les relations qui peuvent avoir lieu entre des objets communicants, les chercheurs se sont inspirés des travaux scientifiques réalisés sur les relations sociales entre les êtres humains et leurs transcriptions potentielles vers des objets. Les travaux de recherche menés par (Fiske et al., 1991), (Hoorens et al., 2000), dans les domaines de la sociologie, et de l'anthropologie sociale sur les comportements sociaux entre les êtres humains, ont formalisé les différentes formes de socialisation humaine et les interactions entre individus selon plusieurs types de relation.

2.1 Approche sociologique de Fiske

L'anthropologue américain Alan Fiske, a étudié la nature des relations humaines et leurs variations interculturelles (Fiske et al. 1991). Il a conçu une théorie des « modèles relationnels » dans laquelle il modélise les interactions sociales entre individus, sur la base de quatre modèles relationnels, qui sont :

• **Partage communautaire (PC)** : Les individus appartiennent à une communauté à laquelle ils contribuent avec ce qu'ils peuvent lui apporter et y prennent ce dont ils ont besoin.

• **Équilibre de concordance (EC)** : Ce modèle est basé sur des relations égalitaires caractérisées par une réciprocité en nature et un échange équilibré en volume.

Une relation est représentée comme un équilibre et chaque individu est préoccupé de l'équilibre de cette relation.

• **Rang d'autorité (RA)** : Les individus sont ordonnés selon certaines dimensions sociales hiérarchisées. Chaque individu peut être caractérisé par un rang d'autorité reflétant son niveau hiérarchique. Les relations entre individus sont ainsi assujetties à leurs rangs d'autorité, souvent asymétriques.

• **Prix de marché (PM)** : La relation est basée sur un principe de proportionnalité de valeurs. Les interactions entre individus sont organisées en fonction d'une échelle commune de valeurs des éléments échangés lors de l'interaction.

Modèle relationnel	Socialité selon la théorie de Fiske	Socialité transposée aux objets communicants
Partage Communautaire (PC)	L'appartenance de chaque individu à une collectivité de partage l'emporte sur toute forme de caractère distinctif individuel.	Le partage des données, services et ressources entre tous les objets d'une communauté est admise de fait par tous, et sans condition.
Équilibre de Concordance (EC)	Les relations entre individus sont caractérisées par une réciprocité et des échanges équilibrés en valeur ou en quantité ou en nature.	Les relations entre objets sont quantifiées et contraintes par un équilibre en quantité ou en valeur des données, des services et des ressources.
Rang d'Autorité (RA)	Les individus sont ordonnés selon une hiérarchie, un statut, qui se traduisent par un classement d'autorité, et des relations asymétriques acceptées.	Les objets sont ordonnés selon un classement hiérarchique de leur rang autorité, imposant une subordination respectée dans leurs relations (maître, esclave).
Prix de Marché (PM)	Les interactions entre individus sont structurées par des coûts économiques, non forcément monétaires, et menées selon un mode d'échange marchand reposant sur une échelle de valeurs.	Les objets délivrent leurs services contre un coût prédéfini, associé à chaque service selon un système de cotation partagé (jeton, bitcoin, ...). Le principe de troc de services, sans monétisation, est admis.

Tableau 1 : Transposition des modèles relationnels sociaux de Fiske vers le monde des objets communicants.

2.2 Transposition du modèle relationnel social de Fiske aux objets communicants

L'approche sociale dans l'analyse des comportements entre objets communicants a été avancée par (Atzori et al., 2012), (Pintus et al., 2015), qui ont proposé un modèle de comportement entre objets communicants en se basant sur le modèle relationnel de Fiske. Le Tableau 1 illustre le modèle de socialité relationnel selon Fiske et sa transposition sur des objets communicants industriels.

Le principe PC permet un partage de ressources, services et de données sans aucune condition ou contrainte.

Le principe EC est employé dans les applications de coopération et de collaboration entre les objets communicants. Dans ce type de relation, la notion de contribution et de l'équilibre de la relation sont mis en œuvre pour mesurer l'apport de chaque objet.

Le principe RA traduit les niveaux hiérarchiques entre des objets communicants dans une relation (un lecteur et des tags RFID, des terminaux maître et esclave...). Les objets échangent des informations d'une manière asymétrique. Le principe PM peut être associé à des objets travaillant ensemble dans le but d'obtenir des avantages mutuels. Dans une telle relation, chaque service d'un objet est associé à un coût, dans une logique de monétisation de service.

2.3 Les Types de relation entre objets

L'interaction sociale entre deux individus reflète le rattachement entre eux et qualifie la relation qui les unie. Le type de relation traduit le lien social entre les interlocuteurs, et l'intensité de parenté qui les lie. Il peut influencer fortement sur la nature de leurs interactions.

Par analogie aux types de relation entre les individus, des travaux ont proposé une classification des types de relation relatives aux objets communicants (Mandler et al., 2015), sur la base des cinq types de relations suivantes:

• **Relation de Co-localisation (CLOC)** : Lorsque des objets sont présents simultanément au même endroit (sur une machine, dans un atelier, une maison, une ville, ...)

• **Relation de Co-travail (CTRA)** : Lorsque des objets coopèrent ensemble dans une même application ou un même processus, pour réaliser une tâche ou un but collectif.

• **Relation de Parenté (PAR)** : Lorsque des objets appartiennent à une même famille (objets similaires, même catégorie, construits dans une même période, par le même fabricant, appartenant à un même lot).

• **Relation de Propriété (PRO)** : Lorsque des objets appartiennent au même propriétaire et interagissent entre

eux. L'objet peut être porté par son propriétaire (une personne, une machine ou un autre objet). Le propriétaire peut stimuler l'interaction des objets.

• **Relation Sociale (SOR)** : Lorsque des objets se rencontrent et entrent en contact, de façon sporadique ou continue, au travers de la rencontre physique de leurs propriétaires respectifs. L'aspect social de l'interaction entre objets décline de l'interaction sociale entre leurs propriétaires.

3 VERS UN OBJET COMMUNICANT INDUSTRIEL SOCIAL

Etablir des interactions sociales exige de nouvelles capacités attachées aux objets communicants (évaluation d'une interaction, position sociale, ...) qui vont au-delà des capacités habituelles des Objets Communicants Industriels classiques (OCI). L'objet doit choisir en premier lieu le type de relation qu'il va établir avec autrui pour pouvoir ensuite interagir selon un modèle relationnel social. Ainsi, pour permettre à un objet communicant de bien choisir le type de relation, nous proposons une hiérarchie des types de relation qui va permettre à l'objet de choisir un type de relation approprié selon des paramètres fixés.

3.1 Hiérarchisation des types de Relation

Plusieurs travaux scientifiques (Procidano et al., 1983) (Zimet et al., 1988), ont analysé les comportements des individus au sein des différentes relations et l'influence des types de relation sur les interactions entre les individus. Souvent les interactions sont établies avec des niveaux de confiance différents selon le type de relation. Pareil au monde des êtres humains, les types de relation peuvent affecter les interactions sociales entre les objets communicants surtout aux niveaux de la confiance entre les partenaires, et de la fiabilité des services (Chen et al., 2016). Un objet ami avec lequel on a eu beaucoup d'interactions auparavant, ou un lien de parenté, est plus fiable qu'un objet inconnu, avec lequel il n'y a jamais eu d'interactions (Ashri et al., 2005).

La relation au sein d'une famille est caractérisée par un niveau de confiance et de soutien très fort entre ses membres, comparativement à une relation entre amis ou individus sans affinité entre eux. (Krackhardt, 1992) montre que plus le lien entre individus est fort, plus les interactions entre ces individus sont solides et sont caractérisées par un niveau de confiance élevé. Ainsi, nous pouvons considérer la relation de parenté (PAR) entre objets communicants, comme la relation la plus forte en termes de fiabilité et de confiance dans les interactions. Les objets en relation de propriété (PRO), ont un lien spécifique entre eux du fait qu'ils appartiennent au même propriétaire. Toutefois, on considère que la relation PRO est caractérisée par un niveau de confiance inférieur à la relation de parenté.

Les objets en relation de Co-travail (CTRA) n'ont généralement aucun lien de parenté, mais la coopération à un travail commun nécessite un certain niveau de

confiance mutuelle. On considère que le niveau de confiance d'une relation CTRA est inférieur à celui de la relation PRO.

L'interaction entre des objets en relation Sociale (SOR) est conditionnée par la rencontre, opportuniste ou planifiée, de leurs propriétaires, qui peuvent eux-mêmes être amis ou non. Etant donné l'incertitude des paramètres, nous considérons que le niveau de confiance entre objets dans une relation SOR est inférieur à celui de la relation CTRA.

Quant à la relation de Co-localisation (CLOC), elle s'établit entre des objets qui se trouvent dans le même environnement de communication ou réseau. C'est la relation la plus implicite entre des objets. Nous accordons le plus faible niveau de confiance à la relation CLOC.

A partir d'une relation CLOC, des objets peuvent décliner ensuite les autres relations PAR, SOR, CTRA, PRO, selon leurs paramètres et caractéristiques. En se basant sur l'analyse précédente de la confiance entre objets, nous proposons une hiérarchie des types de relation entre objets, présentée par ordre croissant dans le Tableau 2.

Type de relation	Rang hiérarchique
CLOC	1
SOR	2
CTRA	3
PRO	4
PAR	5

Tableau 2 : Hiérarchie des relations entre objets

En plus des caractéristiques usuelles (données capteurs, période d'envoi, ...) d'un OCI, un objet communicant industriel social (OCIS) doit posséder des caractéristiques, des attributs et des règles de comportement qui lui permettent d'établir et gérer des interactions sociales. Nous proposons une architecture fonctionnelle d'un OCIS qui tient compte des exigences de l'aspect social des interactions entre objets.

3.2 Architecture d'un objet communicant industriel social

L'interaction sociale entre objets, requiert des caractéristiques et des mécanismes spécifiques associés à chaque objet communicant, qui lui permettent d'établir des relations sociales avec d'autres objets. De nombreux travaux ont défini des architectures d'objet communicant industriel ; avec une vision orientée sécurité entre objets (Zouinkhi, 2011), d'autre orientée objet communicant élément d'un réseau remontant des données RFID (Radio Frequency Identification) ou capteurs vers un site central (Sanchez et al., 2012). (Li et al., 2016) a présenté une structure d'objet communicant industriel avec une vision sociale destinée à partager des statuts et des données de maintenance sur un réseau social de niveau supérieur sans communication directe entre objets. Ces travaux valorisent principalement la communication verticale, sans prendre en compte les mécanismes de relation socio-mimétique entre objets que nous privilégions.

Pour compléter ces approches et ouvrir une nouvelle perspective, nous proposons une architecture d'un Objet Communicant Industriel Social (OCIS), permettant l'invocation et l'échange de services direct entre OCIS, selon le modèle relationnel social de Fiske (Figure 2). Lorsque deux objets sont liés par deux relations ou plus, la relation de rang hiérarchique le plus élevé sera considérée. D'autre part, un objet communicant peut avoir plusieurs relations simultanées avec différents objets. Ainsi, comme indiqué figure 1, un objet peut être en relation sociale avec un autre objet, tout en étant en relation de parenté avec un troisième objet, et de propriété avec un autre objet.

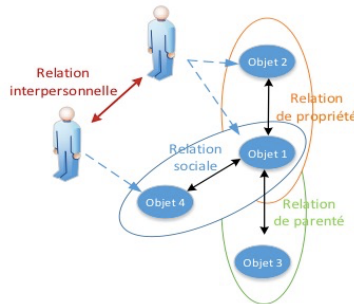


Figure 1 : Relations multiples entre objets

L'établissement d'une relation parmi les 5 types possibles, dépend principalement des caractéristiques et des paramètres de l'objet, auxquels s'ajoutent les « intentions » de l'objet. Celles-ci sont caractérisées par ses besoins pour la réalisation de ses activités (demande de service, partage d'information, ...).

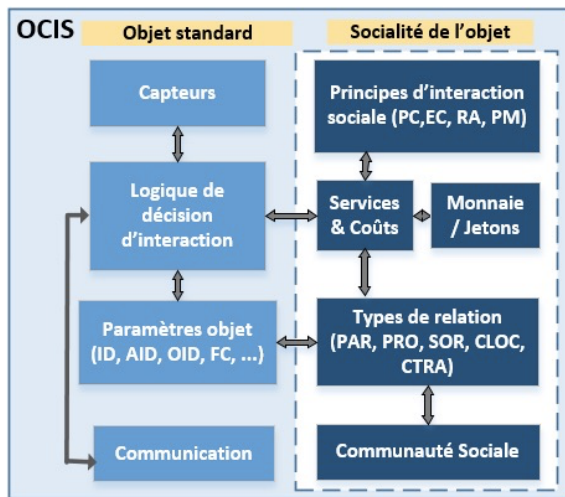


Figure 2 : Architecture fonctionnelle d'un OCIS

La nouvelle architecture proposée présente les éléments d'une architecture classique d'un OCI, complétés par les éléments supports de l'aspect social de l'objet, qui sont :

• **Attributs objet** : Les attributs d'un objet déterminent ses caractéristiques propres permettant de le différencier individuellement de tous les autres objets, et d'influer sur ses comportements avec les autres objets. En plus de l'Identifiant objet (ID), caractéristique de tout OCI, nous ajoutons 3 attributs (OID, FC, AID) paramétrant les liens sociaux de l'objet. Ces attributs entrent en jeu lors d'une

interaction sociale avec un autre objet :

- **Identifiant objet (ID)** : Identifiant unique de l'objet (UUID, MAC, Numéro de Série, ePC, ...), habituellement défini par le fabricant de l'objet.
- **Identifiant propriétaire (OID)**: Code caractéristique du propriétaire de l'objet. Plusieurs objets appartenant au même propriétaire ont le même OID.
- **Code Famille (FC)**: Code représentatif d'un groupement d'objets, possédant des caractéristiques fonctionnelles ou techniques communes. Une famille d'objets partage la même finalité ou le même usage.
- **Identifiant Application (AID)**: Code caractéristique de l'application, de l'activité ou du processus auquel l'objet participe (étape de fabrication, transport, surveillance environnement, ...). Il permet d'identifier les objets qui travaillent ensemble pour la même application ou activité.

• **Services** : Chaque objet est demandeur et/ou fournisseur de services. En plus des paramètres d'appel qui influent sur l'exécution d'un service (demande de température moyenne sur une durée de 1h, ...), nous ajoutons à chaque service les paramètres **coût** et **principe d'interaction** qui conditionnent sa mise en œuvre :

- **Coût d'un service** : Un coût peut être associé à chaque service, afin d'exprimer le « prix » de la réalisation de ce service par l'objet pour autrui. La demande d'un service pourra requérir un autre service de valeur équivalente en retour (relation PM). La métrique de coût des services pourra être définie selon un domaine géographique, un environnement de fonctionnement, une famille d'objets ou une application. Chaque objet pourra posséder un capital assimilable à de la monnaie électronique (ex. *bitcoin*), lui permettant de régler les coûts de services qu'il demande.
- **Principes d'interaction sociale** : Les quatre modèles relationnels sociaux (PC, RA, EC, PM) régissent les principes d'interaction entre objets selon des protocoles d'interaction paramétrés. Ils sont exécutés et agréés par tous les objets en interaction, permettant des demandes et des échanges de services selon une approche socio- mimétique entre OCIS.

• **Communauté sociale** : Au fur et à mesure de ses interactions, chaque objet construit un groupe social composé des objets avec lesquels il est entré en interaction, enrichi des types de relation et des niveaux de confiance correspondants (Chen et al., 2016). Le groupe social est mis à jour à partir d'un historique horodaté des interactions de l'objet qui mémorise chacune de ses interactions (ID, services réalisés, ...). Un OCIS se référera à son groupe social pour entamer une relation avec un objet (objet connu, niveau de confiance, ...).

3.3 Paramétrage des principes d'interaction

L'interaction entre OCS est basée sur les principes d'interaction présentés précédemment. Lors d'une demande de service, les objets doivent s'accorder sur le principe à mettre en œuvre pour la réalisation de ce

service. Nous proposons une formalisation et un paramétrage des quatre principes d'interaction (PC, RA, EC, PM), permettant aux objets de déterminer leur comportement lors des interactions :

- **Principe de partage communautaire (PC)** : Selon ce principe, le plus classique entre objets, un objet offre aux autres ses propres services, ressources ou information sans aucune condition ou contrainte. L'objet doit préciser la liste des services qu'il met à disposition.

- **Principe d'équilibre de concordance (EC)** : Ce principe traduit un certain équilibre en termes d'échange et de contribution entre objets communicants. Pour assurer un équilibre entre les prestations et les contributions entre objets, (Zhang et al., 2014) propose, dans une analyse complémentaire des travaux de Fiske, trois paramètres permettant de quantifier l'équilibre d'une relation :

- **La contribution C** mesure la participation de chaque élément dans l'équilibre d'une relation. La contribution d'un objet communicant peut être mesurée par le nombre de fois que l'objet a fait une demande de service ($C_{request}$), ou qu'il a répondu à une demande d'un autre objet ($C_{response}$).

- **L'équilibre d'interaction ΔC** renseigne sur le degré d'équilibre d'une relation entre objets. Il est égal à la différence des contributions selon la formule suivante :

$$\Delta C = C_{response} - C_{request}$$

- **Principe du rang d'autorité (RA)** : Il traduit une asymétrie de relation entre objets communicants. Les objets peuvent posséder des rangs d'autorité différents, par exemple définis selon un ordre croissant de 0 à 10. L'objet de rang le plus élevé a autorité sur les objets de rangs inférieurs pour lui fournir leurs services, à condition que ces services soient assujettis au principe RA. Dans ce cas, les objets de rang inférieur ont obligation à exécuter les demandes de services vers un objet de rang supérieur. Le Tableau 3 décrit deux services régis selon le principe RA pour un objet de rang d'autorité 3.

Principe	Rang Autorité	Services associés	Description du service
RA	3	Température	Détermination de la température d'objet
		Réglage	Modification d'un paramètre de réglage

Tableau 3 : Services régis selon le principe RA

- **Principe prix de marché (PM)** : Chaque service est quantifié d'un coût qui représente sa valeur (coût du service), ou d'un type de service qui doit être fourni en contrepartie à son exécution (service de contrepartie).

Principe	Services	Paramètre associés	Description du service
PM	Transport	X= 10 (Avoir un crédit de 10 unités)	Transport de produits
	Température	Fournir la pression	Fournir des mesures des objets transportés

Tableau 4 : Exemple de coût des service d'un OCIS.

Au moment de la demande de service auprès d'un objet fournisseur, l'objet demandeur doit disposer d'un capital supérieur au coût du service demandé, ou d'un service correspondant au type de service de contrepartie. Pratiquement, si l'objet ne possède pas de capital suffisant ou de service de contrepartie, un processus d'avoir ou de dette pourrait être mis en œuvre selon le type de relation et le niveau de confiance associé. Chaque objet, selon le principe PM, doit déterminer les services qui vont être assujettis à ce principe et préciser le coût de chacun de ses services ou énumérer les services de contrepartie. Le Tableau 4 détaille les services d'un OCIS, régis par le principe PM.

4 MODELISATION DES INTERACTIONS SOCIALES ENTRE OBJETS

Après avoir exposé le modèle relationnel de Fiske, sa transposition vers des objets communicants industriels, et défini les types de relations qui représentent un cadre d'interaction entre objets, nous allons déterminer les mécanismes d'interaction entre OCIS. Il s'agit de définir un protocole d'interaction réglant les séquences d'interaction sociales entre objets industriels sociaux. Un OCIS doit choisir un type de relation parmi les 5 cités précédemment (PAR, SOR, PRO, CLOC, CTRA) sur la base des caractéristiques des objets partenaires. Une fois la relation établie, les OCIS peuvent entrer en interaction opérationnelle.

4.1 Etablissement des relations entre objets

Le type de relation entre objets communicants traduit le degré de lien entre eux, et renvoi sur le niveau de confiance de leurs relations. Une fois que deux objets se sont identifiés, ils sont par défaut en relation CLOC. Dès lors, d'autres relations peuvent s'établir selon les paramètres *Application_ID*, *Family_Code*, *Owner_ID* des objets. L'algorithme de la Figure 3 décrit la logique de sélection du type de relation opérationnelle finale entre objets, selon une échelle de hiérarchisation définie précédemment.

```

IF (FC1==FC2)
  Relation=PAR
Else IF (OID1==OID2)
  Relation=PRO
Else IF (AID1==AID2)
  Relation=CTRA
Else IF (OD1! =0) AND (OID2! =0)
  Relation=SOR
Else Relation=CLOC
    
```

Figure 3 : Logique de détermination du type de relation

Lorsque deux objets sans aucune caractéristique commune, mais de propriétaires différents, sont en proximité, il s'établit une relation SOR, qui traduit la rencontre sociale des propriétaires ou usagers.

Un type de relation traduit un certain privilège au niveau des interactions, qui est caractérisé par un niveau de confiance, dû au lien et à l'historique entre les objets communicants (relation PAR avec lien de parenté, relation EC avec équilibre parfait, ...). Certains services

peuvent être exclusifs à un type de relation. Par exemple, au sein d'une relation PAR, un objet peut demander une configuration ou un réglage auprès d'un autre objet de la même famille.

4.2 Séquence d'interaction

Un OCIS, initialement isolé, est en découverte permanente de son voisinage. De manière similaire au comportement social humain, un objet qui cherche à satisfaire ses besoins (services, données ...), s'annonce selon un mode *Push* qui est bien approprié à l'autonomie et la mobilité des objets (Breivold, 20011). Une fois qu'il détecte un objet en proximité, il demande son identification (ID). Après l'étape d'identification, l'objet demande les paramètres de l'objet en proximité afin de déterminer le type de relation qu'il peut établir. Chaque objet découvre ensuite les services que peut offrir l'autre objet, les coûts associés, ainsi que les principes d'interaction régissant les demandes de services. La figure 4 représente la séquence générique d'interaction sociale entre objets.

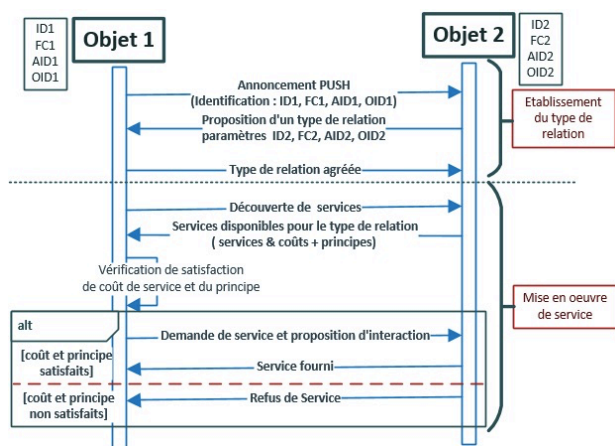


Figure 4 : Séquence d'interaction sociale entre OCIS

Le principe d'interaction (PC, RA, EC, PM) régit la mise en œuvre du service entre objets. L'objet demandeur de service doit satisfaire les contraintes du principe d'interaction fixé par l'objet fournisseur du service. Nous détaillons ci-dessous la mise en œuvre d'un service selon chacun des 4 principes d'interaction :

• **Cas de principe d'interaction PC** : L'échange de service *Communal Sharing* est réalisé sans condition.

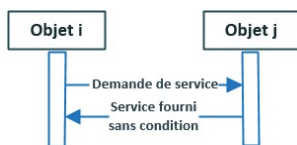


Figure 5 : Echange de service selon le principe PC

L'objet partage ses ressources, services et données avec les objets appartenant à son groupe social sur simple demande (Figure 5). Le principe PC correspond à un partage total de ressources et des services entre objets. Les services et informations ne sont pas sensibles, ni monnayables, et ne relèvent pas d'un niveau de sécurité.

• **Cas de principe d'interaction EC** : L'interaction *Equality Matching* repose sur un équilibre d'échange de services entre objets, qui peut être évalué par le nombre d'appels de service, ou le volume de données échangées précédemment entre les objets.

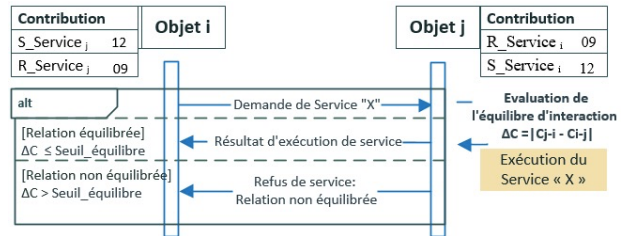


Figure 6 : Echange de service selon le principe EC

L'objet j fournisseur vérifie l'équilibre de la relation avec l'objet i (Figure 6), par le calcul de ΔC , en se basant sur les contributions « C » énumérées de chaque partie, en termes de contribution de données ou de services rendus réciproquement, avant d'offrir ses services.

• **Cas de principe d'interaction PM** : L'interaction *Prix de Marché* réclame un coût ou un type de service de contrepartie que le demandeur doit satisfaire, afin de pouvoir disposer d'un service du fournisseur. Une fois la découverte de service faite, l'objet demandeur connaît les contraintes associées aux services de l'objet fournisseur (principe d'interaction, coût, service de contrepartie ...). Il évalue alors s'il peut prétendre à requérir un service auprès de l'objet fournisseur. La Figure 7 illustre une demande de service selon le principe PM. L'objet i vérifie sa solvabilité d'appel de service auprès de l'objet j. S'il en a la capacité, il s'engage à fournir à l'objet j un service de contrepartie ou payer un coût (Jeton / Bitcoin) pour bénéficier du service pression.

L'objet demandeur peut soit payer/répondre immédiatement à la demande de contrepartie s'il n'a pas de contrainte, on parlera alors de **Paiement à la demande**, soit payer/répondre ultérieurement si sa requête est urgente ou si le paiement/exécution du service de contrepartie est pénalisant pour son fonctionnement actuel. Dans ce cas, on parlera de **Paiement différé**.

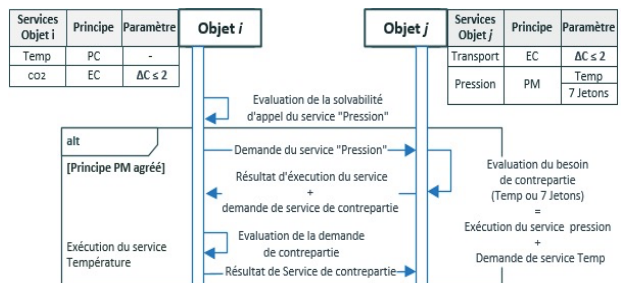


Figure 7 : Demande de service selon le principe PM

Le principe d'interaction PM se place dans une logique de **monétisation des services** des objets communicants (Bonnet et al., 2014). La monétisation de services permet de valoriser les services ou les données échangées entre objets, soit selon une unité de coût (jeton, bitcoin, ...) qui

nécessite l'adhésion commune à un système de référence de **prix de marché**, soit en contrepartie d'un service de valeur équivalente (troc de service), ce qui correspond alors à une situation de **troc de services** entre objets communicants industriels.

• **Cas de principe d'interaction RA** : L'objet demandeur d'un service doit avoir un rang d'autorité plus élevé que l'objet fournisseur, afin que ce dernier délivrer le service. En cas contraire, le service n'est pas réalisé. Après la découverte de services, l'objet i connaît le rang d'autorité de l'objet j (Figure 8). Selon son rang, l'objet i détermine s'il peut requêter un service auprès de l'objet j, par injonction d'autorité. L'objet j exécute le service demandé après vérification de la supériorité d'autorité de l'objet i.

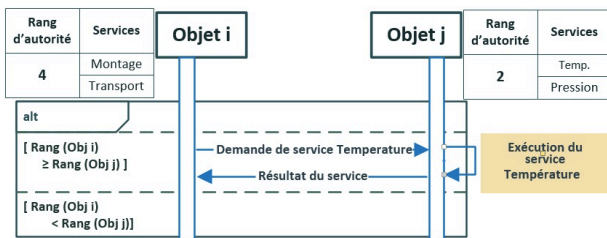


Figure 8 : Echange de service selon le principe RA

Le modèle d'interaction sociale que nous proposons tient compte des relations de service classiquement utilisées dans les systèmes IoT (Yim et al., 2017). Une simple requête suffit pour qu'un objet dispose du service voulu auprès d'un autre objet (principe PC). Nous avons formulé 4 autres principes d'interaction (EC, RA et PM) qui favorisent l'interaction horizontale au sein de communautés d'objets. Le principe EC gère les interactions nécessitant une quantification des contributions lors des échanges entre objets. Le principe RA gère les échanges de service où des exigences d'accessibilité, de latence ou de sécurité sont fortes. Le principe PM considère quant à lui les interactions contraintes et monétisables (énergie, temps CPU, ...).

Des OCIS sont capables de communiquer entre eux de manière ouverte, sans contrainte de réseau prédéfini, permettant ainsi d'améliorer l'interopérabilité et la sélectivité entre objets. Notre modèle vise à unifier les règles d'interaction entre OCIS hétérogènes, et de réguler des interactions selon leurs affinités sociales. Ce modèle permet de créer des communautés d'objets régies par des relations sociales (PAR, SOR, PRO...) au sein desquelles la notion de fiabilité de service basée sur le niveau de confiance entre objets est mise en exergue.

Nous proposons dans le chapitre suivant, un exemple applicatif industriel de notre proposition.

5 APPLICATION

Nous proposons une application de nos propositions dans un cas d'étude de logistique industrielle, de nature transactionnelle non temps réel. Il s'agit de faire transporter des produits hétérogènes par un chariot en respectant des exigences des parties-prenantes chariot et

produits.

Deux produits (OCIS 1 et 2) sont sur un chariot (OCIS 4). Il faut ajouter un autre produit (OCIS3) pour compléter la charge du chariot qui doit les transporter vers un camion (Figure 9). L'ajout de l'objet 3 doit satisfaire aux exigences de transport et de compatibilité de produits.

Des interactions entre les produits et le chariot sont nécessaires à la vérification des exigences individuelles. Les objets 1 et 2 sont construits par le même fournisseur et possèdent les mêmes caractéristiques, ils sont considérés de la même famille et en relation PAR. Chacun d'eux bénéficie des services de l'autre sans aucune exigence selon le principe PC. Les objets 1 et 2 fournissent l'un à l'autre leur service pour le contrôle des caractéristiques des produits afin de garantir la compatibilité.

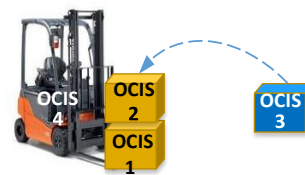


Figure 9 : Approche OCIS pour le transport de produits

L'objet 3 vérifie sa compatibilité avec les objets 1 et 2 avant d'être positionné sur le chariot. Pour cela, il a besoin des services des objets 1 et 2 pour contrôler les valeurs de température et de pression. L'objet 3 doit satisfaire les principes d'interaction exigés par les objets 1 et 2 pour bénéficier de leurs services et faire la vérification de compatibilité de produits. N'ayant pas d'attribut en commun avec les deux objets 1 et 2 (*FC, AID, OID*), l'objet 3 établit une relation sociale *SOR* avec ceux-ci. L'objet 2 exige un rang d'autorité supérieur à 6 pour fournir son service Pression, alors que l'objet 1 fournit son service Température selon le principe PM, dont il exige un coût (10). Sauf dans le cas de relation PAR, il le fournit selon le principe PC. Le chariot exige un coût de 8 selon le principe PM pour le transport d'objets (Tableau 5).

OCIS	Param. Objet	Services	Type relation	Principe	Paramètre de service
Produit 1	FC 1	Température	PAR	PC	
			SOR	PM	Coût = 8
Produit 2	FC 1	Pression	SOR	RA	Rg > 6
Produit 3	FC 2	Taux CO2	SOR	EC	$\Delta C < 2$
Chariot	FC 3	Transport	CTRA	PM	Coût = 10

Tableau 5 : Description des services des OCIS

Les demandes de service entre objets mêlent 4 principes d'interaction sociale (PC, PM, RA, EC) dans un processus totalement distribué. Le principe PM est exploité dans une logique de monétisation des services, dont les coûts associés reflètent à minima les dépenses que l'objet doit assurer pour l'exécution de ses services.

La relation PC permet la diffusion de connaissances (température, pression) dans la communauté d'objets. La priorisation du transport des produits est prise en charge par la relation PM, qui peut être associée à la valeur ajoutée et à l'importance de l'objet à transporter.

La communauté d'objets formée durant le transport de produits est régulée selon les caractéristiques de chaque

OCIS, et permet une gestion douce et consensuelle des finalités individuelles de transport par la typologie des relations sociales admises par chaque partenaire. Ce type de comportement communautaire sera de plus en plus significatif avec la forte croissance des objets nomades et communicants dans les perspectives de l'Industrie 4.0.

6 CONCLUSION

Le paradigme de *Social Internet of Things* (SIoT) propose une nouvelle vision des interactions entre les objets composants l'Internet of Things (IoT). Il vise à mimer les comportements sociaux humains, et à les appliquer aux objets communicants et à apporter de nouvelles solutions dans les interactions et les échanges de services entre entités des systèmes intelligents.

Dans cet article, nous avons présenté le concept SIoT et les caractéristiques des relations sociales, issues des théories sociologiques d'interactions humaines, que nous avons transposé à des objets communicants industriels. Nous avons proposé une architecture nouvelle d'un objet communicant industriel social (OCIS), comme une extension des architectures conventionnelles d'objet communicant industriel, qui permet de gérer la socialité entre objets, au travers de principes d'interaction selon la théorie sociologique de Fiske. Le modèle d'interaction sociale proposé s'appuie sur un modèle relationnel décrivant la globalité des comportements sociaux possibles entre objets industriels, permettant de constituer une communauté d'objets en interaction socio-mimétique. Le concept SIoT offre de nouveaux paradigmes d'interaction comme la monétisation de service entre objets - pouvant contribuer à la valorisation des services, et contribuer aux problématiques d'économie d'énergie. Les notions de confiance et d'équilibre de service s'ouvrent comme de nouveaux axes de recherche à explorer dans le contexte de l'Industrie 4.0.

REFERENCES

- Acatech, (2013) national academy of science and engineering, «Final report of the Industrie 4.0 Working Group», Federal ministry of education and research, Germany.
- Atzori L., Iera A., Morabito G., Nitti M., (2012), «The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization», *Computer Networks* 56 3594–3608.
- Bonnet D., Buvat J., Subrahmanyam K., (2011), «Monetizing the Internet of Things: Extracting Value from the Connectivity Opportunity», *Capgemini cons.*
- Breivold H.P., Sandström K., (2015), «Internet of Things for Industrial Automation – Challenges and Technical Solutions », *IEEE International Conference on Data Science and Data Intensive Systems*, pp 532-539.
- Chen I.R., Bao F., Guo J., (2016), « Trust-Based Service Management for Social Internet of Things Systems», *IEEE Trans. on dependable and secure comp.*, vol. 13.
- Cisco, (2011), <https://www.abavala.com/internet-of-things-estimations-de-traffic-internet-par-cisco>
- Fiske A.P., Haslam N., Fiske S.T., (1991), «Confusing one person with another: what errors reveal about the elementary forms of social relations», *Journal of personality and social psychology*, vol. 60, n°5.
- Fiske A.P., (1992), «The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations », *Psychological Review*.
- Hoorens V., Poortinga Y., (1992), «Behavior in the social context», *International handbook of psychology*, Edition: 1, Chapter: Behavior in the social context, Publisher: Sage, Ed. Pawlik, Rosenzweig, pp.40-53.
- Krackhardt D., (1992), «The strength of strong ties: The importance of philos in organizations», in *Networks and Organizations: Structure, Form, and Action*, vol. 216, Boston, USA: Harvard Business School.
- Li H., Parlakad A.K., (2016), «Social Internet of Industrial Things for Industrial and Manufacturing Assets», 3rd IFAC *Workshop on Advanced Maintenance Engineering, Services and Technology*.
- Mandler B., Barja J., Campista M.E.M., Cagaňová D., Chaouchi H., Zeadally S., Badra M., (2015), «Internet of Things: IoT Infrastructures», Partie 1, Second International Summit, IoT 360°, Rome, Italy.
- Minerva R., Biru A., Rotondi D., (2015), «Towards a definition of the Internet of Things (IoT) », *IEEE Internet of Things Initiative*.
- Nitti M., Girau R., Atzori L., Iera A., Morabito G., (2012), «A Subjective Model for Trustworthiness Evaluation in the Social Internet of Things», 23rd Annual *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*.
- Pintus A., Serra A., Manchinu A., (2015). «Humanizing the Internet of Things, Toward a Human-centered Internet-and- Web of Things», 11th Int. *Conf. on Web Information Systems and Technologies*.
- Procidano M., Heller K., (1983). «Measures of perceived social support from friends and from family: Three validation studies», *Amer. J. Community Psychol.*, vol. 11, no.1, pp. 1–24.
- Sanchez Lopez, T., Ranasinghe, D.C., Harrison, M., McFarlane, D., (2012). «Adding sense to the Internet of Things: an architecture framework for Smart Object systems», *Personal and Ubiquitous Computing*, 16(3).
- Yim H.J., Seo D., Jung H., Back M.K., Kim I., Lee K.C., (2017). «Description and classification for facilitating interoperability of heterogeneous data/events/services in the IoT», *Neurocomputing*, Vol 256, pp 13-22.
- Zhang H., Shen Y., (2014). «A Sociology-based Interaction Relationship Model in IoT», *IEEE Int. Conf. on Progress in Informatics and Computing*, Shanghai, China.
- Zimet G., Dahlem N., Zimet S., Farley G., (1988). «The multidimensional scale of perceived social support», *J. Pers. Assess.*, Vol. 52, no. 1, pp. 30–41.
- Zouinkhi A., (2011), « Contribution à la Modélisation de produit actif communicant, Spécification et Evaluation d'un protocole de communication orienté sécurité des produits », Thèse de doctorat, Université Henri Poincaré, Nancy.

APPROCHE COMPORTEMENTALE POUR UNE SYLVICULTURE COMMUNAUTAIRE DURABLE : APPLICATION A LA FAYA (MALI)

Souleymane DICKO, Daniel ROY,
Sophie HENNEQUIN

LGIPM - Université de Lorraine, Route d'Ars
Laquenexy, F 57078 METZ Cedex
{souleymane.dicko, daniel.roy,
sophie.hennequin}@univ-lorraine.fr

Abdel Kader KEITA

Directeur des Affaires Académiques et de la
Coopération,
Direction Générale de l'Enseignement Supérieur et de la
Recherche Scientifique, B.P 71, Bamako, Mali
abdel-kader.keita@mesrs.ml

RESUME : *Les pays subsahariens se doivent de trouver un juste équilibre entre l'exploitation des ressources naturelles et la préservation de la faune et de la flore. Ainsi, des domaines classés, comme la forêt de la Faya considérée ici, ont été définis par les Etats africains. Ces forêts classées sont soumises à des réglementations strictes au niveau de leur exploitation afin d'assurer la conservation des réserves foncières et écosystémiques mais également la sauvegarde des emplois et des possibilités vivrières et rentières offertes aux populations riveraines. Cependant, les habitudes des différentes parties prenantes couplées au réchauffement climatique génèrent un déboisement important et des feux de brousse impactant fortement la biodiversité et les possibilités offertes. De plus, la concertation entre l'Etat et les différentes communautés exploitant la forêt n'est pas simple. C'est pourquoi, dans ce travail, nous proposons de représenter à l'aide d'une approche systémique basée sur un modèle comportemental les interactions et les besoins des principaux acteurs pour identifier la structuration des différentes activités ainsi que les opportunités d'amélioration en vue d'optimiser les résultats de part et d'autre. Dans cet article, seul le modèle comportemental de premier niveau est présenté. Ce travail est mené en collaboration avec la Direction Nationale des Eaux et Forêts du Mali.*

MOTS-CLES : *approche systémique, besoins des parties prenantes, modèle comportemental, activités sylvicoles communautaires, durabilité.*

1 INTRODUCTION

L'exploitation communautaire des forêts mise en œuvre depuis quelques années dans différents pays africains afin d'améliorer le niveau de vie des populations locales tout en préservant la forêt, ne donne pas toujours les résultats escomptés (German *et al.*, 2010). Cette exploitation concertée de la forêt est pourtant vitale pour permettre le respect du plan d'aménagement territorial et la satisfaction des besoins des différentes communautés (Mbodj, 2014). Dans ce travail, seule la forêt de la Faya, plus vaste forêt classée du Mali, est considérée, l'objectif étant de représenter, en collaboration avec la Direction Nationale des Eaux et Forêts du Mali, les interactions et besoins des différentes parties prenantes afin de mettre en évidence les sources d'amélioration et les principaux points de blocage.

Ainsi, trois grandes approches peuvent être utilisées (Gazull, 2009) : une approche normative qui permet de mettre en évidence les règles de fonctionnement entre entités mais ne permet pas de façon simple de représenter l'évolution dynamique du système, une approche gravitaire qui permet de représenter les flux entre entités mais ne permet pas de représenter parfaitement la prise de décision interactive des entités et une approche comportementale qui permet d'identifier les états objectifs possibles et les événements internes ou externes causant le changement des différentes entités et de leurs interactions. Dans ce travail, nous avons choisi cette dernière approche car nous souhaitons d'une part

identifier les objectifs individuels de chaque partie prenante ; et d'autre part modéliser la dynamique du système. Cette approche associée aux systèmes multi-agents (SMA) nous semble intéressante pour simuler le comportement individuel des acteurs (ou agents) et les relations qu'ils entretiennent entre eux et avec la forêt (Moyaux *et al.*, 2006).

Bousquet et Gautier (1999) présentent deux modes d'intégration des territoires dans l'analyse de la gestion des ressources, où les agents interagissent pour accomplir des tâches et communiquer. Leur modèle reste restrictif car essentiellement orienté sur la population locale de paysans. Gaudieux *et al.* (2016) décrivent le système SIEGMAS (Stakeholders Interactions in Environmental Governance by a Multi-Agent System) afin de simuler les interactions entre parties prenantes dans la gouvernance communautaire des ressources naturelles. Leur modèle ne permet cependant pas de prendre en compte la valorisation de toutes les ressources comme les déchets (Osorio *et al.*, 2016). Purnomo *et al.* (2005) décrivent un modèle de simulation multi-agents d'une forêt gérée par une communauté en Indonésie. Cependant ni les propriétaires privés, ni la totalité des ressources valorisables ne sont traités.

C'est pourquoi, nous proposons de définir un modèle permettant l'identification et la caractérisation des parties prenantes et de leurs interactions afin d'aboutir à la création d'un modèle comportemental qui permettra par la suite d'améliorer la communication et la collaboration autour des différentes activités sylvicoles possibles et des ressources disponibles. L'idée principale est de pouvoir identifier toutes les caractéristiques et états

possibles (même ceux qui pourraient être imprévus par manque d'information que celles-ci soit incomplètes ou partielles ou par anticipation des aléas et modifications) afin d'avoir une approche systémique et intégrée du problème de gestion de la forêt de la Faya. En effet, la dimension sociétale qui est pourtant incontournable, doit être explicitement prise en compte en plus des éléments purement techniques et environnementaux. Dans cet article, nous limiterons notre étude à la définition des parties prenantes suivantes : l'Etat, les populations riveraines, l'exploitant principal et les propriétaires privés ainsi que la forêt de la Faya et son écosystème.

La suite de cet article se décompose de la façon suivante. Dans une deuxième partie, le contexte avec la mise en exergue des principales activités de sylviculture autour de la forêt de la Faya, le principe de concession et les problèmes rencontrés, ainsi que les contraintes seront présentés. Puis, les principales parties prenantes retenues dans ce travail, leur rôle, leurs besoins et interactions seront explicités. Ce qui nous permettra de présenter le modèle comportemental retenu dans une quatrième partie avant de conclure et de proposer quelques perspectives à nos travaux.

2 CONTEXTE

Le Mali, pays africain situé au sud du Sahara, est constitué à 26% par des forêts. La plus vaste est la forêt de la Faya, forêt classée de 80 000 mille hectares, par laquelle passe un bras du Niger et plusieurs marigots dont le plus important porte le nom de la forêt. Située au Nord-Est de Bamako, son climat est de type subsaharien (tropical) avec des galeries forestières, des savanes boisées, arborées et arbustives. La population autour de la forêt vit essentiellement de culture, d'élevage et de cueillette. L'exploitation forestière traditionnelle est sa principale source de revenus. L'inventaire forestier réalisé en 2005 révèle un volume total de bois sur pieds de 1 461 495 m³ répartis entre les formations naturelles (1 372 666 m³) et les plantations (88 829 m³). L'inventaire pastoral fait ressortir des potentialités au niveau des savanes boisées (588kg/ha) et de la savane arborée (665kg/ha) avec une charge moyenne acceptable de 0,166UBT/ha /an. Dans les années 1980, la Faya était caractérisée par une biodiversité importante en flore (espèces de combretacea, ochracea, polygalacea, meliacea, poacea etc.) et en faune (lions, panthères, antilopes de grande taille etc.). Mais suite à l'installation de champs de culture et à la présence importante du cheptel d'éleveurs autres que celui des agropasteurs des villages riverains, le sol se dégrade et appauvrit la forêt en ressources vivantes. Or, la position géographique de la forêt lui permet en principe d'être riche en faune et en flore. Cependant, comme celles-ci sont en nette régression depuis quelques années, le gouvernement a mis en place depuis 2013, une gestion décentralisée de la foresterie via une concession avec un exploitant (la Société de Gestion des Concessions Forestières et Fauniques - SGCF). Des coopératives regroupant les populations

locales sont également définies pour exploiter la forêt avec différentes activités vivrières et rentières mais sous contraintes de respect de l'aménagement territorial prévu.

Cependant, malgré ce système de concession ayant pour objectif d'optimiser l'exploitation foncière, la création d'emplois et la fabrication de produits de qualité issus de la forêt, de nombreux problèmes apparaissent : des feux de brousse (issus principalement d'un mauvais entretien de la forêt), un déboisement conséquent couplé à une exploitation forestière non durable des ressources collectives, une difficulté pour l'Etat de récupérer des informations fiables et donc de contrôler, des moyens réduits (ressources humaines et financières), le sentiment pour les populations locales de subir un traitement différent de l'exploitant et des propriétaires privés, et en règle générale un problème de communication et de collaboration. De plus, les exploitants sont dans l'ensemble peu professionnalisés.

C'est pourquoi, en collaboration avec la Direction Nationale des Eaux et Forêts du Mali, nous allons proposer un cadre permettant une exploitation plus judicieuse des ressources communes en y intégrant des cycles d'économies circulaires (comme la réutilisation des déchets en bois de chauffage, ou en pharmacopée traditionnelle) afin d'améliorer la rentabilité de la sylviculture autour de la forêt de la Faya. Le défi majeur sera de faire comprendre aux diverses parties prenantes que l'exploitation durable de la forêt est dans l'intérêt de tous. Il est à noter que dans cet article, nous présentons uniquement le modèle comportemental de premier niveau des parties prenantes principales : l'Etat, la Faya, les populations locales, l'exploitant et les propriétaires privés (les autres parties prenantes seront intégrées par la suite).

Les contraintes majeures sont les suivantes. Les populations ne disposent pas de ressources suffisantes pour vivre dignement ce qui fait qu'elles ont tendance à consommer trop des ressources disponibles et à détruire la forêt pour leur agriculture vivrière. En ce qui concerne l'Etat, la contrainte principale est son impossibilité à mettre en valeur ses forêts par manque de moyens. Les exploitants souhaitent exploiter la forêt dans toute sa dimension, mais les lois et réglementations les contraignent. Il en va de même pour les propriétaires privés. De plus, l'exploitation de la forêt se fait sur du court-terme sans visibilité sur le long-terme ce qui génère également beaucoup d'abus. Le système de sanctions mis en place par l'Etat ne permet pas de réajuster, ni d'améliorer le niveau de vie des populations ou de préserver les écosystèmes de la forêt. La mise en évidence des cycles d'économies circulaires permettra de faire comprendre aux différentes parties prenantes que celles-ci peuvent profiter doublement des ressources en optimisant leur utilisation et réutilisation (élaboration de projets porteurs pour les femmes en réutilisant certains déchets non utilisés actuellement pour faire des produits cosmétiques ou en créant de l'énergie, etc.) et en ayant des produits et services à moindre coût. En effet, notre idée est d'identifier clairement tous les éléments pouvant améliorer la durabilité de la forêt de

la Faya sans remettre radicalement en cause les activités des populations locales qui sont essentiellement pour elles des activités vivrières. Notre approche n'est donc pas uniquement orientée sur le cycle de vie des produits pouvant être extraits / fabriqués à partir des ressources naturelles disponibles dans la forêt de la Faya mais bien de tenter d'avoir une approche plus globale et à plus long terme pour identifier tous les bénéfices d'un point de vue économique, environnemental et sociétal sur la zone géographique étudiée et selon différents niveaux de participation des acteurs. Par exemple, des déchets (comme des feuilles) peuvent devenir des ressources à valoriser pour augmenter les capitaux manufacturés et naturels. Pour l'exploitant, une approche à plus long terme lui permettra d'asseoir son développement. De même, par exemple, la collaboration plus étroite entre l'Etat et l'exploitant permettra d'améliorer les activités d'écotourisme. La forêt constitue de plus une grande réserve d'oxygène et un stockage de grandes quantités de CO₂ donc, dans le cadre des accords internationaux, des compensations carbone possibles peuvent être mises en exergue avec une industrialisation plus forte de certaines activités sylvicoles.

La rentabilité économique, environnementale et sociale de l'écosystème industriel considéré dépend donc très fortement des liens quelle que soit leur nature entre les parties prenantes qui sont présentées dans ce qui suit.

3 PRINCIPALES PARTIES PRENANTES

De nombreux acteurs participent à la foresterie de la Faya : les populations et municipalités locales, les représentants de l'Etat et l'Etat, les pépiniéristes, les fournisseurs d'engrais (naturels et industriels), les gestionnaires de forêts dont l'exploitant, les propriétaires privés, les transformateurs du bois qui peuvent chercher à maximiser le volume de production et minimiser les coûts globaux logistiques (Ortiz et Thomas, 2003), les transporteurs, les universitaires, les associations, les coopératives, les ONG, les chasseurs, les maîtres de connaissances religieuses et non religieuses, les artisans, etc.

Parmi tous les acteurs cités, la plupart utilise la forêt de façon ponctuelle. C'est pourquoi, dans un premier temps, nous nous intéressons aux populations locales, à l'Etat, à l'exploitant et aux propriétaires privés ainsi qu'à la Faya. Ces acteurs ont chacun des rôles à jouer, entretiennent des liens et interagissent (Figure 1).

L'Etat souhaite réguler les activités, protéger les populations et conserver durablement la forêt et ses ressources. Pour y parvenir, l'Etat élabore des lois et contrôle la bonne application de ces lois. L'Etat établit les conventions et contrats d'exploitation et la délégation d'une partie de la gestion à un ou plusieurs

exploitants. Concernant les zones privées, l'Etat négocie aussi la cogestion et la participation des propriétaires à la collaboration dans le respect des textes en vigueur. Il recouvre aussi les taxes et impôts relatifs à l'exploitation de la forêt. Les fonds récoltés sont réutilisés dans le programme de développement économique et social.

La Faya veut se développer et préserver sa biodiversité (même si celle-ci peut évoluer dans le temps). Elle répond aux besoins des populations riveraines en produits ligneux et non ligneux et doit ainsi être entretenue, reboisée, etc., pour assurer la continuité des services écosystémiques offerts.

Les populations locales souhaitent améliorer leur condition de vie et revenu (besoins de compensation vivrières et rentières). Pour satisfaire ces besoins, elles vont exploiter la forêt en produits ligneux et non ligneux, consommer de façon directe et indirecte (après transformation) les produits offerts par la forêt. Les produits ligneux sont utilisés pour la construction, la menuiserie traditionnelle, l'artisanat, le chauffage, et/ou transformés en charbon de bois, en médicaments (transformation des feuilles, écorces et racines), en produits cosmétiques (comme le karité) et en gris-gris (amulettes porte bonheur traditionnelles). Les produits non ligneux sont des fruits, du miel, des animaux sauvages, etc., qui sont tous de la matière nourricière pour la population. La collaboration entre toutes les parties prenantes est donc prépondérante pour assurer l'exploitation durable de la forêt.

L'exploitant est un investisseur privé dont l'objectif principal est de rentabiliser ses investissements, sous la contrainte des réglementations définies. L'exploitant doit également développer des activités d'écotourisme et des emplois. Il en va de même pour les propriétaires privés.

Dans ce qui suit nous présentons le modèle entités-relations de premier niveau proposé pour représenter notre système de sylviculture communautaire

4 MODELE COMPORTEMENTAL

Les entités principales sont : la Faya (incluant ses composantes privée et publique) fournissant des matières premières, l'Etat, la population (regroupée en coopératives et représentée par une municipalité), la SGCF et les propriétaires privés (Figure 2).

L'Etat agit sur la forêt par un contrôle de l'effectivité des activités autorisées dans la forêt, de façon directe par les gardes forestiers et indirecte par des rapports. L'Etat fait passer un appel d'offre pour le recrutement d'un exploitant pour aider à la gestion de la Faya, puis, initie et signe le contrat de concession avec l'exploitant et informe la population. L'exploitant pourra commencer ses activités.

peuvent être sanctionné en cas de manquement. Il peut selon la gravité de ses infractions être verbalisé,

amendé financièrement, perdre sa concession et/ou ses partenariats. (Figure 4).

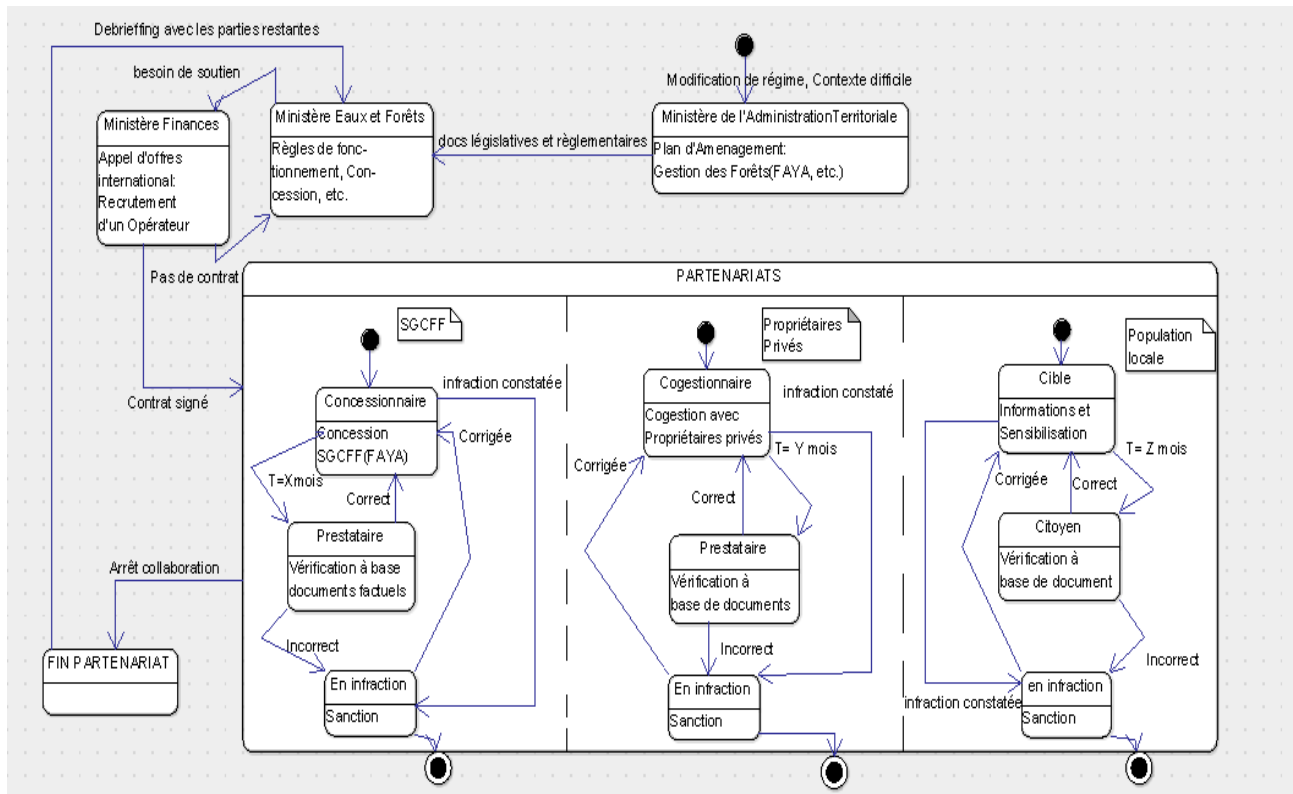


Figure 3 : Diagramme Statechart de l'État

Les populations cultivent dans la forêt, utilisent des produits ligneux et non ligneux.

En retour elles doivent entretenir durablement la forêt.

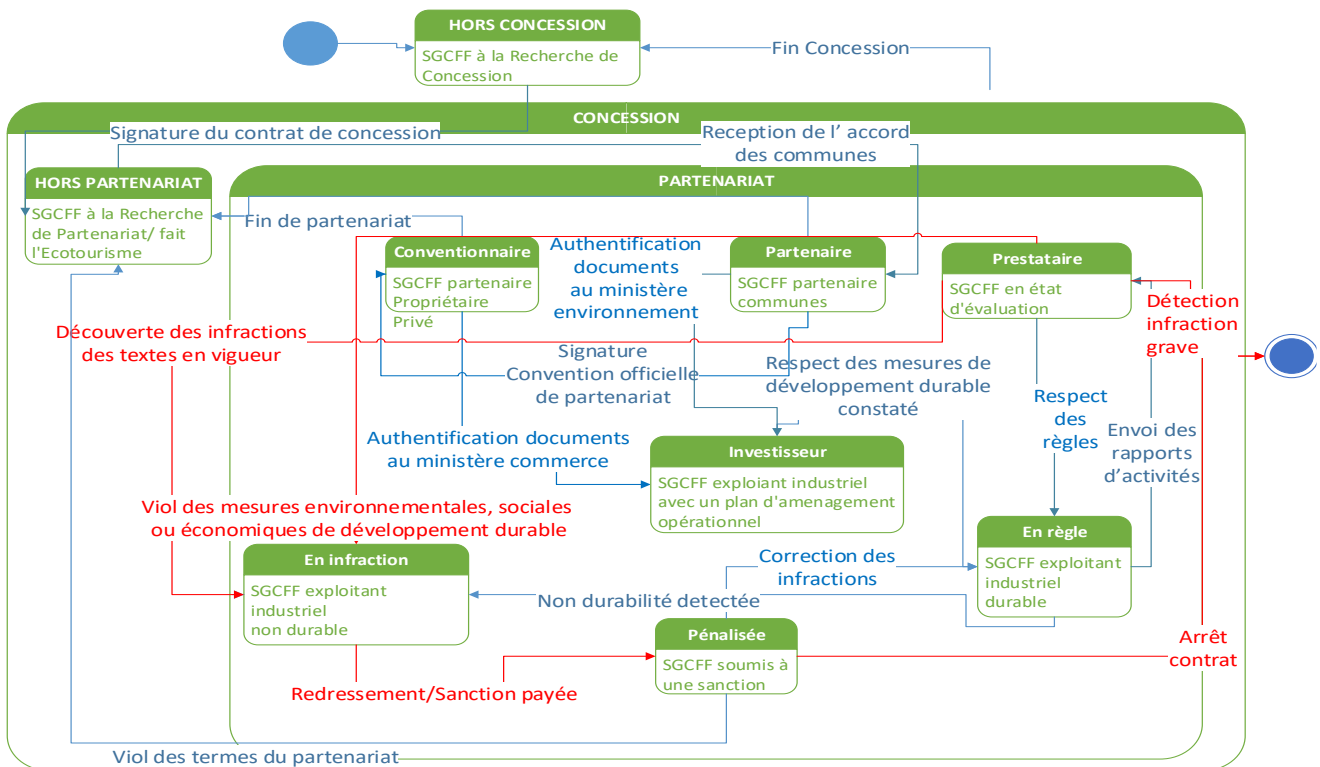


Figure 4 : diagramme Statechart de l'exploitant SGCF

5 CONCLUSIONS

Cette étude constitue l'introduction vers un outil d'aide à la décision qui permettra d'améliorer les profits des différentes parties prenantes dans la gestion de la Faya. L'objectif premier est de formaliser le système afin d'en mesurer précisément les faiblesses et manques. Cet article présente donc le premier pas du processus d'amélioration des activités de foresterie dans le contexte d'un pays en voie de développement avec la prise en compte d'une population riveraine à la fois très dépendante de la forêt et très impactante sur celle-ci. Dans cette optique nous avons établi les statecharts de l'État et de la société d'exploitation.

Pour la suite, nous allons définir les statecharts complets des différentes entités, ainsi que le modèle dynamique de leurs interactions. En parallèle, nous travaillerons étroitement avec la Direction Nationale des Eaux et Forêts du Mali ainsi qu'avec l'exploitant sous convention afin de collecter les données nécessaires au dimensionnement et paramétrage de notre modèle. Ces données collectées seront également discutées lors de réunions avec des représentants de la population locale ainsi qu'avec des clients identifiés des produits issus de la Faya (comme la société exploitant le bois coupé non détaillé dans ce travail). Puis, à partir d'études déjà menées dans d'autres domaines que le génie industriel, nous envisagerons différentes propriétés et phénomènes résultants de modifications sur l'écosystème de la forêt de la Faya et générant de nouveaux comportements des acteurs (comme par exemple le poids de lobbies étrangers pesant sur les activités agricoles, des modifications dans les besoins en produits ligneux et non ligneux, des migrations de population, etc.). Cela nous permettra de mieux identifier et comprendre notre système ainsi que ses différents états.

Le modèle comportemental incluant toutes les dynamiques possibles validé, nous proposerons des améliorations incluant les boucles d'économie circulaire identifiées et les moyens d'une collaboration plus étroite entre les différents acteurs. Pour parvenir à la définition de ces améliorations, nous étudierons le système sous différentes hypothèses et évolutions dynamiques afin d'en augmenter les performances sous un aspect durable, c'est-à-dire en le rendant viable économiquement pour chaque entité. Enfin, via des simulations basées sur un système multi-agents, nous validerons nos solutions et transformerons alors l'outil de simulation en un outil d'aide à la décision pour aider les parties prenantes (principalement l'Etat) à gérer la Faya durablement

REMERCIEMENTS

Les auteurs souhaitent remercier la Direction Nationale des Eaux et Forêts du Mali pour les différentes informations ainsi que l'aide apportée dans le cadre de ce travail ainsi que les correcteurs pour leur retour qui a permis d'améliorer cet article.

Ce travail a été rendu possible grâce au financement de la thèse de doctorat de Souleymane DICKO par l'Ambassade Française au Mali.

REFERENCES

- A Ortiz Virna & A. Thomas, (2003). Méthodes et outils d'optimisation de planification tactique dans une chaîne logistique, *MOSIM'03*, 23 - 25 avril 2003 - Toulouse (France), p. 1-7.
- Bousquet, F. et D. Gautier. (1999). Comparaison de deux approches de modélisation des dynamiques spatiales par simulation multi-agents : les approches "spatiale" et "acteurs", *Cybergeog: European Journal of Geography* <http://cybergeog.revues.org/2389>
- Gaudieux, A., Cazal, J. D., Grondin, D., Sebastien, V. et R. Courdier. (2016). SIEGMAS : Un simulateur multi-agents pour l'étude des territoires de l'Océan Indien. *RIODD 2016*, July 2016, Saint-Étienne, (France).
- Gazull, L. (2009). *Le bassin d'approvisionnement en bois-énergie de Bamako – une approche par un modèle d'interaction spatiale*. PhD Géographie, Université Paris Diderot, 24 Février 2009.
- German, L. A., Karsenty, A. et A.-M. Tiani. (2010). Gouverner les forêts Africaines à l'ère de la mondialisation. *Center for International Forestry Research*, Earthscan – London.
- Mbodj, F. (2014). Gestion des ressources forestières au Mali et revendication des territoires. *Anthropologies et Développement*, p. 43-67, disponible à <http://anthropodev.revues.org/467>
- Moyaux, T., Chaib-Draa, B., et D'Amours, S. (2006). Supply chain management and multi-agent systems: an overview. *Multi-agent-based supply chain management*, p. 1-27.
- Osorio, J., Roy, D., Hennequin, S., Stirnweiss, C. et S. Winckelmuller. (2016). Pilotage durable des déchets ménagers dans un cadre d'économie circulaire. *Journal Européen des Systèmes Automatisés*. 49/6, p.635-657.
- Purnomo, H., Mendoza, G., Prabhu, R., and Yasmi, Y., (2005). Developing Multi-Stakeholder Forest Management Scenarios: A Multi-Agent System Simulation Approach Applied in Indonesia. *Forest Policy and Economics*, 7(4), p. 475-491.

APPROCHE SYSTEMIQUE POUR LE DEVELOPPEMENT DE SYSTEMES EMBARQUES CONNECTES - CONCEPTION ET PRODUCTION D'UN MOOC

C. BARON

LAAS-CNRS, Université de Toulouse, CNRS, INSA
7 avenue du Colonel Roche, 31500 Toulouse, France
claude.baron@laas.fr

B. DANIEL ALLEGRO

Expert Conseil en Penser Système et Ingénierie Système
31620 Castelnau d'Estrétefonds, France
brigitte.daniel.allegro@gmail.com

RESUME: *Un système embarqué connecté intègre des éléments matériels et logiciels et interagit en temps réel avec son environnement. Le développement d'un tel système nécessite une approche systémique, basée sur la pensée systémique et l'ingénierie des systèmes. La pensée systémique démontre la nécessité de suivre une approche pluridisciplinaire et transdisciplinaire (non seulement axée sur la technologie, mais aussi tenant compte des engagements sociétaux ou réglementaires, pour n'en citer que quelques-uns), essentielle pour explorer le domaine des systèmes embarqués. L'ingénierie des systèmes permet d'acquérir une méthode qui fournit les moyens de mettre en place un plan cohérent, intégrant pleinement le développement d'un produit dans son contexte industriel. Cet article reflète le travail d'une équipe d'enseignants, de chercheurs, de consultants, d'experts industriels pour produire un MOOC sur les systèmes embarqués connectés. Les différentes études de cas traitées dans ce MOOC proposent une méthode pour développer des compétences spécifiques dans ce domaine, et pour identifier et résoudre une variété de problèmes qui sont soulevés tout au long du MOOC. Le MOOC a déjà attiré plus de 10 800 apprenants.*

MOTS-CLEFS : *système embarqué ; système connecté ; interdisciplinarité ; MOOC ; approche systémique ; penser système ; ingénierie des systèmes.*

1 INTRODUCTION

Un système embarqué est généralement défini comme un système électronique et informatique autonome, souvent temps réel, spécialisé dans une tâche bien précise. Selon le rapport de l'OPIEC (Observatoire Paritaire Informatique Ingénierie Etudes Conseil) en 2014 (OPIEC, 2014), un Système Embarqué est « un système informatisé, spécialisé et autonome, qui constitue une partie intégrante d'un système plus large ; il combine matériel et logiciel ». En anglais, on parle de « embedded system », ce qui signifie enfoui, traduisant le côté non visible d'un tel équipement. Ses ressources sont généralement limitées par l'hôte qui l'embarque (par exemple limitation en encombrement, masse, autonomie énergétique ou moyen de communication).

Un système connecté est lui-même composé d'autres systèmes ayant des rôles précis : un système d'identification et de captation de données (température extérieure, rythme cardiaque, etc.), un système de transmission de données alimentant une application « intelligente », une interface (comme un smartphone) pour piloter l'application (de Baynast, 2017) En théorie, tout système peut être connecté. On parle d'Internet des Objets ou en anglais d'Internet of Things (IoT).

Selon le même rapport de l'OPIEC, les systèmes embarqués sont au carrefour de plusieurs tendances technologiques de pointe. Ils s'appuient notamment aujourd'hui sur les capacités du « cloud computing » et du « big data » (récupération de données à partir de capteurs embarqués. stockage et exploitation de ces données). Certaines

évolutions relèvent clairement des tendances de marché ; parmi celles-ci, l'Internet des Objets est la tendance la plus clairement identifiée. Les opportunités de marché et les applications des systèmes embarqués connectés sont ainsi nombreuses (par exemple la voiture autonome, l'usine du futur, casque connecté de chirurgien, etc.).

Cependant, des challenges, notamment en termes de sécurité, doivent encore être relevés. D'où l'importance de savoir évaluer la sécurité des systèmes, et de prendre en compte plus largement la Sécurité de Fonctionnement, couvrant également les aspects fiabilité, disponibilité et maintenabilité, pour intégrer ces challenges tout au long du cycle de vie du système : depuis la conception, le développement et l'exploitation du système, voire même jusqu'à son retrait de service.

Selon (OPIEC, 2014), la complexité actuelle des systèmes impose des méthodes de conception et de validation rigoureuses sur lesquelles les acteurs doivent gagner en maturité. C'est à cet objectif que contribue notamment l'approche systémique, qui vise à appréhender les relations des systèmes complexes avec leur environnement, leur fonctionnement et leurs mécanismes pour optimiser leur cycle de vie. Selon Mitchell (Mitchell, 2009), il n'existe pas aujourd'hui de science unique de la complexité ni de théorie unique de la complexité. Mitchell a tenté toutefois d'identifier des propriétés communes aux systèmes complexes à savoir, un comportement collectif complexe, un traitement et une transmission de l'information, une adaptation via l'apprentissage ou l'évolution. Nous considérons que les systèmes embarqués connectés partagent les propriétés des systèmes complexes (la complexité étant considérée au sens de Mitchell).

Cet article témoigne ainsi d'un travail de 18 mois (de janvier 2015 à juin 2016) mené par une équipe d'enseignants, de chercheurs, de consultants et d'industriels, experts dans des disciplines différentes, mais tous convaincus de l'intérêt de l'approche systémique pour la conception de systèmes embarqués connectés. Ce travail a abouti à la mise en ligne d'un MOOC (Massive Open Online Course), le MOOC « Systèmes embarqués et objets connectés » (Baron et al., 2016), hébergé par la plateforme nationale française FUN (France Université Numérique) qui vise à promouvoir l'utilisation des cours en ligne depuis octobre 2013. Cette plateforme est une référence dans le monde francophone. Elle fédère les initiatives des établissements d'enseignements supérieurs qui le souhaitent. Elle répond aux besoins de formation d'un large public, étudiants, salariés, retraités, demandeurs d'emploi, en formation initiale ou en apprentissage tout au long de la vie. Les objectifs, scientifiques et pédagogiques, qui ont motivé notre engagement dans ce projet, sont doubles. Le premier est de partager nos points de vue sur les systèmes embarqués connectés afin de mutualiser entre établissements de formation une présentation générale de ces systèmes dans leurs contextes opérationnels, et de constituer un socle commun qui pose les fondamentaux pédagogiques relatifs à la maîtrise de ces systèmes. Le second objectif est de montrer, en s'appuyant sur des études de cas, comment aborder de tels systèmes, avec méthode et rigueur, grâce à une approche systémique.

En effet, l'enseignement des systèmes embarqués connectés repose essentiellement aujourd'hui sur la proposition de modules d'enseignement disciplinaires, sur la base de disciplines scientifiques classiques telles que l'informatique et réseaux, l'électronique, l'automatique, ou la mécanique. Ce MOOC propose d'une part d'ouvrir à d'autres disciplines (mobilisant d'autres connaissances scientifiques et techniques, mais aussi juridiques, économiques et sociétales) et d'autre part d'intégrer des analyses transverses (comme la sûreté de fonctionnement) permettant ainsi de relever les challenges identifiés dans le rapport de l'OPIEC. De ce fait il est très novateur par le caractère structurant de l'approche systémique qui permet une intégration multi domaines et multi métiers.

Le MOOC a été ouvert deux fois par FUN, en juillet 2016 et en octobre 2017. Il a trouvé un écho important auprès d'un large public, totalisant 4 722 inscriptions en 2016 et 6 530 inscriptions en 2017. Sur la seconde session, 358 attestations avaient été délivrées, soit 5,84% des inscrits, ce qui correspond au taux moyen de suivi des MOOC. Ce taux de délivrance des attestations ne correspond pas forcément à des abandons en cours de route. Les raisons peuvent être en effet multiples : les unes proviennent de l'organisation du MOOC (inscriptions tardives (ainsi sur la seconde session, 439 se sont inscrits après la date de fermeture des QCM), niveau de difficulté, durée de 10 semaines, qualité des supports, etc.), les autres correspondent au profil des apprenants (temps à consacrer au travail personnel requis, disponibilité à assurer sur des dates imposées, situation profes-

sionnelle, niveau d'étude, motivation pour obtenir l'attestation, etc.).

La section 2 de l'article introduit l'approche systémique. La section 3 présente le MOOC ; elle définit la finalité, les objectifs et la structure du MOOC organisée autour de trois études de cas, le public visé ainsi que le mode d'évaluation. Les sections 4, 5 et 6 présentent respectivement en détail chaque étude de cas. La section 7 propose un bilan, notamment une synthèse de la progression pédagogique dans l'étude des cas, ainsi que des retours d'expérience de la session 2016 des pédagogues et apprenants, qui ont conduit à proposer ce MOOC dans un cursus de formation d'ingénieur dès 2017. La section 8 conclut sur l'intérêt de l'approche systémique et les perspectives pédagogiques offertes par le format du MOOC à l'enseignant qui, en contrepartie, adaptera ses pratiques à ce nouveau support.

2 L'APPROCHE SYSTEMIQUE POUR LES SYSTEMES EMBARQUES CONNECTES

Dans le domaine des systèmes embarqués connectés, les échanges d'informations entre une multitude d'agents, virtuels ou réels, dans un contexte fortement évolutif, nécessite aujourd'hui la mise en œuvre d'une approche systémique, en s'appuyant d'une part sur la pensée systémique (ou le Penser Système), consistant à observer et appréhender globalement une situation pour la comprendre (Daniel-Allegro, 2014a) et d'autre part sur l'Ingénierie Système (Meinadier, 2012) pour mettre en œuvre l'ingénierie du système de façon interdisciplinaire et collaborative.

Le guide « Introduction au Penser Système » (Daniel-Allegro, 2014b) de l'Association Française d'Ingénierie Système (AFIS) caractérise ainsi l'attitude du Penseur Système, qui cherche à comprendre des situations complexes, des événements ou des changements de son environnement. Le Penser Système est « la capacité à identifier et à appréhender un tout, aussi bien la structure d'un système que son comportement et ses relations avec son environnement ».

Le guide « Découvrir et Comprendre l'Ingénierie Système » (Meinadier, 2012) de l'AFIS définit l'Ingénierie Système (IS) comme « une approche collaborative et interdisciplinaire, fondée sur la science et l'expérience, qui englobe les activités pour concevoir, développer, faire évoluer et vérifier un ensemble de processus, produits et compétences humaines apportant une solution globalement optimisée à des besoins identifiés et acceptables par l'environnement ». L'ingénierie des systèmes couvre ainsi tous les processus de la norme ISO15288 (ISO/IEC/IEEE 15288, 2015), à savoir les processus contractuels, techniques, de management de projet et supports à l'organisation.

L'approche systémique, intégrant le Penser Système et l'Ingénierie Système, est incontournable pour aborder les systèmes complexes. Elle repose sur une approche collaborative dans un contexte d'entreprise étendue où sont maîtrisées les activités de management de l'ingénierie telles que l'analyse du marché, la compréhension des

missions et des besoins opérationnels, la définition d'une architecture cohérente, ainsi que le management de la qualité, de la traçabilité, des évolutions et de la configuration, de la production, des acquisitions, etc. (Lawson, 2010).

3 PRESENTATION DU MOOC

Cette section précise la finalité et les objectifs du MOOC, décrit comment il est organisé, indique le public ciblé et explique les modes d'évaluation choisis.

3.1 Finalité et objectifs

La finalité de ce MOOC est de promouvoir l'approche systémique en considérant les challenges posés par les systèmes embarqués connectés. Ils sont notamment de savoir appréhender une situation complexe et de caractériser les problèmes qui se posent dès lors que l'on considère l'interaction de tels systèmes avec leur environnement. Les difficultés viennent de l'imbrication de plusieurs secteurs d'activité et de disciplines qu'il est indispensable d'appréhender dans une approche globale. Il faut identifier et caractériser dans l'environnement les facteurs économiques, réglementaires, sociétaux, technologiques influençant plus ou moins fortement la conception et l'utilisation du système dans son cycle de vie.

Il faut aussi rechercher précisément et caractériser l'ensemble des acteurs aussi appelés « parties prenantes » (clients, fournisseurs, contributeurs de différents rôles, détracteurs, opposants, sponsors, organismes de qualification, organismes de santé, de sûreté...) et leur niveau d'intervention dans le cycle de vie du système. L'oubli d'acteurs peut avoir une incidence catastrophique sur le déroulement d'un projet. Si par exemple on ne prend pas en compte les normes environnementales en vigueur, interdisant l'utilisation de produits délétères pour l'environnement, on peut être amené à revoir le système.

Les objectifs du MOOC sont de faire comprendre la nécessité d'intégrer ces différentes disciplines dans le développement des systèmes embarqués connectés et de montrer comment naviguer dans différents niveaux d'abstraction, du plus général au plus détaillé, et d'une modélisation à un produit.

Pour atteindre ces objectifs, le MOOC propose une façon innovante d'aborder ces systèmes par une approche systémique globale et méthodologique, intégrant notamment les aspects socio-économiques aux aspects technico-scientifiques, décloisonnant ainsi les disciplines.

En sus, le MOOC prend le parti d'appuyer la progression pédagogique dans la démarche sur l'analyse d'études de cas réels, issus de domaines d'applications représentatifs, et les met au centre des échanges pour illustrer et appliquer l'approche systémique. Par le biais de ces exemples, il permet de diffuser des connaissances et des pratiques générales sur les systèmes embarqués connectés et d'introduire des pistes pour donner envie aux apprenants d'explorer ou d'approfondir ce domaine.

3.2 Structure

La structure du MOOC a été conçue de façon à favoriser un apprentissage progressif. Elle repose sur trois études de cas illustratifs choisis dans des domaines d'application différents, l'habitat intelligent, l'agriculture et l'aéronautique. Celles-ci permettent l'introduction de différents concepts d'approche systémique. Elles s'appuient sur un même cadre méthodologique générique inspiré de l'ISO 15288 (ISO/IEC/IEEE 15288, 2015) et mettent l'accent sur la partie conception et développement d'un système embarqué connecté.

Ce cadre de référence guide l'apprenant dans le développement des systèmes par le biais d'étapes. Une approche globale d'une situation donnée permet à chaque fois de cerner le périmètre de l'étude pour aborder ainsi de façon cohérente les étapes ultérieures. Les processus techniques guident les étapes d'analyse et de conception préliminaires puis détaillées, d'intégration du produit, de vérification et de validation. Ces étapes intègrent aux processus techniques les processus de management de projet, les processus contractuels (dépôt de brevets, par exemple) ou de support à l'organisation (gestion des infrastructures, de la qualité, etc.)

Chaque étape comporte plusieurs modules pédagogiques constitués chacun d'une vidéo, d'un script et d'un support de présentation associé sous forme de diapositives ainsi que de ressources complémentaires (documents, liens internet, glossaires, etc.).

Au total, 16 interviews, 47 vidéos représentant une durée cumulée de 7h30, une vingtaine d'experts de différentes disciplines spécifiques (comme l'énergie ou les réseaux de communication) ou transverses (comme la sûreté de fonctionnement) ont contribué à sa réalisation.

Chaque étude de cas sert de prétexte pour introduire des concepts ou des pratiques à des niveaux d'approfondissement différents : un détecteur de fumée connecté réalisable chez soi, un système industriel d'aide à la navigation embarqué dans différents types d'engins roulants agricoles, des calculateurs embarqués critiques communicants pour améliorer le confort des passagers dans les avions. Ils ont été conçus spécifiquement pour servir d'illustration au MOOC, sur la base de systèmes existants ou totalement innovants (comme le détecteur de fumée connecté) au moment de la création sur MOOC.

Le choix de domaines d'application différents permet d'ouvrir à des analyses plus spécifiques (exemple de l'interface homme – machine, de l'ergonomie, de la propriété intellectuelle).

Sur le plan des savoir-faire, les cas illustratifs choisis offrent une progression selon trois axes : la maturité des processus, le niveau d'industrialisation et la sûreté de fonctionnement.

1) En ce qui concerne la maturité des processus, nous en distinguons trois niveaux : nous partons d'une première étude de cas ne nécessitant aucune formalisation des processus, puis considérons le cas d'un système commercialisé où les processus sont simplement décrits,

pour enfin aborder le cas où, en avionique, les processus sont formalisés et maîtrisés, faisant l'objet de qualification ou certification par des organismes experts.

2) En ce qui concerne le niveau d'industrialisation, nous démarrons par le cas d'un projet individuel consistant à assembler différents composants du commerce chez soi pour réaliser un prototype de détecteur de fumée connecté, puis considérons le cas d'un projet de système d'aide à la navigation dans un véhicule roulant agricole développé par une PME, pour finir avec le cas d'un projet en avionique mené par un grand groupe industriel.

3) La sûreté de fonctionnement est abordée dans ses différentes composantes (sécurité, disponibilité, fiabilité et maintenabilité / utilisabilité) au travers des trois cas, en mettant l'accent sur certaines composantes selon le cas traité. Par exemple, le cas du détecteur de fumée aborde intuitivement la fiabilité et la maintenabilité, alors que le cas du système critique avionique met l'accent sur la disponibilité et la sécurité.

3.3 Public ciblé

Le MOOC a été conçu pour adresser un large public. Il est en cela parfaitement en cohérence avec les objectifs de diffusion de la plateforme FUN. Il s'adresse notamment à :

- Des techniciens, cadres et ingénieurs de l'industrie,
- Des enseignants / enseignants chercheurs / doctorants et des étudiants en formation initiale,
- Des demandeurs d'emploi en phase de reconversion ou en demande de formation,
- Des retraités,
- Ou toute personne curieuse de découvrir le monde des systèmes embarqués connectés, indépendamment de son niveau d'études.

De fait, le profil des participants à la deuxième session montre que près de 80% des apprenants étaient en activité ou en cours d'études. 80% avaient un niveau d'études supérieur à Bac+2 (54,28% ayant Bac+4/5, 9,39% ayant un niveau supérieur à Bac+5), 9% avaient un niveau d'études inférieur à Bac+2.

3.4 Evaluation

A la fin de chaque étude de cas est proposé un QCM (au total 85 questions) pour tester les connaissances. Il permet à chacun d'évaluer son niveau de compréhension et d'appropriation des concepts abordés dans le cas. Le quizz propose deux essais pour répondre aux questions. Les apprenants ont ainsi la possibilité de revoir certaines vidéos ou de relire les scripts s'ils échouent au premier essai. Une note est donnée à l'issue de chaque QCM. Un exercice final d'évaluation par les pairs peut être proposé en complément ; il consiste alors à traiter une situation transposée des études de cas. A l'issue de ces deux types d'évaluation, la réussite est évaluée et un certificat de suivi délivré, indiquant la progression dans les apprentissages.

4 LE CAS D'UN DETECTEUR DE FUMEE CONNECTE

La première étude de cas considérée appartient au domaine de la maison intelligente. Le système a pour fonctions d'effectuer la détection de fumée dans un logement et d'alerter.

4.1 Mise en situation

La mise en situation est la suivante : la loi n° 2010-238 du 9 mars 2010 rend obligatoire l'installation de détecteurs de fumée dans les lieux d'habitation. Joe, citoyen consciencieux, décide d'en acheter un et de l'installer chez lui... Il décide de se mettre en conformité avec la législation. Il achète et installe un détecteur de fumée à son domicile. Malgré cette précaution sa maison brûle : pendant son absence, un incendie s'est déclenché à cause d'un grille-pain défaillant.

Dans cette mise en situation, le détecteur de fumée a bien fonctionné, mais personne n'a réagi car personne n'a été alerté, et la maison de Joe a quand même brûlé ...

4.2 Problématique traitée et démarche

La question que nous considérons alors est de connecter le détecteur de fumée pour le rendre intelligent, capable d'alerter quelqu'un à distance en cas de détection de fumée. Ainsi, Joe décide de concevoir et fabriquer un détecteur de fumée connecté correspondant à son besoin spécifique, l'alerter à distance en cas de fumée. Il s'interroge avant de se lancer : De quels composants ai-je besoin ? Où me les procurer ou puis-je en fabriquer certains moi-même ? Combien cela va-t-il me coûter ? Ai-je besoin d'outillage spécifique ? Comment m'organiser dans mon travail pour concevoir et assembler mon système ? Comment m'assurer que le détecteur fonctionnera quand il le faut et qu'il ne se déclenchera pas intempestivement sans raison ? Comment être sûr d'avoir pensé à tout dans mon projet ? Est-ce que mon assurance va m'assurer si c'est moi qui ai fabriqué mon détecteur ?

Nous accompagnons Joe dans sa démarche, durant laquelle il va peu à peu s'organiser pour :

- Définir de façon globale les interactions du système embarqué connecté avec son environnement pour en dégager les fonctions qu'il doit remplir (Etape « Approche globale », module « Penser Système »),
- Déterminer les fonctions et la structure (Etape « Analyse préliminaire », module « Analyse fonctionnelle et organique »),
- Étudier plus spécifiquement son ergonomie, son alimentation et les solutions possibles pour le rendre communicant (Etape « Analyse par domaine », modules « Energie, Communication, Ergonomie »),
- Intégrer les différents composants et fabriquer un prototype vérifiable empiriquement (Etape « Synthèse », modules « Découpage fonctionnel organique, intégration » et « Vérification et validation »).

Des pistes d'approfondissement sur l'usine du futur, les enjeux de société, les emplois dans la filière numérique, etc.) sont fournies en fin de cette étude de cas.

4.3 Démonstration de l'intérêt de l'approche systémique

L'objectif pédagogique qui découle de cette étude de cas est de sensibiliser au fait que dans la démarche d'ingénierie système de conception, développement et utilisation d'un produit technique (en l'occurrence ici un système embarqué connecté) il faut prendre en compte un ensemble de paramètres :

- Techniques : technologie utilisable, à quel niveau pour garantir le juste fonctionnement ?
- Economiques : faire ou faire faire ? Quels avantages à développer soi-même ?

Mais aussi :

- Spatio-temporels : organisation et déroulement des phases de conception, fabrication, qualification, déploiement, maintenance, retrait en service,
- Sociétaux : acceptation du produit par son utilisateur comme solution efficace à un problème exprimé,
- Écologiques : gestion de la fin de vie, retrait du produit, recyclage...
- Légaux : critères et processus d'homologation, réglementation, normes et standards applicables.

5 LE CAS D'UN SYSTEME INDUSTRIEL D'AIDE A LA NAVIGATION, EMBARQUE DANS UN VEHICULE AGRICOLE

La seconde étude de cas considérée appartient au domaine agricole.

5.1 Mise en situation

Nous nous plaçons dans la situation d'un employé d'une PME qui veut concevoir, fabriquer et commercialiser un calculateur d'aide au pilotage, embarqué dans un véhicule roulant agricole. Ce système sera vendu pour être intégré dans différents véhicules, indépendamment de la mission agricole (labour, désherbage, tonte, épandage, ...).

La mission du système est de suivre une trajectoire déterminée à l'avance. Pour cela, le calculateur se localise, gère les déplacements du véhicule et détecte également d'éventuels obstacles sur la trajectoire. Pour se localiser, il traite les données de capteurs (un GPS par exemple, qui lui permet de connaître sa position en temps réel). Pour piloter les déplacements, il applique des algorithmes de suivi de trajectoire qui permettent d'envoyer des consignes aux actionneurs (volant, accélérateur, freins). Pour détecter les obstacles, il va par exemple utiliser une caméra ou un télémètre laser. Le système communique avec un opérateur distant afin que celui-ci dispose en permanence d'une vision de l'activité en cours et de l'état de santé du système. L'opérateur peut aussi modifier la cartographie du trajet à réaliser et stopper l'engin à distance en cas de nécessité.

5.2 Problématique traitée et démarche

Nous abordons ici le développement d'un produit industriel tiré par les besoins du marché (« market pull » en anglais). Les processus de développement et de gestion du projet sont caractérisés, encadrés par des standards, les moyens de production sont opérationnels. La sûreté de fonctionnement ne peut plus être abordée intuitivement comme dans le cas du détecteur de fumée. Ses bases sont introduites ici pour être développées plus tard dans le cas 3 sur les aspects sécurité.

Le cheminement suivi amène tout d'abord à représenter le système dans son contexte général, puis à s'interroger sur la meilleure approche pour définir et développer ce système dans son environnement économique.

Nous accompagnons l'employé de PME dans sa démarche durant laquelle il va peu à peu s'organiser pour :

- Introduire la problématique générale des systèmes de pilotage et définir l'ensemble des interactions du système embarqué connecté avec son environnement pour en dégager les fonctions qu'il doit remplir (Etape « Approche globale », module « Penser Système »),
- Déterminer les fonctions et la structure (Etape « Analyse préliminaire », module « Analyse fonctionnelle et organique »),
- Étudier plus spécifiquement l'aide au pilotage (Etape « Analyse par spécialité », modules « Contrôle commande, Géolocalisation, Interface Homme Système »),
- Prendre en compte les analyses transverses (Etape « Transversal », modules « Propriété intellectuelle, Approvisionnement, Obsolescence, Sûreté de Fonctionnement »),
- Étudier les principes d'intégration et de fabrication du système embarqué connecté. Intégrer les différentes disciplines dans l'analyse fonctionnelle et organique du système afin de concevoir un produit répondant aux performances requises et mettre en œuvre une production de petite série (Etape « Synthèse », modules « Découpage fonctionnel organique, Intégration », et « Vérification et Validation »).

Des articles sur les engins agricoles autonomes ou l'Internet des Objets permettent d'approfondir l'étude de cas.

5.3 Démonstration de l'intérêt de l'approche systémique

L'objectif pédagogique qui découle de cette étude de cas est de faire découvrir :

- Le développement d'un système embarqué connecté avec la nécessité de formaliser et de gérer les processus,
- La problématique relative au passage d'un exemplaire unique à une production en petite série,
- L'organisation des ressources au sein de l'entreprise pour développer le produit,

- La nécessité de formaliser et de gérer les processus (incluant maintenance, support, soutien opérationnel, retrait de service, abordant notamment le respect des normes environnementales et le recyclage des composants),
- L'organisation de la qualité, incluant la considération des questions d'homologation.

Le travail en équipe pluridisciplinaire dans cette étude de cas amène celle-ci à partager certains éléments techniques spécifiques d'un domaine de spécialité. Les membres de l'équipe s'enrichissent donc également en élargissant leur domaine de compétence. Nous apprenons ainsi que, sur le plan des interfaces homme/machine, l'utilisateur doit toujours être capable de traiter l'information qu'un système génère ou affiche ; pour cela l'information doit être contextuelle, compréhensible (c'est-à-dire non-ambiguë) et rapidement assimilable. De façon plus spécifique à l'étude de cas, nous apprenons que pour faire de la géolocalisation, le contexte d'utilisation est primordial dans la définition de la solution retenue (on va par exemple éviter une solution basée sur les ondes radios dans un environnement citadin avec des bâtiments pouvant réfléchir ces ondes). Nota : on voit émerger le rôle capital de l'architecte système (intégrateur et chef d'orchestre des processus et des disciplines système) dans un milieu industriel et son positionnement par rapport au chef de projet (garant de la tenue des objectifs de performance du projet). Ces deux rôles distincts peuvent être incarnés par une seule personne dans une petite structure et par des personnes différentes dans une grande structure, nécessitant un travail collaboratif.

6 LE CAS DE CALCULATEURS EMBARQUÉS COMMUNICANTS DANS UN AERONEF

La troisième étude de cas considérée appartient au domaine de l'aéronautique civile.

6.1 Mise en situation

La mise en situation est cette fois la suivante : vous êtes ingénieur dans une société de service. Votre société vous confie une mission de développement d'un système embarqué connecté dans le domaine avionique. Avant le début de cette mission vous avez besoin d'avoir une vue globale sur l'environnement spécifique aux systèmes embarqués avioniques et leurs contraintes de développement.

Le système considéré dans cette étude de cas a pour mission l'amélioration du confort des passagers dans un avion. Le système est complexe et utilise pour assurer sa mission un réseau de capteurs (notamment des accéléromètres), les calculateurs de commande de vol, des traitements numériques des signaux pour préserver le confort des passagers, et des actionneurs de commande des gouvernes (Houati, 2010).

Ce système est critique car il impacte non seulement la sécurité des passagers, mais aussi les capacités de pilotage ainsi que la structure de l'avion. Il devra être tolé-

rant aux erreurs et se prémunir d'un non-respect potentiel des consignes réglementaires, d'une défaillance technique et d'erreurs humaines, qu'elles soient volontaires ou non (EUROCAE, 2009).

Ainsi, contrairement aux cas précédents où le système embarqué était abordé en première analyse par ses fonctionnalités, cette étude de cas est construite à partir de la Sûreté de Fonctionnement qui adresse l'ensemble du cycle de vie du produit en prenant en compte non seulement les environnements de développement mais aussi le contexte opérationnel (ARP4761, 1996).

6.2 Problématique traitée et démarche

Dans cette étude de cas, nous introduisons la problématique de développement des systèmes embarqués communicants dits critiques car ils impactent la sécurité. Les processus de développement et de gestion du projet sont dirigés par des standards et contraints par une réglementation stricte. L'industrialisation est basée sur des composants spécifiques. La sûreté de fonctionnement est primordiale. Celle-ci est traitée dans l'ensemble de ses composantes (sécurité, fiabilité, disponibilité, maintenabilité). Elle s'appuie notamment sur des spécifications et vérifications formelles, et sur la définition d'une architecture répondant aux exigences réglementaires et validée par l'organisme de certification. Elle va conduire tout le développement du système.

Nous accompagnons l'ingénieur de la société de service dans sa démarche, pour l'amener à s'organiser :

- Le "Penser Système" décrit les enjeux, les parties prenantes et commente l'architecture globale du système pour illustrer la complexité globale. Il sensibilise au fait qu'un tel développement est l'affaire de spécialistes et ne s'improvise pas. (Etape « Approche globale », module pédagogique « Penser Système »),
- L'analyse générale (Etape « Analyse générale », modules « Approche méthodologique, Analyse préliminaire ») introduit l'approche méthodologique à mettre en œuvre pour développer un tel système : exigences, fonctionnalités du système à développer, contraintes. Elle introduit également les préoccupations économiques relatives au développement du système. Dans le module d'analyse préliminaire nous recentrons l'étude sur le calculateur de commande de vol électrique, par une analyse des exigences et des contraintes du calculateur en relation avec son environnement.
- Les modules de spécialité abordent la sûreté de fonctionnement, avec un focus sur la sécurité, ainsi que la réglementation, les standards, les labels, la communication, l'énergie. (Etape « Analyses spécifiques », modules « Sûreté de fonctionnement, Sécurité, Réglementation, standards, labels, Communication, Energie »).
- L'étape « Synthèse », modules pédagogiques « Architecture fonctionnelle et organique », aborde les aspects de conception de l'architecture du calculateur en relation avec les choix faits dans les

modules précédents. Nous observons comment les éléments fonctionnels sont mis en œuvre au sein d'un découpage organique. Nous commentons comment l'intégration prend en compte les résultats des analyses précédentes et notamment celles issues de la sûreté de fonctionnement. Pour finir nous traitons des aspects "vérification" et "validation" du produit et de sa mise en œuvre relativement au contexte opérationnel, aux standards, normes et réglementations introduits à l'étape « Synthèse », modules « Intégration, vérification et validation ».

La démarche guidée par l'ingénierie système est dans ce cas très complète. Des approfondissements permettent d'aller plus loin sur le temps de réponse, le management de projet, les normes d'ingénierie, le soutien logistique, la modélisation, la certification, etc.

6.3 Démonstration de l'intérêt de l'approche systémique

L'objectif pédagogique qui découle de cette étude de cas est de comprendre :

- La nécessité de formaliser et de gérer les processus en conformité avec les standards et les organismes de réglementation,
- L'organisation de la qualité,
- Les exigences de certification et de sûreté de fonctionnement, leurs impacts sur le développement des systèmes embarqués critiques.

7 BILAN PEDAGOGIQUE

Cette section propose une synthèse de la progression pédagogique obtenue au fil des études de cas, ainsi que des retours d'expérience de la session 2016.

7.1 L'apprentissage basé sur l'étude de cas

La méthode de formation choisie dans le MOOC, basée sur les études de cas, a favorisé un apprentissage actif des personnes formées grâce à la réflexion individuelle qu'il a suscitée mais aussi les discussions en groupe qui ont pris corps au travers des forums.

Elle a permis de développer chez les apprenants une capacité d'analyse et de résolution de problèmes, un jugement et du sens critique, ainsi qu'une sensibilité aux enjeux technologiques, sociétaux, éthiques, scientifiques et économiques du domaine des systèmes embarqués connectés. A travers une mise en situation réelle d'entreprise, il a permis de développer non seulement le savoir, mais surtout le savoir-faire et le savoir-être des apprenants (Barnes et al., 1994), par une illustration de la posture de penseur système.

Ainsi, tout au long du MOOC, les apprenants ont développé leurs connaissances et savoir-faire, en progressant au fil des études de cas selon trois axes : la maturité des processus, le niveau d'industrialisation et la sûreté de fonctionnement.

Dans le cas du détecteur de fumée, les processus ne sont pas formalisés. Joe doit s'organiser au mieux pour con-

cevoir, réaliser et vérifier ce système. Il s'agit d'un prototype réalisé artisanalement par une seule personne, qui n'a pas vocation à être réalisé en plusieurs exemplaires ni à être commercialisé. Le projet de Joe consiste à choisir et assembler des composants du commerce, son budget pour l'approvisionnement en composant et outillage est réduit. Il n'a pas besoin de compétences pointues. La sûreté de fonctionnement est abordée intuitivement.

Dans le cas du système industriel d'aide à la navigation embarqué dans un véhicule agricole, les processus de développement et de gestion du projet sont caractérisés et encadrés par des standards, les moyens de production sont opérationnels. Le produit est développé et fabriqué en série de façon industrielle, en utilisant les composants du commerce. Le projet est tiré par les besoins du marché. La sûreté de fonctionnement est prise en considération dans ses dimensions de fiabilité et maintenabilité.

Dans le cas du système embarqué communicant critique appartenant au domaine de l'aéronautique, les processus de développement et de gestion du projet sont formalisés, maîtrisés et dirigés par des standards, et contraints par une réglementation stricte. L'industrialisation est basée sur des composants spécifiques. La sûreté de fonctionnement est primordiale et conduit tout le cycle de vie du système.

La Figure 1 synthétise la progression dans les études de cas selon les trois axes définis, la maturité des processus, le niveau d'industrialisation, et le degré de sûreté de fonctionnement.

	Maturité des processus	Niveau d'industrialisation	Sûreté de fonctionnement
CAS 1	Processus non formalisés	Prototype réalisé par assemblage de composants du commerce	Scénarios intuitifs (fiabilité, maintenabilité)
CAS 2	Processus décrits et encadrés par des standards	Développement d'un produit industriel en PME	Sûreté de fonctionnement importante
CAS 3	Processus formalisés et maîtrisés	Industrialisation en série basée sur des composants spécifiques en grand groupe	Sûreté de fonctionnement primordiale (disponibilité, sécurité)

Figure 1 : Progression des cas sur trois axes

7.2 Les savoirs, savoir-faire et savoir-être appris

Le Penser Système a permis à l'apprenant de comprendre la nécessité de disposer d'une approche pluridisciplinaire et transdisciplinaire - pas uniquement centrée sur la technologie mais prenant en compte également les obligations légales, les exigences sociétales, pour ne nommer celles - indispensables pour explorer ce domaine des systèmes embarqués connectés.

La mise en œuvre des processus d'ingénierie système (INCOSE, 2011) a permis à l'apprenant d'acquérir une démarche méthodologique lui donnant les moyens de mettre en place un management de projet cohérent et intégré à un développement de produit.

Les exercices et les forums associés aux études de cas ont permis aux apprenants de révéler et développer des compétences spécifiques pour identifier et résoudre les problèmes tout au long des études de cas.

8 CONCLUSION

L'objectif de ce MOOC était de montrer comment l'approche systémique permet d'appréhender une situation complexe et de caractériser les difficultés qui se posent dès lors que l'on considère l'environnement des systèmes embarqués connectés. Il a notamment soulevé les problèmes d'intégration de diverses disciplines et du management de projet.

Cette expérience a été un succès, durant deux années successives. Quelques commentaires en témoignent : « Les notions abordées étaient intéressantes et accessibles. La diversité des intervenants et des thématiques est un plus », « Un grand merci à l'équipe toulousaine pour ce panorama ambitieux et pour la richesse des interventions proposées ».

L'un des facteurs de réussite tient certainement au fait que l'équipe pédagogique, mise au défi par un tel projet ambitieux et novateur, a également appliqué, pour la conception et la mise en œuvre du MOOC, une approche systémique, coopérative et intégrative de plusieurs compétences et disciplines.

Les choix pédagogiques effectués contribuent également à ce succès. Le format de MOOC a permis un apprentissage actif, par le biais de supports variés et élaborés, de QCM d'auto-évaluation, offrant la possibilité à chacun de progresser à son propre rythme, d'approfondir les ressources complémentaires, et de discuter interactivement avec les autres apprenants et les animateurs du forum. L'étude de cas, ancrée sur la pratique, s'est révélée être un outil stimulant l'apprentissage, tout en dynamisant la formation.

En contrepartie, adopter ce format de MOOC nécessite de remettre en question la pédagogie de l'enseignant. Les apprenants sont en effet plus actifs, mais aussi plus réactifs et plus exigeants. Il faut ainsi répondre à ces exigences par une architecture claire du cours, un vocabulaire précis et homogène, des supports professionnels, des QCM aux questions précises, une exactitude scientifique, la vérification que les sources utilisées soient libres de droit, etc. En outre, un MOOC, une fois produit et mis en ligne, nécessite un accompagnement spécifique des apprenants par l'équipe pédagogique via les forums mais aussi des évolutions régulières, en temps réel pour répondre aux questions soulevées dans les forums, et une capitalisation pour une actualisation du MOOC pour les sessions futures. Il entraîne également une évolution du rôle de l'enseignant, qui doit rythmer à distance et de façon individualisée la progression dans les apprentissages. Il doit également susciter et maintenir intacts l'intérêt et la motivation des apprenants, qui, faute de sollicitations régulières, sont naturellement enclins à se décourager. Ce retour d'expérience positif sur les premières sessions nous a conduit à introduire ce MOOC dans le cursus de formation d'ingénieur de 5^e année à

l'INSA de Toulouse. Cette expérience a permis d'approfondir par un accompagnement en présentiel des sujets d'intérêt pertinents au-delà du MOOC.

REMERCIEMENTS

Les auteurs de l'article remercient l'équipe projet du MOOC, l'équipe d'ingénierie pédagogique et l'équipe de production, ainsi que l'ensemble des contributeurs aux vidéos. Nos remerciements s'adressent également aux services de l'INP et de l'INSA de Toulouse, ainsi que l>IDEX, dont le financement a permis au projet de voir le jour.

REFERENCES

- Arnaud de Baynast, Jacques Lendrevie, Julien Lévy, *Mercator* - 12e éd., Dunod, 2017.
- ARP4761 / ED135, 1996. *Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment*.
- Barnes L. B., Christensen C.R., Hansen A., 1994. *Teaching and the Case Method*. Harvard Business School Press.
- Baron et al., *MOOC Systèmes Embarqués et Objets Connectés*, 2016. Plateforme FUN, <https://www.fun-mooc.fr/courses/course-v1:univ-toulouse+101001+session02/about>
- Daniel-Allegro B., 2014a. *Introduction au Penser Système*. Séminaire AFIS – INSA du 20-11-2014. <http://pensersysteme.fr/>
- Daniel-Allegro B., 2014b. *AFIS Guide d'introduction au Penser Système*. Sous la coordination de B. Daniel Allegro, A. Le Put et J-C. Tucoulou.
- EUROCAE ED-79A, 2009. *Guidelines for development of civil aircraft and systems*.
- Haouati M. S., 2010. *Architectures innovantes de systèmes de commandes de vol*. Thèse de Doctorat Institut National Polytechnique de Toulouse.
- INCOSE, 2011. *Systems Engineering Handbook. A guide for system life cycle processes and activities*. INCOSE -TP- 2003-002.03.2.2
- ISO/IEC/IEEE 15288, 2015. *Systems and software engineering - System life cycle processes*.
- Lawson, H., 2010. *A Journey Through the Systems Landscape*, traduit par Daniel-Allegro B., 2014. Parcours au Pays des Systèmes, College Publications.
- Meinadier J-P., 2012. *AFIS Découvrir et Comprendre l'Ingénierie Système*. Sous la direction de Serge Fiorèse et de Jean-Pierre Meinadier. Ed. Cépadués
- Mitchell M., 2009. *Complexity A guided tour*, Oxford University Press.
- OPIEC (Observatoire Paritaire Informatique Ingénierie Etudes Conseil), 2014. *Etude sur l'évolution des métiers et des besoins en formation pour les systèmes embarqués*, on-line.

APROCHE PROBABILISTE ET MULTIOBJECTIF POUR LA SELECTION DE SCENARIOS DANS UN GRAPHE DE PROJET

M. LACHHAB, T. COUDERT, C. BELER

LGP - Université de Toulouse, INP-ENIT, 47 Avenue d'Azereix – 65000
TARBES – France
majda.lachhab@enit.fr, thierry.coudert@enit.fr, cedrick.beler@enit.fr

RESUME : Dans cet article, une approche probabiliste et multiobjectif est développée afin de comparer, au sens de Pareto et sous incertitude, les scénarios d'un graphe de projet. Chaque scénario est une séquence de tâches et chaque tâche est caractérisée par ses valeurs de coût et de durée. Les incertitudes inhérentes aux critères de coût et de durée sont modélisées par la représentation de l'ensemble des valeurs de ces critères sous forme de distributions de probabilités. Ainsi, ces distributions de probabilités sont prises en compte dans l'évaluation des scénarios de projet et dans le calcul des probabilités de dominance entre ces scénarios. Elles permettent de comparer les solutions en termes de coût et de durée afin de sélectionner celles qui peuvent appartenir au front de Pareto, c'est-à-dire les solutions Pareto-optimales. En effet, le front de Pareto contient toutes les solutions non dominées au sens de Pareto et respectant un seuil d'incertitude exigé par le décideur, c'est-à-dire correspondant à un niveau de risque acceptable. L'approche proposée dans cet article peut être intégrée dans un modèle d'optimisation multiobjectif pour sélectionner les scénarios Pareto-optimaux dans un graphe de projet.

MOTS-CLES : Gestion de projet, Cadre multiobjectif, Incertitude, Approche probabiliste, Pareto-dominance.

1 INTRODUCTION

Les notions de fronts de Pareto et de Pareto optimalité sont souvent abordées pour traiter des problèmes d'optimisation multiobjectif comme dans le domaine de gestion de projet. Dans la littérature, il existe de plus en plus de travaux qui traitent des problématiques liées à l'optimisation multiobjectif dans des projets complexes et, par conséquent, risqués. L'objectif principal de l'optimisation multiobjectif, selon le principe de Pareto-optimalité, est de comparer plusieurs solutions selon différents critères et d'identifier les solutions non dominées. Cet ensemble de solutions forme le front de Pareto où les solutions non dominées sont dites Pareto-optimales.

Notre contribution consiste à proposer une approche probabiliste permettant de comparer, au sens de Pareto et sous incertitude, les scénarios d'un graphe de projet selon les critères de coût et de durée. En effet, l'incertitude dans notre travail est causée par les risques. Elle est modélisée en représentant les valeurs de coût et de durée des tâches du projet sous forme de distributions de probabilités. Ensuite, la durée globale et le coût total de chaque scénario de projet sont calculés sous forme de distributions de probabilités. Donc, l'approche proposée consiste à calculer les probabilités de dominance des scénarios de projets afin de construire le front de Pareto. Ce front de Pareto contient toutes les solutions non dominées au sens de Pareto ayant un seuil d'incertitude fixé par le décideur. Dans de précédents travaux (Lachhab *et al.*, 2016), un outil d'optimisation tenant compte des

incertitudes sur les valeurs de coût et de durée a été proposé. Les incertitudes sont modélisées par des intervalles dont les valeurs sont équiprobables. L'utilisation de distributions de probabilités permet une meilleure adaptation du modèle de projet à un contexte industriel réel.

Cet article est structuré autour de cinq sections. La section 2 décrit quelques techniques permettant la sélection de solutions Pareto-optimales sous incertitude. Ensuite, la définition des coûts et des durées des tâches d'un graphe de projet sous la forme de distributions de probabilités est donnée dans la section 3. L'évaluation de scénarios de projet sous incertitude selon les critères de coût et de durée est détaillée par la suite dans la même section. La section 4 permet de définir les probabilités de Pareto-dominance d'un scénario de projet par rapport à un autre. Cela permet de comparer une solution par rapport à une autre selon deux critères (coût, durée) et de constituer le front de Pareto. Un exemple permettant d'appliquer notre approche probabiliste d'élaboration du front de Pareto dans un espace multiobjectif à deux dimensions est représenté dans la section 5. Enfin, la section 6 permet de conclure et de présenter les perspectives de recherche envisagées pour la suite.

2 ETAT DE L'ART

Les techniques de sélection de solutions Pareto-optimales sont naturellement liées aux problèmes d'optimisation multiobjectif. La spécificité de ces problèmes d'optimisation est la génération d'un ensemble de solutions Pareto-optimales représentant un compromis entre des critères souvent non seulement conflictuels

mais aussi incertains. En effet, la construction du front de Pareto doit tenir compte de ces incertitudes au cours du processus de recherche de solutions Pareto-optimales. Dans ce sens, beaucoup de travaux dans la littérature proposent des techniques d'analyse des solutions Pareto-optimales sous incertitude. Par exemple dans (Mlakar *et al.*, 2014), les auteurs distinguent deux approches différentes pour comparer des solutions sous incertitudes et déterminer les relations de dominance entre elles. La première approche consiste à considérer la valeur approximative et la variance pour chaque critère, les transformer en une seule valeur et comparer ces valeurs uniques. La deuxième approche s'appuie sur le calcul des intervalles de confiance et la comparaison directe des solutions à l'aide de ces intervalles. Les auteurs ont d'ailleurs adopté la dernière technique en proposant de nouvelles relations de Pareto-dominance avec la particularité de comparer des solutions multiobjectif sous contraintes. Dans (Bahri *et al.*, 2013), les auteurs proposent une approche basée sur la théorie des possibilités où les incertitudes sont modélisées par des distributions de possibilités triangulaires. De nouvelles relations de Pareto optimalité ont été proposées et ont conduit à définir des classes pour comparer les solutions. Un autre concept basé sur la comparaison des solutions sous incertitudes à l'aide des intervalles ayant des valeurs continues et distribuées uniformément a été proposé dans (Teich, 2001). Dans cette approche, les solutions sont comparées à partir du calcul des probabilités de dominance entre ces distributions. D'autres travaux considèrent que les incertitudes doivent être prises en compte lors de l'évaluation et la comparaison des solutions dans n'importe quel cadre multiobjectif. Par exemple dans (Das *et al.*, 2009), les auteurs distinguent les approches probabilistes, permettant d'établir des relations probabilistes de dominance à partir de distributions uniformes, et les approches basées sur la logique floue s'appuyant par exemple sur des métriques qui mesurent la qualité de chacune des solutions appartenant au front de Pareto. Dans (Coelho, 2015), l'auteur présente une vue d'ensemble des techniques permettant d'évaluer la probabilité de dominance et qui ont été développées pour l'optimisation multiobjectif basée sur la fiabilité. Dans un autre travail (Hendriks *et al.*, 2011), les auteurs proposent une méthode générale qui tient compte des incertitudes durant l'analyse du front de Pareto. Cette méthode est centrée autour de la dominance non-déterministe entre l'ensemble des solutions et l'utilisation des relations de dominance sur les intervalles. Le travail développé dans cet article propose une approche probabiliste permettant d'évaluer les scénarios et d'élaborer le front de Pareto. Cette approche est basée sur la modélisation des incertitudes avec des distributions de probabilités quelconques. Les incertitudes sont inhérentes aux critères du projet, le coût et la durée. Les solutions Pareto-optimales sont alors distinguées en calculant leurs probabilités de dominance et en les comparant en fonction d'un seuil donné et exigé par le décideur.

3 MODELISATION SOUS INCERTITUDE

Dans cette section, le modèle du graphe de projet considéré dans notre travail est décrit. Les formules des coûts et des durées des tâches associées à un scénario du graphe de projet sont définies et exploitées pour l'évaluation globale et multiobjectif des scénarios sous incertitude.

3.1 Modèle du graphe de projet

Dans cette partie, le graphe de projet basé sur les modèles de graphe ayant été proposés dans (Lachhab *et al.*, 2016) et (Lachhab *et al.*, 2017) est décrit. Ce modèle est un graphe de projet acyclique, orienté et composé d'un ensemble de nœuds et d'un ensemble d'arcs représentant les contraintes de précedence entre ces nœuds. Les nœuds sont des tâches de projet qui peuvent être des activités et/ou des opérateurs logiques de conjonction (AND , \overline{AND}). Un graphe de projet décrit l'ensemble des scénarios possibles représentant les différentes alternatives de choix de conception et de réalisation d'un nouveau système (le livrable). Ces alternatives (OU exclusif) sont formalisées par des arcs divergents et convergents. Les nœuds AND divergents et \overline{AND} convergents permettent de définir un ensemble de sous-séquences pouvant être réalisées en parallèle. Un ensemble de sous-séquences en parallèle forme un sous-projet. Chaque sous projet est donc un sous-graphe qui débute par un nœud de divergence \overline{AND} et qui se termine par un nœud de convergence AND . Les scénarios (solutions) sont des séquences de tâches planifiées au plus tôt en respectant les contraintes de précedence. Le premier et le dernier nœud du graphe de projet sont fictifs car ils décrivent uniquement le début et la fin du projet.

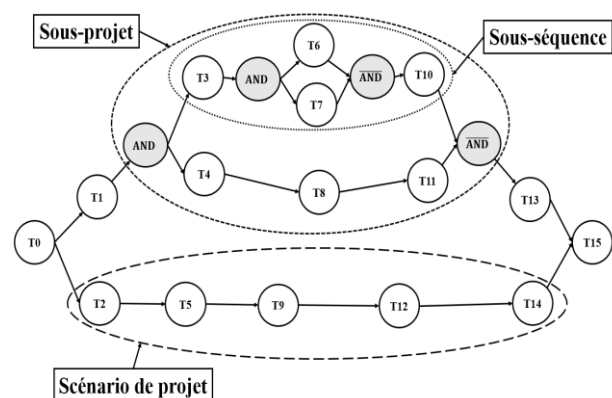


Figure 1 : Exemple de graphe de projet

Un exemple de graphe de projet est donné dans la figure 1. Dans notre approche, les critères associés à chaque nœud tâche du projet correspondent au coût de la tâche et à sa durée. Les coûts et les durées sont des variables aléatoires indépendantes dont les distributions de probabilités représentent l'impact des risques sur les valeurs de coût et de durée pour chaque tâche. L'utilisation des

distributions de probabilité dans notre travail est une façon de modéliser les incertitudes concernant les valeurs des critères de coût et de durée et par conséquent les impacts des risques.

La modélisation formelle des coûts et des durées des tâches sous forme de distribution de probabilités ainsi que l'évaluation globale des scénarios de projet dans un espace multiobjectif à deux dimensions (coût, durée) sont détaillées dans les prochaines sections.

3.2 Durées et coûts des tâches de projet

La modélisation des incertitudes résultant des risques et impactant les coûts et les durées des tâches est faite au moyen de variables aléatoires discrètes et de leurs distributions de probabilités. Ces distributions de probabilités sont définies par un gestionnaire de risques (expert) à partir des probabilités d'occurrence des événements indésirables et la connaissance de leurs impacts sur la durée et les coûts des tâches de projet. Soit, T_i la tâche i appartenant à un scénario du projet. Soit, D_{T_i} et C_{T_i} deux variables aléatoires discrètes représentant respectivement la durée et le coût de la tâche T_i . Les probabilités de distribution de D_{T_i} et de C_{T_i} sont respectivement notées $P_{D_{T_i}}$ et $P_{C_{T_i}}$ avec, $P_{D_{T_i}}(d) = P(D_{T_i} = d)$ la probabilité que $D_{T_i} = d$ et $P_{C_{T_i}}(c) = P(C_{T_i} = c)$ est la probabilité que $C_{T_i} = c$. Soit DD_{T_i} , le domaine de la variable aléatoire D_{T_i} contenant les m valeurs discrètes de D_{T_i} avec des probabilités non nulles et DC_{T_i} le domaine de C_{T_i} contenant les q valeurs discrètes de C_{T_i} ayant des probabilités non nulles tel que :

$$DD_{T_i} = \{d_{i1}, \dots, d_{im}\} / P_{D_{T_i}}(d_{ij}) \neq 0 \quad (1)$$

$$DC_{T_i} = \{c_{i1}, \dots, c_{iq}\} / P_{C_{T_i}}(c_{ij}) \neq 0 \quad (2)$$

La figure 2 donne un exemple de distributions de probabilités pour les coûts et les durées d'une tâche appartenant à un scénario de projet et leurs domaines associés. Dans cette figure sont représentées uniquement les distributions de probabilités relatives à la première tâche T_1 du scénario considéré. Cependant, les distributions de probabilités pour les critères de coût et de durée sont définies avec leurs domaines pour toutes les autres tâches.

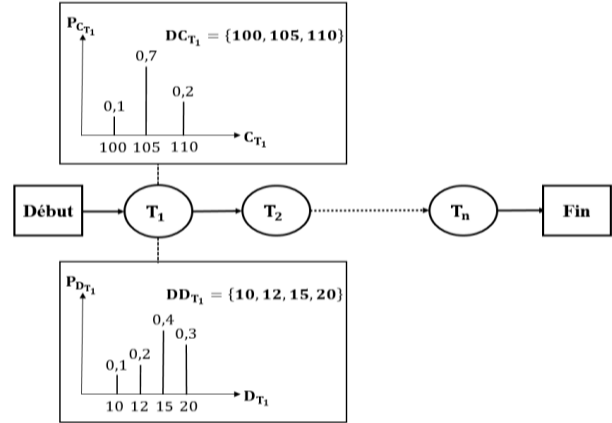


Figure 2 : Exemple de distributions de probabilités pour les coûts et les durées d'une tâche et leurs domaines

3.3 Evaluation multiobjectif d'un scénario de projet

Cette partie consiste à définir la méthode utilisée pour évaluer un scénario de projet en termes de durée globale et de coût total. Cette méthode prend en compte l'impact des risques au travers des distributions de probabilités des durées et des coûts.

Soit S , un scénario du graphe de projet G . La durée globale d'un scénario S est la somme des variables aléatoires correspondant aux durées des tâches et des sous-projets appartenant à ce scénario (idem pour le coût total de S).

Soit $(D_{T_i}, D_{T_{i+1}})$ et $(C_{T_i}, C_{T_{i+1}})$ des variables aléatoires discrètes relatives respectivement aux durées et aux coûts de deux tâches successives (T_i et T_{i+1}) d'un scénario S du projet G . Avec $P_{D_{T_{i+1}}}$ et $P_{C_{T_{i+1}}}$ les probabilités de distribution respectives de $D_{T_{i+1}}$ et de $C_{T_{i+1}}$.

Soit $D_{Sum}^{i+1} = D_{T_i} + D_{T_{i+1}}$ et $C_{Sum}^{i+1} = C_{T_i} + C_{T_{i+1}}$ deux variables aléatoires discrètes correspondant à la somme des variables aléatoires associées aux durées et aux coûts de deux tâches successives d'un scénario S . Soit $P_{D_{Sum}^{i+1}}$ et $P_{C_{Sum}^{i+1}}$ les distributions de probabilités respectives des valeurs des durées et des coûts de D_{Sum}^{i+1} et de C_{Sum}^{i+1} . Soit $DD_{Sum}^{i+1} = DD_{T_i} \times DD_{T_{i+1}}$, le domaine de D_{Sum}^{i+1} calculé en fonction du produit cartésien de DD_{T_i} et de $DD_{T_{i+1}}$, et $DC_{Sum}^{i+1} = DC_{T_i} \times DC_{T_{i+1}}$, le domaine de C_{Sum}^{i+1} calculé en fonction du produit cartésien de DC_{T_i} et de $DC_{T_{i+1}}$, tels que :

$$DD_{T_i} \times DD_{T_{i+1}} = \{(a, b)\} / \forall a \in DD_{T_i}, \forall b \in DD_{T_{i+1}} \quad (3)$$

$$DC_{T_i} \times DC_{T_{i+1}} = \{(x, y)\} / \forall x \in DC_{T_i}, \forall y \in DC_{T_{i+1}} \quad (4)$$

D'où :

$$DD_{Sum}^{i+1} = \bigcup_{\forall (a,b) \in DD_{T_i} \times DD_{T_{i+1}}} (a+b) \quad (5)$$

$$DC_{Sum}^{i+1} = \bigcup_{\forall (x,y) \in DC_{T_i} \times DC_{T_{i+1}}} (x+y) \quad (6)$$

En considérant que D_{T_i} et $D_{T_{i+1}}$ sont deux variables aléatoires discrètes indépendantes (idem pour C_{T_i} et $C_{T_{i+1}}$), la probabilité pour que la somme D_{Sum}^{i+1} soit égale à t est $P_{D_{Sum}^{i+1}}(t)$. La probabilité pour que C_{Sum}^{i+1} soit égale à z est $P_{C_{Sum}^{i+1}}(z)$.

Les probabilités $P_{D_{Sum}^{i+1}}(t)$ et $P_{C_{Sum}^{i+1}}(z)$ sont calculées en utilisant le produit de convolution des distributions de probabilité relatives respectivement aux couples $(D_{T_i}, D_{T_{i+1}})$ et $(C_{T_i}, C_{T_{i+1}})$.

Ainsi, $\forall t \in DD_{Sum}^{i+1}$:

$$P_{D_{Sum}^{i+1}}(t) = P(D_{Sum}^{i+1} = t) = \sum_{d \in DD_{T_{i+1}}} P_{D_{T_i}}(t-d) \times P_{D_{T_{i+1}}}(d) \quad (7)$$

Et, $\forall z \in DC_{Sum}^{i+1}$:

$$P_{C_{Sum}^{i+1}}(z) = P(C_{Sum}^{i+1} = z) = \sum_{c \in DC_{T_{i+1}}} P_{C_{T_i}}(z-c) \times P_{C_{T_{i+1}}}(c) \quad (8)$$

Donc, les probabilités de D_S et C_S (notées respectivement P_{D_S} et P_{C_S}), qui sont deux variables aléatoires discrètes associées respectivement à la durée globale et au coût total du scénario S , composé de la séquence des tâches (T_1, \dots, T_n) , sont calculées en appliquant les formules (7) et (8). Ainsi :

$$\forall t \in DD_S, P_{D_S}(t) = P_{D_{Sum}^n}(t) \quad (9)$$

$$\forall z \in DC_S, P_{C_S}(z) = P_{C_{Sum}^n}(z) \quad (10)$$

Avec DD_S et DC_S , les domaines respectifs de D_S et de C_S , calculés en fonction des formules 5 et 6.

S'il existe des sous-projets contenant par exemple q sous-séquences parallèles et appartenant à la séquence du scénario S , la probabilité du coût total du sous-projet SP va être calculée en sommant les probabilités des coûts de toutes les sous-séquences en réutilisant les formules (8) et (10). Cependant, la durée globale du sous-projet SP est calculée différemment. En effet, la durée globale de SP est une variable aléatoire discrète qui est le maximum des durées de toutes les q sous-séquences parallèles planifiées à leurs dates au plus tôt. Dans la suite, la méthode du calcul de la probabilité de distribution du maximum de q variables aléatoires discrètes est détaillée sur l'exemple de sous-projet de la figure 3.

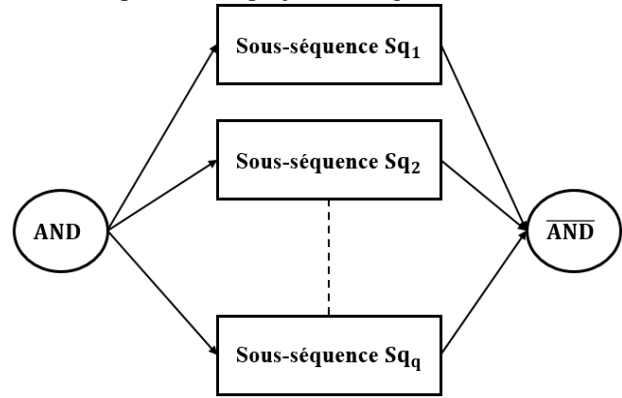


Figure 3 : Exemple de sous-projet comportant q sous-séquences parallèles

Soit les notations suivantes :

- DSq_k est une variable aléatoire discrète représentant la durée de la sous-séquence Sq_k ,
- d_{ik} est une valeur i de la durée DSq_k ,
- P_{DSq_k} est la probabilité de distribution de DSq_k ,
- $P_{DSq_k}(d_{ik})$ est la probabilité que DSq_k prenne la valeur d_{ik} ,
- DD_{DSq_k} est le domaine de la variable aléatoire DSq_k contenant les q valeurs discrètes de DSq_k tel que :

$$DD_{DSq_k} = \{d_{ik}\} / P_{DSq_k}(d_{ik}) \neq 0 \quad (11)$$

La durée d'un sous-projet SP est la durée totale des q sous-séquences parallèles (planifiées au plus tôt). La durée de SP est définie par le maximum des durées des q sous-séquences du sous-projet SP , notée par D_{Max}^{SP} , tel que :

$$D_{Max}^{SP} = \text{Max}(DSq_1, \dots, DSq_q) \quad (12)$$

Le maximum des variables aléatoires discrètes D_{Max}^{SP} , est une variable aléatoire discrète. DD_{Max}^{SP} est le domaine des valeurs de D_{Max}^{SP} . Le domaine DD_{Max}^{SP} est défini par :

$$DD_{Max}^{SP} = \bigcup_{k=1}^q DDSq_k \quad (13)$$

$P_{D_{Max}^{SP}}$ est la distribution de probabilité de D_{Max}^{SP} .

Dans le cas où le sous-projet SP contient deux sous-séquences parallèles. La probabilité pour que D_{Max}^{SP} soit égale à t est définie par :

$$\begin{aligned} P_{D_{Max}^{SP}}(t) &= P_{DSq_1}(DSq_1 = t) \times P_{DSq_2}(DSq_2 \leq t) \\ &\quad + P_{DSq_2}(DSq_2 = t) \times P_{DSq_1}(DSq_1 \leq t) \\ &\quad - P_{DSq_1}(DSq_1 = t) \times P_{DSq_2}(DSq_2 = t) \end{aligned} \quad (14)$$

Avec, $\forall k \in \{1, \dots, q\}$:

$$P_{DSq_k}(DSq_k \leq t) = \sum_{\substack{d \in DDSq_k \\ d \leq t}} P_{DSq_k}(DSq_k = d) \quad (15)$$

Pour q sous-séquences en parallèle, l'expression de

$P_{D_{Max}^{SP}}(t)$ devient : $\forall t \in D_{Max}^{SP}$,

$$\begin{aligned} P_{D_{Max}^{SP}}(t) &= \\ &\sum_{\substack{\forall DSq_k \\ k \in \{1, \dots, q\}}} [P_{DSq_k}(DSq_k = t) \times \prod_{\substack{\forall DSq_j \\ j \in \{1, \dots, q\} \\ j \neq k}} P_{DSq_j}(DSq_j \leq t)] \\ &- \prod_{\substack{\forall DSq_k \\ k \in \{1, \dots, q\}}} P_{DSq_k}(DSq_k = t) \end{aligned} \quad (16)$$

4 PROBABILITE DE PARETO-DOMINANCE

Dans la section précédente, les formules permettant de calculer les distributions de probabilités de la durée globale et du coût total d'un scénario ont été définies. Dans cette section, la probabilité pour qu'un scénario domine un autre scénario au sens de Pareto en termes de coût et de durée est définie dans la sous-section 4.1. Ensuite, ces scénarios vont être comparés afin de construire le front de Pareto selon les exigences du décideur à propos des incertitudes acceptées.

4.1 Pareto-dominance entre deux scénarios

Dans cette partie, la probabilité de dominance entre deux scénarios dans le cas multiobjectif est définie.

Soit S_i et S_j deux scénarios de projet à comparer pour déterminer lequel des deux scénarios domine l'autre au sens de Pareto, par rapport aux critères de coût et de durée, dans un environnement incertain. Soit D_{S_i} la durée globale associée au scénario S_i . D_{S_i} est une variable aléatoire discrète représentée par la distribution de probabilité $P_{D_{S_i}}$. Les distributions de probabilités $P_{D_{S_i}}$ et $P_{D_{S_j}}$ associées respectivement aux durées des scénarios S_i et S_j ont une plage de valeurs comprise entre les valeurs nominales et maximales respectives des durées pour chaque scénario, notées $D_{S_i}^-$, $D_{S_j}^-$ et $D_{S_i}^+$, $D_{S_j}^+$ (cf. figure 4).

Un scénario S_i domine un scénario S_j au sens de Pareto (noté $S_i \prec S_j$) si, pour tout critère O , $O_{S_i} \leq O_{S_j}$ et il existe au moins un critère O tel que $O_{S_i} < O_{S_j}$ (avec O_{S_i} , la valeur du critère O pour le scénario S_i). Nous devons donc évaluer ces inégalités selon leur probabilité. Prenons le cas du critère durée de réalisation d'un scénario i (D_{S_i}).

La probabilité pour que D_{S_i} soit strictement inférieure à D_{S_j} est calculée par :

$$\begin{aligned} P(D_{S_i} < D_{S_j}) &= P(D_{S_i}^- \leq x < D_{S_j}^-) + \\ &P(D_{S_i}^+ \leq y < D_{S_j}^+) - P(D_{S_i}^- \leq x < D_{S_j}^-) \times \\ &P(D_{S_i}^+ \leq y < D_{S_j}^+) + P(D_{S_j}^- \leq x < y \leq D_{S_i}^+) \end{aligned} \quad (17)$$

D'où,

$$\begin{aligned} P(D_{S_i} < D_{S_j}) &= \sum_{x \in [D_{S_i}^-, D_{S_j}^-]} P(D_{S_i} = x) + \\ &\sum_{y \in [D_{S_i}^+, D_{S_j}^+]} P(D_{S_j} = y) - \sum_{x \in [D_{S_i}^-, D_{S_j}^-]} P(D_{S_i} = x) \times \\ &\sum_{y \in [D_{S_i}^+, D_{S_j}^+]} P(D_{S_j} = y) + \\ &\sum_{\substack{(x,y) \in [D_{S_j}^-, D_{S_i}^+]^2 \\ x < y}} P(D_{S_i} = x) \times P(D_{S_j} = y) \end{aligned} \quad (18)$$

La probabilité que D_{S_i} soit égale à D_{S_j} est donnée par :

$$P(D_{S_i} = D_{S_j}) = \sum_{\substack{(x,y) \in [D_{S_i}^-, D_{S_i}^+] \times [D_{S_j}^-, D_{S_j}^+] \\ x=y}} P(D_{S_i} = x) \times P(D_{S_j} = y) \quad (19)$$

Ces probabilités ont été définies en considérant les différents cas possibles de représentation des distributions de probabilités des durées de deux scénarios. La figure 4 donne l'exemple de représentation de deux distributions de probabilités selon le critère de durée pour le cas où $D_{S_i}^- < D_{S_j}^- < D_{S_i}^+ < D_{S_j}^+$.

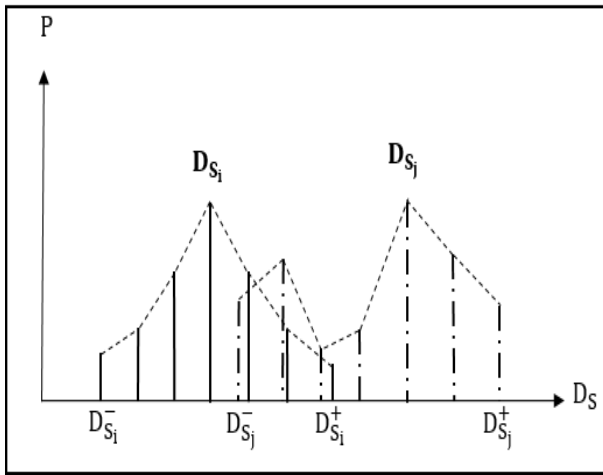


Figure 4 : Exemple de représentation de deux distributions de probabilités des durées de scénarios

Par analogie, Soit C_{S_i} la durée globale associée au scénario S_i . C_{S_i} est une variable aléatoire discrète représentée par la distribution de probabilité $P_{C_{S_i}}$. Les distributions de probabilités $P_{C_{S_i}}$ et $P_{C_{S_j}}$ associées respectivement aux durées des scénarios S_i et S_j ont une plage de valeurs comprises entre les valeurs nominales et maximales respectives des durées pour chaque scénario, notées $C_{S_i}^-$, $C_{S_j}^-$ et $C_{S_i}^+$, $C_{S_j}^+$.

La probabilité que C_{S_i} soit strictement inférieure à C_{S_j} est calculée par :

$$P(C_{S_i} < C_{S_j}) = P(C_{S_i}^- \leq x < C_{S_j}^-) + P(C_{S_i}^+ \leq y < C_{S_j}^+) - P(C_{S_i}^- \leq x < C_{S_j}^-) \times P(C_{S_i}^+ \leq y < C_{S_j}^+) + P(C_{S_i}^- \leq x < y \leq C_{S_i}^+) \quad (20)$$

D'où,

$$P(C_{S_i} < C_{S_j}) = \sum_{x \in [C_{S_i}^-, C_{S_j}^-]} P(C_{S_i} = x) + \sum_{y \in [C_{S_i}^+, C_{S_j}^+]} P(C_{S_j} = y) - \sum_{x \in [C_{S_i}^-, C_{S_j}^-]} P(C_{S_i} = x) \times \sum_{y \in [C_{S_i}^+, C_{S_j}^+]} P(C_{S_j} = y) + \sum_{\substack{(x,y) \in [C_{S_j}^-, C_{S_i}^+]^2 \\ x < y}} P(C_{S_i} = x) \times P(C_{S_j} = y) \quad (21)$$

La probabilité que C_{S_i} soit égale à C_{S_j} est donnée par :

$$P(C_{S_i} = C_{S_j}) = \sum_{\substack{(x,y) \in [C_{S_i}^-, C_{S_i}^+] \times [C_{S_j}^-, C_{S_j}^+] \\ x=y}} P(C_{S_i} = x) \times P(C_{S_j} = y) \quad (22)$$

Définition 1. Probabilité de dominance dans le cas multiobjectif : $S_i(C_{S_i}, D_{S_i})$ domine $S_j(C_{S_j}, D_{S_j})$ au

sens de Pareto (noté, $S_i \prec S_j$) si :

$$\begin{aligned} & (C_{S_i} < C_{S_j}) \text{ et } (D_{S_i} < D_{S_j}) \quad \underline{\text{OU}} \\ & (C_{S_i} < C_{S_j}) \text{ et } (D_{S_i} = D_{S_j}) \quad \underline{\text{OU}} \\ & (C_{S_i} = C_{S_j}) \text{ et } (D_{S_i} < D_{S_j}). \end{aligned}$$

Ainsi, la probabilité que $S_i \prec S_j$ au sens de Pareto est donnée par :

$$P(S_i \prec S_j) = P(C_{S_i} < C_{S_j}) \times P(D_{S_i} < D_{S_j}) + P(C_{S_i} < C_{S_j}) \times P(D_{S_i} = D_{S_j}) + P(C_{S_i} = C_{S_j}) \times P(D_{S_i} < D_{S_j}) \quad (23)$$

L'équation 23 peut être écrite aussi sous la forme suivante :

$$P(S_i \prec S_j) = P(C_{S_i} < C_{S_j}) \times [P(D_{S_i} < D_{S_j}) + P(D_{S_i} = D_{S_j})] + P(C_{S_i} = C_{S_j}) \times P(D_{S_i} < D_{S_j}) \quad (24)$$

4.2 Exigence de Pareto-dominance entre scénarios

Dans la sous-section précédente, la probabilité pour qu'un scénario de projet domine un autre scénario a été définie. Cette probabilité doit être calculée pour tous les

scénarios deux à deux et dans les deux sens car $P(S_i \prec S_j) \neq P(S_j \prec S_i)$, $\forall i \in \{1, \dots, n\}$ et $\forall j \neq i$.

Une fois que toutes les probabilités sont calculées, le décideur fixe un seuil de tolérance d'incertitude, noté θ . Ce seuil permet de sélectionner les solutions Pareto-optimales en termes de coût et de durée avec un certain niveau d'acceptation des risques représentés par des incertitudes sur les critères d'évaluation du projet.

Définition 2. Exigences de Pareto-dominance dans le cas multiobjectif : Pour garder une solution S_i sur le front de Pareto (FP), il faut qu'elle soit dominée au maximum avec le seuil θ par les autres solutions S_j , telles que $(S_j \neq S_i)$. C'est-à-dire :

$$\forall i \in \{1, \dots, n\} \text{ et } \forall j \neq i : \max_j P(S_j \prec S_i) \leq \theta \quad (25)$$

Dans la prochaine section, un exemple de graphe de projet est considéré afin d'appliquer la méthode probabiliste proposée dans la section 4 et les formules de probabilités définies dans la section 3.

5 APPLICATION

Soit G , le modèle de graphe de projet considéré (figure 5). T_0 et T_{10} sont deux tâches fictives indiquant le début et la fin du projet.

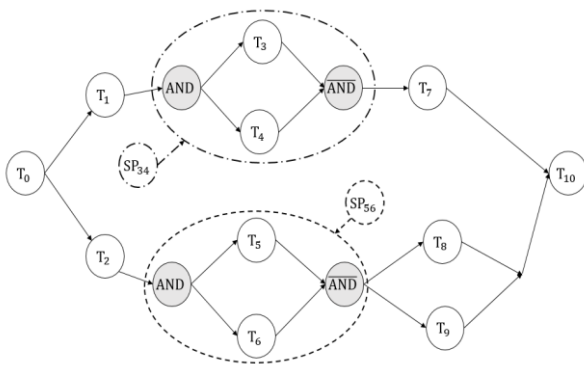


Figure 5 : Exemple de graphe de projet

Dans ce graphe, il y a trois scénarios (solutions) possibles : $S_1 = \{T_1, SP_{34}, T_7\}$, $S_2 = \{T_2, SP_{56}, T_8\}$ et $S_3 = \{T_2, SP_{56}, T_9\}$. Pour chaque tâche du projet $\{T_1, \dots, T_9\}$ sont associées deux distributions de probabilités relatives aux critères de coût et de durée. Les distributions de probabilités associées aux sous-projets SP_{34} et SP_{56} sont calculées à l'aide des formules 8 et 16. Ensuite, les distributions de probabilités associées aux valeurs de coût et de durée des trois scénarios S_1, S_2 et S_3 sont calculées à l'aide des formules 7, 8, 9 et 10 (cf. sous-section 3.3). Les

probabilités de dominance entre S_1, S_2 et S_3 sont enfin calculées en fonction des formules 24, 18, 19, 21 et 22 (cf. sous-section 4.1). Les résultats de calcul des probabilités de dominance entre les scénarios sont représentés dans le tableau 1. Le tableau 1 indique aussi les valeurs du maximum de l'inégalité 25 (cf. sous-section 4.2). En effet, chaque scénario est comparé en terme de Pareto-dominance avec les deux autres scénarios en calculant les probabilités pour qu'il soit dominé par les autres scénarios. Le maximum des valeurs de probabilités pour que chaque scénario soit dominé par les autres et ensuite calculé afin de le comparer avec la valeur du seuil θ défini dans la sous-section 4.2.

Solution	Probabilités de dominance	Max
S_1	$P(S_2 \prec S_1) = 0,00051$ $P(S_3 \prec S_1) = 0,000013$	0,00051
S_2	$P(S_1 \prec S_2) = 0,94$ $P(S_3 \prec S_2) = 0,19$	0,95
S_3	$P(S_1 \prec S_3) = 0,99$ $P(S_2 \prec S_3) = 0,51$	0,99

Tableau 1 : Valeurs des probabilités de dominance entre S_1, S_2 et S_3

Il reste à définir les exigences de dominance et le front de Pareto associé (cf. figure 6). Pour cela, le décideur doit fixer la valeur du seuil θ . Par exemple :

- $\theta = 1$, le front de Pareto est : $FP = \{S_1, S_2, S_3\}$. Ce qui implique une exigence très faible de dominance. Une très grande incertitude est tolérée. Le front de Pareto comporte toutes les solutions et un niveau de risque très élevé est accepté (aucune solution n'est suffisamment dominée pour être retirée).
- $\theta = 0,95$, le front de Pareto est : $FP = \{S_1, S_2\}$. Ce qui implique une exigence faible de dominance (grande incertitude).
- $\theta = 0,2$, le front de Pareto est : $FP = \{S_1\}$. Ce qui implique une exigence moyenne de dominance. L'incertitude tolérée est de niveau moyen. Un niveau de risque moyen est acceptable.
- $\theta = 0$, le front de Pareto est : $FP = \emptyset$. Ce qui implique une exigence stricte (trop forte) de dominance (incertitude nulle). Ce cas n'est pas réaliste car toutes les solutions sont éliminées systématiquement.

La figure 6 représente les trois scénarios avec leurs intervalles de valeurs de critères coût et durée. Dans notre exemple illustré sur la figure 6, la solution S_1 est

une solution qui peut être moins chère et susceptible d'être réalisée avec une durée plus courte par rapport aux autres solutions. Cependant, vu la non uniformité des distributions de probabilités des valeurs de coûts et de durées pour chaque scénario, les solutions S_2 et S_3 pourraient s'avérer plus intéressantes (les probabilités nous le montrent). Si le décideur fixe le seuil θ avec une valeur faible, cela signifie qu'il cherche à avoir moins de solutions sur le front de Pareto en acceptant moins d'incertitudes. Seules les solutions dont la probabilité de dominance est suffisante sont conservées. Le risque de choisir une solution moins intéressante vis-à-vis des critères est donc plus faible. Au contraire, un seuil θ plus élevé donnera un front de Pareto comportant plus de solutions. Le panel est plus large pour faire un choix, mais certaines solutions sont plus risquées dans le sens où elles peuvent s'avérer ne pas être les meilleures vis-à-vis des critères. Dans tous les cas, c'est au décideur de faire le choix final de la solution à mettre en œuvre en fonction des deux critères (coût et durée) et du niveau de risque qu'il est prêt à accepter.

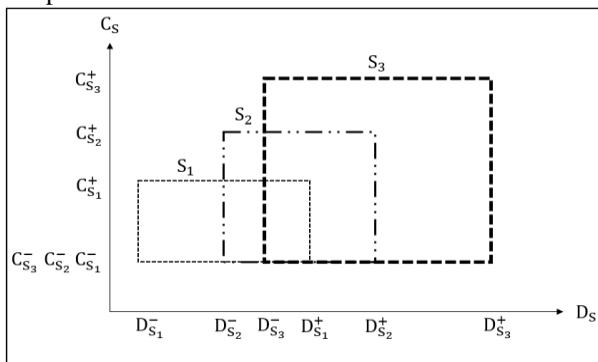


Figure 6 : Front de Pareto de S_1 , S_2 et S_3

6 CONCLUSION

Dans cet article, une approche probabiliste et multiobjectif a été proposée pour comparer, au sens de Pareto et sous incertitude, les scénarios d'un graphe de projet. L'approche permet de construire un front de Pareto en tenant compte des probabilités des valeurs de coût et de durée pour chaque scénario ainsi que du niveau d'incertitude acceptée par le décideur. Les probabilités représentent les impacts des risques sur le projet. Nos futurs travaux consistent à incorporer cette méthode probabiliste dans un algorithme multiobjectif à base de colonie de fourmis initialement proposé dans (Lachhab *et al.*, 2016). Cet algorithme permet de sélectionner des scénarios Pareto-optimaux dans un graphe de projet complexe. Le but de cette intégration est d'adapter l'algorithme afin de le rendre plus réaliste vis-à-vis de la prise en compte des différents risques potentiels et de leurs impacts sur le coût et la durée d'un scénario de projet.

REFERENCES

- Bahri, O., B.-A. Nahla and E.-G. Talbi, 2013. A Possibilistic framework for solving multi-objective problems under uncertainty: Definition of new Pareto optimality. *The IEEE 27th International Symposium on Parallel and Distributed Processing Workshops and PhD Forum (IPDPSW'2013)*, Cambridge, MA, USA, p. 405-414.
- Coelho, R.-F., 2015. Probabilistic Dominance in Multiobjective Reliability-Based Optimization: Theory and Implementation. *IEEE Transactions on Evolutionary Computation*, vol. 19.
- Das, S., S.-R. Chowdhury, B.-K. Panigrahi, S. Pattnaik and S. Das, 2009. Multi-objective Optimization with Uncertainty: Probabilistic and Fuzzy Approaches. *The World Congress on Nature & Biologically Inspired Computing*, Coimbatore, Inde.
- Hendriks, M., M. Geilein and T. Basten, 2011. Pareto Analysis with Uncertainty. *IEEE/IFIP 9th International Conference on Embedded and Ubiquitous Computing*, Melbourne, Australie.
- Lachhab, M., C. Béler, E.-L. Solano Charris and T. Coudert, 2017. Towards an Integration of System Engineering and Project Risk Management Processes for a Decision Aiding Purpose. *The 20th World Congress of the International Federation of Automatic Control (IFAC World Congress'2017)*, Toulouse, France, vol. 50, p. 7266-7271.
- Lachhab, M., T. Coudert and C. Béler, 2016. Scenario selection optimization in System Engineering Projects under uncertainty: A multi-objective Ant Colony method based on a learning mechanism. *The IEEE International Conference on Industrial Engineering and Engineering Management (IEEM'2016)*, Bali, Indonésie, p. 1235-1239.
- Mlakar, M., T. Tušar and B. Filipič, 2014. Comparing Solutions under Uncertainty in Multiobjective Optimization. *Mathematical Problems in Engineering*, vol. 2014.
- Teich, J., 2001. Pareto-Front Exploration with Uncertain Objectives. *Evolutionary Multi-Criterion Optimization. EMO 2001. Lecture Notes in Computer Science*, Berlin, Allemagne, vol. 1993, p. 314-328.

AVEZ-VOUS IDENTIFIE TOUTES LES PARTIES PRENANTES ?

Karla Itzel GOMEZ SOTELO, Xin YI, Claude BARON, Philippe ESTEBAN
 LAAS-CNRS, Université de Toulouse, CNRS, INSA
 UPS, Toulouse, France.
 {kgomezso, xyi, claude.baron,
 philippe.esteban}@laas.fr

Citlalih Yollohtli Alejandra GUTIERREZ ESTRADA, Luis de Jesús LAREDO VELAZQUEZ
 Instituto Tecnológico de Toluca
 Toluca, Estado de México – Mexico
 cgutierrez@toluca.tecnm.mx;
 luislaredovelazquez@gmail.com

RESUME: *Il existe aujourd'hui une compétition féroce entre les entreprises pour concevoir et livrer des produits sur le marché dès que possible; celles-ci cherchent ainsi à développer plus rapidement que jamais des systèmes qui doivent parfaitement répondre à des besoins de clients toujours plus exigeants. Les chercheurs en ingénierie système, en gestion de projet et en définition de concept, entre autres disciplines, ont compris qu'identifier précisément et exhaustivement les parties prenantes dès le début d'un projet de conception d'un nouveau système est essentiel pour bien définir les besoins et livrer au final un système qui réponde au mieux aux diverses attentes. L'objectif de cet article est ainsi d'aider les équipes d'analyse et de conception à déterminer les différentes parties prenantes d'un projet de développement d'un système technique. Notre proposition consiste en une typologie des parties prenantes permettant de faciliter leur identification et une table pour pondérer leur contribution au projet en fonction de leur importance. La proposition, conforme à la norme ISO / IEC / IEEE 15288 (2015), a été expérimentée dans trois études de cas, aboutissant à une première validation de celle-ci et à des opportunités d'amélioration futures.*

MOTS-CLEFS: *Ingénierie système; conception; gestion de projet; partie prenante; besoins.*

1 INTRODUCTION

Dans la compétition économique actuelle, les entreprises font la course pour mettre sur le marché avant les autres de nouveaux produits (Kiritani et Ohashi, 2015). Dans cet objectif, les entreprises doivent surmonter plusieurs difficultés. Blanchard et Fabrycky (2011) soutiennent que définir le problème est la tâche la plus difficile lors de la conceptualisation d'un nouveau produit. Si le problème n'est pas bien défini, les conséquences peuvent être désastreuses et la solution rejetée (Faisandier, 2012). Ce dernier ajoute que face à un nouveau problème, les individus ou organisations pensent immédiatement aux solutions sans investir suffisamment de temps pour définir le problème, ce qui peut conduire à l'échec du projet. En fait, l'une des raisons majeures d'échec des projets est une mauvaise définition des besoins, résultant en un produit délivré qui ne correspond pas aux besoins du client (Kiritani et Ohashi, 2015). Par conséquent, il est essentiel de recueillir et d'exprimer correctement les besoins des clients (ie. de façon non ambiguë) (Kiritani et Ohashi, 2015) car ceux-ci seront ensuite transformés en exigences pour le produit. Que ces besoins soient mesurables et qu'il soit donc possible de les exprimer qualitativement ou quantitativement est un atout supplémentaire (Faisandier, 2012; Kossiakoff et al., 2011). Pour (Blanchard et Fabrycky, 2011), la voix du client doit être entendue. Cependant, au-delà de la seule voix du client, les voix de toutes les parties prenantes¹ doivent

aussi se faire entendre afin que le produit soit un succès commercial. En effet, identifier les besoins exacts et complets des parties prenantes est essentiel et « Comment trouver ces besoins ? » reste LA difficulté dans un projet de développement de nouveau système. Si nous poursuivons le raisonnement, dans la mesure où les besoins sont exprimés par les différentes parties prenantes, il est par conséquent indispensable d'impliquer les parties prenantes dès le début du projet pour éliciter l'ensemble le plus complet des besoins (Faisandier; 2012) (SEBoK, 2017). Ainsi, la question clef devient « Qui sont les acteurs ? » et « Avez-vous envisagé toutes les parties prenantes ? »

Cet article apporte une contribution méthodologique et pragmatique à l'identification et l'analyse des parties prenantes impliquées dans un projet de développement du nouveau système. Il propose une typologie de ces parties prenantes facilitant ainsi leur identification, ainsi qu'une table de pondération des parties prenantes en fonction de leur contribution au projet permettant ainsi de mesurer l'influence des parties prenantes (et de leurs besoins exprimés) sur le projet. Les propositions de l'article sont conformes aux standards, notamment à ISO/IEC/IEEE 15288 (2015).

La section 2 contextualise la problématique en ingénierie système, gestion de projet et définition de concept. La section 3 présente quelques contributions historiques majeures. La section 4 propose une typologie et une pondération des parties prenantes. La section 5 présente trois études de cas et discute les contributions. La section 6 conclut et indique quelques perspectives de recherche.

¹Une « partie ayant un droit, une part ou une prérogative qui fait que le système ou certaines de ses propriétés doivent satisfaire les besoins ou les attentes de cette partie » (ISO/IEC/IEEE, 2017)

2 CONTEXTUALISATION DE LA PROBLEMATIQUE

Cette section souligne l'importance de l'identification des parties prenantes en ingénierie système, en gestion de projet et en définition de concept.

2.1 Ingénierie système

L'ingénierie système peut être décrite comme une « *approche coopérative interdisciplinaire pour le développement progressif et la vérification d'une solution pour le système, équilibrée sur l'ensemble de son cycle de vie, satisfaisant aux attentes d'un client et acceptable par tous* » (ISO/IEC/IEEE, 2017). Le cycle de vie de cette solution (ou système) est illustré en Figure 1.

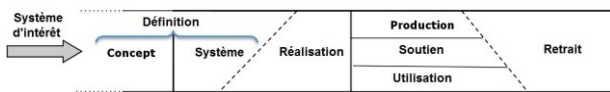


Figure 1. Cycle de vie générique et ses différentes phases (SEBoK, 2017)

Chaque phase a ses propres objectifs; par exemple, selon SEBoK (2017), les objectifs de la phase Concept sont: a) identifier les besoins des parties prenantes, b) évaluer les concepts alternatifs, et c) recommander des solutions possibles.

Selon l'ISO/IEC/IEEE 15288 (2015), quatre types de processus différents se déroulent à chaque cycle de vie du système: les processus d'entreprise, les processus contractuels, les processus de gestion de projet et les processus techniques. Cet article s'intéresse aux processus techniques qui transforment les besoins des parties prenantes en un produit devant satisfaire le client.

Le premier processus technique qui intervient (au début de la phase de définition du concept) est le processus de *Définition des besoins des parties prenantes*. Ce processus comprend des activités et des tâches; parmi elles, l'activité *Préparer la définition des besoins et exigences des parties prenantes* contient la tâche *Identifier les parties prenantes qui ont un intérêt au système tout au long de son cycle de vie*. Le travail de recherche présenté dans cet article se concentre sur cette tâche spécifique.

En conclusion, l'ingénierie système recommande d'identifier les parties prenantes ou les catégories de parties prenantes à travers le cycle de vie du système (ISO/IEC/IEEE 15288, 2015). Cette identification doit avoir lieu très tôt, au début de la phase de définition du concept (Faisandier, 2012; SEBoK, 2017).

2.2 Gestion de projet

Selon le PMI (2017), la vue interne d'un projet peut être perçue comme le cycle de vie du projet. Il y a cinq phases dans le cycle de vie du projet (cf. Figure 2).

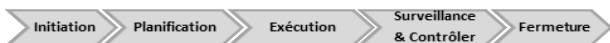


Figure 2. Cycle de vie du projet (PMI, 2017)

Dans la phase d'*Initiation* se déroule le processus d'*Identification des parties prenantes* (cf. Table 1).

Identification des parties prenantes		
Entrées	Outils	Sorties
1. Charte du projet	1. Jugement d'expert	1. Registre des parties prenantes
2. Documents commerciaux	2. Collecte de données	2. Demande de modification
3. Plan de gestion du projet	3. Analyse des données	3. Mises à jour du plan de gestion du projet
4. Documents du projet	4. Représentation des données	4. Mise à jour des documents du projet
5. Accords	5. Réunions	
6. Facteurs environnementaux d'entreprise		
7. Actifs de processus organisationnels		

Table 1. Entrées, outils et sorties du processus d'identification des parties prenantes (PMI, 2017)

Les *entrées* sont les documents précédemment créés, nécessaires pour exécuter le processus; les *outils* correspondent aux outils et techniques pouvant être utilisés selon les besoins du projet; les *sorties* sont les documents qui en résultent. Les outils permettent de s'adapter à des projets particuliers; selon les besoins du projet ils peuvent être utilisés ou non. Chaque outil inclut différentes techniques comme des questionnaires ou des techniques de brainstorming pour la collecte de données. L'analyse de données comprend des techniques d'analyse des parties prenantes comme lister les parties prenantes, leurs postes dans l'organisation, leurs rôles sur le projet, leurs attentes, leurs attitudes, leurs enjeux (pouvoir, influence, intérêt ou impact, droits légaux ou moraux, propriété, connaissances, contribution, modèle de saillance, etc.).

En conclusion, la gestion de projet recommande que l'identification, la classification et l'évaluation des parties prenantes impliquées se déroulent au début du projet, en appliquant les outils et les techniques qui sont les plus appropriés pour chaque projet.

2.3 Définition de concept

La définition de concept est l'une des premières étapes dans le développement du système (cf. Figure 1). Les ingénieurs doivent en effet comprendre les besoins des parties prenantes avant que le concept soit défini. Ainsi, la définition des besoins dépend étroitement de l'identification des parties prenantes.

Freeman (1984) définit les parties prenantes comme « *tout groupe ou individu qui peut affecter ou est affecté par les réalisations des objectifs organisationnels* » (Donaldson et Preston, 1995). Clarkson (1995) classe les parties prenantes en deux grandes catégories, *primaires* et *secondaires*. Laplume et al. (2008) classent les parties prenantes en *internes* et *externes*. (Wikstrom et al., 2010) ont mis en évidence que l'implication des parties prenantes au plus tôt dans la définition du projet est l'un des fondements de la création de valeur.

Les parties prenantes incluent bien évidemment le client mais pas seulement. Un projet ne peut bien sûr pas démarrer sans clients, cependant, les autres parties prenantes comme les consultants, ingénieurs et gestionnaires de projet, doivent également participer et interagir (Aapaoja et al., 2013). En effet, la définition de concept est une activité fortement pluri et interdisciplinaire qui implique une coopération étroite avec les clients, les concepteurs et ingénieurs (Wang et al., 2002).

Par conséquent, il est également crucial dans ce domaine de bien identifier les parties prenantes. On commence par définir le problème, on analyse les besoins, on en fait la synthèse et on évalue les idées résultant d'un brainstorming pour produire un concept (Wodehouse et Ion, 2010) (cf. Figure 3). Ce concept, représentant les idées initiales, peut se matérialiser sous différentes formes, comme un résultat de simulation, une image ou un graphique.

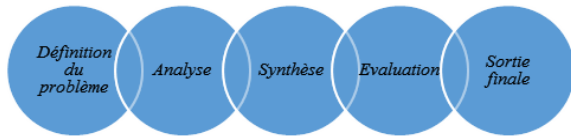


Figure 3. Processus de définition de concept (Wodehouse et Ion, 2010)

Dans la phase de *Définition du problème*, les clients expriment leurs besoins; plus tard les ingénieurs transformeront ceux-ci en exigences (Aapaoja et al., 2013). Ces dernières années, l'identification des parties prenantes a attiré beaucoup d'attention. En effet, cette étape a une influence forte sur le développement d'un concept, étant donné que les besoins auxquels le système délivré devra répondre proviennent de celles-ci. Ainsi, identifier complètement les véritables parties prenantes qui sont impliquées dans la conception du système est une question majeure.

En conclusion de cette analyse, nous pouvons dire que les domaines de l'ingénierie système, la gestion de projet et la conception s'accordent sur le fait que l'identification des parties prenantes et leur participation sont indispensables lors de la définition des besoins et exigences pour la bonne conception du futur système. Lorsque les parties prenantes collaborent avec l'équipe d'analyse et de conception, elles aident à clarifier ce qu'elles attendent vraiment du système.

3 CONTRIBUTIONS HISTORIQUES

Au fil des époques, certains auteurs majeurs ont tenté de répondre à cette question: « Quelles sont les parties prenantes à examiner lors de la définition des besoins et des exigences lors du développement d'un nouveau système ? » Quelques propositions intéressantes sont présentées ci-dessous.

3.1 Freeman et Reed, 1983

En 1983, Freeman et Reed ont fait une étude historique sur cette question. Ils ont constaté qu'avant les années 1970, étaient considérées comme parties prenantes seulement les actionnaires, employés, clients, fournisseurs, financeurs et l'entreprise. Plus tard, dans le milieu des années 1970, les chercheurs ont commencé à penser à *faire participer les parties prenantes* au lieu de simplement étudier leur *influence* sur le projet.

Ces auteurs proposent deux processus pour analyser les parties prenantes. Le premier processus est une méthode systématique d'identification des parties prenantes et d'évaluation de l'efficacité des stratégies organisation-

nelles; le second est une méthode systématique pour l'analyse de l'importance relative des parties prenantes, leur potentiel de coopération et leur menace concurrentielle. Ces processus aboutissent à la Table 2 qui donne quelques exemples de parties prenantes et de leurs différents enjeux de pouvoir.

Enjeu de pouvoir	Formel ou Votant	Économique	Politique
Capitaux propres	Actionnaires Administrateurs Intérêts minoritaires		Actionnaires dissidents
Économique		Fournisseurs Détenteurs de titres Clients Syndicats	Gouvernements locaux Gouvernements étrangers Groupes de consommateurs Syndicats
Influenceurs	Gouvernement SEC ² Directeurs externes	EPA ² /OSHA ²	Nader's Raiders ³ Gouvernement Associations professionnelles

Table 2. Parties prenantes selon (Freeman et Reed, 1983)

3.2 Mitchell, Agle et Wood, 1997

Mitchell et al. (1997) proposent une identification des parties prenantes basée sur trois attributs: pouvoir, légitimité et urgence.

Pouvoir: une relation entre les acteurs sociaux, dans lequel un acteur social, A, peut obtenir d'un autre acteur social, B, de faire quelque chose que B n'aurait pas fait autrement.

Légitimité: une perception généralisée ou des hypothèses que les actions d'une entité sont souhaitables, ou appropriées dans le cadre de certains systèmes de normes, de valeurs ou de croyance, construits socialement.

Urgence: le degré avec lequel les parties prenantes demandent une attention immédiate.

Chaque partie prenante peut avoir un, deux ou trois attributs; si une partie prenante n'a aucun attribut, cela signifie qu'elle n'est pas une partie prenante. Ainsi, la théorie de (Mitchell et al., 1997) permet de distinguer quelles sont les véritables parties prenantes. Il en résulte une classification en huit types de parties prenantes; la Figure 4 les identifie en fonction des attributs.

Mitchell et al. (1997) proposent également d'appliquer le concept de *saillance* dans le but d'expliquer à qui et à quoi il faut faire attention. La *saillance*, définie comme « le degré de priorité que donnent les managers aux demandes en concurrence des parties prenantes », est utilisée pour montrer que « le pouvoir gagne en autorité par la légitimité, et gagne en exercice par l'urgence ».

En concordance avec cette proposition, on peut dire que les parties prenantes identifiées ont différents niveaux d'importance selon leur niveau de saillance.

² United States Securities and Exchange Commission, United States Environmental Protection Agency, United States Occupational Safety and Health Administration

³ Associé de Ralph Nader et d'organisations dirigées par lui, qui participe à l'activisme juridique, politique et social de Ralph Nader.

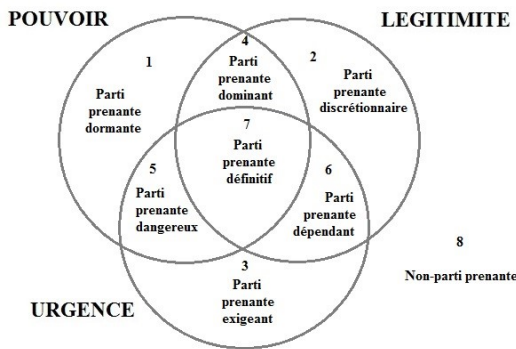


Figure 4. Typologie des parties prenantes (Mitchell et al., 1997)

3.3 Sharp, Finkelstein et Galal, 1999

Contrairement à la plupart des auteurs ((Faisandier, 2012) ou (SEBoK, 2017) pour ne citer qu’eux) qui mettent l’accent sur les relations entre le système et les parties prenantes, Sharp et al. (1999) proposent pour identifier les parties prenantes une approche axée sur les interactions entre les différents types de parties prenantes selon leur rôle, la nature de leurs relations et leurs interactions.

Ces auteurs proposent un ensemble de parties prenantes auquel ils réfèrent comme « *parties prenantes de base* » qui sont les:

- Utilisateurs*: personnes, groupes ou sociétés qui vont interagir et contrôler le système directement;
- Développeurs*: équipe d’analyse et de conception impliquée dans le processus d’ingénierie des exigences;
- Législateurs*: organismes professionnels, administrations, syndicats, représentants légaux, cadres de la sécurité, auditeurs d’assurance qualité, ... qui peuvent élaborer des directives qui auront une incidence sur le développement et/ou l’exploitation du système;
- Décideurs*: structures décisionnelles associées à l’utilisation ou au développement du système, comme des gestionnaires ou des contrôleurs financiers.

Sharp et al. (1999) suggèrent: a) d’identifier exactement ce qui se passe dans une situation de travail, d’observer les pratiques existantes, d’étudier le plan de travail correspondant à un nouveau rôle, d’utiliser des techniques participatives comme des enquêtes contextuelles par exemple pour impliquer les individus; b) de considérer que les parties prenantes peuvent être internes à l’équipe, internes à l’organisation, ou externes; c) de considérer le cycle complet d’activités pour chaque mission; d) de prendre en compte le cycle de vie complet du système lors de l’identification des parties prenantes « *de base* ». Remarquons que cette dernière recommandation est également présente dans Faisandier (2012) et le SEBoK (2017), et qu’elle est conforme à ce qui est décrit dans l’ISO/IEC/IEEE 15288 (2015) dans le processus de *Définition des besoins et exigences des parties prenantes*, plus précisément dans la tâche *Identifier les parties prenantes qui ont un intérêt dans le système tout au long de son cycle de vie*.

Sharp et al. (1999) proposent également le processus suivant afin d’identifier systématiquement les parties prenantes: 1) identifier tous les rôles spécifiques au sein du groupe de parties prenantes *de base*; 2) identifier les parties prenantes *fournisseur* pour chacun des rôles *de base*; 3) identifier les parties prenantes *client* pour chaque rôle *de base*; 4) identifier les parties prenantes *satellite* pour chacun des rôles *de base*; 5) répéter les étapes 1 à 4 pour chacun des groupes de parties prenantes identifiées dans les étapes 2 à 4 (cf. Figure 5).

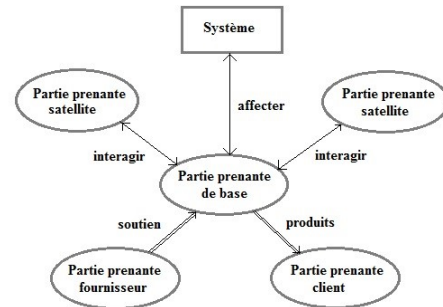


Figure 5. Relations entre parties prenantes (Sharp et al., 1999)

3.4 Glinz et Wieringa, 2007

Glinz et Wieringa (2007) soulignent que la première activité d’ingénierie des exigences doit être de déterminer qui sont les acteurs et quelle est leur importance. Les auteurs s’intéressent alors à l’identification des rôles des parties prenantes, en recherchant les personnes ou organismes qui: a) ont un intérêt actif dans le système parce qu’ils utilisent ou sont impliqués dans le processus de développement du système; b) doivent gérer, entretenir ou exploiter le système; c) sont impliqués dans le projet comme analyste, développeur, architecte, ingénieur qualité ou gestionnaire; d) sont responsables du processus ou de l’activité que le système est censé soutenir ou automatiser; e) ont un intérêt financier; e) sont des régulateurs du système; f) sont concernés par le système dans un sens négatif.

Affirmant que les parties prenantes ont des degrés d’importances différents, les auteurs proposent que les rôles des parties prenantes soient priorisés. Cette priorisation repose sur l’évaluation des risques encourus en ignorant ou en négligeant une partie prenante (cf. Table 3)

Rôle des parties prenantes	Conséquences du système si l’on néglige les parties prenantes	Risque encouru
Critique	Pourrait tuer le projet	Haut
Rôle du maire	Aurait un impact important et négatif sur le système	Moyen
Rôle du mineur	Aurait un impact marginal sur le système	Faible

Table 3. Priorisation des rôles des parties prenantes (Glinz et Wieringa, 2007)

3.5 Jepsen and Eskerod, 2009

Jepsen et Eskerod (2009) expliquent que la liste des parties prenantes est générée en pensant aux contributions nécessaires au projet et qui pourrait/devrait fournir ces contributions ou serait influencé autrement par le projet.

Les auteurs proposent de suivre cette démarche pour analyser les parties prenantes: a) identifier les parties prenantes (importantes); b) caractériser les parties prenantes en soulignant leurs contributions, leurs attentes concernant les récompenses pour ces contributions⁴ et leur pouvoir par rapport au projet; c) décider de la stratégie à utiliser pour influencer chaque partie prenante.

Comme le temps est une contrainte importante dans les projets, Jepsen et Eskerod (2009) proposent d'évaluer le *niveau de pouvoir* des parties prenantes avant d'éliciter les besoins, afin de prioriser d'entrée les parties prenantes et de consacrer majoritairement le temps à éliciter les besoins de celles qui comptent le plus. Les auteurs recommandent d'estimer le *niveau de pouvoir* selon un barème comportant quatre niveaux: aucun, faible, moyen et élevé. Ensuite, lors de l'élicitation des besoins, le chef de projet doit identifier où se trouve la *valeur* en déterminant les attentes et les bénéfiques souhaités par chaque partie prenante.

3.6 Turner and Zolin, 2012

Turner et Zolin (2012) proposent huit types différents de parties prenantes ayant des intérêts différents dans le projet: propriétaire ou investisseurs, responsable ou sponsor du projet, consommateurs, opérateurs ou utilisateurs, chef de projet et équipe projet, fournisseurs et public.

4 PROPOSITIONS: TYPOLOGIE ET PONDERATION DES PARTIES PRENANTES

Cette section propose une typologie des parties prenantes, selon leur rôle dans le projet, visant à faciliter l'identification de celles-ci, ainsi qu'une table permettant de pondérer les parties prenantes intervenant selon leur influence dans le processus décisionnel. Ces propositions sont complètement conformes à l'ISO/IEC/IEEE 15288 (2015). Elles sont accompagnées par une fiche qui guide leur application, décrite à la fin de cette section.

Cet article explique en détail comment *Identifier et pondérer les parties prenantes*, première étape d'une méthodologie plus complète pour transformer les besoins en exigences des parties prenantes, proposée par Gomez, Sotelo et al. (2017). La pondération des parties prenantes sera notamment utilisée dans l'étape 4 de la méthodologie: *Etablissement de l'ordre de priorité des besoins des parties prenantes (SkN)*. La Figure 6 présente la méthodologie; elle ne sera cependant pas développée ici.

4.1 Typologie des parties prenantes

La proposition repose principalement sur les huit types de parties prenantes suggérés par Turner et Zolin (2012).

⁴ Une contribution est le soutien nécessaire de la part des parties prenantes. "Les contributions ne sont pas seulement des livrables mais aussi comment les parties prenantes prennent part aux travaux du projet en participant aux réunions et en prenant des décisions" (Jepsen et Eskerod, 2009).

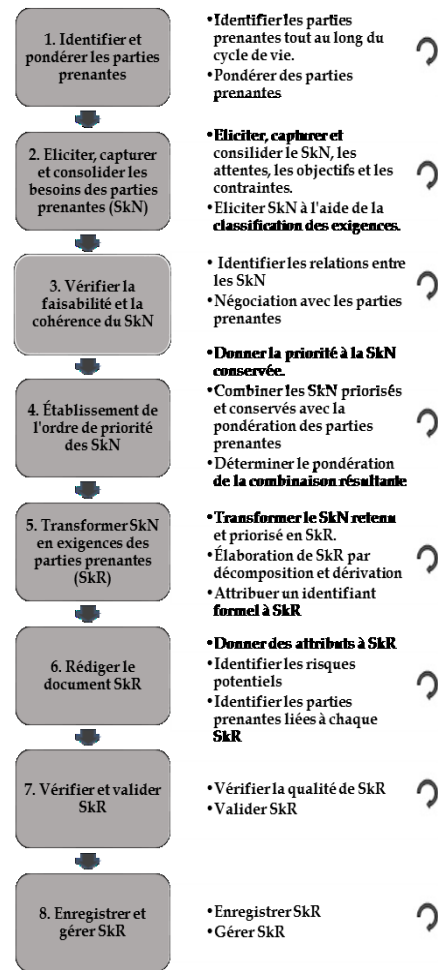


Figure 6 Méthodologie pour transformer les besoins en exigences des parties prenantes (Gomez Sotelo et al., 2017)

Elle intègre également: a) les contributions de Sharp et al. (1999) et Glinz et Wieringa (2007) quant aux parties prenantes de type législateurs ou régulateurs; b) les recommandations de Faisandier (2012) de considérer également comme parties prenantes les personnes qui ne désirent pas le système ou qui sont mal intentionnées par rapport au projet, et de considérer l'ensemble du cycle de vie; et c) les résultats de Glinz et Wieringa (2007) qui considèrent aussi les personnes qui sont affectées négativement par le système ou qui pourraient tuer le système (cf. Table 4).

Types de parties prenantes	Description
1. Législateurs et régulateurs	Organismes professionnels, agences gouvernementaux, syndicats, représentants légaux, responsables de la sécurité, auditeurs de l'assurance de la qualité, etc. peuvent produire des lignes directrices pour l'exploitation qui auront une incidence sur le développement et / ou l'exploitation du système.
2. Propriétaire ou investisseurs	Qui paie, par, pour son exploitation et en tire des revenus.
3. Chef de projet ou sponsor du projet	Avant le projet, identifier le besoin d'un nouveau système et les avantages potentiels qu'il apportera
4. Consommateurs	Acheter le produit ou le service que le nouveau système produit
5. Opérateurs / utilisateurs	Qui exploitent le système au nom du propriétaire
6. Chef de projet et équipe de projet	Qui dirige l'équipe du projet et qui exécute le projet
7. Fournisseur principal	Direction de l'entrepreneur principal

8. Autres fournisseurs	Personnes ou groupes qui fournissent des biens, des matériaux, des travaux ou des services
9. Public	Le public concerné par les impacts environnementaux et sociaux du système, il voudra savoir comment ses impôts ont été dépensés
10. Personnes qui ne veulent pas du système ou malveillantes	Personnes qui sont touchées par le système dans un sens négatif ou qui ont l'intention malveillante de représenter un risque ou une menace possible pour le système.

Table 4. Typologies des parties prenantes

4.2 Pondération des parties prenantes

Comme nous l'avons souligné, plusieurs auteurs affirment qu'il est nécessaire de faire la différence entre les parties prenantes selon leur importance:

- Glinz et Wieringa (2007) établissent que les parties prenantes peuvent avoir différents niveaux d'importance;
- Freeman et Reed (1983) suggèrent qu'une fois que les parties prenantes sont identifiées, il est nécessaire d'évaluer leur importance relative;
- Jepsen et Eskerod (2009) recommandent l'identification des parties prenantes (importantes) et l'estimation de leur niveau de pouvoir.

Toutefois, cette différence doit être identifiée et traduite objectivement afin d'obtenir des résultats fiables. Une solution consiste à pondérer les intervenants.

La stratégie que nous proposons pour cela est basée sur le niveau de *saillance* (Mitchell et al., 1997), déjà utilisé notamment en gestion de projet pour classer et analyser l'impact potentiel des acteurs d'un projet. La proposition est que chaque partie prenante soit évaluée selon ses trois attributs (pouvoir, légitimité et urgence), résultant en l'obtention d'une *valeur* totale pour chaque partie prenante permettant de déterminer son importance relative dans les processus de prise de décision. Pour réduire la subjectivité lors de cette pondération, nous proposons d'utiliser des valeurs numériques liées au contexte des parties prenantes à l'intérieur de l'organisation (cf. Table 5). En ligne se trouvent les trois attributs (pouvoir, légitimité et urgence); dans la colonne de gauche, les valeurs de poids proposées pour chaque attribut; et dans le corps de la table, les descriptions des contextes spécifiques des parties prenantes. Cette table est utile pour la pondération des parties prenantes, du fait de sa simplicité d'utilisation.

Comme on peut le voir, les valeurs de *pouvoir* (de 0,00 à 3,00) sont supérieures aux valeurs de *légitimité* (de 0,00 à 1,50), et les valeurs de *légitimité* sont elles-mêmes plus élevées que les valeurs d'*urgence* (de 0,00 à 0,75). Cette proposition de barème pour les valeurs des poids repose sur l'hypothèse de Mitchell et al. (1997) que « *le pouvoir gagne en autorité par la légitimité, et gagne en exercice par l'urgence* » (le choix des valeurs est arbitraire). Avec les valeurs proposées, la valeur maximale pour une partie prenante est ainsi de 5,25; si la valeur égale zéro, la partie prenante ne doit pas être prise en compte. Cette affectation de valeurs permet ainsi la détection des véritables parties prenantes et celles qui n'influent pas.

La proposition intègre la possibilité que plusieurs parties prenantes peuvent avoir le même *poids* et donc avoir la même importance relative, étant donné que cette valeur *totale* est obtenue par addition des valeurs obtenues pour chaque attribut.

Valeurs	Attributs
Pouvoir	
<i>La partie prenante...</i>	
0.00	...ne peut pas influencer ou prendre des décisions et déterminer les actions à suivre
1.00	...peut parfois influencer les décisions et déterminer les actions à suivre
2.00	...peut parfois prendre des décisions et / ou déterminer les actions à suivre
3.00	...peut prendre des décisions et / ou déterminer les actions à suivre
Légitimité	
<i>La partie prenante...</i>	
0.00	...n'a pas de réclamation légale, morale ou présumée sur l'entreprise et les groupes qui ont la capacité d'influencer le comportement, la direction, le processus ou le résultat de l'entreprise
0.50	...a une petite réclamation légale, morale ou présumée envers l'entreprise et les groupes qui ont la capacité d'influencer le comportement, la direction, le processus ou le résultat de l'entreprise
1.00	...a des prétentions légales, morales ou présumées sur l'entreprise et les groupes qui ont la capacité d'influencer le comportement, la direction, le processus ou le résultat de l'entreprise
1.50	...a une réclamation totale, légale, morale ou présumée sur l'entreprise et les groupes qui ont la capacité d'influencer le comportement, la direction, le processus ou le résultat de l'entreprise
Urgence	
<i>La partie prenante...</i>	
0.00	...retard dans la gestion de la réclamation acceptable et / ou ne demandent pas une attention immédiate
0.25	...retard dans la gestion de la réclamation parfois acceptable et / ou demande parfois une attention immédiate
0.50	...retard dans la gestion de la réclamation presque jamais acceptable et / ou demande presque immédiatement une attention
0.75	...retard dans la gestion de la réclamation presque inacceptable et / ou demande immédiatement une attention

Table 5. Table de pondération des parties prenantes

4.3 Fiche support

Pour aider à appliquer les propositions de cet article, nous avons élaboré un formulaire. Un extrait de ce formulaire (de taille trop importante pour être montré dans son intégralité) est présenté en Figure 7. Il contient les champs suivants: en haut, les différentes étapes du cycle de vie de système (sur la figure apparaît seulement l'étape 'Concept'); en colonne les attributs (pouvoir, légitimité et urgence) et la *valeur totale*; à gauche, les différents types de parties prenantes.

Nos recommandations pour appliquer les propositions et utiliser le formulaire sont:

- Identifier les parties prenantes selon la typologie proposée et les écrire dans la colonne parties prenantes du formulaire;
- Choisir une partie prenante identifiée et lire chacune des quatre descriptions de l'attribut *pouvoir* indiquées en Table 5;
- En s'appuyant sur les connaissances et l'expérience, sélectionner la description qui correspond le mieux à la partie prenante; une seule description doit être choisie;
- Cocher la case correspondante de la description choisie.
- Répéter les étapes 2 à 4 avec les attributs *légitimité* et *urgence*.
- Obtenir la *valeur totale* pour chaque partie prenante en additionnant les valeurs de *puissance*, de *légitimité* et d'*urgence*; noter cette valeur dans la colonne *T*. Répéter les étapes 2 à 6 jusqu'à ce que toutes les parties prenantes soient affectées d'un poids.

Étapes du cycle de vie et leurs objectifs	Concept												
	Identifier les besoins des partis prenantes												
	Évaluer les concepts de rechange Recommander des solutions possibles												
Types de parties prenantes	Parties prenantes	Pouvoir			Legitimité			Urgence			T		
		0	1	2	3	0,0	0,5	1,0	1,5	0,00		0,25	0,50
1. Législateurs et régulateurs													
2. Propriétaire ou investisseurs													
3. Chef de projet ou sponsor du projet													
4. Consommateurs													
5. Opérateurs / utilisateurs													
6. Chef de projet et équipe de projet													
7. Fournisseur principal													
8. Autres fournisseurs													
9. Public													
10. Personnes qui ne veulent pas du système ou malveillantes													

Figure 7. Extrait du formulaire

5 EXPERIMENTATION DE LA PROPOSITION

Trois expériences ont été menées par des étudiants sur des projets différents à l’Institut Technologique de Toluca au Mexique. L’objectif de ces études de cas était d’évaluer si la typologie et la table de pondération étaient utiles pour identifier les parties prenantes et leur importance relative.

5.1 Système d’information pour surveiller les arrêts dans une chaîne de production

Dans une entreprise fabriquant des emballages industriels et pour la restauration, les étudiants devaient déterminer les parties prenantes associées à un nouveau système d’information qui permettrait de surveiller les arrêts dans les chaînes de production. L’objectif du système est de mesurer, contrôler et améliorer le niveau de production. Grâce à la méthodologie proposée, les étudiants ont pu mesurer l’importance d’impliquer les parties prenantes dans les premières étapes de la conception du système et ont été guidés pour trouver l’ensemble des parties prenantes. Dans cette étude de cas, il a été constaté que les opérateurs de la ligne de production appartenaient à la fois à la catégorie 5 *Opérateurs/utilisateurs* et à 10 *Personnes qui ne veulent pas du système*. Au moment de cette détection, la direction a décidé de prendre des mesures pour diminuer la résistance au changement de la part des opérateurs au moment de la mise en œuvre du nouveau système.

5.2 Application web sur un téléphone mobile pour des achats en ligne

L’objectif de cette étude était de concevoir un système pour aider les acheteurs lors de l’achat de vêtements en ligne, en comparant pour eux les prix d’un produit donné dans différents magasins d’une manière simple et rapide. La typologie proposée a permis aux élèves de définir les parties prenantes et leur valeur. Ils ont ainsi pu déterminer que les consommateurs obtenaient une valeur totale maximale, ce qui est cohérent avec la nature du système. Les élèves ont ainsi pris des décisions comme par

exemple faire une étude de marché préalable afin de connaître en profondeur les besoins des consommateurs.

5.3 Générateur électrique sans fil

Le projet consistait ici à concevoir un générateur électrique produisant l’énergie nécessaire pour charger un téléphone mobile. Dans ce cas d’étude encore la typologie des parties prenantes s’est révélée précieuse car elle a procuré une aide réelle aux étudiants. Une fois déterminés les poids des parties prenantes, les étudiants se sont rendu compte que plusieurs d’entre elles avaient des *valeurs totales* élevées. De ce fait, leurs différents besoins devaient être considérés au même niveau d’importance; en effet, les besoins de l’investisseur étaient orientés vers la rentabilité et les parts de marché, le *fournisseur principal* s’inquiétait de la disponibilité de composants de qualité et le *chef de projet* souhaitait une éventuelle analyse comparative afin de proposer un produit à valeur ajoutée.

5.4 Avantages et possibilités d’amélioration

Après avoir expérimenté nos propositions dans trois études de cas, les étudiants ont répondu à deux questions: 1) Quels sont les intérêts des propositions ? et 2) Quels sont les problèmes rencontrés lors de l’application des propositions ? Les élèves ont considéré que la typologie des parties prenantes a vraiment aidé à l’identification de celles-ci, et que la table de pondération a permis d’évaluer objectivement leur importance relative. Toutefois, ces études de cas ont été des projets d’étudiants, ce qui signifie que c’étaient des projets relativement simples avec quelques parties prenantes seulement qui, dans certains cas, ont joué plusieurs rôles. Ceci a d’ailleurs induit de la confusion pour distinguer le *type* de chaque partie prenante. Une autre difficulté qu’ils ont rencontrée est inhérente à leur expérience réduite en conception de systèmes; ils auraient ainsi aimé avoir un peu de formation préalable afin de mieux comprendre le vocabulaire, et davantage d’instructions pour appliquer la méthode.

On peut dire qu’à travers les études de cas, une première validation de la typologie proposée a été réalisée; les différents types de parties prenantes proposés ont permis de les identifier. Une observation importante est qu’une partie prenante peut appartenir à plusieurs catégories en même temps et avoir des poids différent. Par exemple, en considérant l’étude de cas décrite dans la sous-section 5.1, l’opérateur peut avoir une pondération de 1,50 dans la catégorie *Opérateur/utilisateur*, et en même temps une pondération de 2,25 dans la catégorie *Personne qui ne veut pas du système*; dans ce cas nous recommandons de conserver la pondération la plus élevée.

La validation des pondérations a été effectuée en comparant les valeurs obtenues pour chaque partie prenante dans l’application de notre proposition, par rapport à les valeurs obtenues subjectivement sur la base de l’expérience des participants.

Les valeurs proposées pour les attributs pourraient être analysées dans des études futures, afin de détecter si les coefficients de pondération sont adéquats ou peuvent être améliorés d'une façon ou d'une autre.

6 CONCLUSION

L'identification des parties prenantes est une activité essentielle (et exigée par ISO/IEC/IEEE 15288, 2015) dans le développement de systèmes car elle conditionne la définition des besoins des systèmes, et de ce fait, tout le développement ultérieur du système et la réussite du projet. Cet article vise à procurer un support méthodologique pragmatique à la bonne réalisation de cette activité, par la proposition d'une typologie des parties prenantes et une table de pondération de celles-ci selon leur importance dans le projet. Les premières expérimentations dans de petits projets prouvent que les propositions apportent une aide utile. Un travail complémentaire de clarification du vocabulaire et de formation pour l'utilisation du formulaire support pourra être fait en prolongement de ce premier travail. Une application des propositions sur des projets plus importants permettra d'avoir une meilleure évaluation de leur intérêt.

REFERENCES

- Aapaoja, A., Haapasalo, H. and P. Söderström, 2013. Early stakeholder involvement in the project definition phase: case renovation. *ISRN Industrial Engineering*, 2013.
- Blanchard, B. S., and W. J. Fabrycky, 2011. *Systems Engineering and Analysis*. 5th ed. Prentice Hall International Series in Industrial and Systems Engineering. Boston, Prentice Hall.
- Clarkson, M. E., 1995. A stakeholder framework for analyzing and evaluating corporate social performance. *Academy of management review*, 20(1), pp.92-117.
- Donaldson, T. and L. E. Preston, 1995. The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of management Review*, 20(1), pp.65-91.
- Faisandier, A., 2012. *Systems Opportunities and Requirements*, vol.2. Engineering and Architecting Multidisciplinary Systems. France, Sinergy'Com.
- Freeman, R. E., and D. L. Reed, 1983. Stockholders and Stakeholders: A New Perspective on Corporate Governance. *California Management Review*, 25(3), 88-106.
- Freeman, R. E., 1984. *Strategic management: A stakeholder approach*. Cambridge university press.
- Glinz, M., and R. Wieringa, 2007. Guest Editors' Introduction: Stakeholders in Requirements Engineering. *IEEE Software*, 24(2), 18-20.
- Gómez Sotelo, K. I., C. Baron, P. Esteban and C.Y.A. Gutiérrez Estrada, 2017. Metodología para transformar necesidades en requisitos, integrando Ingeniería de Sistemas, Calidad y Pensamiento Esbelto. *ECORFAN. Revista de Aplicaciones de la Ingeniería*. 1(2), 42 – 51.
- ISO/IEC/IEEE, 2015. *Systems and Software Engineering -- System Life Cycle Processes*. Geneva, Switzerland: ISO/IEC/IEEE 15288:2015.
- ISO/IEC/IEEE, 2017. *Systems and Software Engineering -- Vocabulary (SEVocab)*. Geneva, Switzerland: ISO/IEC/IEEE 24765:2017.
- Jepsen, A. L., and P. Eskerod, 2009. Stakeholder analysis in projects: Challenges in using current guidelines in the real world. *International Journal of Project Management*, 27(4), 335-343.
- Kiritani, K., and M. Ohashi, 2015. The Success or Failure of the Requirements Definition and Study of the Causation of the Quantity of Trust Existence Between Stakeholders. *Procedia Computer Science*, 64, 153-160.
- Kossiakoff, A., W. N. Sweet, S. J. Seymour, and S. M. Biemer, 2011. *Systems engineering: principles and practice* (2nd ed). Hoboken, N.J: Wiley-Interscience.
- Mitchell, R. K., B. R. Agle and D. J. Wood, 1997. Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review*, 22(4), 853-886
- Laplume, A. O., K. Sonpar, and R. A. Litz, 2008. Stakeholder theory: Reviewing a theory that moves us. *Journal of management*, 34(6), 1152-1189.
- Project Management Institute (PMI) 2018. *Guide to the Project Management Body of Knowledge (PMBOK)*. S.l.: Project Management Inst.
- SEBoK contributors, 2017. *Guide to the Systems Engineering Body of Knowledge (SEBoK)*. SEBoK. Retrieved from [http://sebokwiki.org/w/index.php?title=Guide_to_the_Systems_Engineering_Body_of_Knowledge_\(SEBoK\)&oldid=53122](http://sebokwiki.org/w/index.php?title=Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK)&oldid=53122).
- Sharp, H., A. Finkelstein, and G. Galal, 1999. Stakeholder identification in the requirements engineering process. *IEEE*. 387-391.
- Turner, R., and R. Zolin, 2012. Forecasting Success on Large Projects: Developing Reliable Scales to Predict Multiple Perspectives by Multiple Stakeholders Over Multiple Time Frames. *Project Management Journal*, 43(5), 87-99.
- Wodehouse, A. J. and W. J. Ion, 2010. Information use in conceptual design: Existing taxonomies and new approaches. *International Journal of Design*, 4(3).
- Wang, L., W. Shen, H. Xie, J. Neelamkavil, and A. Pardasani, 2002. Collaborative conceptual design—state of the art and future trends. *Computer-Aided Design*, 34(13), 981-996.
- Wikström, K., K. Artto, J. Kujala, and J. Söderlund, 2010. Business models in project business. *International Journal of Project Management*, 28(8), 832-84.

Bi-objective MIP formulation for the optimization of maintenance planning on French military aircraft operations

Franco Peschiera, Nicolas Dupin, Olga Battaïa, Alain Haït

ISAE-SUPAERO, Université de Toulouse, 31055 Toulouse cedex 4 - France
 {franco.peschiera,nicolas.dupin,olga.battaia,alain.hait}@isae-supaeero.fr

ABSTRACT: *A specific Flight and Maintenance Planning problem is presented. In this problem, preventive maintenance operations are scheduled for military aircraft along with the assignment of regular missions. The quality of the solutions is measured via a bi-objective function that smooths both maintenance operations and aircraft unavailability among time periods. A real world dataset provided by the French Air Force is used to test an exact optimization approach based on a Mixed Integer Programming model. Mono-objective computations are solved to optimality for medium size instances, allowing to compute exactly the Pareto frontier of the bi-objective optimization problems. The tests show that these two objectives do not lead to the same optimal solution, but very good compromise solutions can be found thanks to the bi-objective optimization.*

KEYWORDS: *optimization, Flight and Maintenance Planning, aircraft maintenance, Mixed Integer Programming, bi-objective optimization*

1 Introduction

Flight and Maintenance Planning (FMP) problems are raised in civilian or military operational applications. They schedule maintenances of aircraft taking into account the implications of maintenance immobilizations for the demand in flight hours, for a conjoint optimization among maintenance dates and flight hours.

The specific FMP of this paper comes from the French Air Force's operations. It has two specificities regarding the classical FMP considered in the literature. On one hand, the fleet of aircraft is heterogeneous, with different standards, capacities and retrofits. On the other hand, there is a will to align two objectives: the smoothing of the maintenance operations and the minimization of the global unavailability of aircraft to be able to add new missions. This last point calls for a bi-objective optimization.

This paper investigates how to solve such new optimization problem, and how to take good compromise decisions if these objectives are antagonistic.

The article is structured as follows:

Section 2 presents a detailed description of the problem. An analysis of the previous work on FMP is given in section 3. Section 4 presents a new MIP formulation for the considered problem. Section 5 introduces the real case study. Section 6 discusses the obtained results. Finally, section 7 provides conclusions and directions of further work.

2 Problem statement

The problem consists in assigning resources to predefined tasks and scheduling periodic preventive maintenances for these same resources. For the seek of simplicity, the following convention has been adopted: aircraft are called resources, missions are called tasks.

2.1 Tasks

There is a fixed set of $j \in \mathcal{J}$ tasks to be accomplished over an horizon of time divided into $t \in \mathcal{T}$ discrete periods. For their execution, these tasks require the assignment of a specific number of resources R_j each period of time the task is active. The start and end periods for each task are known and a task is considered active between its start and end period.

During each period, tasks consume an amount of time equal to H_j from each of its assigned resources.

The total number of resources being used at any given time in a specific task should be equal to R_j .

Each task requires one and only one type of resource which, in addition, should comply with additional task requirements.

2.2 Resources

There is a set $i \in \mathcal{I}$ of available resources that are assigned to tasks in order to accomplish them. Each resource can only be assigned to a single task in any given period. These resources suffer from wear and

tear and require regular maintenance operations during their lifetime. The need for maintenance is calculated based on two indicators.

The first one is called "remaining elapsed time" (or ret_{it}). It expresses the amount of time (measured in time periods) after which the resource cannot be used anymore and has to undergo a maintenance operation. Its value is calculated for each resource i and each time period t . In a similar way, "remaining usage time" (or rut_{it}) is used to measure the amount of time that the resource i can be used before needing a maintenance operation at any given period t .

At any given period, including at the start of the planning horizon, each resource has a specific status given by remaining usage time and remaining elapsed time.

2.3 Maintenances

Maintenance operations are the process by which resources that have reached a limit in some indicator can return to a state where they can continue to be used in tasks.

Each maintenance operation has a fix duration of M periods.

After a maintenance operation, a resource restores its remaining elapsed time and remaining usage time to their max values E and H respectively.

2.4 Possible states

As a summary, the following are the possible logical states that a resource can be in: assigned to a task (see 2.1); under maintenance (see 2.3); available.

2.5 Time

In planning tasks and maintenances, it is important to take into account the initial and end state of each resource. This initial state can be a maintenance or an assigned task. If a resource is under maintenance, it needs to continue in this state for its remaining maintenance time. Tasks' assignments should be taken into account in a similar manner.

The remaining used and elapsed times are assigned to each resource at the beginning of the planning horizon.

Finally, the state of each resource at the end of the planning horizon, its remaining (elapsed, used) time, needs to be defined.

2.6 Objectives

In this study, the following objectives are considered.

Given that the creation of new tasks and the duration of maintenance are considered stochastic in real-life, one basic goal is to maximize the robustness of the planning by having the greatest amount of available resources at every period of the planning horizon.

Given the limited amount of maintenance capacity and its cost, another goal is to smooth as much as possible the number of resources under maintenance over the planning horizon.

As it will be shown in the mathematical formulation, these objectives are quite related one with the other: the more resources are in maintenance in a given period, the more unavailable resources will be had.

3 Related work

The Military Flight and Maintenance Planning problem considered assigns missions and schedules maintenance operations to aircraft. In this problem, aircraft are needed to comply with a series of planned missions while at the same time needing to receive frequent preventive maintenances in order to be ready for new missions.

It is a variant of the Flight and Maintenance Planning problem, where maintenance operations and flights are scheduled to commercial aircraft. The main differences are the fact that flights are not modeled geographically: all aircraft are assumed to exit the main depot and return to the same depot at the end of the flight.

Other differences include the following.

Firstly, the objective function's motivation is on reducing the load on the maintenance facilities and guaranteeing the availability of the fleet instead of reducing the cost. Also, the maintenance operations take considerably longer and the rules that govern the maintenances are different.

The temporal scope of the problem is also larger, making the normal planning horizon and each period larger too: years instead of months and months instead of days, respectively.

In the following, the studies on the Military Flight and Maintenance Planning are discussed.

In Kozanidis (2008), a model for the Hellenic Air Force is presented. It assigns flight hours to aircraft and schedules flexible maintenance periods based on a "remaining usage time" rule. The objectives are to maximize this "remaining usage time" for the aircraft as well as to smooth the availability of aircraft during the planning horizon. Instances of 24 aircraft and 6 monthly periods were solved with Mixed Integer Programming (MIP) formulations and heuristics.

In Cho (2011), the objective is to smooth the maintenance operations. Randomly generated instances of 15 aircraft and 520 bi-daily periods were solved with MIP formulations. The final state of aircraft is taken into account in order to distribute as much as possible the "remaining usage time" of aircraft.

In Verhoeff et al. (2015), a model for the Royal Netherlands Air Force is developed. Here, three criteria for improving the planning robustness are presented and applied in a MIP model: availability, serviceability and sustainability.

In Gavranis and Kozanidis (2015), an exact algorithm is presented and tested on instances of 6 periods and between 10 and 200 aircraft for the special case of maximizing the residual flight time availability or total remaining usage time.

Both Kozanidis (2008) and Cho (2011) develop heuristic alternatives to the MIP formulation for solving larger problem instances. In Steiner (2006), a constructive, multi-stage heuristic approach is described to obtain balanced maintenances for the Swiss Air Force. Both calendar-based maintenances actions (CBMA) and usage-based maintenance actions (UBMA) are used. Finally, in Kozanidis et al. (2014), two heuristics are presented to solve large instances of the problem by taking consecutive decisions period by period with the help of mathematical models at each step.

It should be noted that previous formulations only consider an aggregate demand in flight hours for all aircraft. The aircraft are assumed to be homogeneous and interchangeable. The use of aggregate demand makes it impossible to take into account the task requirements for aircraft.

Lastly, in De Chastellux (2016) a mission-centered modeled is presented. Here, cycles, composed of one maintenance followed by mission assignments were used. Instances of 6 - 70 aircraft and 60 monthly periods were solved. Close to optimal solutions for small to medium instances using MIP formulations and heuristics based on these MIP formulations were found.

In the following, an alternative model is presented. This model is then tested in a case study.

4 Optimization model

4.1 Sets

\mathcal{T}_j	time periods $t \in \mathcal{T}$ in which task j is active.
\mathcal{J}_t	tasks $j \in \mathcal{J}$ to be realized in period t .
\mathcal{I}_j	resources $i \in \mathcal{I}$ that can be assigned to task j .
\mathcal{O}_i	tasks $j \in \mathcal{J}$ for which resource i can be used.
\mathcal{T}_t^s	time periods $t' \in \mathcal{T}$ such that $t' \in \{\max\{1, t - M + 1\}, \dots, t\}$.

4.2 Parameters

H_j	amount of resource time required by task j .
R_j	number of resources required by task j .
M	maintenance duration in number of periods.
E	remaining elapsed time after a maintenance.
H	remaining usage time after a maintenance.
W_1	weight of the first objective in the objective function.
W_2	weight of the second objective in the objective function.
N_t	number of resources in already-planned maintenances in period t at the beginning of the planning horizon.
D_t	number of resources to be assigned in total in period t .
Rut_i^{Init}	remaining usage time for resource i at the start of the planning horizon.
Ret_i^{Init}	remaining elapsed time for resource i at the start of the planning horizon.
Ret_{sum}^{Init}	sum of remaining elapsed times at the start of the planning horizon.
Rut_{sum}^{Init}	sum of remaining elapsed time at the start of the planning horizon.

4.3 Variables

The following decision variables define a solution.

a_{jti}	=1 if task $j \in \mathcal{J}$ in period $t \in \mathcal{T}_j$ is realized with resource $i \in \mathcal{I}_j$, 0 otherwise.
m_{it}	=1 if resource $i \in \mathcal{I}$ starts a maintenance operation in period $t \in \mathcal{T}$, 0 otherwise.
rut_{it}	remaining usage time (continuous) for resource $i \in \mathcal{I}$ at the end of period $t \in \mathcal{T}$.
ret_{it}	remaining elapsed time (integer) for resource $i \in \mathcal{I}$ at the end of period $t \in \mathcal{T}$.
u_{max}	maximal number (integer) of unavailable resources in any period.
m_{max}	maximal number (integer) of resources in maintenance in any period.

Note that a_{jti} and m_{it} are initially set up to 0 for all resources already in maintenance at the beginning of the planning horizon for the remaining time periods of maintenance. The remaining usage time for each resource at the beginning of the planning horizon is used to initialize rut_{i0} .

4.4 Constraints

The objective is to simultaneously minimize the maximum number of maintenances and the maximum number of unavailable aircraft.

$$\text{Min } W_1 m_{max} + W_2 u_{max} \quad (1)$$

where weights W_1 and W_2 are chosen by the decision maker. W_1 penalizes the maximum number of aircraft in maintenance among all periods. W_2 penalizes the maximum number of unavailable aircraft among all periods. The following constraints are used in the model:

$$\sum_{t' \in \mathcal{T}_i^s} \sum_{i \in \mathcal{I}} m_{it'} + N_t \leq m_{max} \quad t \in \mathcal{T} \quad (2)$$

$$\sum_{t' \in \mathcal{T}_i^s} \sum_{i \in \mathcal{I}} m_{it'} + N_t + D_t \leq u_{max} \quad t \in \mathcal{T} \quad (3)$$

$$\sum_{i \in \mathcal{I}_j} a_{jti} = R_j \quad j \in \mathcal{J}, t \in \mathcal{T}_j \quad (4)$$

$$\sum_{t' \in \mathcal{T}_i^s} m_{it'} + \sum_{j \in \mathcal{J}_t \cap \mathcal{O}_i} a_{jti} \leq 1 \quad t \in \mathcal{T}, i \in \mathcal{I} \quad (5)$$

Maintenance capacity is controlled by (2). The number of unavailable resources is defined by (3). Tasks' resource requirements are defined by (4). Constraints (5) guarantee that a resource can be used only for one task or maintenance operation at the same period.

$$\begin{aligned} rut_{it} &\leq rut_{it-1} + Hm_{it} \\ &- \sum_{j \in \mathcal{J}_t \cap \mathcal{O}_i} a_{jti} H_j \\ &t = 1, \dots, \mathcal{T}, i \in \mathcal{I} \end{aligned} \quad (6)$$

$$rut_{i0} = Rut_i^{Init} \quad i \in \mathcal{I} \quad (7)$$

$$rut_{it} \geq Hm_{it} \quad t \in \mathcal{T}, i \in \mathcal{I} \quad (8)$$

$$rut_{it} \in [0, H] \quad t \in \mathcal{T}, i \in \mathcal{I} \quad (9)$$

$$\begin{aligned} ret_{it} &\leq ret_{it-1} - 1 \\ &+ Em_{it} \\ &t = 1, \dots, \mathcal{T}, i \in \mathcal{I} \end{aligned} \quad (10)$$

$$ret_{i0} = Ret_i^{Init} \quad i \in \mathcal{I} \quad (11)$$

$$ret_{it} \geq Em_{it} \quad t \in \mathcal{T}, i \in \mathcal{I} \quad (12)$$

$$ret_{it} \in [0, E] \quad t \in \mathcal{T}, i \in \mathcal{I} \quad (13)$$

$$\sum_{i \in \mathcal{I}} ret_{it} \geq Ret_{sum}^{Init} \quad t = |\mathcal{T}| \quad (14)$$

$$\sum_{i \in \mathcal{I}} rut_{it} \geq Rut_{sum}^{Init} \quad t = |\mathcal{T}| \quad (15)$$

The remaining usage time is defined by (6)-(7) and its limits by (8)-(9). Similarly, the remaining elapsed time is defined by (10)-(11) and its limits by (12)-(13). Finally, constraints 14 and 15 guarantee that

id	$ \mathcal{J} $	$ \mathcal{T} $	assign	vars	cons	nonzeros
L0	9	11	310	6132	6081	37395
L1	9	21	650	12512	12399	80535
L2	9	31	990	18892	18719	123685
L3	9	41	1249	24984	25021	165971
L4	10	11	530	6449	6092	38346
L5	10	21	1070	13119	12420	82356
L6	10	31	1610	19789	18750	126376
L7	10	41	2069	26171	25062	169532
L8	11	11	1080	7270	6329	41266
L9	11	21	2120	14750	12809	87856

Table 1 – Instances used in the experiments with indicators of size: "id" is the instance; $|\mathcal{T}|$ is the number of periods; "assign" is the number of assignments of resources to tasks; $|\mathcal{J}|$ is the number of tasks; "vars" the number of variables; "cons" is the number of constraints; and "nonzeros" is the number of non zero values in the matrix.

resources have, globally, the same amount of remaining used and elapsed times at the beginning and at the end of the planning horizon.

5 Datasets

The dataset being used was provided by the French Air Force and already used in De Chastellux et al. (2017). Based on the information provided, a number of instances were created as subsets of the main dataset in order to run different experiments.

Table 1 shows the characteristics of each instance. The main differences between instances are in the number of tasks and the number of periods.

The criteria for the creation of the instances were the following:

1. The original instance was the largest instance where an integer solution was found in less than 3600 seconds: it contains 11 tasks and 21 periods in total.
2. We took out the task with the biggest number of required resources.
3. We increased the number of periods in multiples of 10 until no integer solutions can be found in 3600 seconds.
4. We repeated step 2.

5.1 Tasks

The dataset includes 11 tasks. Most of them demand between 1 and 8 resources per period, which is equal

to one month. The most loaded and second most loaded tasks demand 50 and 20 resources respectively per period. Candidate resources are calculated based on each task's requirements matching the candidate's capabilities.

Some tasks have a limited duration, meaning they start and end during the planning horizon. Others are active during the whole planning horizon.

5.2 Maintenances

All maintenance operations last exactly 6 periods. Each maintenance sets the remaining usage time of a resource to 1000 hours. After it, a resource can be without maintenance for a maximum of 60 consecutive periods. The nominative capacity for maintenance is 18 resources at any given time.

5.3 Resources

There are 127 resources in total. Each one has its own functionalities and past assignments.

6 Results

6.1 Numerical tests

All tests were run on a 64 bit Ubuntu 16.04 workstation with 16GB of RAM and Intel i7-6700HQ CPU 2.60GHz x 8 processor. Several experiments were done with the MIP formulation on problem instances presented in table 1.

Table 2 gives an overview of the obtained results. As can be seen, smaller instances are solved up to optimality while medium-size ones are stopped after one hour.

For this calculation, both weights on the objective function were set at 1. A sensibility analysis of these two parameters is provided in section 6.2.

Table 3 shows the details of the solving process for each instance. As it can be seen, cuts do not improve significantly the relaxation (the value of "bound cuts" is very close to the root solution). In fact, the original linear relaxation (root) is quite similar to the final lower bound (bound). This suggests that further improvement in modelization would help to reduce the gap faster.

An example of a tighter modelization could be to replace the constraints that count the remaining usage time and elapsed time for tighter equivalents.

The greatest number of cuts used was of 'Implied bound' and 'Mixed integer rounding', depending on the instance. Since these are two generic-type cuts, it can be concluded that the solver has not detected

id	objective	gap (%)	time (s)	bound
L0	62.0	0.0	0.7	62.0
L1	63.0	0.0	68.7	63.0
L2	63.0	1.6	3600.1	62.0
L3	64.0	3.6	3603.9	61.7
L4	82.0	0.0	0.9	82.0
L5	83.0	0.0	144.0	83.0
L6	83.0	1.2	3600.1	82.0
L7	84.0	2.6	3609.1	81.8
L8	139.0	0.0	530.6	139.0
L9	149.0	6.1	3600.0	139.9

Table 2 – Information on the solution of each tested instance: "id" is the instance identifier; "objective" is the value of the objective function; "gap" is the percentage gap between the integer solution and the linear relaxation; "time (s)" is the solution time or 3600; and "bound" is the greatest lower bound found from the linear relaxation.

id	root	b. cuts	bound	cuts (#)	cuts (s)
L0	56.0	61.9	62.0	37	0.7
L1	56.0	63.0	63.0	494	20.3
L2	61.4	62.0	62.0	982	25.6
L3	60.6	60.7	61.7	1149	42.5
L4	81.8	82.0	82.0	49	0.9
L5	82.0	83.0	83.0	752	20.3
L6	81.4	82.0	82.0	862	68.8
L7	80.6	80.7	81.8	1489	65.1
L8	136.8	137.8	139.0	285	6.9
L9	138.2	139.6	139.9	1157	310.8

Table 3 – Solution details on the progress of the lower bound: "id" is the name of the instance; "root" is the relaxation before cuts; "b. cuts" is the relaxation after the cuts; "bound" is the relaxation at the time limit; "cuts (#)" is the number of cuts; and "cuts (s)" is the time the cuts took in seconds.

any particular structure as part of this problem.

Table 4 shows the progress of the integer solution for each instance. For more detail on this progress, figure 1 shows an example of progress for instance with id=L7. It can be seen that initial solutions are quite far from the optimal ones due to the use of heuristics at the beginning of the solving process. In fact, for two rather small instances the solver was able to find an optimal solution before starting to branch.

The obtained results show that the branching is not providing much improvement in increasing the lower bound or finding better integer solutions. The quantity of symmetries in the model, represented as candidates for each task, could be a reason the difficulty in solving large instances. It may exist a very large number of possible candidates to be potentially as-

id	first	sol. cuts	last
L0	1550.0	62.0	62.0
L1	1558.0	99.0	63.0
L2	1532.0	67.0	63.0
L3	1532.0	137.0	64.0
L4	1544.0	82.0	82.0
L5	1552.0	847.0	83.0
L6	1552.0	95.0	83.0
L7	1552.0	856.0	84.0
L8	1594.0	147.0	139.0
L9	149.0	-	149.0

Table 4 – Solution details on the progress of the integer solution: "id" is the name of the instance; "first" is the first solution integer found; "sol. cuts" is the solution after the cuts; "last" is the last solution found at the time limit.

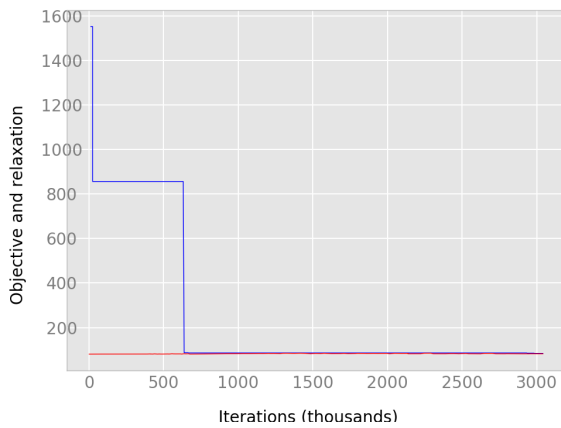


Figure 1 – Solving progress of the integer solution (in blue) and the lower bound (red) along the solver iterations (x axis). Instance id=L7.

signed for each task with respect to the requirements of resources for each task. This reduces the branching capacity of the solver, which tries to explore many nodes that are almost equivalent in reality.

An example of an alternative to break these symmetries could be to slightly change the objective function so it can measure more gradually the smoothness of the objective and help get better bounds. Another alternative could be to assign priorities to some combinations of resource-task or to limit the number of candidates that can be assigned to each task by some heuristic that clusters resources with tasks. Finally, arbitrarily preassigning (fixing) the final state (in terms of remaining usage time) to each resource could also help break those symmetries without important loss on the objective function value.

Figures 2 and 3 show an example distribution of maintenance operations and unavailable numbers of re-

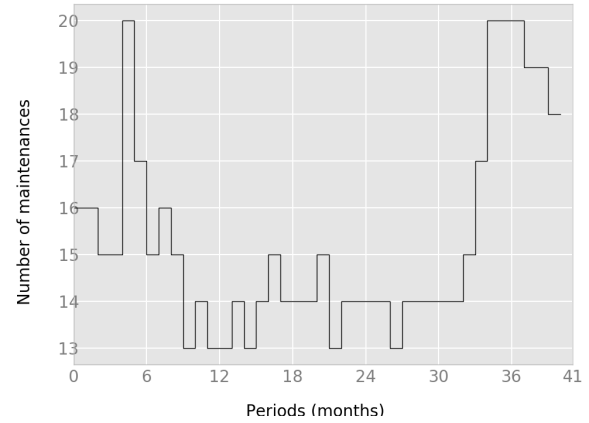


Figure 2 – Solution for instance id=L7: number of resources under maintenance at each period.

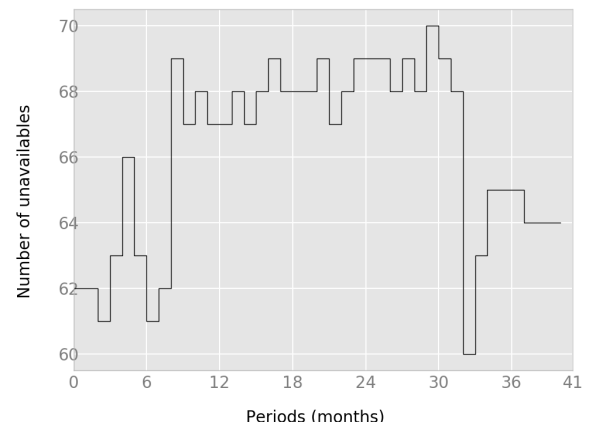


Figure 3 – Solution for instance id=L7: number of unavailable resources at each period.

sources per period, respectively.

6.2 Bi-objective analysis

The further calculation was done in order to analyze the impact of the weights W_1 and W_2 on the solution. In it, both weights were changed between 0 and 1 with pace of 0.1 under the condition that $W_1 + W_2 = 1$. This was done for all previously defined instances. This method does not a priori guarantee to compute exactly Pareto Fronts: optimal weighted computations can only prove that some points guarantee supported solutions. However, a posteriori analysis can conclude that the bi-objective convex relaxation guarantees the completeness of the Pareto Front. By doing this analysis, exact extreme points can be computed (and thus the Nadir and ideal points), and some intermediate points found for all the intermediate integer values. Table 5 presents the obtained results of this analysis. It can be seen that all instances have a

small number of non-dominated solutions.

id	# points	First	Last
L0	3	(18, 45)	(15, 47)
L1	3	(20, 46)	(15, 48)
L2	2	(16, 47)	(15, 48)
L3	3	(19, 45)	(15, 49)
L4	3	(18, 65)	(15, 67)
L5	3	(20, 66)	(15, 68)
L6	2	(16, 67)	(15, 68)
L7	4	(19, 66)	(16, 69)
L8	2	(21, 118)	(18, 121)
L9	3	(24, 121)	(19, 123)

Table 5 – Non-dominated points found on each instance after 3600 seconds. All instances were analyzed: "First" and "Last" indicate the two extremes of the calculated Pareto frontier; "# points" indicates the number of Pareto optimal points found. Not all experiments were solved up to optimality.

An individual case study is illustrated in Figure 4, where an optimal solution was found in each experiment. Table 6 shows the obtained results. The instance used consisted of 21 periods and 10 tasks corresponding to instance with id=L5. Computations W10 and W09 prove that (15, 68) is an extreme point of the Pareto Front. Computations W00 and W01 prove that (20, 66) is the other extreme point of the Pareto Front. This implies the ideal and Nadir points are, respectively, (15, 66) and (20, 68). Computation W03 proves that (16, 67) is also a non dominated point. Having all the discrete points in $\llbracket 66, 68 \rrbracket$ for the m_{max} objective, it can be concluded that no other non dominated points exists. In this case, weighted computations prove the completeness of the convex relaxation, which is also exactly the Pareto front of non-dominated solutions: no unsupported solution exists in this case.

exp	maint	unavailable	W_1	W_2
W00	20	66	0.0	1.0
W01	20	66	0.1	0.9
W02	20	66	0.2	0.8
W03	16	67	0.3	0.7
W04	16	67	0.4	0.6
W05	15	68	0.5	0.5
W06	15	68	0.6	0.4
W07	15	68	0.7	0.3
W08	15	68	0.8	0.2
W09	15	68	0.9	0.1
W10	15	69	1.0	0.0

Table 6 – Values of u_{max} (unavailable) and m_{max} (maint) for different combinations of W_1 and W_2 . Instance id=L5. All experiments were solved to optimality.

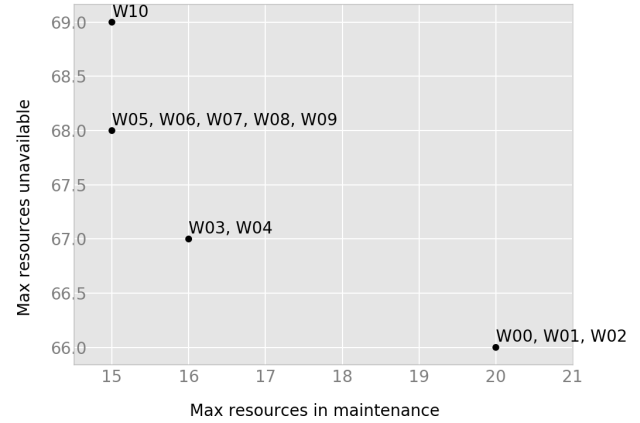


Figure 4 – Pareto diagram for instance id=L5.

7 Conclusions and further work

This paper proposed a new MIP formulation to address a specific FMP with an heterogeneous fleet of aircraft. Two objective functions were chosen to address the French Air Force requirements. The first one was to smooth maintenance operations and the second one was to increase aircraft availability. The characteristics of MIP convergence were analyzed on real-world instances. Mono-objective problem instances were solved to optimality for medium size instances. Then, exact Pareto frontiers for bi-objective formulation were calculated for these instances. The use of these two objectives does not lead to the same optimal solution but very good compromise solutions can be found due to the bi-objective optimization. For larger instances, difficulties in finding optimal solutions occurred because of multiple symmetries and a poor quality of MIP generic primal heuristics.

The first perspective is to improve the resolution of large size instances. Heuristics could be implemented, but also mathematical programming can be improved with symmetry breaking or MIP reformulations. A matheuristic method derived from the existing MIP formulation can also be used to improve the performance.

For the bi-objective optimization for large size instances, more advanced bi-objective methods such as ϵ -constraints or Two-Phases-Method should be investigated for the computation of larger Pareto frontiers. Lastly, additional constraints from the real world application can be incorporated in the model.

References

Cho, P. (2011). *Optimal Scheduling of Fighter Aircraft Maintenance*, PhD thesis, Massachusetts Institute of Technology.

- De Chastellux, P. (2016). *Planification de la maintenance des avions de chasse*, Master's thesis, ENSTA ParisTech.
- De Chastellux, P., Dupin, N. and Bazot, P. (2017). Planification optimisée de maintenances d'aéronefs militaires, *18ème édition du congrès ROADEF2017*.
- Gavranis, A. and Kozanidis, G. (2015). An exact solution algorithm for maximizing the fleet availability of a unit of aircraft subject to flight and maintenance requirements, *European Journal of Operational Research* **242**(2): 631–643.
- Kozanidis, G. (2008). A Multiobjective Model for Maximizing Fleet Availability under the Presence of Flight and Maintenance Requirements, *Journal of Advanced Transportation* **43**(2): 155–182.
- Kozanidis, G., Gavranis, A. and Liberopoulos, G. (2014). Heuristics for flight and maintenance planning of mission aircraft, *Annals of Operations Research* **221**(1): 211–238.
- Steiner, A. (2006). A heuristic method for aircraft maintenance scheduling under various constraints, *Proceedings of the 6th Swiss Transport Research Conference. Ancona, Italy*.
- Verhoeff, M., Verhagen, W. J. C. and Curran, R. (2015). Maximizing operational readiness in military aviation by optimizing flight and maintenance planning, *Transportation Research Procedia* **10**(July): 941–950.

Caractérisation de services industriels en réponse à appels d'offres

Delphine GUILLON^{1,2}, Élise VAREILLES¹, Michel ALDANONDO¹,
Eric VILLENEUVE² et Christophe MERLO²

¹Université de Toulouse, IMT Mines Albi
Campus Jarlard
81013 Albi cedex 09 - France
prenom.nom@mines-albi.fr

²ESTIA Recherche
92, allée Théodore Monod
64210 Bidart - France
p.nom@estia.fr

RÉSUMÉ : *En réponse à appel d'offres, de nombreuses entreprises proposent des services. Dans cet article, nous nous intéressons à la question suivante : Comment modéliser les services industriels pour pouvoir les élaborer facilement lors d'une réponse à appel d'offres ? Un modèle de configuration d'offres-produit est présenté, puis une réflexion sur l'extension de ce modèle aux offres-service est menée. Suite à une étude de la littérature pour définir les services et identifier leurs caractéristiques, l'impact de ces caractéristiques sur le modèle d'offre-produit est analysé. De nouvelles caractéristiques sont également introduites. Ce travail doit permettre de proposer une typologie qui servira de base à l'adaptation du modèle d'offres-produit aux services.*

MOTS-CLÉS : *services industriels, réponse à appel d'offres, modélisation, caractérisation de services*

1 INTRODUCTION

En réponse à appel d'offres, les soumissionnaires doivent faire face à un environnement de plus en plus complexe. Les concurrents sont nombreux et les clients exigeants. Pour augmenter leurs chiffres d'affaires et rester compétitives, les entreprises soumissionnaires doivent proposer des offres à la fois conformes aux exigences clients, attractives (en termes de solution proposée, coût et délai), et réalisables (en termes de processus de réalisation, ressources allouées et risques encourus). Face à ce constat, le projet OPERA¹ vise à proposer aux soumissionnaires un outil d'aide à la définition d'offres exploitant leurs connaissances et bonnes pratiques.

Suite à plusieurs entretiens avec les quatre entreprises partenaires du projet, il a été identifié que la grande majorité d'une offre est construite en réutilisant des éléments connus, déjà utilisés et caractérisés. Environ 10% d'une offre nécessite un travail d'ingénierie pour concevoir de nouveaux éléments. L'élaboration d'offre correspond donc majoritairement à une situation de conception routinière (Chandrasekaran 1990), voire même de configuration (Mittal et Frayman 1989).

Deux types d'entreprises sont impliquées dans le projet OPERA : deux sont des entreprises manufacturières et deux sont des sociétés de conseil. Nous distinguons donc les offres-produit, qui nécessitent l'élaboration

d'un produit, des offres-service, qui nécessitent l'élaboration d'un service. De nombreux articles traitent de la configuration de produits, et de la configuration concurrente de produits et du processus de réalisation associé (Abeille et al. 2010, Robin et al. 2010, Sylla et al. 2017). Dans (Guillon et al. 2017), nous avons proposé un modèle générique d'élaboration d'offres-produit, qui permet de configurer un produit et son processus de réalisation, en utilisant des informations caractérisant le contexte de l'offre. Cependant, la représentation des offres-service à l'aide de ce modèle questionne. Si certains auteurs (Felfernig et al. 2014) soulignent l'intérêt de la configuration de services, peu détaillent comment l'appliquer. Pourtant, le marché des services est en pleine croissance (Chae 2012) et les besoins industriels sont importants.

Les travaux menés à travers le projet OPERA visent donc à étendre le modèle générique de configuration d'offres-produit aux offres-service. Dans cet article, nous menons une réflexion préalable à un modèle d'offre-service. L'extension de ce modèle n'est pas immédiate car services et produits sont très différents (Grönroos 2000). Parmi les différences notables, la présence ou non d'une nomenclature semble particulièrement critique. Pour un produit, le soumissionnaire doit toujours configurer une nomenclature de composants physiques, même minimale. Pour un service, un questionnement émerge autour de l'existence d'une telle nomenclature et, si elle existe, sur sa définition et sa représentation.

1. En savoir plus sur le projet OPERA : <https://research-gi.mines-albi.fr/display/OPERA>

La diversité des services industriels est très vaste, ce qui pourrait impliquer une typologie de services associée à une typologie de modèles génériques. Dans cet article, nous proposons d'identifier et discuter les caractéristiques permettant de mettre en place une telle typologie de services.

Nos travaux étant liés au projet OPERA, cet article s'intéressera uniquement aux services business-to-business (B2B) dans un cadre industriel. Ces services recouvrent, entre autres, les services suivants : études de calculs thermiques, conseil en organisation, formation, développement de site internet, paramétrage logiciel, maintenance et location d'équipements industriels. L'édition logiciel est hors périmètre car le logiciel est considéré comme un produit.

La suite de cet article est structurée de la façon suivante. La deuxième section présente le modèle générique d'offres-produit que nous souhaitons étendre aux offres-service. La troisième section est consacrée à une revue de littérature sur la définition et la description de services. La quatrième section introduit les préalables à un modèle d'offres-service. La dernière section est consacrée à l'étude de caractéristiques des services issues de la littérature et présente de nouvelles caractéristiques, permettant de poser les bases d'une typologie de services.

2 MODÈLE GÉNÉRIQUE D'OFFRES-PRODUIT

Cette section introduit le modèle générique de configuration d'offres-produit que nous souhaitons étendre aux offres-service.

Configuration. Le soumissionnaire doit choisir, parmi un ensemble de composants connus, caractérisés et chiffrés, ceux qui correspondent aux besoins du client (Mittal et Frayman 1989).

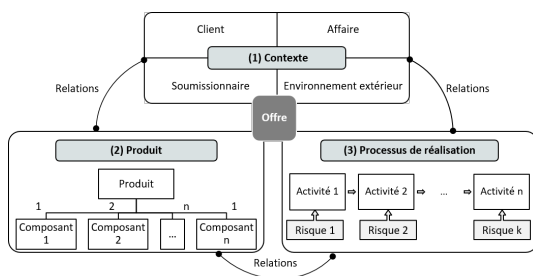


FIGURE 1 – Modèle d'offres-produit

Dans (Guillon et al. 2017), un premier modèle de configuration d'offres-produit, illustré par la figure 1, a été proposé. Pour élaborer une offre pertinente et faire du retour d'expérience sur les cas similaires, le soumissionnaire doit caractériser (1) le contexte de l'offre, (2) la nomenclature du produit et (3) le processus de réalisation du produit. Des indicateurs caractérisant le

produit et le processus de réalisation permettront de comparer différentes offres. Il peut s'agir d'indicateurs financiers (coût, prix, marge), d'indicateurs de délai, de charge et/ou de durée, mais aussi d'indicateurs de confiance (Sylla et al. 2017).

(1) Contexte. Quatre types de données caractérisent le contexte de l'offre. Elles caractérisent :

- le client (a-t-on déjà travaillé avec lui ? est-il stratégique ?),
- l'affaire (est-ce un marché public ou privé ?),
- le soumissionnaire (a-t-on un carnet de commande rempli ? l'atelier est-il chargé ?),
- l'environnement extérieur (est-ce qu'un concurrent important est présent sur le marché ?).

(2) Produit. Pour décrire un produit, le soumissionnaire a le choix entre deux approches : une description fonctionnelle des besoins, attentes et fonctions ou une description hiérarchique des composants. Nous proposons que le soumissionnaire configure la nomenclature du produit (Aldanondo et Vareilles 2008).

(3) Processus de réalisation. Que le soumissionnaire ait choisi une description fonctionnelle ou bien une description hiérarchique des composants, il est important de définir le processus de réalisation, décrit comme un ensemble d'opérations à réaliser pour produire et livrer le produit (Pitiot et al. 2014). Dans (Aldanondo et Vareilles 2008), les auteurs proposent également une configuration concurrente produit / processus. Dans le cadre du projet OPERA, la configuration du processus de réalisation s'appuie sur un processus générique constitué des opérations suivantes : études, approvisionnement, fabrication et livraison. Le processus de réalisation n'est pas défini de façon très détaillée car l'étude se fait en avant-vente et ne requiert donc pas d'approfondir ce point. L'objectif, lors de la définition du processus de réalisation en réponse à appel d'offres, est d'identifier les activités clés pour évaluer les coûts liés au processus, fournir une date de livraison pertinente, et identifier les risques-projet clés.

Dans nos travaux (Guillon et al. 2017), le modèle d'offres-produit sera exploité avec une approche de type problèmes de satisfaction de contraintes (CSP) (Montanari 1974).

Cette section a donc permis d'introduire le modèle générique d'offres-produit. Nous explorerons à présent la possibilité d'étendre ce modèle générique à tout ou partie des offres-service.

3 DÉFINITIONS ET DESCRIPTIONS D'UN SERVICE

Au cours de ces dernières années, le secteur tertiaire a pris de plus en plus d'ampleur : l'économie globale est passée d'une vision orientée produits à une vision

orientée service, et dans la plupart des pays développés la majorité des revenus provient des services (Chae 2012). Aujourd'hui, l'essor des services est tel que les entreprises du secteur tertiaire répondent de plus en plus à des appels d'offres. Cette section est donc consacrée aux travaux existants sur la définition et la description de services. Tout d'abord, nous définissons ce qu'est un service, pour dans un deuxième temps présenter les travaux existants sur la description de services.

3.1 Définitions d'un service

Plusieurs auteurs ont défini les services dans la littérature. Nous en présentons quelques-uns ci-après.

Le service comme un processus. Pour Grönroos, un service est un processus constitué de plusieurs activités plus ou moins intangibles, qui souvent résultent d'interactions entre le client et les employés et/ou des ressources physiques et/ou des biens et/ou des systèmes du fournisseur de service, et qui sont fournies comme solutions aux problèmes du client. Il propose, plutôt que d'essayer de définir ce qu'est un service, de définir les caractéristiques d'un service. L'une des trois caractéristiques fondamentales de la majorité des services est qu'ils sont des processus constitués d'activités ou de séries d'activités plutôt que de composants (Grönroos 2000). Il définit un service comme un processus, par opposition au produit qui est *a thing*. Pour (Carlborg et Kindström 2014) également, un service est souvent défini comme un processus.

Le service comme une capacité. L'INSEE définit un service de la façon suivante : « Une activité de service se caractérise essentiellement par la mise à disposition d'une capacité technique ou intellectuelle. A la différence d'une activité industrielle, elle ne peut pas être décrite par les seules caractéristiques d'un bien tangible acquis par le client. » (INSEE 2017)

Dimensions d'un service Dans (Bullinger et al. 2003), un service est caractérisé par trois dimensions différentes : (1) une dimension structure, qui détermine la capacité et la volonté de fournir le service en question, (2) une dimension processus, et (3) une dimension résultat, qui comprend une définition du contenu du service et un plan structuré des services.

Ces définitions font ressortir des aspects intéressants et variés dans la définition de service, sur lesquels nous nous appuierons par la suite.

3.2 Descriptions d'un service

L'intérêt de la configuration de services est le même que pour la configuration de produits : trouver l'équilibre entre la personnalisation pour répondre aux besoins variés des clients, et les bénéfices de la standar-

disation (Felfernig et al. 2014). L'intérêt de la configuration de masse appliquée aux services a également été discuté (Heiskala et al. 2005). Après avoir listé les avantages et défis liés à la configuration de masse pour les services, ils concluent que les avantages sont les mêmes que pour les produits. Pourtant, les résultats sur la customisation de masse pour les biens ne sont peut être pas directement applicables aux services (Felfernig et al. 2014). Gérer à la fois customisation et standardisation pour les entreprises de service est donc réellement un défi (Løkkegaard et al. 2016).

Cette section a donc pour objectif d'explorer ce problème à partir d'une revue de la littérature.

Décomposition d'un service en éléments. Plusieurs auteurs traitent de la décomposition d'un service en éléments. « Un service est une combinaison d'éléments physiques et non physiques, intégrés selon des configurations spécifiques à chaque client » (traduit de l'anglais) (Carlborg et Kindström 2014).

Dans (Goldstein et al. 2002), les auteurs expliquent que, « du point de vue du fournisseur de services, concevoir un service signifie définir une combinaison appropriée de composants physiques et non physiques ». Ils précisent que « les composants d'un service sont souvent des éléments non physiques, mais plutôt une combinaison de processus, compétences humaines et matériels » (traduit de l'anglais).

Dans (Ki Moon et al. 2009), la notion de familles de services est explorée. Ils proposent d'étendre les méthodes utilisées sur les familles de produits aux familles de services. Une famille de services est définie comme un ensemble de services basés sur un ensemble de processus, activités, objets et caractéristiques communes. Pour délivrer un service, un ensemble de modules de services, eux-même composés d'un ensemble de composants de services, seront « assemblés ».

Dans (Bullinger et al. 2003), les auteurs conseillent, pour la dimension résultat d'un service, si celui-ci est complexe, de le diviser en services partiels. De plus, il serait alors préférable que ces services soient structurés de façon modulaire.

Modularité des services. Dans (Løkkegaard et al. 2016), l'intérêt des services modulaires, qui ont les mêmes avantages que les produits modulaires, est mis en avant. Dans (Böttcher et Klingner 2011), les auteurs proposent une méthode pour décomposer un service B2B en modules, et définissent un module de service de la façon suivante (traduit de l'anglais) : « Un *service module* offre une fonctionnalité bien définie via des interfaces décrites précisément. Un *service module* peut être utilisé pour décomposer et peut donc être lui-même une partie d'un *service module* de plus haut niveau. » Cette définition est très proche de celle des composants pour un produit. Ils ajoutent : « Cette

décomposition permet une configuration spécifique pour un client donné, car le client peut assembler une offre de service à partir d'un ensemble de modules de services donnés. » Cette définition est similaire à ce qui se fait en configuration de produits, où un client peut assembler son produit à partir d'un ensemble de composants donnés.

Cependant, ce modèle n'est pas directement applicable à notre cas d'étude. Nous ne cherchons pas exactement à modulariser des services, mais plutôt à les modéliser, qu'ils soient modulaires ou non. Nous ne voulons pas modifier des services existants pour les faire rentrer dans un modèle de « nomenclature de services », mais plutôt trouver un modèle pour les services existant. Un produit, modulaire ou non, peut être décomposé en composants, liés entre eux par une architecture, appelée nomenclature. Nous cherchons à identifier si l'on peut faire de même pour les services. Les approches processus, éléments et modulaires laissent supposer qu'un service pourrait être représenté par une structure arborescente.

4 PRÉALABLES À UN MODÈLE D'OFFRES-SERVICE

En réponse à appel d'offres, le travail porte essentiellement sur le chiffrage des livrables, puis, dans certains cas, sur l'identification des charges et l'estimation d'un délai de livraison. Pour une offre-produit, le soumissionnaire cherche principalement à chiffrer ce produit, d'où la définition d'une nomenclature. De plus, les choix techniques faits sur les composants pourront intéresser le client. Pour une offre-service, le soumissionnaire chiffre principalement un processus, et non plus les composants d'un produit. Les ressources employées sur le processus de réalisation sont valorisées : ressources essentiellement humaines, mais aussi matérielles dans certains cas. Les achats et coûts liés à un déplacement sont des coûts liés au processus de réalisation. Contrairement à un produit où les coûts liés aux matières premières et aux composants sont importants, pour les services ceux-ci sont souvent nuls, (excepté pour la maintenance, où des composants physiques sont alors achetés pour le client).

De plus, la structuration de l'offre dépend souvent de la demande de chiffrage du client, contrairement aux produits où la nomenclature est intrinsèque à la famille du produit et indépendante du client. Pour des services, la structuration des lots de travail pourra donc être très différente d'un client à l'autre, pour un travail assez similaire.

L'application à un service du modèle d'offre-produit présenté dans la figure 1 soulève plusieurs questions.

(1) Contexte. Tout d'abord, la partie contexte de l'offre reste inchangée. Que ce soit pour un produit

ou un service, le soumissionnaire doit caractériser le contexte pour choisir l'offre la plus pertinente à proposer au client.

(2) Produit ou Service. Ensuite, la partie décrivant le produit proposé par le soumissionnaire est difficilement transférable à un service. En effet, comme évoqué précédemment, l'existence même d'une nomenclature de service est questionnée. Nous avons vu dans la section 3.1 que certains auteurs définissent un service comme un processus. Cela signifierait que, pour élaborer un service, le soumissionnaire peut se contenter d'élaborer le processus de réalisation. Dans le modèle d'offres-produit, cela signifierait que la partie (2) produit du modèle serait supprimée et que rien ne la remplacerait (car même pour un produit nous définissons le processus de réalisation dès l'avant-vente), ou que cette partie serait remplacée par un WBS (Work Breakdown Structure), comme en projet. Nous devons donc déterminer si cette partie doit être supprimée ou adaptée aux services. Si l'adaptation est choisie, il convient d'identifier un formalisme permettant de décrire un service et les informations qui lui sont propres.

(3) Processus de réalisation. Ensuite, comme dit précédemment, le processus de réalisation est particulièrement important pour un service. Le choix des ressources et la bonne identification des durées de chaque tâche sont essentiels. Le modèle d'offres-service comprendra donc une partie processus et risques.

La suite de cet article est consacrée aux différences entre produits et services. En effet, s'il est impossible d'appliquer directement la configuration de produits à la configuration de services, lister les caractéristiques qui distinguent un produit d'un service doit permettre de comprendre comment étendre le modèle d'offres. La revue de littérature présentée ici est principalement issue du marketing, qui est un domaine où la notion de service a été largement étudiée.

5 CARACTÉRISTIQUES SUSCEPTIBLES D'IMPACTER UNE TYPOLOGIE DE SERVICES

Le service est souvent défini par opposition au produit dans la littérature (Edvardsson et al. 2005). S'il est si difficile de définir ce qu'est un service, il est possible de lister ses principales caractéristiques, souvent opposées à celles des produits. Historiquement, quatre caractéristiques ont été identifiées (termes anglophones) : intangibilité (**I**ntangibility), hétérogénéité (**H**eterogeneity), non séparabilité de la production et de la consommation (**I**nseparability) et périssabilité (**P**erishability). Ces quatre caractéristiques sont parfois regroupées sous le terme IHIP (Edvardsson et al. 2005, Felfernig et al. 2014, Lovelock et Gummesson 2004).

Bien que ces caractéristiques se retrouvent beaucoup dans la littérature, elles ont aussi été critiquées au cours des dernières années (Edvardsson et al. 2005, Lovelock et Gummesson 2004). En effet, il semble que les caractéristiques IHIP ne puissent pas s'appliquer à tous les types de services. Nous détaillerons ces critiques dans la section suivante. Dans (Lovelock et Gummesson 2004), les auteurs proposent donc soit (1) d'abandonner la distinction produit/service et d'arrêter de les déclarer comme deux domaines différents, (2) de trouver un nouveau paradigme pour définir les services, (3) de travailler sur des typologies de services. L'option (1) n'est pas applicable en configuration. En effet, comme cela a été évoqué précédemment, il est impossible de traiter produits et services de manière identique car nous ne pouvons définir tous les services avec une nomenclature physique. L'option (2) est celle conservée par Lovelock et Gummesson. Ils proposent de définir un service par l'absence de transfert de propriété. Cette caractéristique et son possible impact sur notre modèle seront discutés dans la section suivante. Enfin, l'option (3) semble convenir le mieux à la configuration de services. En effet, la diversité de services étant très vaste, nous pensons pouvoir définir une typologie de services, avec pour chaque type de service un modèle générique d'élaboration d'offres.

Nous proposons donc de discuter des caractéristiques permettant d'établir une telle typologie de services. Les caractéristiques IHIP et l'absence de transfert de propriété seront discutées dans la section 5.1, pour identifier si non seulement elles permettent de distinguer un produit d'un service, mais aussi si elles permettent de distinguer différents types de services. Puis nous proposerons de nouvelles caractéristiques dans la section 5.2.

5.1 Caractéristiques d'un service issues de la littérature

Dans cette sous-section, nous présentons chacune des quatre caractéristiques IHIP, ainsi que la caractéristique d'absence de transfert de propriété, nous discutons de leurs limites, puis nous questionnons leurs impacts sur le modèle d'offres.

5.1.1 Intangibilité

Définition de la caractéristique et discussion. L'intangibilité est la première caractéristique d'un service, la plus citée mais surtout la plus critique et celle dont découle toutes les autres (Bateson 1979). Cette intangibilité est définie à l'origine par opposition aux produits qui sont, eux, tangibles, c'est à dire palpables et matériels (Shostack 1977). En effet, une prestation intellectuelle est intangible. La définition de l'INSEE dans la sous-section 3.1 souligne cette intangibilité. Bateson et Bery distinguent l'intangibilité physique

et l'intangibilité mentale (Bateson 1979, Berry 1980). L'intangibilité physique désigne l'incapacité des cinq sens à saisir un service, alors que l'intangibilité mentale désigne la difficulté de se faire une représentation mentale du service. Dans (Bielen et Sempels 2004), les auteurs proposent une échelle de mesure bidimensionnelle du concept d'intangibilité, basée sur ces deux intangibilités.

L'applicabilité de cette caractéristique à l'ensemble des services peut être questionnée. En effet, si nous nous intéressons au livrable comme valeur ajoutée de la prestation de service, l'intangibilité n'est pas toujours si évidente. Par exemple, pour une étude de calcul ou un audit en organisation, le livrable est un rapport écrit, bien tangible. Pour une formation au contraire, on peut considérer que le livrable est la connaissance acquise par les personnes formées, un savoir-faire complètement intangible. Le concept d'intangibilité ne semble donc pas toujours entièrement applicable.

Impact sur le modèle générique d'offres. Cette caractéristique est essentielle car c'est elle qui remet en question la partie clef de configuration d'offres-produit : la nomenclature. Si lors d'une réponse à appel d'offres pour un produit, le soumissionnaire doit toujours décrire les caractéristiques d'un bien tangible et identifier les composants d'une nomenclature physique, les descripteurs d'un service en réponse à appel d'offres restent à identifier. De plus, cette intangibilité « entraîne une difficulté pour le consommateur à évaluer l'offre et ses caractéristiques (attributs, prix, qualité...) » (Bielen et Sempels 2004). Elle entraîne également pour le soumissionnaire une difficulté à mettre un prix sur un service. Malgré tout, si un service est décomposé en activités comme dans un WBS ou en éléments suivant une décomposition par lots, une nomenclature intangible et non plus physique pourrait être présente. Par contre, cette caractéristique d'intangibilité est importante pour différencier différents types de services et aboutir à une typologie.

5.1.2 Hétérogénéité

Définition et Discussion. L'hétérogénéité concerne le potentiel de haute variabilité dans la performance des services. L'hétérogénéité d'un service est due aux ressources humaines, par opposition à la répétabilité d'une machine. La qualité et l'essence d'un service peut varier d'un intervenant à l'autre et d'un client à l'autre. La performance d'un service pour un même intervenant peut elle aussi varier d'un jour à l'autre. Cette caractéristique entraîne donc une difficulté de standardisation (Zeithaml et al. 1985).

L'hétérogénéité s'applique bien aux prestations de conseil en organisation ou de formation. Par exemple, une prestation de formation est fortement liée à l'in-

tervenant et son état d'esprit lors du déroulement des activités. Cependant, pour une étude de calcul, si la qualité de rédaction du rapport dépend de l'ingénieur, les résultats des calculs, eux, sont indépendants de la personne qui les réalise.

Impact sur le modèle. Cette caractéristique a un fort impact sur la qualité du service délivré, mais n'a pas d'impact sur le modèle générique d'offres.

5.1.3 Non séparabilité

Définition et discussion. La caractéristique de non séparabilité à plusieurs facettes, toutes liées mais avec quelques nuances. Tout d'abord, par définition, cela signifie que la production et la consommation du service sont simultanées (Parasuraman et al. 1985, Edvardsson et al. 2005, Lovelock et Gummesson 2004). Dans (Grönroos 2000), la majeure partie des services ont trois caractéristiques de base :

- La première a été citée dans la section 3.1 : le service défini comme un processus.
- La deuxième est celle qui vient d'être citée : les services sont au moins dans une certaine mesure produits et consommés simultanément.
- Enfin, la troisième est la suivante : le client participe au processus de production du service, au moins dans une certaine mesure.

Cette implication du client dans le processus de réalisation (en formation par exemple) découle de la non séparabilité (Heiskala et al. 2005). Grönroos ajoute que la valeur ajoutée d'un service est le résultat des interactions client-fournisseur, par opposition au produit, où la valeur ajoutée est créée au sein de l'usine. Selon lui, (traduit de l'anglais) « pour les services, les interactions entre le client et le fournisseur sont souvent présentes et d'importance, bien que les parties impliquées n'en soient pas toujours conscientes » (Grönroos 2000).

Dans le cas d'une offre-service portant sur une étude de calcul, celle-ci est d'abord produite par le fournisseur, chez le fournisseur, puis transmise au client, qui pourra alors la « consommer », l'exploiter. Le client ne participe pas au processus de réalisation. Il est impliqué uniquement au début et à la fin, comme pour un produit. Au début, il va fournir au soumissionnaire ses spécifications, puis il sera recontacté seulement à la fin du processus pour la livraison du service : le rapport d'étude. Le concept d'inséparabilité ne semble donc pas applicable. La formation est bien, dans une certaine mesure, produite et consommée simultanément puisque l'exécution de la formation participe à la fois de la production par l'intervenant et de la consommation par le client. Le client intervient bien dans la production du service car il est impliqué dans le processus de réalisation : il est présent (et indispensable) pendant la formation. La caractéristique d'inséparabilité s'applique donc bien.

Impact sur le modèle. Cette caractéristique ne modifie pas intrinsèquement le modèle d'offres. Dans la partie (3) Processus de réalisation de celui-ci, des activités impliquant le client pourront apparaître.

5.1.4 Périssabilité

Définition et discussion. La périssabilité signifie que les services ne peuvent être conservés. Si la ressource n'est pas utilisée à un instant t , elle ne pourra pas être récupérée, utilisée plus tard. Un service a une consommation immédiate, contrairement à certains produits qui ont une durée de vie de plusieurs années après livraison. Ils ont une durée éphémère et ne peuvent être stockés. Un service de location par exemple, n'est pas pérenne car le client ne jouit du service que pendant une durée limitée, contrairement à l'achat d'un produit, où le client jouit de la pérennité du bien. Les entreprises de services ont donc souvent une problématique de synchronisation de la demande client et de leurs capacités (Zeithaml et al. 1985).

Si l'on s'intéresse au livrable de l'étude de calcul précédente, celui-ci (le rapport d'étude) peut être stocké chez le client, et réutilisé par celui-ci quand il en aura besoin. De même, dans le cas d'une formation en ligne, comme un MOOC, celui-ci pourra être « stocké » puis réutilisé ultérieurement. Ces services ne semblent donc pas périssables.

Impact sur le modèle. Cette caractéristique n'a pas d'impact sur le modèle d'offres.

5.1.5 Transfert de propriété

Définition et discussion. Lovelock et Gummesson proposent de caractériser le service par l'absence de transfert de propriété (Lovelock et Gummesson 2004).

Impact sur le modèle. Cette caractéristique n'a aucun impact sur le modèle d'offres.

Seule l'intangibilité a un impact sur le modèle générique d'élaboration d'offres et sera conservée comme critère pour établir une typologie de services. Les caractéristiques d'hétérogénéité, de non séparabilité, de périssabilité et d'absence de transfert de propriété n'ont pas d'impact sur le modèle d'offres. Cependant, d'autres caractéristiques peuvent permettre de différencier des services. Nous en proposons dans la section suivante.

5.2 Identification de nouvelles caractéristiques

5.2.1 Obligation de moyens / de résultats

Définition et discussion. Les exigences du client peuvent porter uniquement sur le résultat final du processus de réalisation (le livrable), ou bien également

sur le déroulement du processus de réalisation. Dans le premier cas (obligation de résultat), on se rapproche de l'attente du client dans le cas de certains produits. Peu importe comment celui-ci a été fabriqué et assemblé, l'attente première du client porte sur le produit lui-même. C'est le cas, par exemple, des études de calculs. Peu importe pour le client que l'étude ait été faite dans un pays ou un autre, il se focalise sur le fait que les calculs soient réalisés et que le rapport soit clair et exploitable. Dans le deuxième cas (obligation de moyens), le client est intéressé par le déroulement du processus de réalisation lui-même. Par exemple, pour une formation, la durée, le lieu de la formation et les qualifications de l'intervenant sont autant de critères qui vont intéresser le client potentiel. Mais le client pourrait aussi n'être intéressé que par le résultat : un diplôme qu'il souhaite acquérir.

Impact sur le modèle. Si le client est intéressé par le déroulement du processus, celui-ci sera mieux défini. En effet, alors que dans le premier cas, le processus est défini uniquement pour le fournisseur, dans le deuxième cas, une attention plus particulière doit être portée aux points du processus qui intéressent le client. Notre modèle d'offre-services devra être capable de modéliser une offre selon ces deux points de vue.

5.2.2 Aspect temporel

Définition et discussion. L'aspect déroulement d'un service est important. Le service peut se dérouler une seule fois, de façon unitaire, ou de façon itérative. Par exemple, le client pourra passer un appel d'offre pour une unique étude de calcul sur une pièce donnée, ou bien pour avoir à sa disposition un pool d'ingénieurs ayant les compétences pour faire autant d'études de calcul que nécessaire pendant une durée donnée. Dans le cas de la maintenance, le processus est également itératif.

Impact sur le modèle. Cette caractéristique peut avoir deux sortes d'impacts sur le modèle générique d'offres. Elle peut impacter la partie (3) Processus de réalisation, car celui-ci pourra être plus complexe, avec plusieurs processus par exemple, ou un processus qui est répété plusieurs fois. Cette caractéristique peut aussi impacter la partie (2) Service de l'offre car des éléments sur les ressources mises à disposition du client, de façon générale et indépendamment d'une partie précise du processus de réalisation, pourront être nécessaires.

5.2.3 Incertitude

Définition et discussion. Le soumissionnaire peut s'engager sur des services plus ou moins incertains. Cette incertitude peut se situer sur trois niveaux. Il peut s'agir d'incertitude temporelle. Dans le cas de la maintenance par exemple, le soumissionnaire peut

s'engager sur une période donnée à intervenir en cas de problème. Il peut avoir identifié un éventail d'activités, de processus possibles à effectuer, mais ne sait pas exactement combien de fois ni quand il va devoir intervenir. La deuxième incertitude peut être liée aux ressources. Enfin, l'incertitude peut se situer sur les activités à réaliser elles-mêmes.

Impact sur le modèle. Cette incertitude a également un fort impact sur le modèle générique d'offres. Les parties (2) Service et (3) Processus de réalisation sont impactées. Dans le cas d'une incertitude temporelle, on peut imaginer que un à plusieurs processus sont définis, et que l'incertitude réside sur le nombre et le quand de chaque processus. Dans le cas d'une incertitude sur les ressources, le modèle générique d'offre n'est pas modifié. Enfin, dans le cas d'une incertitude structurelle, c'est la partie (2) Service et (3) Processus de réalisation qui peuvent être impactées.

5.3 Synthèse

Compte-tenu de ces caractéristiques, rien n'interdit la modélisation d'un service par une représentation structurelle corrélée au processus de réalisation. Certaines de ces caractéristiques peuvent influencer sur cette structure.

6 CONCLUSION

Cet article s'intéresse aux services industriels en réponse à appels d'offres. Un modèle de configuration d'offres-produit est proposé, puis une réflexion sur l'extension de ce modèle aux offres-service est menée. Dans le modèle d'offres-produit, la partie (1) contexte et la partie (3) processus de réalisation ont aussi lieu d'être pour les offres-service. La partie (2) nomenclature du produit est elle questionnée. Les quatre caractéristiques du service issues de la littérature (intangibilité, hétérogénéité, non séparabilité et périssabilité) n'ont pas toutes des impacts sur le modèle d'offres et ne suffisent pas pour bâtir une nouvelle typologie de services. De nouvelles caractéristiques sont proposées (attente du client, aspect temporel et incertitude), et leur identification permet de poser des premières pistes pour l'extension du modèle d'offres aux services.

Par la suite, une typologie de services industriels basée sur ces caractéristiques sera proposée, ainsi qu'une adaptation du modèle d'offres-produit aux offres-service.

REMERCIEMENTS

Nous remercions l'ensemble des partenaires du projet ANR OPERA (n° CE10-0010) pour leur implication.

RÉFÉRENCES

- Abeille, J., T. Coudert, É. Vareilles, L. Geneste, M. Aldanondo et T. Roux (2010). « Formalization of an Integrated System/Project Design Framework: First Models and Processes ». In : *Complex Systems Design & Management: Proceedings of the First International Conference on Complex System Design & Management CSDM 2010*. Sous la dir. de M. Aiguier, F. Bretaudeau et D. Krob. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 207–217.
- Aldanondo, M. et É. Vareilles (2008). « Configuration for mass customization: How to extend product configuration towards requirements and process configuration ». In : *Journal of Intelligent Manufacturing* 19.5, p. 521–535.
- Bateson, J. E. (1979). *Why we need services marketing*. Division of Research, Graduate School of Business Administration, Harvard University.
- Berry, L. L. (1980). « Services marketing is different ». In : *Business* 30.3, p. 24–29.
- Bielen, F. et C. Sempels (2004). « Proposition d'une échelle de mesure bidimensionnelle du concept d'intangibilité ». In : *Recherche et Applications en Marketing* 19.3, p. 21–44.
- Böttcher, M. et S. Klingner (2011). « Providing a method for composing modular B2B services ». In : *Journal of Business & Industrial Marketing* 26.5, p. 320–331.
- Bullinger, H. J., K. P. Fähnrich et T. Meiren (2003). « Service engineering - Methodical development of new service products ». In : *International Journal of Production Economics* 85.3, p. 275–287.
- Carlborg, P. et D. Kindström (2014). « Service process modularization and modular strategies ». In : *Journal of Business & Industrial Marketing* 29.4, p. 313–323.
- Chae, B. K. (2012). « An evolutionary framework for service innovation: Insights of complexity theory for service science ». In : *International Journal of Production Economics* 135.2, p. 813–822.
- Chandrasekaran, B. (1990). « Design problem solving: A task analysis ». In : *AI magazine* 11.4, p. 59.
- Edvardsson, B., A. Gustafsson et I. Roos (2005). *Service portraits in service research: a critical review*. T. 16. 1, p. 107–121.
- Felfernig, A., L. Hotz, C. Bagley et J. Tiihonen (2014). *Knowledge-Based Configuration: From Research to Business Cases*.
- Goldstein, S. M., R. Johnston, J. Duffy et J. Rao (2002). « The service concept: The missing link in service design research? ». In : *Journal of Operations Management* 20.2, p. 121–134.
- Grönroos, C. (2000). *Service management and marketing: customer management in service competition*. John Wiley & Sons.
- Guillon, D., A. Sylla, É. Vareilles, M. Aldanondo, É. Villeneuve, C. Merlo, T. Coudert et L. Geneste (2017). « Customer Supplier Relation: towards a Constraint-Based Model for Bids ». In : *Industrial Engineering and Engineering Management*. Singapore.
- Heiskala, M., K.-S. Paloheimo et J. Tiihonen (2005). « Mass Customisation of Services : Benefits and Challenges of Configurable Services ». In : *Frontiers of e-Business Research (FeBR)*, p. 206–221.
- INSEE (2017). *Définitions – Services*. URL : <https://www.insee.fr/fr/metadonnees/definition/c1161> (visité le 02/10/2017).
- Ki Moon, S., T. W. Simpson, J. Shu et S. R. Kumara (2009). « Service representation for capturing and reusing design knowledge in product and service families using object-oriented concepts and an ontology ». In : *Journal of Engineering Design* 20.4, p. 413–431.
- Løkkegaard, M., N. H. Mortensen et T. C. McAlone (2016). « Towards a framework for modular service design synthesis ». In : *Research in Engineering Design* 27.3, p. 237–249.
- Lovelock, C. et E. Gummesson (2004). « Whither Services Marketing? ». In : *Journal of Service Research* 7.1, p. 20–41.
- Mittal, S. et F. Frayman (1989). « Towards a Generic Model of Configuration Tasks ». In : *IJCAI*. T. 89, p. 1395–1401.
- Montanari, U. (1974). « Networks of constraints: Fundamental properties and applications to picture processing ». In : *Information Sciences* 7. Supplement C, p. 95–132.
- Parasuraman, A., V. A. Zeithaml et L. L. Berry (1985). « A Conceptual Model of Service Quality and Its Implications for Future Research ». In : *Journal of Marketing* 49.4, p. 41–50.
- Pitiot, P., M. Aldanondo et E. Vareilles (2014). « Concurrent product configuration and process planning: Some optimization experimental results ». In : *Computers in Industry* 65.4, p. 610–621.
- Robin, V., C. Merlo, G. Pol et P. Girard (2010). « Management of a Design System in a Collaborative Design Environment Using PEGASE ». In : *Modelling and Management of Engineering Processes*, p. 189–200.
- Shostack, G. L. (1977). « Breaking free from product marketing ». In : *The Journal of Marketing* 41.2, p. 73–80.
- Sylla, A., É. Vareilles, M. Aldanondo, T. Coudert, L. Geneste et K. Kirytopoulos (2017). « Customer/Supplier Relationship: reducing Uncertainties in Commercial Offers thanks to Readiness, Risk and Confidence Considerations ». In : *Advances on Mechanics, Design Engineering and Manufacturing*. Springer, p. 1115–1122.
- Zeithaml, V., A. Parasuraman et L. Berry (1985). « Problems and strategies in services marketing ». In : *The Journal of Marketing* 49.2, p. 33–46.

CENTRES D'APPEL D'URGENCE FRANÇAIS (SAMU) : UNE DEMARCHE DIRIGEE PAR LES MODELES POUR L'EXAMEN DE LEUR FONCTIONNEMENT

E.PETITDEMANGE, F.FONTANILI, E.LAMINE, M.LAURAS

CGI / IMT Mines Albi - Campus Jarlard 81000 Albi cedex 01

Epetitde@mines-albi.fr, mlauras@mines-albi.fr, fontanil@mines-albi.fr, lamine@mines-albi.fr

RESUME : *En France, les Services d'Aide Médicale Urgente (SAMU) gèrent les appels d'urgence de chaque département à travers un Centre de Réception et de Régulation des Appels (CRRA). Leur mission est d'assurer une écoute médicale permanente aux demandes d'urgence médicale et d'y répondre dans les meilleurs délais avec la réponse la mieux adaptée. Ces centres jouent un rôle majeur dans la gestion des urgences pré-hospitalières, mais sont-ils vraiment efficaces ? Comment fonctionnent-ils ? Est-il possible d'améliorer leur performance ? Pour répondre à ces questions, trois hypothèses sont émises : (1) Les pratiques des SAMU sont standardisées d'un département à un autre. (2) Tous les appels sont décrochés et traités. (3) Les ressources sont dimensionnées pour prendre en charge tous les appels. A l'aide de la modélisation des processus de deux SAMU pilotes ainsi que l'étude de données sur les appels entrants, ces hypothèses seront infirmées. En effet, d'un SAMU à un autre, des différences logicielles, matérielles et fonctionnelles sont observables. De plus, certains appels ne sont pas décrochés. Cet article propose d'établir une démarche dirigée par les modèles pour le diagnostic du fonctionnement des SAMU afin de proposer ensuite des solutions permettant d'améliorer leur performance.*

MOTS-CLES : *SAMU, Centre d'appel d'urgence, Modélisation, Performance, Evaluation, Diagnostic*

1 INTRODUCTION

Depuis la loi N°86-11 du 6 janvier 1986¹ chaque département français est tenu de disposer d'un service d'aide médicale urgente (SAMU). Les missions du SAMU sont de répondre par des moyens médicaux aux situations d'urgence qui leur sont transmises². Il existe au moins deux possibilités pour joindre un SAMU, quel que soit le lieu où l'appelant se trouve et quel que soit son opérateur : en composant le numéro national d'appel d'urgence (le 15) ou alors le numéro européen (le 112). Les appels via ces numéros sont automatiquement routés vers le SAMU départemental le plus proche. Chaque SAMU assure une écoute permanente des besoins en urgence et doit, si la situation l'exige, envoyer le moyen le plus adapté dans le délai le plus court³. En plus de cette mission de veille de l'urgence, chaque SAMU a une mission de formation à l'urgence⁴

De ce fait, un SAMU se compose d'un Centre de Réception et de Régulation des Appels (CRRA, d'un centre de formation des soins d'urgence mais aussi d'une ou plusieurs Structures Mobiles d'Urgence et de Réanimation (SMUR) qui sont réparties dans le département pour assurer la meilleure couverture possible du territoire. Chaque SMUR dispose d'une à plusieurs équipes médicales mobiles ainsi que de différents vecteurs de transport: véhicule léger, ambulance de réanimation, hélicoptère.

La communauté scientifique s'est déjà intéressée à l'optimisation de l'efficacité des SAMU. La littérature est très riche sur les problèmes de déploiement des ambulances (Iskander 1989) (Ingolfsson et al. 2003). Mais (Aringhieri et al. 2017) montre qu'il existe un manque dans la littérature avant l'envoi des moyens. Toutefois quelques études ce sont intéressées au traitement des appels d'urgence. En effet, (Lamine et al. 2015) proposent une analyse d'un centre d'appel avec un modèle de simulation à événements discrets couplé à une approche d'exploration de processus (process mining). Dans (Aboueljine et al. 2014), les auteurs utilisent aussi la simulation à événements discrets et proposent des scénarios d'améliorations d'un SAMU. Ils évaluent l'impact de différents scénarios (nombre de ressources, déploiement du SMUR, redéploiement multi-période, processus

¹ (Loi n° 86-11 du 6 janvier 1986 relative à l'aide médicale urgente et aux transports sanitaires)

² Selon le (Code de la santé publique - Article L6311-1)

³ Selon le (Code de la santé publique - Article R6311-2)

⁴ (Code de la santé publique - Article R6311-5)

du déploiement du SMUR) sur une prise en charge du patient en moins de vingt minutes. C'est une étude qui est plutôt axée sur le SMUR et l'envoi de moyens. Ces deux études ne s'intéressent qu'à un seul SAMU, sans aucune forme de collaboration et sans aucune comparaison avec d'autres SAMU.

La problématique de ce travail consiste à établir un diagnostic de deux SAMU distincts selon trois hypothèses de départ. La première est que tous les SAMU fonctionnent de la même manière, quel que soit le département. La deuxième est que tous les appels d'urgence sont pris en charge, en accord avec la mission gouvernementale des SAMU. La troisième est que la répartition des ressources humaines et matérielles est suffisante et adaptée au respect de la mission des SAMU. Pour parvenir à vérifier ces hypothèses et à mettre en lumière certains problèmes, nous proposons d'adopter une démarche scientifique basée sur une intégration avisée de méthodes et outils issus du génie industriel tels que la modélisation de processus et l'évaluation de performance. Le périmètre d'étude sera restreint au CRRA.

Cet article se décompose en quatre parties. Dans la partie 2, il s'agit d'établir un état des lieux sur le fonctionnement et l'organisation d'un SAMU et plus particulièrement de son CRRA.

La partie 3 est dédiée à la modélisation du processus de gestion d'un appel d'urgence. Cette modélisation se base sur les observations effectuées dans deux SAMU pilotes, le SAMU 31 de la Haute Garonne et le Samu 12 de l'Aveyron. Une attention particulière sera portée sur les différents états d'un appel entrant au SAMU dans le but d'évaluer l'efficacité du dispositif global. La quatrième partie décrit l'évaluation des hypothèses faites précédemment au vue de l'analyse du fonctionnement de ces deux SAMU et des appels d'urgence qui y parviennent. Cette analyse met en lumière des problèmes de fonctionnement communs aux deux SAMU et la pertinence des indicateurs de performance actuellement en place.

Enfin, la dernière partie permet de conclure l'article par une synthèse et des perspectives sur la poursuite des travaux.

2 ETAT DES LIEUX DU FONCTIONNEMENT DU SAMU

Le SAMU a pour mission d'assurer une écoute médicale permanente de manière à répondre avec les moyens adaptés aux appels d'urgences, dans les délais les plus courts possibles. Il est aussi en charge de l'organisation de l'accueil des patients qui nécessitent une hospitalisation. De plus, il doit assurer une mission de formation auprès du personnel des urgences.

2.1 Les acteurs du SAMU

Le SAMU fait intervenir plusieurs acteurs. Le personnel du CRRA, qui comprend les ARM (Assistants de Régulation Médicale) et les MR (Médecins Régulateurs). Les ARM prennent les appels, effectuent un premier tri et

une prise d'informations sur le patient avant de transférer l'appel à un MR. Le rôle du MR est d'effectuer un diagnostic médical pour évaluer si l'envoi de moyens est nécessaire. Les moyens pouvant être engagés sont variés : une équipe SMUR, le SDIS (Service Départementale d'Incendie et de Secours), une ambulance privée, un médecin généraliste, la police ou la gendarmerie. Tous ces acteurs peuvent être amenés à interagir avec le SAMU. Le fait que le Samu soit médicalisé est une particularité européenne qui ne se retrouve pas dans tous les pays. En effet (Belaidi et al. 2009) identifie les différents pays dans lesquels le médecin est présent sur le processus de traitement des appels. Dans le modèle anglo-saxon des moyens sont envoyés mais l'appel n'est traité par un médecin qu'une fois arrivé au service d'urgence.

2.2 Les centres d'appels

Dans la suite de cet article, nous nous intéresserons uniquement au CRRA. Pour se faire, il est intéressant de modéliser le processus de traitement des appels d'urgences. On considère plusieurs types d'appel au SAMU (voir Figure 1) : les appels entrants et les appels sortants. Les appels entrants sont les appels qui viennent de l'extérieur et qui arrivent au CRRA tandis que les appels sortants sont les appels émis par le CRRA en interne ou vers l'extérieur. Le périmètre de l'étude est restreint aux appels entrants. Parmi ceux-ci, on distingue les appels rejetés, les appels perdus et les appels décrochés. Ces derniers se décomposent en deux types : les appels médicaux et appels non-médicaux. En fonction de l'état du patient, un Dossier de Régulation (DR) ou un Dossier de Régulation Médicale (DRM) est ouvert. Il permet d'effectuer le suivi de l'appel lors de la réception des bilans médicaux. Certains appels ne nécessitent qu'un dossier de régulation sans l'intervention de moyens médicaux, d'autres n'auront même pas de dossier créé car l'urgence à traiter n'est pas du ressort du SAMU mais d'une autre entité, SDIS, gendarmerie, urgence vétérinaire etc.

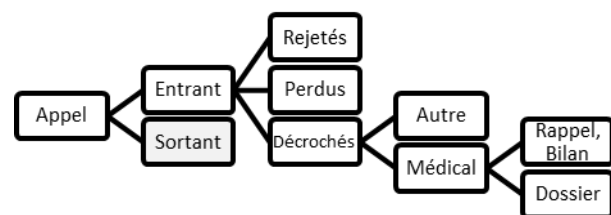


Figure 1: Typologie des appels d'urgence

Lorsqu'un appel arrive au SAMU, il passe par plusieurs états. L'association SAMU et Urgences de France (SUdF) définit ces différents états (Samu Urgence de France and Société Française de médecine d'Urgence 2015):

- Un *appel entrant* est considéré comme un appel initié par une personne extérieure vers le centre d'appel. Il aboutit à l'autocommutateur qui diffuse un message d'accueil.
- Un *appel rejeté* est un appel qui sort du système de manière automatisé sans possibilité d'être décroché ni

avec une densité de population de 33 hab./km² et plutôt rural avec des zones de moyenne montagne qui rend l'accès aux lieux d'interventions compliqué dans certains cas.

3.1 Modélisation des processus

Le processus de la prise en charge d'un appel d'urgence a été modélisé en faisant l'hypothèse que seuls les appels entrants sont pris en compte. Dans la gestion d'un appel d'urgence, on trouve quatre macro-tâches pouvant être décomposées ainsi (voir figures 3 et 4) :

- *L'évaluation de l'appel*, où l'ARM décroche l'appel pour prendre les informations de base sur le patient : nom, prénom, adresse, numéro de téléphone, raison de l'appel. Cela lui permet d'effectuer une rapide analyse de la gravité en se basant essentiellement sur son expérience. Il peut déclencher un envoi de moyen dit « départ réflexe » dans les cas d'urgences vitales, sinon il transfère l'appel au MR pour réguler l'appel.
- *La régulation de l'appel* concerne le médecin régulateur qui va pouvoir analyser la situation et diagnostiquer le patient par téléphone. Il pourra ensuite décider des mesures à prendre : conseil médical ou envoi de moyens.
- *L'engagement des moyens* peut être décidé par le MR après qu'il ait régulé l'appel, en envoyant des équipes de soin sur place. Plusieurs types de moyens peuvent être engagés : SMUR, pompiers, ou ambulance.
- *L'orientation vers un établissement de santé* peut arriver quand le patient a besoin d'être transporté dans un hôpital après l'intervention de l'équipe médicale mobile. C'est aussi le rôle du SAMU de gérer ce transfert.

Si l'on compare les deux processus globaux de Rodez (cf. figure 4) et Toulouse (cf. figure 3), on constate une première différence qui est la présence d'une ressource supplémentaire dans l'organisation à Toulouse. En effet, une personne est affectée à temps plein sur la gestion des moyens. C'est un ARM qui ne décroche pas d'appel mais qui gère les réservations d'ambulances, l'envoi des équipes SMUR et le contact avec les pompiers pour en-

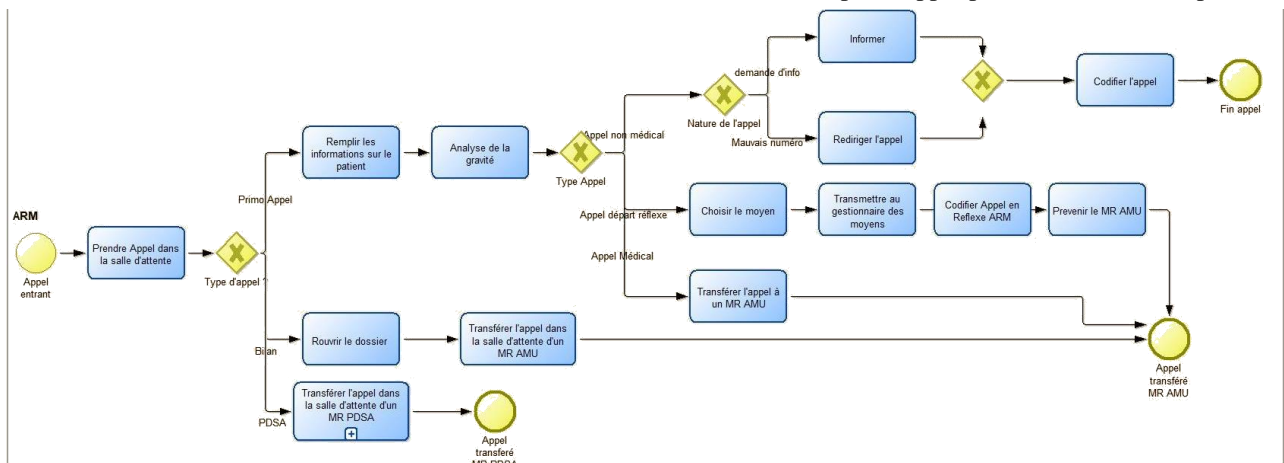


Figure 3 : Processus d'évaluation de l'appel à Toulouse

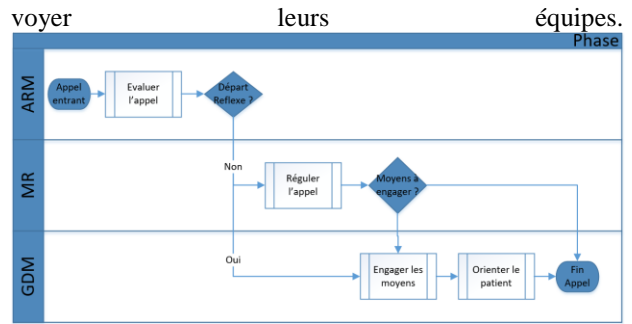


Figure 4 : Processus de la prise d'un appel d'urgence à Toulouse

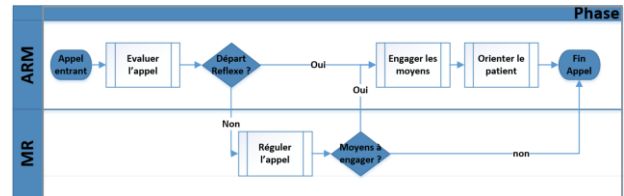


Figure 5 : Processus de la prise d'un appel d'urgence à Rodez

Si l'on regarde de plus près les processus d'évaluation de l'appel à Rodez (figure 6) et à Toulouse (figure 5), d'autres différences sont observables. En effet, il apparaît que les ARM de Toulouse décrochent aussi les appels de la permanence des soins afin de les transférer à un médecin généraliste en régulation, tandis qu'à Rodez une différence est faite entre les ARM de l'aide médicale urgente et ceux de la permanence des soins. Les deux sont gérés de manière indépendante. La manière de récupérer les bilans d'interventions est complètement informatisée à Toulouse, alors que l'utilisation du papier est encore très présente à Rodez. La gestion des appels en attente est différente. En effet, à Toulouse, il y a une visibilité sur les appels en attentes. On sait depuis combien de temps ils attendent. Pour chaque appel, on sait aussi s'il est en lien avec un dossier médical déjà ouvert. Les ARM peuvent choisir l'appel qu'ils souhaitent décrocher. Il n'y a pas de critères officiels pour la priorisation du décroché, intuitivement les appels qui ont attendus le plus longtemps sont décrochés en priorité. Ce qui souligne l'intérêt de modéliser d'une manière fine les différents états d'un appel. Cela permet en effet de voir combien de temps un appel peut rester en attente pour

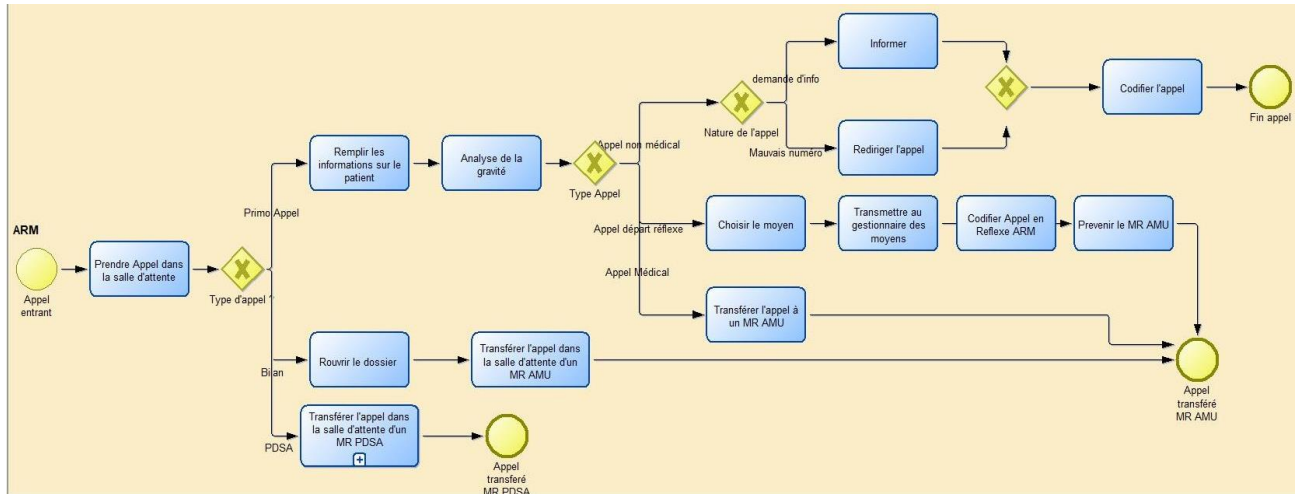


Figure 6 : Processus d'évaluation de l'appel à Rodez

définir un indicateur de temps maximum d'attente ou de nombre d'appels pris avant un certain temps. A Rodez, le logiciel de téléphonie ne permet pas la visualisation de l'attente des appels. On ne sait pas combien d'appels sont en attente d'être décrochés ni depuis combien de

temps ils attendent. De plus, il n'est pas possible de choisir l'appel que l'on souhaite décrocher : le premier appel à avoir été mis en attente sera le premier appel décroché.

3.2 Profil des appels

L'étude des données d'appel sur une journée dans chaque SAMU permet de représenter les appels entrants et les appels décrochés afin de mettre en lumière ceux qui ne sont pas décrochés. Le profil d'une journée de week-end pour les appels entrants de l'aide médicale urgente à Toulouse est représenté sur la figure 6. Ce sont des journées significatives car les appels sont plus nombreux que durant la semaine. On constate une augmentation des appels non décrochés entre 11h00 et 13h00. 44 à 59% des appels ne sont pas décrochés, ce qui signifie qu'ils ne sont pas pris en charges, qu'ils sont perdus

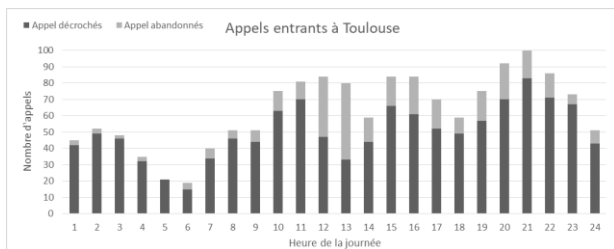


Figure 6 : Appels Entrants et décrochés à Toulouse

Un autre créneau est critique, entre 19h00 et 20h00, puisque là aussi, 24% des appels ne sont pas décrochés. Ce n'est que durant la nuit entre minuit et 5h00 du matin que les appels sont presque tous décrochés. Pour Rodez (figure 7), les appels sont presque tous décrochés. Toutefois il convient de mettre en parallèle la volumétrie des appels. A Toulouse, les appels entrants varient entre 10 et 90 appels par heure alors qu'à Rodez les appels varient entre 1 et 15 appels par jour.

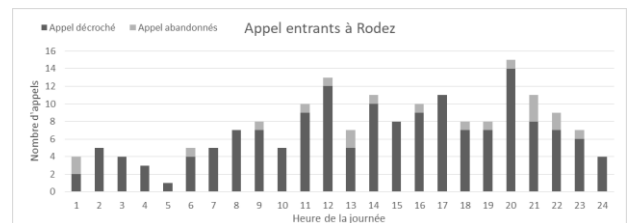


Figure 7: Appels entrants et décrochés à Rodez

Parmi les appels décrochés, nous avons voulu voir en combien de temps ces appels étaient décrochés. Ainsi une distinction est faite entre les appels décrochés avant et après 60 secondes d'attente.

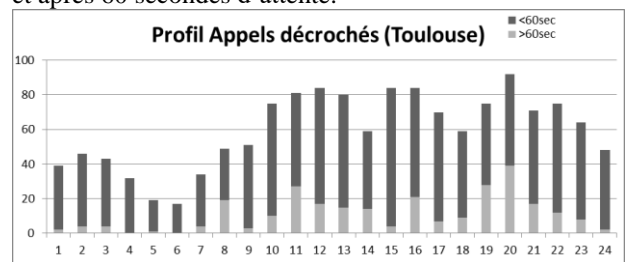


Figure 8 : Appels décrochés avant et après 60 secondes à Toulouse

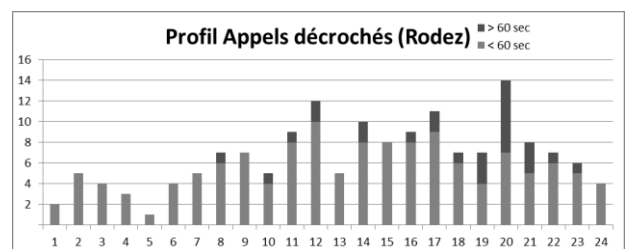


Figure 9 : Appels décrochés avant et après 1min à Rodez

Le taux des appels décrochés après 60 secondes atteint 42% des appels à 19h00 pour Toulouse (figure 8). A Rodez, entre 19h00 et 21h00 le taux des appels décrochés après une minute varie entre 38% et 50% (figure 9).

3.3 Gestion des ressources

La manière dont sont gérées les ressources humaines varient d'un SAMU à l'autre, tant en terme d'horaires, d'effectif, que de rôle. En effet, le SAMU de Rodez a mis en place deux quarts de travail, un de jour et un de nuit, avec une rotation des ressources. Toulouse a trois plages horaires : de 6h à 13h42, de 12h18 à 20h et de 20h à 6h avec deux rotations des ressources. Le nombre d'ARM est de 8 pour les créneaux de 6h à 20h et de 6 de 20h à 6h. Les médecins sont trois durant la journée et deux durant le quart de nuit. A Rodez, l'affectation des ressources est fixe puisqu'en tout temps il y a un ARM et un MR. Il est intéressant de mettre en parallèle le nombre de ressources en poste avec la courbe des appels entrants (voir figure 10). En effet, on peut voir que l'affectation des ressources varie très peu en fonction de l'heure de la journée et qu'elle ne semble pas corrélée avec la variation journalière de la quantité d'appel. On retrouve le même phénomène à Rodez.

Il pourrait être intéressant de définir un indicateur de productivité des ARM et des MR en fonction de l'heure de la journée et de la prévision du nombre d'appel prévu. Cela permettrait de voir les moments où des ressources sont manquantes ou en surplus.

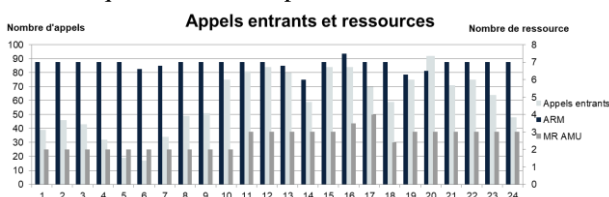


Figure 10 : Lien entre appels entrant et affectation des ressources

4 VERIFICATION DES HYPOTHESES

Toutes ces observations nous permettent de vérifier les hypothèses qui ont été émises au début.

4.1 Hypothèse 1 : Des SAMU aux fonctionnements standardisés

La modélisation des processus des deux SAMU montre des différences de fonctionnement entre les deux établissements. On trouve des différences :

- de gestion des ressources humaines : Toulouse a spécialisé un ARM pour la gestion des moyens alors que Rodez n'a pas d'ARM spécifique pour cette tâche;
- de logiciels : les deux établissements ne sont pas équipés avec le même niveau de logiciel pour la régulation médicale. Les systèmes de téléphonie sont aussi différents. Certaines tâches sont automatisées à Toulouse uniquement. C'est le cas de la vérification de la disponi-

bilité des ambulances qui a Rodez doit être fait en appelant les compagnies d'ambulances une à une;

- organisationnelles : Toulouse n'a pas différencié ses ARM entre la PDSA et l'AMU, alors qu'à Rodez les rôles sont spécifiques.

Ainsi, l'hypothèse de standardisation n'est pas vérifiée. Ces observations peuvent être complétées par les études de (Lamine et al. 2015) et de (Aboueljineane 2014) où deux autres SAMU sont étudiés et des différences sont aussi constatées notamment au niveau des logiciels de régulation et de téléphonie. D'un SAMU à l'autre les pratiques et les moyens diffèrent ce qui ne permet pas d'assurer une homogénéité de la prise en charge des urgences sur le territoire français.

4.2 Hypothèse 2 : Des SAMU efficaces

Dans cette hypothèse, nous voulions vérifier que tous les appels étaient pris en charge afin de respecter la mission des SAMU. L'analyse du profil des appels sur une journée montre rapidement que tout au long de la journée des appels ne sont pas décrochés, surtout aux heures de pointes comme celle du changement d'équipe aux alentours de 19h. On constate alors que presque la moitié des appels ne sont pas décrochés. De plus, si l'on considère les données fournies à la SAE (Statistique Annuelle des Établissements de santé) (Recherche - SAE Diffusion), on peut calculer le nombre d'appel décroché en moins de 60 secondes sur l'ensemble des appels décrochés en se basant sur les données de l'année 2016. La Qualité de service est défini par Samu et Urgence de France comme étant le rapport entre le nombre d'appel décroché avant x secondes et l'ensemble des appels décrochés (voir équation 1) (Samu Urgence de France and Société Française de médecine d'Urgence 2015). Ces derniers ont fixé une valeur cible à 99% des appels qui doivent être décrochés. En utilisant ces données, on arrive à une qualité de service de 84% pour Toulouse et de 74% pour Rodez, soit bien en dessous de la valeur cible. Au-delà du fait que l'indicateur de qualité de service n'est pas respectée cela reflète que tous les appels ne sont pas pris en charge rapidement. De plus, si on regarde cet indicateur pour toute la région Occitanie, on constate qu'aucun des SAMU n'atteint l'objectif de 99%. La pertinence de cet indicateur est contestable et semble difficilement atteignable dans l'état actuel des SAMU. De nouveaux indicateurs de performances seront définis dans la suite de ces travaux.

Équation 1: Formule de la Qualité de Service Selon SUdF

$$\text{Qualité de Service} \quad QS_x = \frac{nb \text{ d'appels décrochés en } x \text{ sec}}{\text{ensemble des appels décrochés}}$$

4.3 Hypothèse 3 : Une affectation des ressources efficiente

La dernière hypothèse concerne la bonne affectation des ressources. La comparaison du nombre de personnels affectés dans une journée avec la saisonnalité des appels

vient infirmer cette hypothèse. De plus, il s'agit de prendre en compte une saisonnalité hebdomadaire, le nombre d'appels augmente durant la fin de semaine, et même une saisonnalité annuelle en fonction des épidémies. Or, les ressources sont attribuées sur un nombre moyen d'appel par jour lissé sur l'année sans prendre en compte l'effet de saisonnalité. De ce fait une affectation fixe des ressources ne permet pas en l'état actuel de répondre aux appels d'urgence lors des pics d'activités. Il faudrait proposer une affectation des ressources qui puisse être adaptée à la saisonnalité des appels. C'est ce que nous travaillerons dans la suite de ces travaux.

5 CONCLUSION ET PERSPECTIVES

Dans cet article, nous avons émis des hypothèses de départ sur le fonctionnement des SAMU. A travers une démarche de modélisation, nous avons montré que les services téléphoniques d'urgence français n'avaient pas un fonctionnement unique et qu'il pouvait y avoir des différences de fonctionnement d'un département à un autre. Ces informations devraient être complétées en ajoutant plus de SAMU à l'étude afin d'observer si des caractéristiques particulières influent sur les modes de traitement des appels selon les départements. Cela permettrait de créer des catégories de SAMU. Plusieurs types de différences ont été identifiées : fonctionnelles, informatiques, ou de ressources humaines. Ensuite, nous avons pu constater que l'efficacité des centres d'appel n'était pas optimale car tous les appels n'étaient pas décrochés. Certaines tranches horaires sont sujettes à des pics d'abandons d'appels. Enfin, l'affectation des ressources se base sur une moyenne des appels. Donc il arrive qu'il n'y ait pas assez de personnes pour décrocher lors des pics d'appels. L'affectation des ressources ne prend pas en compte la variabilité du nombre d'appels dans la journée.

La littérature montre que les études faites jusqu'à présent sur les centres d'appels des SAMU ne concernent qu'un SAMU pris isolément. Travailler avec des SAMU de la même région permet d'envisager des scénarios de collaborations entre les SAMU.

Enfin, nous avons démontré que, dans un contexte d'augmentation des appels, l'affectation des ressources ne permet pas de décrocher tous les appels. Le fonctionnement des SAMU peut être mieux évalué et amélioré. Pour la suite du travail nous comptons utiliser la simulation à événement discret afin de proposer des scénarios d'amélioration du fonctionnement et de collaboration des SAMU.

REMERCIEMENTS

Nous tenons à remercier le Dr. Pierre Rodriguez (chef du service du SAMU 12), le Dr Laurent Gout (médecin régulateur au SAMU 31) et le Dr Vincent Bounes (chef du service du SAMU 31) pour leur collaboration active et

leur partage de connaissance sur le fonctionnement du SAMU.

REFERENCES

- Aboueljinnane L (2014) Evaluation et amélioration des performances des Systèmes d'Aide Médicale Urgente : application au SAMU du département du Val de Marne. phdthesis. (Ecole Centrale Paris).
- Aboueljinnane L, Sahin E, Jemai Z, Marty J (2014) A simulation study to improve the performance of an emergency medical service: Application to the French Val-de-Marne department. *Simul. Model. Pract. Theory* 47(Supplement C):46–59.
- Anon Code de la santé publique - Article L6311-1
- Anon Code de la santé publique - Article R6311-2
- Anon Code de la santé publique - Article R6311-5
- Anon Loi n° 86-11 du 6 janvier 1986 relative à l'aide médicale urgente et aux transports sanitaires
- Anon Recherche - SAE Diffusion. Retrieved (January 23, 2018), <https://www.sae-diffusion.sante.gouv.fr/sae-diffusion/recherche.htm>.
- Aringhieri R, Bruni ME, Khodaparasti S, van Essen JT (2017) Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Comput. Oper. Res.* 78:349–368.
- Belaïdi A, Besombes B, Marcon E, Guinet A (2009) Toward a Decision Support Tool for Emergency Networks in France. *Intell. Patient Manag. Studies in Computational Intelligence.* (Springer, Berlin, Heidelberg), 25–37.
- Gill P, Stewart K, Treasure E, Chadwick B (2008) Methods of data collection in qualitative research: interviews and focus groups. *Br. Dent. J.* 204(6):291.
- Ingolfsson A, Erkut E, Budge S (2003) Simulation of single start station for Edmonton EMS. *J. Oper. Res. Soc.* 54(7):736–746.
- Iskander WH (1989) Simulation Modeling For Emergency Medical Service Systems. 1989 Winter Simul. Conf. Proc. 1107–1111.
- Jones M (1999) An introduction to qualitative methods for health professionals. *J. R. Soc. Med.* 92(12):657–657.
- Lamine E, Fontanili F, Mascolo MD, Pingaud H (2015) Improving the Management of an Emergency Call Service by Combining Process Mining and Discrete Event Simulation Approaches. *Risks Resil. Collab. Netw. IFIP Advances in Information and Communication Technology.* (Springer, Cham), 535–546.
- Lima MA de QV, Maciel PRM, Silva B, Guimarães AP (2014) Performability evaluation of emergency call center. *Perform. Eval.* 80:27–42.
- Rajsiri V, Lorré JP, Bénaben F, Pingaud H (2010) Knowledge-based system for collaborative process specification. *Comput. Ind.* 61(2):161–175.

Samu Urgence de France, Société Française de médecine d'Urgence (2015) Samu Centre 15 référentiel et guide d'évaluation. <http://www.samu-de->

france.fr/documents/actus/155/802/sfmusudf_referentiel_samu_2015.pdf.

Column Generation for the Kidney Exchange Problem

Lucie Pansart, Hadrien Cambazard, Nicolas Catusse

Gautier Stauffer

Univ. Grenoble Alpes, CNRS, Grenoble INP*, G-SCOP

The Centre of Excellence in Supply Chain (CESIT)

38000 Grenoble, France

KEDGE Business School

33000 Bordeaux France

{firstname.name}@grenoble-inp.fr

gautier.stauffer@kedgebs.com

ABSTRACT: *The Kidney Exchange Problem (KEP) aims at finding the best exchanges in a barter market where agents are patients with a willing but incompatible donor. (Abraham et al. 2007) introduced a natural (exponential) integer programming formulation called the cycle formulation that they could solve efficiently by a branch-and-price approach. Recently, several countries allowed for the participation of altruistic donors in the exchanges. The corresponding variant of KEP is harder to solve as the pricing problem becomes \mathcal{NP} -complete. In this work, we study and experiment a column generation approach that takes into account altruistic donors. We use advanced techniques to circumvent the \mathcal{NP} -hardness of the pricing problem and show that the corresponding method can provide excellent guaranteed feasible solutions in a small amount of time.*

KEYWORDS: *Kidney exchange problem, column generation, integer programming, color coding*

1 INTRODUCTION

One in every thousand Europeans is facing end-stage renal disease for which the most effective treatment is kidney transplant. For such a patient, waiting for a deceased donor can be tedious while living donation is a viable alternative. Of course, in order to avoid certain excesses, living donation conditions are very strict and are defined by specific laws, for example, only a close relative is authorized to donate his or her kidney to a specific patient. Living donation is faced with a major hurdle in the potential incompatibilities that can occur between the patient and his or her donor: it is risky to perform an incompatible transplant as it would significantly reduce the patient's immune defenses. The use of paired donation programs can lead to higher graft survival rates (NHS 2017). A paired donation program makes exchanges between incompatible patient-donor pairs (a patient and a close relative willing to donate one kidney but medically incompatible with the patient). If a pair is included in an exchange, the donor gives one kidney to another – compatible – patient and the patient receives a kidney from another – compatible – donor of the exchange. Paired donation programs can also involve altruistic donors defined as singletons who anonymously and charitably give a kidney. Programs performing paired donations grow continuously, sometimes aiming at encompassing several countries, as their effectiveness is strongly dependent on the number of participants. Emerging in-

ternational and European programs require to tackle pools of the order of thousands of patients. The problem of finding the "best" exchanges in a program is a challenging mathematical problem, in particular the question of developing efficient algorithms capable of tackling a large number of patients and donors at a time.

We focus in this study on algorithms providing good upper and lower bounds in order to be able to assess the quality of our results. (Abraham et al. 2007) developed an exponential integer programming formulation, the cycle formulation, for the problem containing only patient-donor pairs. They solve the model with a branch-and-price algorithm that performs well. This model yet fails to accommodate with the addition of altruistic donors. Informally (see section 2.2 for details), the linear relaxation of the cycle formulation is solved with a technique referred to as column generation, which iteratively add well-chosen variables. Identifying a variable to add is known as the pricing problem and is \mathcal{NP} -complete in the presence of altruistic donors (Plaut et al. 2016b) in contrast with (Abraham et al. 2007). In this paper, we propose to combine local search and a randomized method, called color coding, to solve the pricing problem efficiently. We also use a relaxation of the pricing problem to compute an upper bound even if the pricing step is not solved optimally.

The rest of this section introduces the problem, its modeling, and a short literature review. In section 2, we present the integer programming formulation used

*Institute of Engineering Univ. Grenoble Alpes

in this article as well as the column generation algorithm. The next two sections describe the algorithms used to solve the pricing problem (section 3) and the Kidney Exchange Problem (section 4). Finally experimental results are reported in section 5.

1.1 Definition of the problem

Paired donation. A basic Kidney Paired Donation Program (KPD) involves incompatible patient-donor pairs. It creates exchanges between some of the pairs such that the patient of a pair receives the kidney from the donor of another pair. Figure 1 shows the simplest case involving only two pairs.

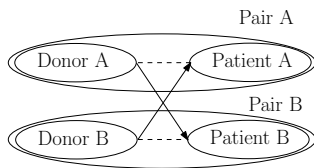


Figure 1 – An exchange of size 2. Dashed lines show incompatibility and arrows show the exchange.

A KPD can create several exchanges among the pool of pairs, called *cycles of donation*, of any size. However, the withdrawal of a donor after the transplant of his or her patient means that in another pair a donor already gave a kidney while his or her patient still waits for a transplant. To avoid this unacceptable situation, all operations of a cycle must be executed simultaneously. As a cycle requires twice as many operating rooms and surgical teams than pairs – one for each patient and each donor –, in practice cycles must be of a limited size $k \in \mathbb{N}$, typically $k = 2, 3, 4$.

Domino paired donation. When KPD involves *altruistic donors*, these latter create *chains of donation*. In this case, called *Domino Paired Donation*, the simultaneity is still desirable as the withdrawal of a donor breaks an exchange, but it is no more required since the critical situation described in the previous paragraph will never happen. Thus, some donors can be *bridge donors*: their patient will receive a kidney before they give their own. The last donor of a chain can either give a kidney to the *waiting list* or be considered later as an altruistic donor. As a result, there is no consensus on the need of a limit on the size of a chain, but when it exists, this limit is usually greater than k . In this paper we consider chains with at most $l \in \mathbb{N}$ transplants. Figure 2 illustrates the main types of chains.

Model. The Kidney Exchange Problem (KEP) is the problem of finding the *best* exchanges in a KPD, where *best* is defined by hospitals or transplant agencies.

Let us define P to be the set of patient-donor

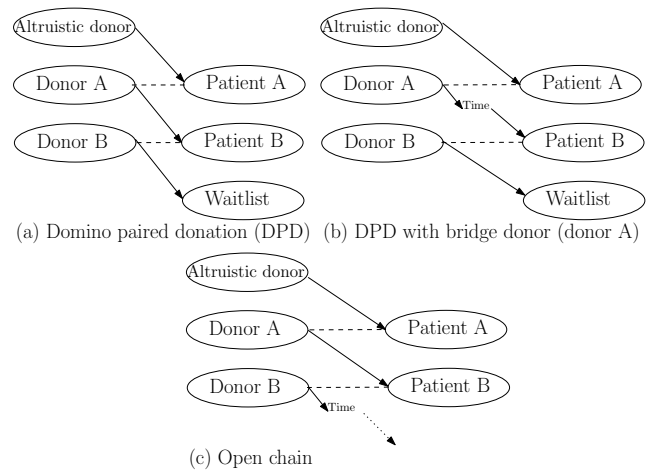


Figure 2 – The main types of chains. Dashed lines show incompatibility and arrows show the exchange. The dotted line shows a potential future exchange, not known at this stage of the program.

pairs and N the set of altruistic donors. The compatibility graph of a KEP is the directed graph $G = (V = P \cup N, A)$, where $A = \{(uv) \text{ if donor of } u \text{ is compatible with patient of } v, \forall u \in V, v \in P\}$. We assume we are given a weight function $w : A \rightarrow \mathbb{R}^+$. w_{uv} represents the medical benefit of transplanting the kidney of donor u to patient v . A simple chain (resp. cycle) is a chain (resp. a cycle) with no repeated vertices (except the starting and ending vertices). We use the term **valid chain** to refer to a simple chain in G initiated by an altruistic donor with at most l arcs and **valid cycle** to refer to a simple cycle in G of length at most k . Chains and cycles must be simple since a donor (resp. a patient) can give (resp. receive) only one kidney. The term **valid path** refers to either a valid cycle or chain. The weight w_p of any path p is defined as the sum of its arcs weights. Figure 3 illustrates an example of the model. Formally, the KEP is: given G, w, k and l , find a set of disjoint valid paths of maximum weight in G . Note that this solution does not necessarily cover the maximum number of vertices since it depends on the weight function.

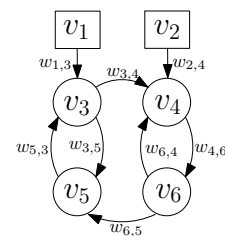


Figure 3 – An example of compatibility graph G . Squares represent altruistic donors and circles represent incompatible patient-donor pairs. Many solutions are possible, depending on l and k , e.g.: {chain $(v_1v_3v_5)$ and chain $(v_2v_4v_6)$ } or {chain $(v_1v_3v_5)$ and cycle (v_4v_6) } or cycle $(v_3v_4v_6v_5)$...

1.2 Literature review

Since we focus on integer programming approaches in this paper, the following section will mainly review literature on this topic. Other approaches and variants were studied these past years. The reader can find comprehensive literature reviews in (Mak-Hau 2017) and (Gentry et al. 2011).

The idea of kidney exchange programs was introduced by (Rapaport 1986) and developed for the first time in 1991 in South Korea (Kwak et al. 1999). Later on, the Netherlands, the UK and Spain started their program, followed by many others, including France and the USA. This problem is currently an important research topic in Europe (see (COST 2016)).

The KEP was proved \mathcal{NP} -complete by (Abraham et al. 2007) when $k > 2$ even in the case of KPD with pairs only (without altruistic donors). This basic case was widely studied and several integer programs were proposed. Standard formulations are exponential so branch-and-price approaches were developed by (Abraham et al. 2007) and (Klimentova et al. 2014), while (Constantino et al. 2013), (Mak-Hau 2017) and (Dickerson et al. 2016) introduced alternative compact formulations.

When chains are allowed, the problem is more complex. Indeed, the linear relaxation is harder to compute because the pricing problem is \mathcal{NP} -complete (Plaut et al. 2016b). To the best of our knowledge, no correct branch-and-price algorithm including altruistic donors was published (previous propositions of (Glorie et al. 2014) and (Plaut et al. 2016a) turned out to be incorrect, as the authors explained in (Plaut et al. 2016b)). Finally, (Anderson et al. 2015) described algorithms using integer programming formulations based on the traveling salesman problem. Compact formulations either do not scale up to large instances or provide poor upper bounds (Mak-Hau 2017). As we hope to construct international programs, including several hundreds or thousands of patients, we focus on the natural exponential formulation.

2 INTEGER PROGRAMMING

We introduce in this section an integer program directly adapted from the cycle formulation of (Abraham et al. 2007) and the framework of the column generation.

2.1 Path formulation

Let us define \mathcal{P} as the set of all valid paths in G . The decision variables of the *Path Formulation* (PF) are:

$$\forall p \in \mathcal{P}, x_p = \begin{cases} 1 & \text{if path } p \text{ is chosen in the exchange} \\ 0 & \text{otherwise} \end{cases}$$

$$z^* = \max \sum_{p \in \mathcal{P}} w_p x_p \quad (1)$$

$$\text{s.t. } \sum_{p \in \mathcal{P}: v \in p} x_p \leq 1 \quad \forall v \in V \quad (2)$$

$$x_p \in \{0, 1\} \quad \forall p \in \mathcal{P} \quad (3)$$

Constraints (2) impose that each vertex is taken at most once, namely donors and patients are involved in zero or one exchange. The objective (1) is to maximize the overall weight of exchanges. The linear relaxation of (PF) is noted (PF_{LP}) and composed of equations (1), (2) and :

$$x_p \geq 0 \quad \forall p \in \mathcal{P} \quad (4)$$

2.2 Column Generation

In (PF) the number of variables grows exponentially with k and l . To tackle this problem, a prevailing solution is to use a branch-and-price algorithm which mixes brand-and-bound with column generation. Column generation is used to solve the linear relaxation of (PF) as explained below.

2.2.1 Principle

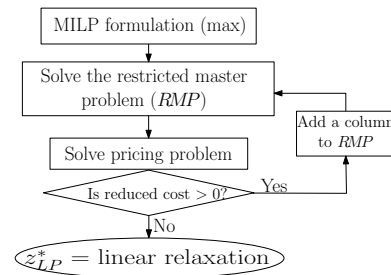


Figure 4 – Framework of the column generation

The general framework of the column generation is shown in figure 4. It is a form of the simplex algorithm; indeed, the simplex algorithm does not need all variables to be explicitly included in the model: the required variables can be iteratively added, until it is proved that no more variable is required. Thus, the linear relaxation of (PF) can be solved starting from a subset of valid paths – this problem is called the *restricted master problem* –, and this subset grows as we generate new variables (*i.e.*, new columns). Generating a new variable aims at improving the objective value of the current basis in the simplex algorithm. Finding such a column is called the *pricing problem* or *slave problem*. In our case, the pricing problem is the following: find a valid path of positive reduced cost or prove none exists. Thus, the pricing problem is to find a valid path of maximal reduced cost. The reduced cost of a path p is given by $rc_p = w_p - \sum_{v \in p} \alpha_v$

where α_v is the dual value of vertex v associated to constraints (2) in the linear relaxation of (PF).

2.2.2 Complexity of the pricing problem

Suppose we are given G , w , l and $a_v \forall v \in V$.

Cycles. The *cycle pricing problem* is the problem of finding a valid cycle of maximal reduced cost in G . It can be solved in polynomial time in $|V|$ with a modified Bellman-Ford algorithm (Glorie et al. 2012).

Chains. The *chain pricing problem* is the problem of finding a valid chain of maximal reduced cost in G . It is \mathcal{NP} -complete (Plaut et al. 2016b).

2.2.3 Our algorithm

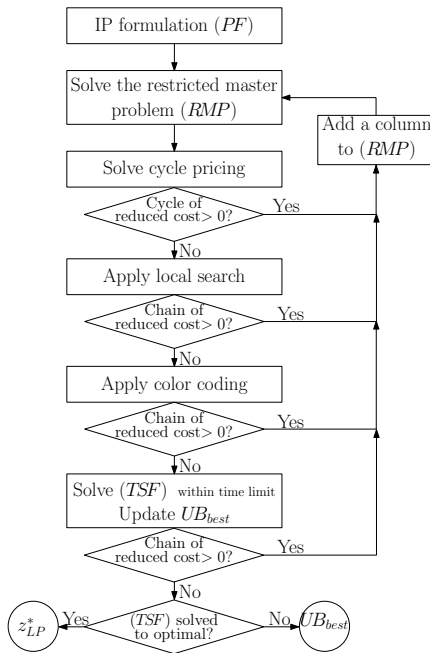


Figure 5 – The column generation algorithm

As the ChPP is \mathcal{NP} -complete, some efforts have to be made to solve this problem. In the next section, we will present different ways to solve it: an exact method – an integer programming formulation called Time-Staged Formulation (TSF) – and two non-exact methods – a local search heuristic and a randomized method called color coding (Alon et al. 1995).

We solve the pricing problem with algorithms of increasing complexity, as shown in figure 5. We try to solve it with the fastest methods first, which can be sufficient to find lots of columns. When no column is found with these algorithms, the exact method is used, solving (TSF) within a small time limit. As we will see in section 4.1, this provides a new upper bound on z^* and possibly a new column or the proof of the optimality of z_{LP}^* . The best upper bound

UB_{best} is kept all along the course of the algorithm and is updated when (TSF) is solved.

3 SOLVING THE PRICING PROBLEM

In this section we study how to solve the ChPP which is the difficult part of the pricing problem. We construct a graph $G' = (V' = V \cup \{s, t\}, A')$ where s is a source and t a sink. We construct A' with A , A_s the set of arcs from the source to all altruistic donors and A_t the set of arcs from all pairs to the sink, namely $A_s = \{(sv) : \forall v \in N\}$, $A_t = \{(vt) : \forall v \in P\}$ and $A' = A \cup A_s \cup A_t$. Given the dual values $\alpha_v \forall v \in V$, we define the arc function:

$$rc : A' \rightarrow \mathbb{R}$$

$$a \rightarrow rc_a = \begin{cases} -\alpha_v & \text{if } a = (sv) \in A_s \\ 0 & \text{if } a = (vt) \in A_t \\ w_{uv} - \alpha_v & \text{if } a = (uv) \in A \end{cases}$$

The reduced cost of a path, defined previously as $rc_p = w_p - \sum_{v \in p} \alpha_v$ is also $rc_p = \sum_{(uv) \in p} rc_{uv}$. We look for a valid chain p^* of maximal reduced cost: $rc_{p^*} = rc^* = \max_{p \in \mathcal{P}} rc_p$. It is a simple longest path problem for which a dynamic programming approach adapted from (Held & Karp 1962) can be developed, but does not scale up to the instances we aim at solving. In fact, it is enough for the column generation algorithm to find a valid chain of positive (not maximal) reduced cost, so the use of heuristic methods is reasonable. However, to prove that the column generation finished, an exact method is necessary. Thus, we present in this section two non-exact methods and an integer programming to handle the ChPP.

3.1 Color coding (CD)

(Alon et al. 1995) proposed a randomized algorithm to find simple paths of a given length, called *color coding*. We use this method to find a solution of ChPP. It follows two steps:

1. randomly color each vertex $v \in V$ with a color $c_v \in \{1, 2, \dots, C\}$, where $C \in \mathbb{N}$ is the fixed number of colors. Note that s and t are not colored.
2. solve the problem of finding a *colorful chain* – i.e., a valid chain using each color at most once – of maximal reduced cost.

Step 2 is solved with a dynamic program, defined by:

$$g^*(C, v) = \max_{\substack{u \in V' : c_u \in C \setminus \{c_v\} \\ (uv) \in A'}} \{g^*(C \setminus \{c_v\}, u) + rc_{uv}\}$$

$g^*(C, v)$ is the maximal reduced cost of a chain starting at s , visiting exactly one vertex of each color of C and finishing in v . The maximal reduced cost of a colorful chain is given by $\max_{\substack{C \subseteq \{1, \dots, C\} \\ |C| \leq l+1}} g^*(C, v)$

This dynamic programming has a space complexity of $\mathcal{O}(|V|2^C)$ and a time complexity of $\mathcal{O}(|V|^22^C)$.

In order to get an optimal chain p^* by color coding, step 1 must, by chance, color p^* with different colors. In this case, p^* would be colorful and be returned by step 2. The probability that this event occurs is $Pr[p^* \text{ is colorful}] = \frac{\# \text{ colorful chains}}{\# \text{ colored chains}} = \frac{C!/(C-l)!}{C^l} = \frac{C!}{(C-l)!C^l} = \rho$. To increase the chance of coloring an optimal chain with different colors, the algorithm must be repeated several times.

Let X be a random variable denoting the number of iterations needed for p^* to be colorful for the first time. X follows a geometric distribution, so the probability that this event occurs before the i^{th} iteration is $Pr[X \leq i] = 1 - (1 - \rho)^i$. We want to repeat the algorithm enough times to guarantee a high probability of having a colorful optimal chain. Let $\epsilon \in [0, 1]$, we seek i such that $Pr[X \leq i] = 1 - \epsilon \Leftrightarrow i = \frac{\ln(\epsilon)}{\ln(1-\rho)}$.

Choice of the parameters. The efficiency of the color coding algorithm is determined by C , the number of colors, and t , the number of iterations. As they grow, the probability of finding the optimal chain grows with them, as well as the computation time. For given l and ϵ (the maximal allowed probability to fail), their values are chosen to minimize the theoretical complexity. More precisely, let $\bar{C} \in \mathbb{N}$ be the maximal authorized number of colors (for computational reasons), we choose:

$$(C, t) = \arg \min_{C \in [l, \dots, \bar{C}], t \in \mathbb{N}} \left\{ t \times 2^C \text{ s.t. } t = \frac{\ln(\epsilon)}{\ln(1-\rho)} \right\}$$

The steps 1 and 2 are then repeated t times and the best colorful valid path is returned with a probability of $1 - \epsilon$ to be optimal.

3.2 Local search (LS)

C and t grow as l does to guarantee the wanted probability of success making the color coding suffering from long computation times. In this case, the leading alternative is the use of heuristics. We developed a local search which starts with a random chain and try to improve it with three kinds of movements. They aim at increasing the reduced cost of the chain, keeping it valid, by inserting, removing or exchanging 1, 2 or 3 vertices. The local search stops when it reaches a local minimum, *i.e.*, when no movements improves the current path.

3.3 Integer programming (TSF)

The methods described above run fast but when they find no valid chain of positive reduced cost, no conclusion can be made. If such a chain exists, we want

to find it; if it does not exist, we want to prove it. Thus, an exact method must be implemented to address this issue. We state the problem as a time-staged integer programming formulation (TSF). We define $I = \{0, 1, \dots, l+1\}$, $I^* = \{1, \dots, l+1\}$ and the decision variables: $\forall (uv) \in A'$, $i \in I$

$$r_{uv}^i = \begin{cases} 1 & \text{if } (uv) \text{ is the } i^{\text{th}} \text{ arc in the path} \\ 0 & \text{otherwise} \end{cases}$$

$$rc^* = \max \sum_{i \in I} \sum_{(uv) \in A'} rc_{uv} r_{uv}^i \quad (5)$$

$$\text{s.t.} \quad \sum_{v \in N} r_{sv}^0 = 1 \quad (6)$$

$$\sum_{i \in I \setminus \{0,1\}} \sum_{u \in P} r_{ut}^i = 1 \quad (7)$$

$$\forall v \in V, \forall i \in I : \sum_{\substack{u \in V' : \\ uv \in A'}} r_{uv}^i = \sum_{\substack{u \in V' : \\ uv \in A'}} r_{vu}^{i+1} \quad (8)$$

$$\forall i \in I : \sum_{(uv) \in A'} r_{uv}^i \leq 1 \quad (9)$$

$$\forall v \in V : \sum_{i \in I} \sum_{\substack{u \in V' : \\ uv \in A'}} r_{uv}^i \leq 1 \quad (10)$$

$$\forall uv \in A', \forall i \in I : r_{uv}^i \in \{0, 1\} \quad (11)$$

Constraints (6) and (7) ensure that the first chosen arc leaves the source and that the sink is reached. Constraints (8) are flow constraints guaranteeing that if a vertex is reached by the i^{th} arc then it is left with the $i+1^{\text{th}}$. Constraint (9) imposes that only one arc is taken at stage i and constraint (10) forbids a vertex to be taken more than once. Note that the solution given by (TSF) may contain $l+2$ arcs but two of them are fictive (one from the source and one to the sink).

4 SOLVING THE KEP

The following section describes how the column generation can be further exploited to provide a feasible guaranteed solution to (PF) without the use of a complete branch-and-price method.

4.1 Dual and lagrangian bounds

Define (PF_β) as the linear program composed of equations (1), (2), (4) and the redundant constraint:

$$\sum_{p \in \mathcal{P}} x_p \leq \frac{|V|}{2} \quad (12)$$

This constraint is valid for every KEP. Indeed it is stating that the number of paths is at most half of the number of pairs, which is implied by the fact that a path involves at least two vertices. Let β be the

dual value associated with constraints (12). The dual program of (PF_β) is :

$$(D) \quad \min \sum_{v \in V} \alpha_v + \frac{|V|}{2} \beta \quad (13)$$

$$s.t. \quad \sum_{v \in p} \alpha_v + \beta \geq w_p \quad \forall p \in \mathcal{P} \quad (14)$$

$$\alpha_v \geq 0 \quad \forall v \in V \quad (15)$$

$$\beta \geq 0 \quad (16)$$

Solving optimally the i^{th} pricing problem provides $rc^* = \max_{p \in \mathcal{P}} (w_p - \sum_{v \in p} \alpha_v^i) \geq 0$. Thus, setting β^i to rc^* provides a feasible solution (α^i, β) of (D) , where α^i is the vector of dual values of the i^{th} restricted master problem. By weak duality, any feasible solution (α, β) produces an upper bound of z_{LP}^* so that $\sum_{v \in V} \alpha_v + \beta \frac{|V|}{2} \geq z_{LP}^*$. Furthermore, as the restricted master problem is solved optimally, we know that $\sum_{v \in V} \alpha_v^i = z_{LP}^i$. Thus, $z_{LP}^i + rc^* \frac{|V|}{2} = UB_D^i$ is a dual (upper) bound of z_{LP}^* at iteration i .

When rc^* cannot be computed (recall that the pricing problem is \mathcal{NP} -complete), we compute rc_{LP}^* the value of the linear relaxation of (TSF) . It is an upper bound of rc^* so $z_{LP}^i + rc_{LP}^* \frac{|P|}{2} = UB^i$ is also an upper bound of z_{LP}^* .

Consequently, these bounds are upper bounds of the optimal value of (PF) : $z^* \leq UB_D^i \leq UB^i$. Note that when $rc^* = 0$ the dual bound is equal to the primal value, which proves that (PF_{LP}) is solved optimally.

4.2 Solving of (PF)

Even if z_{LP}^* was not computed within the time limit, the algorithm always returns an upper bound UB_{best} of z^* , as well as a set $\mathcal{P}' \subseteq \mathcal{P}$ of valid paths. When the column generation stops, we solve the integer program (PF) restricted to \mathcal{P}' thus providing a feasible solution \underline{z} for the problem: this is a lower bound. The quality of this feasible can be assessed against the upper bound UB_{best} . As we will see in section 5, it seems unnecessary to implement the branch-and-price since the gap between the upper and the lower bounds is small for most instances, ensuring that the feasible solution is close to the optimal solution.

4.3 Pre-processing

The size of the graph can be reduced to improve the efficiency of all the algorithms. We use a simple filtering method that removes arcs that cannot belong to any valid path, because they are "too far" from other vertices. Let d_{uv} be the shortest distance (in number of arcs) between u and v , $\forall u, v \in V$.

$R_N = \{(uv) \in A : \forall r \in N \ d_{ru} \geq l\}$ is the set of arcs that cannot belong to a valid chain.

$R_V = \{(uv) \in A : d_{vu} \geq k\}$ is the set of arcs that cannot belong to a valid cycle. $R = R_N \cap R_V$ is the set of arcs that cannot belong to any valid path and thus $A = A \setminus R$.

5 NUMERICAL EXPERIMENTS

5.1 Benchmark

Instances were generated with an online data set generator¹. We choose the parameters in order to vary δ the density of the compatibility graph, $|P|$ the number of patient-donor pairs and γ the percentage of altruistic donors in relation to $|P|$. These parameters take the following values: $\delta = 5\%, 10\%$, $\gamma = 1\%, 5\%, 10\%$ and $|P| = 100, 200, 400$, leading to 18 classes of instances, each of them containing 10 instances. The probability of success at finding the optimal chain by color coding was set to 90%.

5.2 Experimental setup

We used the CPLEX Java API (version 12.6) to solve the different linear programs. Experiments were performed on an Intel®Xeon®CPU E5-2440 v2 @ 1.90GHz processor and 32 GB of RAM, with a memory limit of 8 GB of RAM. The time limit was set to 900 seconds. The time limit to solve (TSF) was set to 90 seconds.

5.3 Results

In this section, we use the following notations (average and number out of 10).

- CPU: average time in seconds to compute z_{LP}^*
- #LR: number of instances where z_{LP}^* is computed
- #OPT: (resp. %OPT) number (resp. percentage) of instances where z^* is computed
- GAP: average gap between upper and lower bounds for (PF)

Table 1 presents the results for $k = 3$ and $l = 4$. It shows that the computation time increases as the number of pairs in the instance grows. As expected, this increase is stronger when the graph is denser. The results also demonstrate that the path formulation provides a linear relaxation of high quality. Small gaps between lower and upper bounds ($< 0.3\%$ in average for $|P| = 100, 200, 400$) are reported and the proof of optimality is done for more than 35% of the instances within 2 minutes of computation time (15 seconds in average). This supports the idea that implementing the complete branch-and-price algorithm

¹Available at <http://www.dcs.gla.ac.uk/~jamest/kidney-webapp/#/generator>

δ	γ	$ P $	CPU	# LR	# OPT	GAP
5%	1%	100	0.1	10	10	0%
		200	1.2		4	0.47%
		400	25.7		2	0.44%
		800	784		9	4.46%
	5%	100	0.3	10	7	0.30%
		200	3.6		3	0.20%
		400	78		2	0.28%
		800	-		0	77.01%
	10%	100	0.3	10	8	0.10%
		200	4.9		7	0.13%
		400	115		5	0.05%
		800	-		0	80.34%
10%	1%	100	0.4	10	5	0.57%
		200	6.9		1	0.41%
		400	124		1	0.14%
		800	-		0	83.87%
	5%	100	1.3	10	4	0.55%
		200	17.3		0	0.37%
		400	300		0	0.44%
		800	-		0	85.45%
	10%	100	2.1	10	4	0.38%
		200	31		1	0.17%
		400	587		0	0.2%
		800	-		0	86.43%

Table 1 – $k = 3$ and $l = 4$

may not be necessary. We also tested the algorithm for instances where $|P| = 800$, but z_{LP}^* was calculated for only 9 instances (with $\delta = 5\%$ and $\gamma = 1\%$). For all the other instances the final gap was 81.8% in average, but this high gap is due to the inability to compute the linear relaxation, as the lower gap of 4.46% demonstrates. Figure 6 shows the evolution of the upper bound (UB_{best}) and the lower bound (value of the restricted master problem) as the columns are added for a specific instance with 800 pairs, a density of 5% and 5% of altruistic donors. The upper bound is not calculated when cycles only are added.

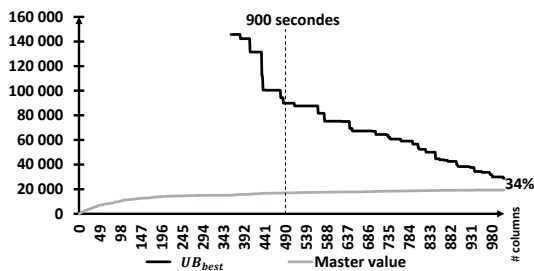


Figure 6 – Evolution of the value of (RMP) and UB_{best} with a time limit of 1 hour for $|P| = 800$, $\delta = 5\%$ and $\gamma = 5\%$

Computation times are higher for $l = 7$ and $l = 10$, but the algorithms behaves similarly. Due to the difficulty of the pricing problem when l is that big, the linear relaxation z_{LP}^* may not be computed within the

	$l = 7$		$l = 10$	
	GAP	% OPT	GAP	% OPT
z_{LP}^*	0.4%	34%	0.4%	44%
no z_{LP}^*	25.9%	-	29.2%	-

Table 2 – Average gap and percentage of solved instances depending on the computation of z_{LP}^*

time limit. Figure 7 outlines the difficulty to compute z_{LP}^* by showing the decreasing number of instances where it is computed and the increasing time to obtain it. We summarized in table 2 the average gap and the percentage of solved instances depending on the computation (or not) of the linear relaxation.

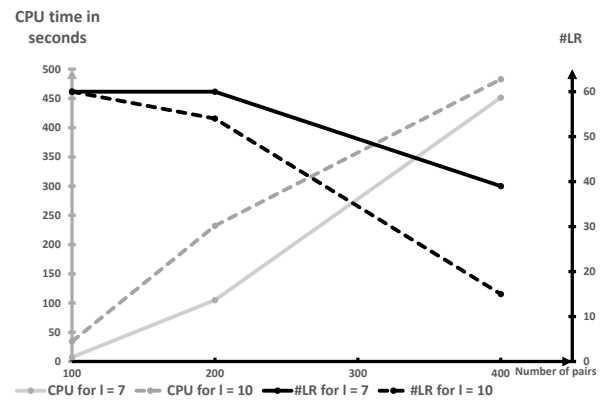


Figure 7 – Number of instances (out of 60) where z_{LP}^* is computed (in black) and average CPU time to compute it (in grey), for $l = 7$ (solid lines) and $l = 10$ (dashed lines).

l	Cycles	Chains	CyPP	LS	CD	(TSF)
4	161	129	0.08	0.05	0.31	1.06
7	263	174		0.25	4.69	3.21
10	354	149		0.64	-	5.65

Table 3 – Number of columns added and average time of pricing algorithms (in seconds)

In average, over all the instances, solving the pricing problems uses more than 90% of the total computation time. Table 3 details the average times, for one iteration, of the different pricing algorithms as l grows. This highlights that the bottleneck remains the computation of each chain pricing problem.

These results show that our algorithm is efficient to find good lower and upper bounds. When the column generation ends before the time limit, the gap is almost (and sometimes is) zero. However, when the time limit is reached before computing the linear relaxation, the gap may deteriorate, because the best dual bound kept all along the column generation algorithm is poor. The improvement of this dual bound – *e.g.*, by updating $\frac{|V|}{2}$ which is a poor upper

bound on the number of paths in a solution or thanks to a stronger integer programming formulation than (*TSF*) – could provide a better final gap even when the linear relaxation is not provided.

6 CONCLUSION

In this paper we studied a column generation algorithm for the kidney exchange problem. It uses the exponential formulation derived from the formulation of (Abraham et al. 2007), which was never strongly exploited up to now to handle altruistic donors.

The results show an encouraging track to explore, as the column generation algorithm provides integer solutions of good (guaranteed) quality in a reasonable amount of time and without the need of branching. However, the solved instances are smaller than the real or expected pools in the USA or in Europe, the target being several hundreds, nay thousands of patients. Further work needs to be done to stabilize the column generation algorithm and thus reduce the number of iterations, which is an important limitation noticed in the results, as well as the need to speed up the pricing step. This includes work on the dual bound described in the previous section as well as development of new algorithms for the chain pricing problem, *e.g.*, derandomization of the color coding, stronger integer formulations, more efficient heuristics or dynamic programming approach with filtering to overcome memory issues.

REFERENCES

- Abraham, D. J., Blum, A. & Sandholm, T. (2007). Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges, *Proceedings of the 8th ACM conference on Electronic commerce*, ACM, pp. 295–304.
- Alon, N., Yuster, R. & Zwick, U. (1995). Color-coding, *Journal of the ACM (JACM)* **42**(4): 844–856.
- Anderson, R., Ashlagi, I., Gamarnik, D. & Roth, A. E. (2015). Finding long chains in kidney exchange using the traveling salesman problem, *Proceedings of the National Academy of Sciences* **112**(3): 663–668.
- Constantino, M., Klimentova, X., Viana, A. & Rais, A. (2013). New insights on integer-programming models for the kidney exchange problem, *European Journal of Operational Research* **231**(1): 57–68.
- COST (2016). http://www.cost.eu/COST_Actions/ca/CA15210.
- Dickerson, J. P., Manlove, D. F., Plaut, B., Sandholm, T. & Trimble, J. (2016). Position-indexed formulations for kidney exchange, *Proceedings of the 2016 ACM Conference on Economics and Computation*, ACM, pp. 25–42.
- Gentry, S. E., Montgomery, R. A. & Segev, D. L. (2011). Kidney paired donation: fundamentals, limitations, and expansions, *American Journal of Kidney Diseases* **57**(1): 144–151.
- Glorie, K. M., van de Klundert, J. J. & Wagelmans, A. P. (2014). Kidney exchange with long chains: An efficient pricing algorithm for clearing barter exchanges with branch-and-price, *Manufacturing & Service Operations Management* **16**(4): 498–512.
- Glorie, K., Wagelmans, A. & van de Klundert, J. (2012). Iterative branch-and-price for large multi-criteria kidney exchange, *Econometric institute report* **11**.
- Held, M. & Karp, R. M. (1962). A dynamic programming approach to sequencing problems, *Journal of the Society for Industrial and Applied Mathematics* **10**(1): 196–210.
- Klimentova, X., Alvelos, F. & Viana, A. (2014). A new branch-and-price approach for the kidney exchange problem, *International Conference on Computational Science and Its Applications*, Springer, pp. 237–252.
- Kwak, J., Kwon, O., Lee, K. S., Kang, C. M., Park, H. Y. & Kim, J. (1999). Exchange-donor program in renal transplantation: a single-center experience, *Transplantation proceedings*, Vol. 31, Elsevier, pp. 344–345.
- Mak-Hau, V. (2017). On the kidney exchange problem: cardinality constrained cycle and chain problems on directed graphs: a survey of integer programming approaches, *Journal of Combinatorial Optimization* **33**(1): 35–59.
- NHS (2017). Annual report on living donor kidney transplantation.
- Plaut, B., Dickerson, J. P. & Sandholm, T. (2016a). Fast optimal clearing of capped-chain barter exchanges, *AAAI Conference on Artificial Intelligence (AAAI)*.
- Plaut, B., Dickerson, J. P. & Sandholm, T. (2016b). Hardness of the pricing problem for chains in barter exchanges, *arXiv preprint arXiv:1606.00117*.
- Rapaport, F. T. (1986). The case for a living emotionally related international kidney donor exchange registry., *Transplantation proceedings*, Vol. 18, pp. 5–9.

CONCEPTION D'UN SYSTEME DE VERIFICATION DE LA DIAGNOSTICABILITE PAR MODEL CHECKING A PARTIR DU MODELE DU SYSTEME

M. CHANKATE, A. PHILIPPOT, V. CARRE-MENETRIER.

Université de Reims Champagne-Ardenne, CReSTIC (EA3804), Reims, France

mehdi.chankate.emsi.marrakech@gmail.com

alexandre.philippot@univ-reims.fr

veronique.carre@univ-reims.fr

P. MARANGE

Université de Lorraine, CRAN, UMR 7039, Campus Sciences, BP 70239, 54506 Vandoeuvre-lès-Nancy, France

CNRS, CRAN, UMR 7039, France

pascale.marange@univ-lorraine.fr

RESUME : Dans ce papier, nous nous intéressons à la vérification de la diagnosticabilité par model-checking des Systèmes à Événements Discrets (SED) modélisés à base d'automates à états finis. Nous proposons une méthode basée sur un algorithme permettant l'analyse de la diagnosticabilité du SED à partir de son modèle de base. Celui-ci constitue la donnée d'entrée de notre méthode. Une condition de la diagnosticabilité étant la présence d'un cycle indéterminé dans un système, nous proposons de rechercher l'ensemble des cycles puis de les analyser. Pour cela, notre méthodologie se découpe en 3 phases. Une phase de modélisation et instanciation de modèles, une phase de recherche de cycles et enfin une phase d'analyse de la non-diagnosticabilité. Nous illustrons notre approche autour d'une étude expérimentale.

MOTS-CLES : Diagnostic, Diagnosticabilité, Automate à états finis, Systèmes à Événements Discrets.

1 INTRODUCTION

Ces dernières années, les travaux autour du diagnostic ont pris de l'essor dans le monde académique et industriel du fait de l'augmentation de la complexité des systèmes, mais aussi des coûts d'interventions de maintenance. Pour améliorer la disponibilité et la fiabilité des installations, il est nécessaire de développer des approches systématiques de diagnostic afin de détecter et d'isoler des fautes. Par ailleurs, il est devenu primordial de mettre en place des approches permettant d'évaluer les performances de ces méthodes en termes de détection, localisation et identification d'une faute dans un délai fini. On parle alors de diagnosticabilité. Parmi les approches de diagnostic, la littérature a montré un intérêt particulier autour des approches à base de modèles pour le diagnostic des Systèmes à Événements Discrets (SEDs) (Cassandra and Lafortune, 1999).

La notion de diagnostiqueur (ou diagnostoser), observateur du système, est apparue dans les travaux de (Sampath et al., 1995). C'est également dans ces travaux que la notion de « diagnosticabilité » est apparue, permettant de garantir le diagnostic d'une faute dans un délai fini. Elle apparaît comme une information primordiale sur le système et sur le travail à réaliser sur l'observateur de fautes. En d'autres termes, il se pose la question de la nécessité de construire un diagnostiqueur si le système n'est pas diagnosticable à certains types de fautes. Différentes évolutions de la diagnosticabilité sont retrouvées dans la littérature notamment en fonction de

la structure du système et de sa modélisation (Contant et al., 2004), (Debouk et al., 2000), (Qiu and Kumar, 2006). Mais quelle que soit cette diagnosticabilité, il est primordial de pouvoir évaluer sa capacité. (Sampath et al.1995) se base sur la construction d'un diagnostiqueur alors que (Jiang et al., 2001) construit un modèle vérificateur (par une méthode dite du twin-plant). D'autres méthodes, par la construction d'un automate vérificateur non déterministe dans (Yoo, et al., 2002) ou par le test du vide dans un automate de Büchi dans (Cassez and Tripakis, 2008), ont été traitées comme des approches centralisées. Dans le cadre des structures décentralisées, il y a très peu d'algorithmes concernant la vérification de diagnosticabilité locale et/ou modulaire. Dans (Pencolé, 2004), un vérificateur local est construit pour chaque sous-système. Dans (Boussif and Ghazel, 2017) une variante du diagnostiqueur classique est mise en place, en séparant explicitement les états normaux de ceux fautifs dans chaque nœud du diagnostiqueur.

La plupart de ces approches mettent en avant des algorithmes de résolution performants dans la mesure où le système reste de taille raisonnable et que la partition de fautes est très faible (souvent une seule faute considérée). En effet, en outre le problème d'explosion combinatoire de l'espace d'états que l'on pourrait retrouver sur le diagnostiqueur du système, la complexité combinatoire de l'algorithme d'évaluation en est un. Les différentes approches font très souvent appel à des modèles intermédiaires pouvant rendre la résolution parfois complexe.

L'approche présentée dans ce papier se base sur une méthode de vérification de la diagnosticabilité à travers l'utilisation du Model-Checking. Cette proposition permet notamment l'analyse de la diagnosticabilité à partir du modèle global du système sans construction de modèles intermédiaires. Nous montrons qu'une telle méthode permet aussi la diagnosticabilité d'un système caractérisé par l'apparition de plusieurs types de fautes. Nous nous intéressons dans ce travail à déterminer une solution dans une architecture centralisée afin de prouver la faisabilité de l'approche. Nos travaux futurs se focaliseront sur l'étude de la diagnosticabilité dans une architecture distribuée.

Dans ce qui suit nous rappelons dans un premier temps quelques notions sur la modélisation des SEDs et les opérations associées, la définition de la diagnosticabilité et les deux principales méthodes de diagnosticabilité de la littérature. La section 3 se focalise sur la méthode de vérification de la diagnosticabilité par Model-Checking que nous proposons. Une étude expérimentale illustre l'algorithme en section 4. Enfin nous concluons le papier avec des perspectives d'amélioration de l'algorithme.

2 PRELIMINAIRES

Un des moyens formels pour étudier le comportement logique des SED est basé sur la théorie des langages et les automates. Dans ces travaux, l'aspect temporel du système est ignoré, seuls l'ordre et l'identité des événements générés par le système sont pris en compte. Par conséquent, un modèle logique où seules les séquences d'événements sont manipulées, est parfaitement adapté à ce type de problème.

2.1 Modélisation des SEDs

Soit un système modélisé par un automate à états finis $G = (Q, \Sigma, \delta, q_0)$ avec Q l'ensemble des états composant le modèle, Σ l'ensemble fini des événements avec deux sous-ensembles Σ_o : l'ensemble des événements observables, et Σ_{uo} : celui des événements non observables tel que $\Sigma = \Sigma_o \cup \Sigma_{uo}$. δ est la relation de transition $(q, \sigma, q') \in Q \times \Sigma \times Q$ avec $q' \in \delta(q, \sigma)$ et enfin $q_0 \in Q$ est l'état initial.

On peut dénoter par $\Sigma_f \subseteq \Sigma_{uo}$ l'ensemble des événements fautifs à diagnostiquer. L'ensemble des fautes Σ_f forme une partition de m sous-ensembles de fautes $\Sigma_f = \Sigma_{f1} \cup \Sigma_{f2} \cup \dots \cup \Sigma_{fm}$ notée $\Pi_f = \{\Sigma_{f1}, \dots, \Sigma_{fm}\}$, avec $m \geq 1$.

Un diagnostiqueur étant un modèle des événements observables du système, il convient alors de définir la fonction de projection sur un modèle. On définit donc la fonction de projection sur l'ensemble des événements observables par $P: \Sigma^* \rightarrow \Sigma_o^*$, i.e., P retire tous les événements non observables d'une séquence d'événements donnés. De manière générale, $P(\sigma) = \sigma$ si

$\sigma \in \Sigma_o$ et $P(\sigma) = \epsilon$ si $\sigma \in \Sigma_{uo}$; $P(s.\sigma) = P(s)P(\sigma)$ avec $s \in \Sigma^*$ et $\sigma \in \Sigma$. Par déploiement, la fonction de projection aux langages est $P(L) = \{P(s)/s \in L\}$ avec L le langage régulier clos par préfixe $L \subseteq \Sigma^*$ produit par le modèle G . La figure 1 donne l'effet de la projection P avec : $\Sigma = \{a, b, c, uo, f1\}$, $\Sigma_o = \{a, b, c\}$, $\Sigma_{uo} = \{uo, f1\}$, $\Sigma_f = \{f1\}$

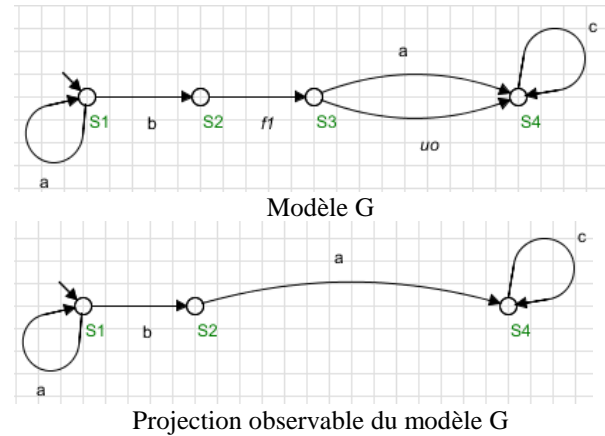


Figure 1. Exemple de la projection observable du modèle G

Par ailleurs, une fonction de composition synchrone de deux modèles G_1 et G_2 est définie par $G_1 \parallel G_2 = (Q_1 \times Q_2, \Sigma_1 \cup \Sigma_2, \delta, q_1 \times q_2)$. Elle permet notamment d'obtenir l'évolution du modèle du système global.

2.2 Définition de la Diagnosticabilité

La notion de diagnosticabilité d'un système a été proposée sur la base de la construction d'un diagnostiqueur global dans (Sampath, 1995). Cette vérification consiste à évaluer les deux conditions nécessaires suivantes pour que le diagnostiqueur soit diagnosticable :

C1. Il existe au moins un état du diagnostiqueur pour lequel le diagnostiqueur décide avec certitude l'occurrence d'une faute appartenant à une partition Π_{fi} .

C2. Il ne doit pas y avoir de cycles dits indéterminés pour lesquels le diagnostiqueur est incapable de décider avec certitude l'occurrence d'une faute appartenant à une partition Π_{fi} .

Ainsi, il est possible de savoir si un système est diagnosticable avec certitude, incertitude ou ambiguïté. Ci-après nous rappelons les notions relatives à une faute (Fi-certain, Fi-incertain, Fi-ambigu).

Définition 1. (Fi-certain dans G)

Un état q est Fi-certain si toute trace de q_0 à q contient obligatoirement la faute Fi.

Définition 2. (Fi-incertain dans G)

Un état du système est Fi-incertain s'il y a deux traces s_1 et s_2 partant de q_0 jusqu'à respectivement q_1 et q_2 et ayant la même projection observable tel que s_1 contienne Fi alors que s_2 ne contienne pas Fi et que dans le système

initial G , les traces s_1 et s_2 mènent vers deux états différents (q_1, q_2).

Définition 3. (Fi-ambigu dans G)

Un état du système est Fi-ambigu s'il y a deux traces s_1 et s_2 partant de q_0 jusqu'à q_1 et ayant la même projection observable tel que s_1 contienne F_i alors que s_2 ne contienne pas F_i et que dans le système initial G , les traces s_1 et s_2 mènent vers un même état (q_1).

De plus, il convient d'admettre les deux hypothèses suivantes :

- L'automate G est vivant c'est à dire qu'à partir de chaque état q de G , il existe au moins une transition franchissable ;
- Il n'y a pas de cycle contenant uniquement des événements non-observables.

L'absence de cycles d'événements inobservables permet de garantir que l'on dispose d'observations régulières pour observer les effets d'une faute. $L(G)$ est donc diagnosticable par rapport à la projection $P : \Sigma \rightarrow \Sigma_o$ et par rapport à la partition Π_f de Σ_f , si la condition suivante est satisfaite (éq. 1):

$$(\forall i \in \Pi_f)(\exists ni \in \mathbb{N})(\forall s \in \Psi(\Sigma_{fi}))(\forall t \in L/s) : [|t| \geq ni \\ \Rightarrow \omega \in P^{-1}L(G) [P(st)] \Rightarrow \Sigma_{fi} \in \omega] \quad (1)$$

Avec :

- $\Psi(\Sigma_{fi})$ l'ensemble des traces se terminant par une faute de Σ_{fi} ;
- L/s l'ensemble des séquences suffixes de s dans $L(G)$;
- $|t|$ la longueur de la séquence t ;
- $P^{-1}L(G) [P(u)]$ toutes les traces v dans $L(G)$ telles que $P(v)=u$.

$\Psi(\Sigma_f)$ est l'ensemble des séquences finies d'événements qui se terminent par un événement fautif (de Σ_{fi}). Cette définition signifie qu'un langage L est diagnosticable si et seulement si, pour toute séquence 's' se terminant par un événement fautif et toute suite d'événements 't' de 's' suffisamment longue ($|t| \geq ni$), alors toute autre séquence d'événements ' ω ' ayant la même projection observable que 's.t ($P(s.t) = P(\omega)$)', contient nécessairement un événement fautif de Σ_f .

Propriété 1. (Cycle dans G)

Soit CL l'ensemble des cycles cl possibles dans un système G tel que : ($q \in Q, \sigma \in \Sigma$)

$$cl_1 = (q_0, \sigma_0, q_1, \sigma_1, \dots, q_n, \sigma_n, q_m) \text{ avec } m[0, n] \\ cl_2 = (q_0, \sigma'_0, q'_1, \sigma'_1, \dots, q'_p, \sigma'_p, q'_o) \text{ avec } o \in [0, p]$$

Pour vérifier la diagnosticabilité dans le modèle G du système il convient de vérifier la non-présence de deux cas de figure :

i. Le système G est non diagnosticable s'il existe un cycle cl_1 dans G avec F_i -certain, un autre cycle cl_2 ne contenant pas la faute F_i et ayant la projection observable (éq. 2).

$$(\exists cl_1 \subseteq CL / \exists (q_j, \sigma_j) \in L : \sigma_j \in F_i) (\exists cl_2 \subseteq CL / \forall (q_k, \sigma_k) \in L : \sigma_k \in \Sigma_o) / P(cl_1) = P(cl_2) \text{ avec } \{j \in [0, n], k \in [0, p]\} \quad (\text{éq. 2}).$$

ii. Le système G est non diagnosticable s'il existe un cycle cl_1 dans G avec F_i -certain, un autre cycle cl_2 avec F_j -certain ($F_i \neq F_j$) et cl_1 et cl_2 partageant la même projection observable (éq. 3)

$$(\exists cl_1 \subseteq CL / \exists (q_j, \sigma_j) \in L : \sigma_j \in F_i) (\exists cl_2 \subseteq CL / \exists (q_k, \sigma_k) \in L : \sigma_k \in F_{s \neq i}) / P(cl_1) = P(cl_2) \text{ avec } \{j \in [0, n], k \in [0, p]\} \quad (\text{éq. 3}).$$

2.3 Analyse de la diagnosticabilité selon Sampath et Jiang

(Sampath et al., 1995) et (Jiang et al., 2001) ont proposé chacun un algorithme de vérification de la diagnosticabilité d'un système G (fig. 2), sur le modèle du diagnostiqueur.

Dans la première approche, l'auteur effectue une projection du système G pour obtenir un modèle observateur G' déterministe. A partir de là, et en associant un label à chaque état (N : normal, F_i : Fautif i), un diagnostoser G_d est obtenu. L'algorithme de recherche de cycles indéterminés est appliqué sur G_d . L'algorithme de vérification est exponentiel par rapport au nombre d'états du système G et doublement exponentiel par rapport au nombre de types de fautes. Le risque d'explosion combinatoire peut apparaître dans cette étape.

Dans la seconde approche la construction du modèle intermédiaire « vérificateur » de (Jiang et al., 2001) est obtenue par composition synchrone du modèle observateur labellisé G_o avec lui-même « twin-plant » ($G_v = G_o || G_o$). Cette approche présente donc un risque d'explosion de l'espace d'états dès la construction du vérificateur. L'algorithme de vérification consiste alors à chercher l'absence ou non de cycle cl avec ($F_i \neq F_j$) dans G_v . Il est donc polynomial par rapport au nombre d'états de G et exponentiel par rapport au nombre de types de fautes.

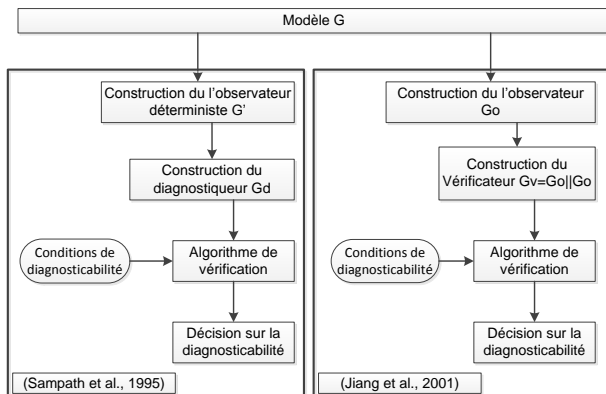


Figure 2. Etapes des approches de Sampath et Jiang

L'approche proposée consiste à éviter la construction de modèles intermédiaires et par ailleurs les compositions d'automates en utilisant des instances de classe du modèle G . La première instantiation G_0 permettra notamment de rechercher un cycle indéterminé et de la comparer son observation à un autre cycle indéterminé issu de la seconde instantiation G_1 . Ainsi, sans composition parallèle, le Model-Checker permettra de vérifier par équivalence en termes de parcours de graphe les conditions de diagnosticabilité (Fig. 3). L'approche est détaillée dans la partie 3.2 mais auparavant il convient de rappeler quelques principes sur ce dernier.

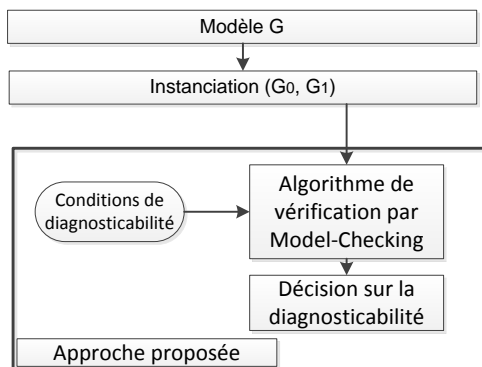


Figure 3. Structure de notre approche

2.4 Principe du Model-Checking

Le Model-Checking est une méthode algorithmique formelle généralement mis en opposition avec le test et la simulation (Schnoebelen *et al.*, 1999). Au même titre que l'« Automated Theorem Proving » basée sur le raisonnement automatisé et qui tente de prouver un ensemble limité de théorèmes mathématiques par des programmes informatiques, le Model-Checking permet de vérifier formellement qu'un système à états fini modélisé satisfait une spécification logique. Ces méthodes peuvent être automatisées et parcourir exhaustivement tous les comportements mais impliquent que le modèle et la spécification soient correctement exprimés. Le Model-Checking nécessite l'utilisation de :

- Un modèle M du système,
- Une spécification φ des propriétés à vérifier,

- Un algorithme vérifiant automatiquement si le modèle satisfait la formule noté $M \models \varphi$.

Le modèle considéré peut être exprimé au moyen d'un graphe orienté de type automates à états (Kripke, 1963). La spécification peut se formuler en logique temporelle (LTL : Logique Temporelle Linéaire, CTL : Computation Tree Logic ou structure de Kripke) (Clarke, Emerson and Sistler, 1986). LTL et CTL sont des logiques modales où la notion de vérité dépend de l'évolution du monde, c'est-à-dire qu'une proposition peut-être, à un moment, fausse puis, plus tard, devenir vraie. Elles sont définies sur un ensemble de propositions atomiques et combinées par un certain nombre de connecteurs logiques : non, ou, et, implications, vrai et faux (\neg , \vee , \wedge , \Rightarrow , \Leftrightarrow , *true*, *false*), ainsi que d'autres opérateurs temporels appelés « modalités » ou « quantificateurs ». L'ensemble permettant d'exprimer des formules exprimant l'atteignabilité, la sécurité ou la vivacité d'un système (E: possibly exists a path, A: Inevitably for all paths, F: some state in a path, and G: all states in a path).

3 VERIFICATION DE LA DIAGNOSTICABILITE PAR MODEL CHECKING

Afin d'éviter la construction de modèle intermédiaire, nous proposons d'exploiter la recherche de propriétés sur le modèle G par Model-Checking.

Une condition de la diagnosticabilité définit dans (Sampath, 1995) étant la présence d'un cycle indéterminé à base des événements non observables dans le système G , nous proposons de rechercher l'ensemble des cycles (avec ou sans faute), le choix des cycles revient au Model-Checker, car pour l'instant aucune condition de choix n'est imposée dans la recherche des cycles. Dans les travaux futurs nous prévoyons un choix spécifique à cette recherche notamment sur la seconde instantiation G_1 . Une fois les deux cycles trouvés, une comparaison est mise en place, il faut alors analyser à la fois si le cycle possède une faute dans sa séquence, mais aussi si celui-ci est distinguable des autres avec certitude. La Fig. 4 illustre les différentes étapes de la méthodologie proposée :

- Modélisation et Instantiation du système
- Recherche de deux cycles
- Analyse de la non-diagnosticabilité

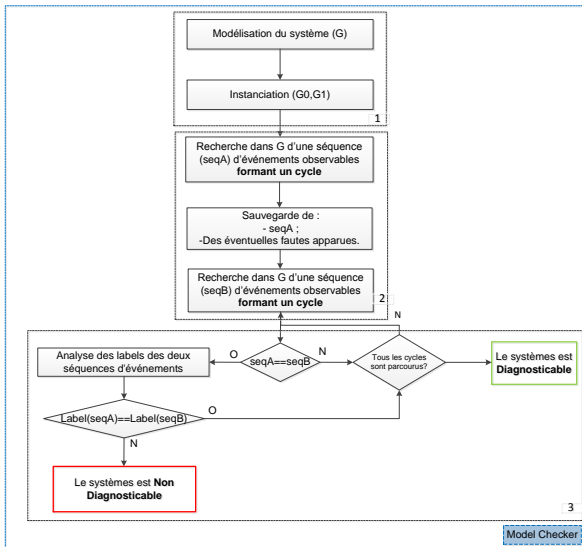


Figure 4. Logigramme de la solution proposée.

3.1 Modélisation et Instanciation

L'étape de modélisation consiste à donner une interprétation abstraite du système. Au même titre que (Sampath, 1995) ou (Jiang et al., 2001), ce modèle G est représenté comme à automate à états finis comme définit dans la section 2. Cependant, afin de décomposer l'information contenue dans une transition, nous allons interpréter chaque événement par un ensemble de labels $\{P[i], E[j], F[k]\}$ avec :

- $P[i]$: mémorisation de l'état atteint i ,
- $E[j]$: mémorisation de l'événement et de son observation ($j=0$: événement non-observable, $j>0$: événement observable),
- $F[k]$: mémorisation du type de faute ($k=0$: aucune faute, $k>0$: faute f_k)

Afin de chercher différents cycles et de pouvoir les comparer par la suite, nous proposons d'instancier le modèle G deux fois (G_0 et G_1). Ces deux automates sont donc identiques mais n'évolueront pas en même temps dans notre vérification.

3.2 Recherche d'un cycle dans G_0 et G_1

Le model-checker va commencer par parcourir le modèle G_0 en enregistrant les séquences d'événements à partir de son état initial. C'est cet enregistrement qui va lui permettre de vérifier à chaque occurrence si un cycle d'événements apparaît ou non. Lorsque celui-ci trouve une séquence observable (seqA) avec un cycle cl_1 , une sauvegarde de ses informations est effectuée à travers la mémorisation des états visités et des fautes apparues. La seconde recherche peut alors débuter, mais cette fois-ci sur le second modèle G_1 . De la même façon, cette séquence observable (seqB) avec un cycle cl_2 est sauvegardée et l'ensemble peut être analysé.

3.3 Analyse de la non-diagnosticabilité

L'analyse commence par vérifier si les deux séquences sont identiques en termes d'événements observables. Si les 2 séquences ne disposent pas de la même observation $P(cl_1) \neq P(cl_2)$, il convient alors au model-checker de rechercher l'existence dans G_1 d'un autre cycle qui posséderait la même observation. S'il n'en trouve pas, alors le système sera considéré comme diagnostiquable sous les hypothèses de la section 2.2.

Dans le cas où les 2 séquences observables sont identiques $P(cl_1) = P(cl_2)$, les deux labels doivent alors être analysés. Si les deux labels sont égaux $(Label(seqA)) = (Label(seqB))$, cela signifie qu'on est en train d'analyser la même trace/chemin et qu'il faut rechercher alors une autre séquence observable avec cycle. Par contre, si les deux labels sont différents $(Label(seqA)) \neq (Label(seqB))$, alors cela signifie que deux séquences disposent de la même observation mais possèdent pour l'une une faute et pour l'autre non ou une faute différente. Le model-checker répondra donc par la non-diagnosticabilité du système.

Cette phase d'analyse se base sur la mémorisation des fautes (ou non) du cycle cl_1 et des fautes (ou non) du cycle cl_2 . Pour cela on effectue deux opérations : le produit ($Pdct_i$) et la somme (Smc_i) de toutes les fautes apparues dans les cycles tel que :

$$Pdct_1 = \prod_{i=1}^m (F_i) / (F_i \in Cycle1)$$

$$Smc_1 = \sum_{i=1}^m (F_i) / (F_i \in Cycle1)$$

$$Pdct_2 = \prod_{i=1}^m (F_i) / (F_i \in Cycle2)$$

$$Smc_2 = \sum_{i=1}^m (F_i) / (F_i \in Cycle2)$$

Les produits ($Pdct_1, Pdct_2$) et les sommes (Smc_1, Smc_2) jouent le rôle de décision sur le label de faute. Comparer Smc_1 à Smc_2 et $Pdct_1$ à $Pdct_2$ pour arriver à un des deux cas possibles :

- Si $(Pdct_1 = Pdct_2)$ **ET** $(Smc_1 = Smc_2)$ alors les labels de faute sont les mêmes entre les deux cycles trouvés et par conséquent les deux cycles ont la même ou les mêmes partitions de fautes,
- Si $(Pdct_1 \neq Pdct_2)$ **OU** $(Smc_1 \neq Smc_2)$ alors les deux cycles ont deux partitions de fautes différentes et donc le système G est Non-diagnosticable.

Cette recherche de 2 cycles dans 2 instances du modèle G n'est pas une composition synchrone à l'image de (Jiang et al., 2001) mais implique au pire le parcours

d'autant d'états pour la recherche de cycles par le model-checker.

4 EVALUATION EXPERIMENTALE

Pour évaluer et comparer notre approche, nous prenons l'exemple illustré par la figure 5 tiré des travaux de (Sampath et al., 1995). Les sous-ensembles des événements sont $\Sigma_o = \{a, b, c\}$, $\Sigma_{uo} = \{u0, f_1\}$, $\Sigma_f = \{f_1\}$.

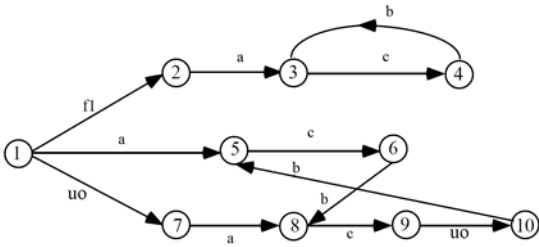


Figure 5. Modèle G de l'exemple

4.1 Diagnosticabilité de (Sampath et al., 1995)

Pour construire le diagnostiqueur de l'exemple, nous appliquons la projection observable pour obtenir l'observateur G' (fig. 6), puis nous rendons déterministe et nous labélisons les états par (N, F) pour obtenir le diagnostiqueur de la figure 7.

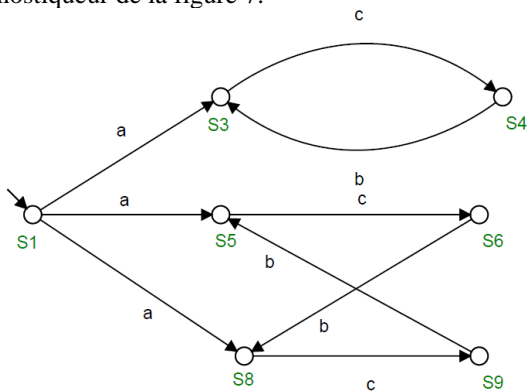


Figure 6. Observateur G' de G

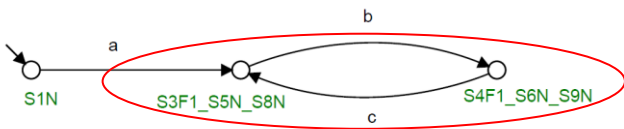


Figure 7. Diagnostiqueur G_d de G

D'après l'analyse de la figure 7, nous remarquons qu'il existe un cycle caractérisé par l'incertitude à cause de la présence d'une boucle avec l'étiquette (N et F). La faute F1 est indéterminée et par conséquent le système G est non diagnosticable.

4.2 Diagnosticabilité de (Jiang et al., 2001)

La méthode de (Jiang et al., 2001) consiste à construire dans un premier temps le pseudo-diagnostiqueur G_o

labélisé (fig. 8). Dans un second temps, le vérificateur est obtenu en utilisant le « twin plant » du pseudo-diagnostiqueur ($G_v = G_o || G_o$) illustré en figure 9.

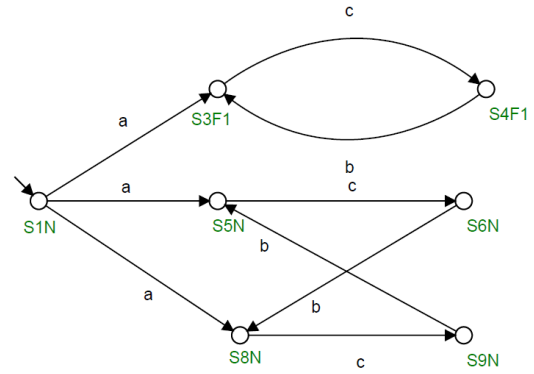


Figure 8. Pseudo-diagnostiqueur G_o de G

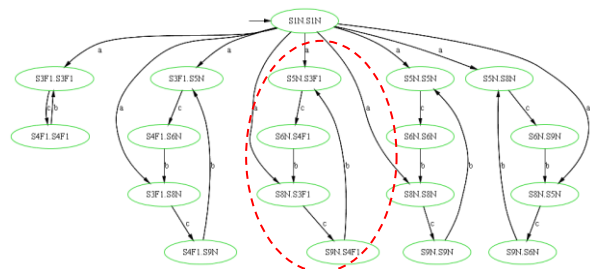


Figure 9. Vérificateur G_v de G

D'après le vérificateur G_v, nous constatons la présence d'un cycle incertain « offending cycle », et par conséquent le modèle G est bien non diagnosticable.

4.3 Diagnosticabilité par Model-Checking

Pour illustrer notre approche par Model-Checking, nous avons utilisé le model-checker UppAal (Behrmann et al., 2006). Pour la première étape de modélisation, il convient d'interpréter les événements du modèle initial par un ensemble de labels {P[i], E[j], F[k]}. Il en résulte la table 1. Le modèle est ensuite obtenu sous la forme de la figure 10. Celui-ci est alors instancié 2 fois dans le model-checker (G0 et G1).

Type d'événement	Evénement	Indice
Observables	a	E=1
	b	E=2
	c	E=3
Non Observables	u0	E=0
	f1	E=0 et F=1

Table 1. Affectation événement-indice

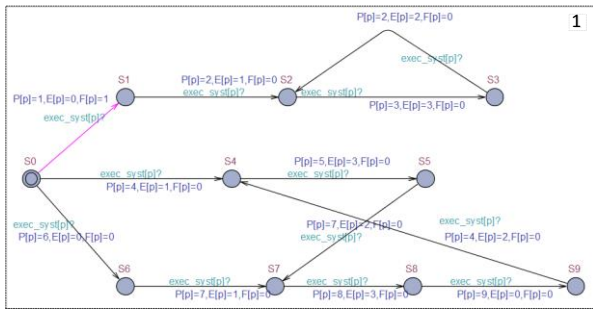


Figure 10. Modélisation sous Uppaal du modèle G

Afin d'exécuter automatiquement la deuxième étape de recherche de cycles de l'algorithme proposé dans la figure 4, un modèle de synchronisation est nécessaire (fig. 11). Cet automate commence par la recherche d'un premier cycle dans G_0 par une boucle :

- L'évolution de l'automate G_0 par synchronisation (*exec_var[]*),
- Appel d'une fonction de recherche de cycle avec mémorisation de la séquence (*evol()*).

Une fois le cycle trouvé (*cycle == 1*) et sauvegardé (*save_init()*), l'automate passe à la recherche d'un second cycle sur G_1 par une deuxième boucle sur le même principe de la première (zone 2 de la figure 11). La troisième étape d'analyse (zone 3 de la figure 11) consiste à appeler une fonction de calcul des équations 4 et 5 (*cal_pdt_somme()*) qui permettra ensuite de comparer l'observation des cycles et l'analyse des labels de fautes par une fonction (*test_NDiag()*). Si la variable *NDiag==1*, alors cela signifie que nous avons deux cycles avec la même projection observable mais qu'ils n'ont pas le même label de faute.

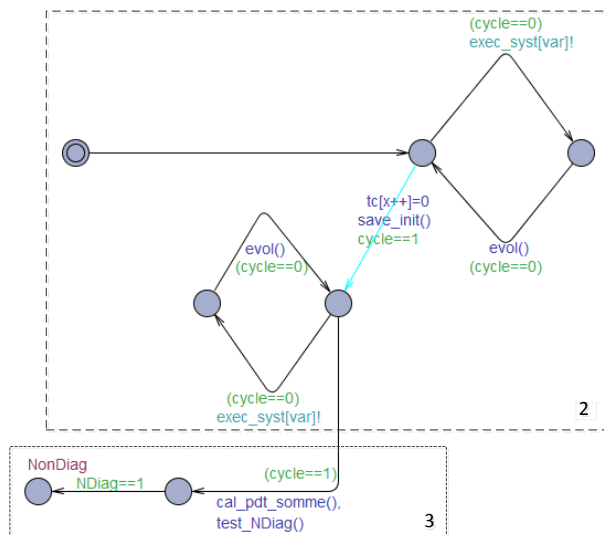


Figure 11. Modèle automate à états de synchronisation et contrôle (Synch)

Pour effectuer la vérification, nous avons implanté nos modèles dans le model-checker UppAal et nous avons vérifié s'il existait un chemin permettant d'atteindre la localité *Nondiag* du modèle de la figure 11. Cette propriété est notée dans la logique CTL : $EF \text{ Nondiag}$.

Dans l'exemple étudié, le model-checker nous retourne une trace vérifiant la possibilité d'atteindre la localité *NonDiag*. Cette trace nous permet de connaître les 2 cycles qui ont la même séquence d'événements observables mais qui ne remontent pas la même détection de fautes. Le modèle G est donc bien non-diagnosticable, ce qui est cohérent avec les deux autres approches.

4.4 Discussion

Concernant l'analyse de la diagnosticabilité, nous remarquons que les deux approches nécessitent la construction d'un modèle intermédiaire alors que l'approche proposée dans ce papier ne l'exige pas. La présence de plusieurs types de fautes rend la vérification de la diagnosticabilité plus complexe pour les approches classiques, par contre notre approche n'est pas influencée. L'originalité de notre travail est d'utiliser le model-checking comme un outil pour vérifier l'atteignabilité, et de travailler sur le modèle G directement.

Dans les approches de Sampath et Jiang, la complexité se trouve dans la modélisation du modèle intermédiaire et peut engendrer une explosion combinatoire. Dans notre cas, cette complexité ne se trouve pas dans la modélisation mais dans la recherche des cycles dans G_0 et G_1 . La complexité de recherche est liée au nombre de cycles dans G . Actuellement nous n'avons pas évalué cette complexité et cela fera l'objet de travaux futurs.

5 CONCLUSION

Cet article présente une formulation de la vérification de la diagnosticabilité dans le cadre du model-checking. Nous avons proposé un algorithme pour vérifier la diagnosticabilité directement à partir du modèle du système. Cet algorithme permet notamment de tenir compte de plusieurs fautes.

Ce début de travail de recherche a permis une première comparaison avec deux approches reconnues dans la littérature. Cette étude devra être poursuivie pour montrer la possibilité de passer sur des systèmes de plus grande taille et pour comparer nos approches en termes de temps de calcul et d'espace d'états explorés.

Une amélioration à court terme est de diminuer l'espace d'états explorés de notre algorithme en vérifiant l'équivalence des séquences d'événements dès le début de la recherche du second cycle. Ainsi, le modèle ne parcourra que les séquences équivalentes en termes d'événements observables. A plus long terme, nous aimerions utiliser les contre-exemples remontés par le model-checker pour proposer des ajouts d'informations sur le système et ainsi proposer une aide à la conception du diagnostiqueur.

REFERENCES

- Behrmann, G., David, A., Larsen, K.G., Hakansson, J., Petterson, P., Yi, W., Hendriks, M.: Uppaal 4.0. In: Third International Conference on the Quantitative Evaluation of Systems - (QEST'06). pp. 125–126 (Sept 2006)
- Boussif, A., and Ghazel, M. (2017). A diagnoser-based approach for intermittent fault diagnosis of discrete-event systems. In 2017 American Control Conference (ACC), pp. 3860–3867.
- Cassandra C.G. and Lafortune S. (1999). Introduction to discrete event systems. Kluwer Academic Publishers, Dordrecht, 1999.
- Cassez, F., and Tripakis, S. (2008). Diagnostic des systèmes temporisés. 145--176.
- Contant, O., Lafortune, S., and Teneketzis, D. (2004). Diagnosis of modular discrete event systems1. IFAC Proc. Vol. 37, 327–332.
- Debouk, R., Lafortune, S., and Teneketzis, D. (2000). Coordinated Decentralized Protocols for Failure Diagnosis of Discrete Event Systems. Discrete Event Dyn. Syst. 10, 33–86.
- Jiang, S., Huang, Z., Chandra, V., and Kumar, R. (2001). A polynomial algorithm for testing diagnosability of discrete-event systems. IEEE Trans. Autom. Control 46, 1318–1321.
- Pencolé, Y. (2004). Diagnosability Analysis of Distributed Discrete Event Systems. In Proceedings of the 16th European Conference on Artificial Intelligence, (Amsterdam, The Netherlands, The Netherlands: IOS Press), pp. 38–42.
- Qiu, W., and Kumar, R. (2006). Decentralized failure diagnosis of discrete event systems. IEEE Trans. Syst. Man Cybern. - Part Syst. Hum. 36, 384–395.
- Sampath, M., Sengupta, R., Lafortune, S., Sinnamohideen, K., and Teneketzis, D. (1995). Diagnosability of discrete-event systems. IEEE Trans. Autom. Control 40, 1555–1575.
- T. Yoo and S. Lafortune. Polynomial-time verification of diagnosability of partially-observed discrete-event systems. IEEE Trans. on Automatic Control, 47(9):1491–1495, 2002.

CONTRIBUTION DES ARCHITECTURES DE CONTROLE HOLONIQUES A L'INDUSTRIE 4.0

Olivier CARDIN

LUNAM Université, Université de Nantes, LS2N UMR
CNRS 6004, 2 avenue du Pr Jean Rouxel, 44470
Carquefou, France
olivier.cardin@ls2n.fr

William DERIGENT

Research Centre for Automatic Control of Nancy,
CNRS UMR 7039, Campus Sciences, BP 70239,
Vandoeuvre-lès-Nancy Cedex, 54506, France
william.derigent@univ-lorraine.fr

Damien TRENTESAUX

LAMIH UMR CNRS 8201, University of Valenciennes
and Hainaut-Cambrésis, UVHC Valenciennes, France
damien.trentesaux@univ-valenciennes.fr

RESUME : *La flexibilité revendiquée par les systèmes de production de prochaine génération induit une profonde modification du comportement et du cœur même des systèmes de contrôle. L'hyper-connectivité et les capacités de gestion des données ciblées par le paradigme Industrie 4.0 permettent l'émergence de systèmes de contrôle plus flexibles et réactifs, basés sur la coopération d'entités autonomes et connectées dans le processus de décision. Au cours des 20 dernières années, le paradigme holonique est devenu le paradigme central de ces évolutions et a lui-même évolué. Cette contribution vise à souligner les évolutions conceptuelles dans l'application du paradigme holonique dans les architectures de contrôle des systèmes de production et à mettre en évidence à la fois l'intérêt que revêtent ces architectures pour l'application du paradigme Industrie 4.0 mais également les tendances de recherche actuelles dans ce domaine.*

MOTS-CLES : *Systèmes de production, Pilotage industriel, Holonic manufacturing systems, Architecture de contrôle, Industrie 4.0.*

1 INTRODUCTION

L'évolution des systèmes industriels vers ce que l'on appelle Industrie 4.0 est principalement basée sur le développement de ressources, produits et acteurs hautement connectés tout au long des processus industriels. Cette connectivité accrue permet de générer un flux constant d'informations diffusé par et disponible pour tous les acteurs. Elle ouvre de nombreuses opportunités pour améliorer l'efficacité des systèmes industriels, dans de nombreuses phases du cycle de vie des produits. On trouvera par exemple des bénéfices directs en :

- *Logistique* : les informations transmises immédiatement à l'ensemble de la chaîne logistique doit permettre une synchronisation constante entre les acteurs et faciliter l'adaptation aux changements (Barreto et al., 2017) ;
- *Production* : l'utilisation d'objets communicants comme éléments de la chaîne de production permet d'envisager le recueil en ligne d'informations relatifs à l'état du système productif. Il est alors possible d'ajuster le comportement de l'atelier en temps réel en raison de conditions ou de changements anormaux (Mrugalska and Wyrwicka, 2017) ;

- *Maintenance* : ces informations permettent d'envisager des capacités novatrices de surveillance et de gestion des connaissances basées sur la simulation (Bokrantz et al., 2017).

Cette transversalité horizontale tout au long du cycle de vie est très bien exprimée dans le modèle RAMI4.0 (*Reference Architecture Model for Industry 4.0*) illustré Figure 1 (Zezulka et al., 2016). Dans ce framework à trois dimensions, les aspects multi-échelle de l'approche (de l'atelier à la chaîne logistique par exemple) ainsi que les différents points de vue sur les gains potentiels de l'application d'Industrie 4.0 montrent une intégration dynamique verticale des données beaucoup plus intense qu'actuellement et illustre les nombreux défis restants encore à relever (Barbosa et al., 2016).

Industrie 4.0 suppose ainsi une adaptabilité temps-réel de l'ensemble de l'entreprise aux changements de conditions externes (production, relation clients/fournisseurs, etc.). Pour permettre une utilisation efficace des données disponibles pour la conduite de ces processus industriels orientés Industrie 4.0, il devient alors nécessaire de définir des systèmes informatiques d'entreprises dotés d'architectures de contrôle connectées, interopérables, flexibles et réactives. L'une des approches (orientée bottom-up) possibles est alors d'utiliser et d'intégrer verticalement les architectures de contrôle holoniques (Holo-

nic Control Architectures – HCA) développées initialement pour le contrôle de la production.

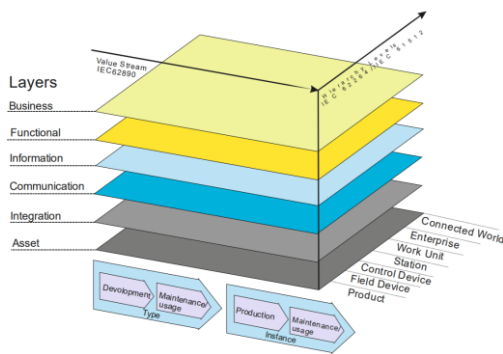


Figure 1. Modèle RAMI4.0 (Zezulka et al., 2016)

Au cours des 20 dernières années, les HCA ont été largement étudiées et développées. Elles sont devenues de plus en plus efficaces vis-à-vis de ces caractéristiques et leur utilisation au niveau industriel commence à se répandre (Blanc et al., 2008; Le Mortellec et al., 2013).

Cet article a pour objectif de donner un aperçu de l'évolution des HCA au cours des deux dernières décennies et de la manière dont elles peuvent contribuer à la diffusion des technologies de l'Industrie 4.0. Les perspectives de recherche scientifique actuelles sont présentées dans une dernière section.

2 FONDAMENTAUX DES ARCHITECTURES DE CONTRÔLE HOLONIQUES

Une HCA est une architecture composée de holons, appelée holarchie. Un holon est une entité décisionnelle (avec entrées et sorties) potentiellement composée récursivement d'un ensemble de holons de sous-niveau. Un holon fait en même temps partie d'une organisation plus large composée de holons de niveau supérieur (cette dualité est appelée effet Janus (Koestler, 1978)). Il est important de noter qu'un holon est aussi composé d'une partie physique associée à une partie numérique (qui peut être modélisée en tant qu'agent numérique, avatar, jumeau numérique). Enfin, les holons sont des entités capables de communiquer avec leurs « pairs » ou d'autres entités externes à l'holarchie. La prise de décision, élément clé d'une architecture holonique, est un énorme sujet de recherche en soi. Selon notre expérience dans le cadre applicatif du génie industriel, nous suggérons que décider est l'activité qui consiste à réduire un ensemble de possibilités. Une notion proche est le concept de « degré de liberté » qu'une décision réduit. De ce point de vue, les activités décisionnelles classiques peuvent être identifiées: choix (ensemble réduit à un singleton, par exemple un holon choisit une vitesse de convoyage) ou classement (ordre d'intégration dans l'ensemble : par exemple un holon décide d'une suite de ressources à visiter). Une décision se prend également soit sur un domaine discret (par exemple, sur un ensemble de produits à réaliser) soit sur un domaine continu (par exemple, sur un domaine de vitesse de rotation ou de consommation d'énergie admissible).

Étant donné l'importance des aspects physiques et matériels au sein d'HCA, décider n'est « qu'une » activité

parmi un ensemble plus large définissant un processus que nous avons appelé processus décisionnel. Prolongeant les idées de base de (Simon, 1996), un processus décisionnel se compose de plusieurs activités : suivi, déclenchement, conception de décisions possibles, évaluation a priori des décisions, décision, application et évaluation a posteriori des décisions. Étant donné l'aspect récursif d'une holarchie, ce processus est également récursif et peut être implémenté dans différentes couches de holons responsables de leur mise en œuvre. Par exemple, un holon en charge d'un contrôle qualité doit décider ou non d'une alerte à un opérateur de supervision. Cette activité est en fait une activité de déclenchement par dépassement d'un seuil d'alerte de non qualité au sein d'un processus décisionnel qui vise à contrôler le taux de non qualité. Cette activité, si on l'analyse plus finement, peut elle-même être décomposée en un processus décisionnel géré par des holons de niveau inférieur visant à décider, à leur niveau, par une stratégie d'apprentissage, du meilleur seuil de déclenchement afin d'éviter soit une sur-réaction en sollicitant trop fréquemment l'opérateur soit une sous-réaction qui conduit à un taux de non qualité trop élevé pour les clients.

Les holons, en tant qu'entités décisionnelles doivent ainsi disposer d'un certain degré d'autonomie pour construire leurs décisions. Dans ce contexte, l'**autonomie** est définie comme le degré de liberté disponible pour chaque holon par rapport à sa capacité de décision, quel que soit son niveau hiérarchique au sein de l'holarchie. Elle peut se traduire par exemple sous la forme d'un espace de recherche lors de l'utilisation d'outils d'optimisation, les contraintes sur cet espace limitant cette autonomie. Le niveau d'autonomie de chaque holon peut être défini pendant la phase de conception de l'HCA par le concepteur lui-même, mais il peut également être ajusté par un holon de niveau supérieur avec une application à un niveau inférieur pendant la phase d'exploitation. Par exemple, un holon superviseur décide de restreindre l'ensemble de holons ressources possibles parmi lesquels des holons produits de niveau inférieur doivent choisir car une opération de maintenance doit être effectuée sur l'un de ces holons ressources. La communication entre holons permet également d'ajuster l'autonomie des holons. Par exemple un protocole de négociation directe (de type contract-net) peut conduire un ensemble de holons à améliorer la qualité de leurs décisions au travers d'un mécanisme d'enchères élargissant les espaces de recherche pour éviter les optimums locaux. On retrouve ici la vision classique en optimisation par les contraintes : une décision est soit sous contraintes soit sur les contraintes.

En raison de ces caractéristiques, on s'attend à ce qu'un comportement émergent se produise au niveau de l'HCA. De notre point de vue, un **comportement émergent** est l'observation d'une propriété à un niveau global d'un HCA qui n'a pas été explicitement intégrée (programmée) dans les holons composant cette HCA. Par exemple, l'utilisation d'algorithmes de champs potentiels attractifs / répulsifs peut conduire les holons produits à éviter naturellement les holons ressource actuellement

indisponibles pour cause de panne et à les sélectionner à nouveau après la fin de la maintenance, sans détailler ce comportement de manière explicite dans le HCA. Dans la littérature, plusieurs propriétés émergentes attendues sont régulièrement décrites en utilisant la notation self-*, par exemple : self-adaptation, self-reconfiguration, etc. (Leitão et al., 2016).

3 EVOLUTION DES ARCHITECTURES

Quelques études ont déjà proposé un parallèle entre Industrie 4.0 et architectures de contrôle (Meissner et al., 2017). Les critères généralement utilisés concernent la dualité hiérarchie/hétéarchie et leur hybridation et cherchent à montrer la performance de chacune de ces propositions. Notre proposition ici est de proposer un historique plus précis des évolutions, en prenant notamment en compte les dimensions outil et applications.

3.1 Bénéfices et évolutions des architectures de référence

La mise en place d'un système de contrôle basé sur un ensemble d'entités communicantes autonomes est une tâche qui nécessite de nombreuses compétences de la part des développeurs. Comme nous l'avons déjà dit, plusieurs travaux de recherche parallèles ont étudié les architectures de contrôle les plus efficaces d'une manière générale avec un point de vue de la performance, mais le processus de développement n'est pas obligatoirement amélioré de la même manière.

Cependant, compte tenu du développement des systèmes de production holoniques (Holon Manufacturing Systems – HMS), les systèmes ciblés permettent de réduire la complexité du processus. En effet, les systèmes de production évoluent généralement dans un environnement semi-structuré et gèrent un ensemble de données naturellement structurées. De ce fait, il est possible d'orienter et de faciliter la construction de l'architecture en spécialisant a priori certains des holons qui seront intégrés. Le rôle de l'architecture de référence est d'exprimer le rôle et les relations entre ces holons prédéfinis afin que toute l'architecture puisse être structurée en fonction de la référence.

3.2 Architectures génériques

La première architecture de référence décrite dans la littérature provient d'un projet IMS appelé Holonic Manufacturing Systems en 1996 (Van Brussel et al., 1998). Son acronyme est PROSA (Figure 2), en tant qu'acronyme des 4 holons qui composent l'architecture :

- Holons Produits, en charge de la gestion des connaissances de production;
- Holons Ressources, en charge de la gestion des connaissances du processus;
- Holons Ordre, en charge de la gestion de l'exécution du processus;
- Holons Staff, non représentés sur la Figure 2, agissant en tant que conseiller global pour l'ensemble de l'architecture.

Cette architecture est la plus référencée en littérature scientifique (plus de 1700 à la fin de 2017) et constitue souvent la base des architectures émergentes comme étant la plus générique.

Cette architecture a été la première d'une série, développée pour différents domaines (à savoir la production, la logistique, la maintenance, etc.). Par exemple, PROSIS (produit, ressource, ordre, simulation pour la structure isoarchique) a été conçue pour offrir un paradigme d'organisation différent en utilisant le concept d'isoarchie (Pujo et al., 2009). Une isoarchie est une architecture ne contenant pas de liens hiérarchiques de subordination entre holons. Au même niveau de décision, toutes les différentes entités décisionnaires ont exactement le même pouvoir dans le mécanisme de prise de décision. Ainsi, cela correspond à l'absence totale de hiérarchie, à la fois fonctionnelle et temporelle. Dans l'architecture PROSIS, le Staff Holon, inutile dans un tel contexte isoarchique, a été remplacé par un Holon Simulation qui vise à simuler l'évolution du système de production à partir de son état actuel, obtenu via l'écoute active et l'analyse des interactions entre tous les autres holons.

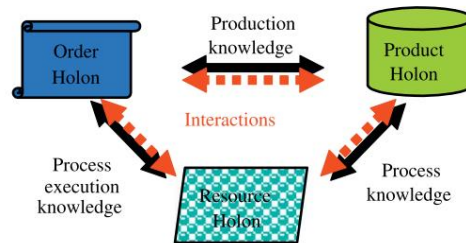


Figure 2. Représentation simplifiée de l'architecture PROSA

3.3 Architectures orientées multi-agents

Parallèlement, la question de la mise en œuvre des architectures dans une perspective de système de contrôle a été étudiée et plusieurs architectures ont été dérivées de PROSA afin de bénéficier des théories et des outils développés dans le domaine multi-agents.

Parmi ceux-ci, HCBA (Holon Component Based Architecture) (Chirn and McFarlane, 2000) est la première architecture basée sur une fusion de différents concepts issus du développement de composants, du système multi-agents (MAS) et du HMS. Le but d'une telle fusion est de développer une architecture hautement décentralisée, construite à partir de composants coopératifs et intelligents autonomes et modulaires, capables de gérer rapidement les différents changements, en mettant l'accent sur la reconfigurabilité du système. HCBA est composé de 2 types de composants dans le système de production, composant ressource et composant produit. Le composant ressource est composé d'une partie physique et d'une partie de contrôle virtuel. La partie physique est dédiée à l'exécution des opérations, tandis que sa partie contrôle gère l'exécution de l'opération au niveau des ressources, le processus de décision lié aux ressources, et la communication avec les autres composants pour la négociation. Les ressources sont en charge de l'ordonnancement des opérations, tout en recherchant l'optimisation de leur utilisation. Le composant produit est également composé

d'une partie physique et d'une partie informationnelle. Sa partie physique peut représenter des matériaux, des pièces, des palettes etc. De plus, la partie informationnelle est responsable du programme de production, y compris le contrôle du routage, le contrôle du processus, la prise de décision et l'information de production. La partie information est composée d'agents virtuels avec des rôles spécifiques. Chaque Composant Produit est composé d'un Coordinateur de Produit créant des agents WIP (Work in Process). Les deux sont en charge de l'exécution des commandes, mais à différents niveaux. Le Coordinateur Produit assure le suivi de production d'un lot alors que les agents WIP sont en charge du suivi de production d'une pièce individuelle. Par conséquent, les agents WIP négocient avec la communauté des ressources pour définir le traitement des pièces dans l'atelier. Ces négociations sont réalisées dans un objectif fixé par le coordinateur de produit (Figure 3).

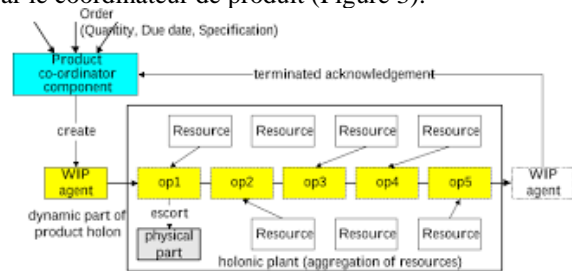


Figure 3. Structure de l'architecture HCBA

De la même manière, Delegate MAS (D-MAS) est un modèle architectural qui permet à un agent de déléguer une responsabilité à un essaim d'agents légers pour soutenir cet agent dans l'accroissement de ses fonctions (Verstraete et al., 2008). L'agent émetteur peut déléguer plusieurs responsabilités, chacune appliquant le modèle D-MAS. L'agent peut utiliser une combinaison de systèmes multi-agents délégués pour gérer une seule responsabilité. Le délégué MAS peut également fournir des services à d'autres agents. D-MAS est une description plus générique d'une approche inspirée par le comportement de colonies de fourmis. Le modèle D-MAS se compose de trois éléments : l'agent, la fourmi et l'environnement (Figure 4).

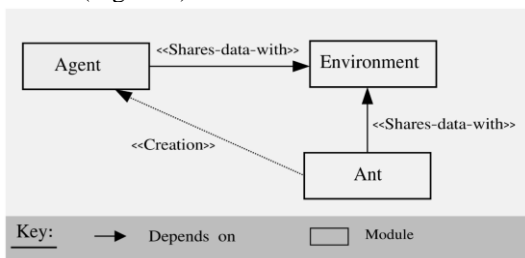


Figure 4. Vue architecturale de D-MAS (Verstraete et al., 2008)

3.4 Extensions des architectures holoniques

Au cours des années, le champs applicatif des HCA s'est étendu et a permis de faire évoluer les concepts et ainsi accroître leur généralité.

En premier lieu, ADACOR (ADaptive holonic COntrol aRchitecture) est une architecture de référence holonique pour les systèmes de production distribués, proposée par

(Leitão and Restivo 2006). ADACOR est une HCA décentralisée qui considère également la centralisation pour tendre vers une optimisation globale du système. Les holons appartiennent aux classes suivantes : Holons Produit (ProdH), Holons Tâches (TH), Holons Opérations (OpH) et Holons Superviseurs (SupH), interconnectés suivant la représentation proposée dans la Figure 5.

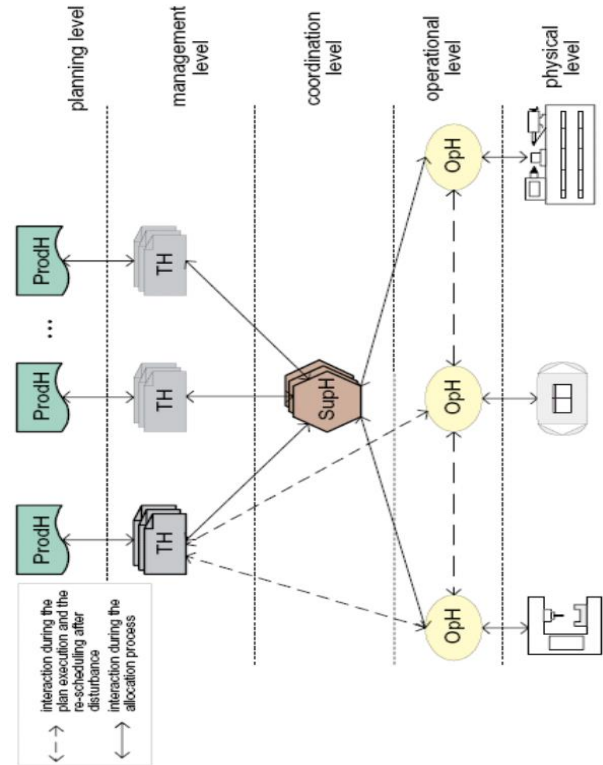


Figure 5. Répartition des holons dans ADACOR (Leitão and Restivo, 2006)

En exécution normale, l'architecture ADACOR maintient le système de production à l'état stationnaire, où les holons sont organisés selon une structure hiérarchique, les OH suivant les plannings d'optimisation proposés par les SupH pour les TH. Lorsqu'un problème survient (retard, défaillance de la machine, ...), le système global entre dans un état transitoire, caractérisé par la réorganisation des holons nécessaires pour réagir à la perturbation. Pour ce faire, ADACOR utilise un mécanisme d'étalement semblable à la phéromone pour diffuser l'information globalement. Grâce à celui-ci, ADACOR introduit la possibilité de changer dynamiquement la holarchie entre l'état stationnaire et le transitoire.

Pour étendre l'adaptabilité de PROSA, une extension aux systèmes hybrides dynamiques, c'est-à-dire des systèmes spécifiques où les comportements discrets et continus doivent être pris en compte dans le contrôle du système, a été présentée comme l'architecture de référence H2CM (Holonc Hybrid Control Model) (Indriago et al., 2016). Cette architecture est basée sur les trois holons de base de PROSA sans le Staff. Deux différences principales avec le PROSA classique apparaissent : (i) Chaque ressource est liée avec un ordre et un produit tout au long de sa vie. L'Holon Ordre est en charge de la surveillance de la ressource tandis que le Holon Produit est en charge de la gamme à appliquer sur le produit réel. Le contenu et

les objectifs des holons Ordre et Produit sont en constante évolution, mais la structure reste indéfiniment la même ; (ii) Un lien récursif clair est présent sur le Holon Ressource. En effet, chaque ressource composée peut être décomposée de manière fractale en une ou plusieurs holarchies, comprenant une ou plusieurs ressources et leurs holons ordre et produit associés. Les relations d'agrégation créées ici peuvent être modifiées tout au long du fonctionnement du système ; Les holarchies peuvent être créées et détruites en ligne.

L'architecture SURFER (Le Mortellec et al., 2013) représente une adaptation intéressante des systèmes de contrôle pour la maintenance et la surveillance de systèmes hautement complexes (à savoir les trains dans ce cas). Le modèle architectural holonique générique proposé pour le diagnostic est présenté dans la Figure 6. Ce modèle est composé de structures de diagnostic récursives, y compris des sous-systèmes et leurs méthodes de diagnostic associées. Chaque système diagnostiqué est composé d'une partie contrôle et d'une partie contrôlée, qui opèrent dans un contexte évolutif. Au niveau le plus bas de la structure holonique, la partie contrôlée est typiquement composée d'éléments physiques (par exemple, capteurs, commutateurs et actionneurs) qui sont liés à des contraintes mécaniques et électriques.

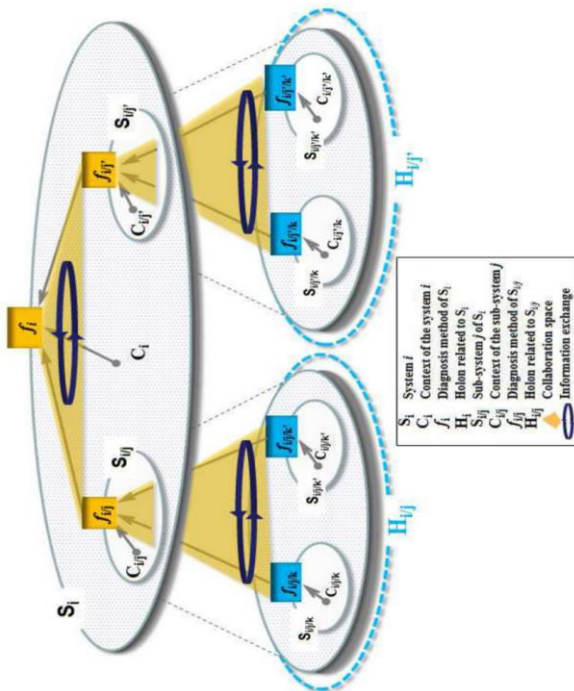


Figure 6. Bases de l'architecture SURFER

3.5 HCA orientées services et cloud

Une des tendances dans le développement des HCA suit l'évolution des technologies basées sur le cloud. Dans cette tendance mondiale, une HCA de référence a été développée pour profiter de certaines normes basées sur le web telles que les services web. L'architecture du SoHMS (Gamboa Quintanilla et al., 2016) est principalement basée sur les principes et concepts introduits par PROSA (Van Brussel et al 1998), combinant certains concepts d'interaction de HCBA (Chirn and McFarlane,

2000) et d'ADACOR (Leitão and Restivo, 2006). Cette proposition utilise le produit (PH), les ressources (RH) et les ordres (OH) de PROSA, et le concept de directory facilitator (DF) des systèmes multi-agents. Même si les concepts de base restent proches des concepts originaux, leur comportement a été adapté pour faire des services (Manufacturing Services - Mservices) le principal élément d'interaction, orienté vers les activités de planification et d'ordonnancement. La Figure 7 montre un diagramme de classes de l'architecture détaillant les relations et les échanges de données entre acteurs. Du point de vue des services, un nouvel élément ajouté à l'architecture est le SIL (Service Interface Layer), utilisé comme interface entre les descriptions de service et leurs méthodes d'implémentation en atelier. En raison de son caractère individuel et propriétaire, chaque ressource possède une instance SIL, contenant toutes les informations sur la manière d'implémenter un service au niveau le plus bas.

3.6 HCA dynamiques

La définition d'HCA dynamiques est probablement l'une des tendances actuelles les plus prometteuses dans la littérature (Cardin et al., 2015). Il postule que le comportement du système peut être modifié dynamiquement afin de s'adapter aux changements de l'environnement et ainsi réduire les états transitoires et la perte de performance associée. Dans le domaine de la production, ORCA (Pach et al., 2014) a été l'une des premières HCA dynamiques formalisées dans la littérature (Figure 8).



Figure 7. Structure du SoHMS (Gamboa Quintanilla et al., 2016)

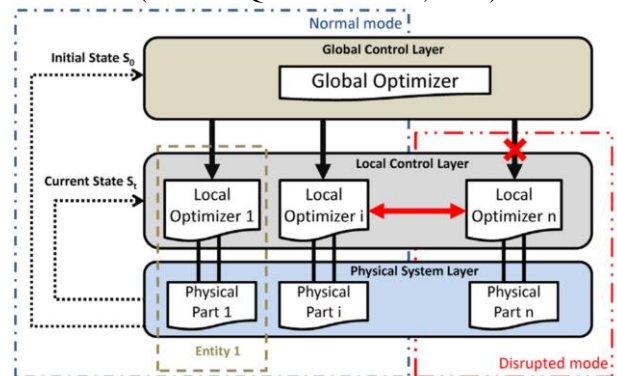


Figure 8. Organisation globale de ORCA (Pach et al., 2014)

Une évolution du mécanisme ADACOR a également été présentée dans (Barbosa et al., 2015) sous le nom ADACOR². L'objectif est de laisser le système évoluer dynamiquement à travers des configurations découvertes

en ligne, et pas seulement entre un état stationnaire et un état transitoire (Figure 9). Le reste de l'architecture est néanmoins assez similaire à ADACOR.

La dernière architecture en date est appelée POLLUX (Jimenez et al., 2017). La nouveauté principale est centrée sur le mécanisme d'adaptation de l'architecture, grâce à des paramètres de gouvernance qui élargissent ou contraignent le comportement des holons de bas niveau concernant les perturbations observées par le niveau supérieur, l'idée étant de trouver la "meilleure" architecture qui convient aux perturbations détectées.

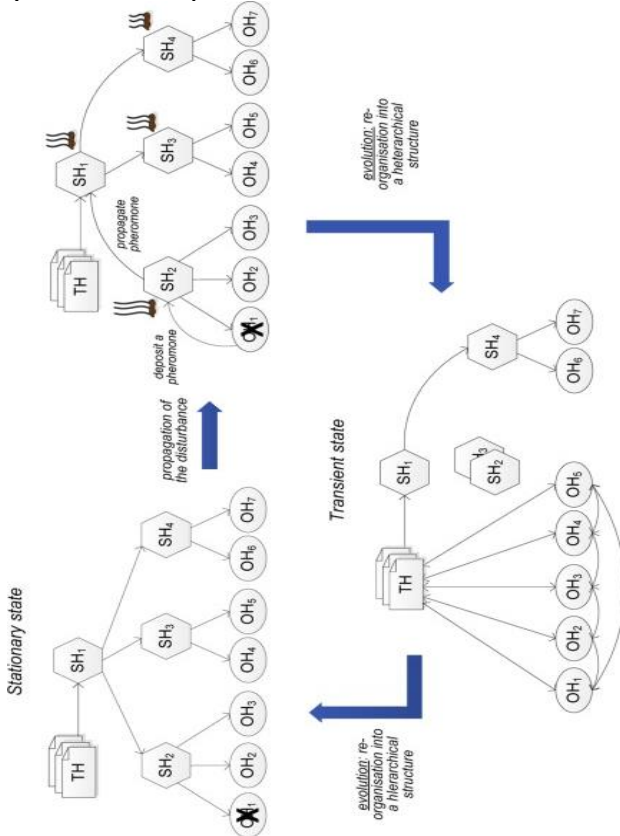


Figure 9. Evolution ADACOR² (Barbosa et al., 2015)

4 SYNTHÈSE ET PERSPECTIVES SCIENTIFIQUES

Une synthèse récente des travaux relatifs à Industrie 4.0 (Hermann et al., 2016) a montré que les principaux axes de recherche sont :

1. Les Systèmes Cyber-Physiques ;
2. L'Internet des Objets ;
3. L'Usine Intelligente et Connectée ;
4. L'Internet des Services ;
5. Le Produit Intelligent.

Cette synthèse met bien en avant les intégrations à la fois horizontale et verticale des données au sein des architectures de contrôle. Au travers de l'analyse présentée dans les sections précédentes, nous avons répertorié dans la Figure 10 les différents axes d'évolution des HCA au fil du temps afin de mettre en évidence les différentes tendances actuellement rencontrées et leur convergence avec les thématiques Industrie 4.0. On peut ainsi noter

un focus particulier sur les notions de flexibilité et d'adaptabilité des HCA lors des premières années, spécifiquement pilotées par le cœur d'application qui restait le contrôle de la production. Au fur et à mesure de l'élargissement des champs d'application potentiels, de nouvelles tendances sont apparues en termes de connectivité et d'ouverture à une intégration verticale, ce qui se trouve être en cohérence totale avec les enjeux présentés dans le modèle RAMI4.0 (Figure 1).

Néanmoins, certains défis supplémentaires doivent être résolus afin d'envisager leur adoption par le secteur industriel dans un contexte industrie 4.0. Nous listons ci-dessous quelques-uns de ces défis.

Comportements émergents / adaptabilité : dans la vision de l'industrie 4.0, les systèmes devraient être capables de réagir à des événements imprévus et de proposer de nouveaux comportements. Les architectures actuelles intègrent des mécanismes de réaction, basés sur des entités décisionnelles locales, parfois coordonnées par des entités globales décisionnelles. Un problème scientifique important consiste à ajuster correctement le niveau d'autonomie des entités décisionnelles locales par rapport au niveau d'autorité des entités décisionnelles globales. La notion d'agrégation inhérentes aux holarchies en général et aux HCA en particulier permet de propager des décisions sur plusieurs niveaux tout en laissant de la flexibilité aux niveaux hiérarchiquement inférieurs, ce qui constitue une preuve de faisabilité de l'application de ces concepts (Jimenez et al., 2017).

Garantie de performance : un comportement émergent implique généralement un manque de garantie sur la performance attendue du système. Cette manque de garantie est donc une problématique allant de pair avec l'augmentation de la réactivité des architectures de contrôle orientées Industrie 4.0. Les HCA dynamiques comme ADACOR², POLLUX ou ORCA sont intéressantes à ce niveau, car elles intègrent un module d'ordonnement optimal utilisé en état normal, couplé à des capacités réactives, exécutées lorsqu'une perturbation se produit. Lorsque cela se produit, le HCA peut modifier sa propre organisation pour minimiser l'impact de cette perturbation. De telles architectures garantissent que les performances du système de fabrication sont optimales dans un état normal, mais pas toujours en mode dégradé. En fait, en raison du niveau d'autonomie et d'adaptabilité laissé aux entités décisionnelles locales, il pourrait être difficile d'assurer une performance minimale du système dans toutes les situations dégradées. Cela dépend certainement du nombre et des types de turbulences survenant sur le système de production. Une étude comparative de chaque architecture holonique dynamique peut alors être un défi très intéressant à explorer. Pour garantir la performance des architectures holoniques, les preuves formelles constituent une autre possibilité de recherche intéressante. Cette possibilité est également intéressante en ce qui concerne les considérations éthiques, étant donné que les preuves du comportement sont les premiers pas vers la détermination des responsabilités en cas d'incident.

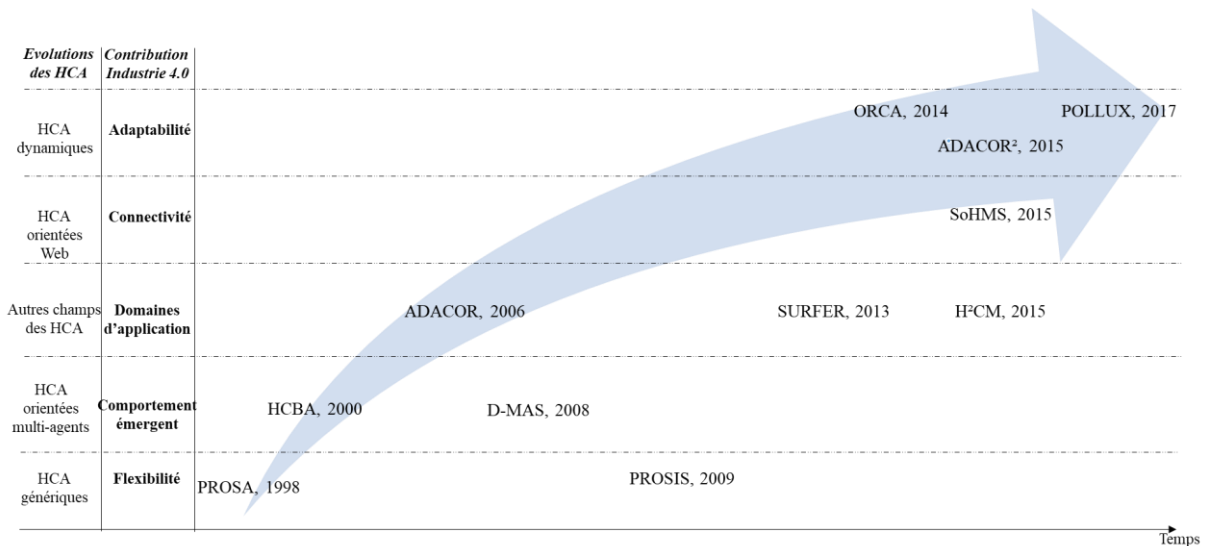


Figure 10. Evolutions des HCA au cours du temps et leurs contributions respectives à Industrie 4.0

Enfin, la génération du code lui-même est un aspect important, et une ingénierie dirigée par les modèles du logiciel est probablement une réponse en adéquation avec les HCA présentées précédemment.

Durabilité / gestion de l'énergie : les questions de durabilité constituent l'un des piliers de l'industrie 4.0. Les HCA peuvent être très bénéfiques pour plusieurs d'entre eux. Par exemple, la gestion de l'énergie à travers la notion de smart grid nécessite un contrôle massivement distribué, évolutif et coordonné pour lequel une perspective holonique sera pertinente. Le contrôle de l'activité de production des systèmes de production ou de logistique sera également affecté par la disponibilité de l'énergie due aux sources d'énergie renouvelables. Ces types de contraintes pourraient bénéficier d'une HCA dynamique afin de s'adapter dynamiquement aux changements fréquents de l'environnement. Une dernière orientation repose sur la connectivité des éléments de consommation, ce qui donnera un accès complet à leur consommation en temps réel et prévue pour une prise de décision en temps réel.

Big Data : un aspect important de l'industrie 4.0 est l'utilisation massive de capteurs pour le suivi des processus de production. Cependant, ceux-ci produisent une quantité importante de données qui ne peuvent pas être traitées par les architectures fortement centralisées actuelles (de type cloud). Certaines fonctions de pré-traitement devraient être intégrées et les HCA apportent des solutions dans ce sens, afin de transformer les données en informations pertinentes pour les holons, et même en connaissances de production quel que soit le niveau.

Intégration de l'humain dans la boucle : dans un atelier orienté Industrie 4.0, l'humain est partie intégrante à la fois du côté opératoire mais également au niveau de la prise de décision. Au niveau opératoire, les HCA permettent une reconfigurabilité de l'architecture et du système et une implémentation du concept Plug&Produce notamment via l'orientation-services (Gamboa Quintanilla et al., 2016). Il est encore néanmoins nécessaire de pouvoir garantir un comportement éthique de l'HCA

intégrant des humains et des composants physiques, potentiellement dangereux (Trentesaux and Rault, 2017). Concernant la prise de décision, les HCA permettent d'ouvrir des interfaces entre l'humain et le logiciel grâce à des avatars intelligents, les holons se comportant alors comme des digital twins des éléments du système.

5 CONCLUSION

Cet article vise à souligner l'évolution des architectures de contrôle holoniques au cours des 20 dernières années en rapport avec ce qui concerne le développement du paradigme de l'industrie 4.0. Cette étude a montré que plusieurs tendances principales sont apparues ces dernières années pour faire face aux objectifs d'adaptabilité et de flexibilité des systèmes industriels. Des détails ont également été fournis sur les HCA de référence disponibles dans la littérature, et enfin quelques éléments sur les perspectives de recherche qui peuvent être attendues dans ce domaine au cours des prochaines années.

Cette analyse montre une grande cohérence entre l'évolution de HCA et les objectifs de l'industrie 4.0, parmi lesquels:

- Le concept de base des HCA qui permet une autonomie et des comportements émergents des holons correspond aux orientations de l'Industrie 4.0;
- L'extension des domaines industriels d'applications des HCA, afin d'intégrer toutes sortes de systèmes dans le système industriel de nouvelle génération orienté Industrie 4.0;
- L'hyper-connectivité des HCA, notamment dans la perspective cloud, qui vise à tirer parti des nouveaux équipements et exigences de l'industrie 4.0 pour les différentes machines et organisations;
- L'adaptabilité des systèmes industriels face à l'évolution de leur environnement est favorisée par l'évolution dynamique des HCA, encore à

l'étude mais constituant une opportunité prometteuse.

REFERENCES

- Barbosa, J., Leitão, P., Adam, E., Trentesaux, D., 2015. Dynamic self-organization in holonic multi-agent manufacturing systems: The ADACOR evolution. *Computers in Industry* 66, 99–111.
- Barbosa, J., Leitão, P., Trentesaux, D., Colombo, A.W., Karnouskos, S., 2016. Cross benefits from cyber-physical systems and intelligent products for future smart industries, in: 2016 IEEE 14th International Conference on Industrial Informatics (INDIN). Presented at the 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), pp. 504–509.
- Barreto, L., Amaral, A., Pereira, T., 2017. Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing, Manufacturing Engineering Society International Conference 2017, MESIC 2017, 28-30 June 2017, Vigo (Pontevedra), Spain* 13, 1245–1252.
- Blanc, P., Demongodin, I., Castagna, P., 2008. A holonic approach for manufacturing execution system design: An industrial application. *Engineering Applications of Artificial Intelligence* 21, 315–330.
- Bokrantz, J., Skoogh, A., Berlin, C., Stahre, J., 2017. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. *International Journal of Production Economics* 191, 154–169.
- Cardin, O., Trentesaux, D., Thomas, A., Castagna, P., Berger, T., El-Haouzi, H.B., 2015. Coupling predictive scheduling and reactive control in manufacturing hybrid control architectures: state of the art and future challenges. *J Intell Manuf* 1–15.
- Chirn, J.-L., McFarlane, D.C., 2000. A holonic component-based approach to reconfigurable manufacturing control architecture, in: 11th International Workshop on Database and Expert Systems Applications Proceedings. pp. 219–223.
- Gamboa Quintanilla, F., Cardin, O., L'Anton, A., Castagna, P., 2016. A modeling framework for manufacturing services in Service-oriented Holonic Manufacturing Systems. *Engineering Applications of Artificial Intelligence* 55, 26–36.
- Hermann, M., Pentek, T., Otto, B., 2016. Design principles for industrie 4.0 scenarios, in: *System Sciences (HICSS)*, 2016 49th Hawaii International Conference on. IEEE, pp. 3928–3937.
- Indriago, C., Cardin, O., Rakoto, N., Castagna, P., Chacòn, E., 2016. H2CM: A holonic architecture for flexible hybrid control systems. *Computers in Industry* 77, 15–28.
- Jimenez, J.-F., Bekrar, A., Zambrano-Rey, G., Trentesaux, D., Leitão, P., 2017. Pollux: a dynamic hybrid control architecture for flexible job shop systems. *International Journal of Production Research* 55, 4229–4247.
- Koestler, A., 1978. *Janus: A Summing Up*. Vintage Books.
- Le Mortellec, A., Clarhaut, J., Sallez, Y., Berger, T., Trentesaux, D., 2013. Embedded holonic fault diagnosis of complex transportation systems. *Engineering Applications of Artificial Intelligence* 26, 227–240.
- Leitão, P., Karnouskos, S., Ribeiro, L., Lee, J., Strasser, T., Colombo, A.W., 2016. Smart Agents in Industrial Cyber-Physical Systems. *Proceedings of the IEEE* 104, 1086–1101.
- Leitão, P., Restivo, F., 2006. ADACOR: A holonic architecture for agile and adaptive manufacturing control. *Computers in Industry* 57, 121–130.
- Meissner, H., Ilse, R., Aurich, J.C., 2017. Analysis of Control Architectures in the Context of Industry 4.0. *Procedia CIRP, 10th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '16*. [Edited by: Roberto Teti, Manager Editor: Dorian M. D'Addona] 62, 165–169.
- Mrugalska, B., Wyrwicka, M.K., 2017. Towards Lean Production in Industry 4.0. *Procedia Engineering, 7th International Conference on Engineering, Project, and Production Management* 182, 466–473.
- Pach, C., Berger, T., Bonte, T., Trentesaux, D., 2014. ORCA-FMS: a dynamic architecture for the optimized and reactive control of flexible manufacturing scheduling. *Computers in Industry* 65, 706–720.
- Pujo, P., Broissin, N., Ounnar, F., 2009. PROSIS: An isoarchic structure for HMS control. *Engineering Applications of Artificial Intelligence, Distributed Control of Production Systems* 22, 1034–1045.
- Simon, H.A., 1996. *The Sciences of the Artificial*. MIT Press.
- Trentesaux, D., Rault, R., 2017. Ethical Behaviour of Autonomous Non-Military Cyber-Physical Systems. Presented at the XIX International Conference on Complex systems: control and modeling problems, Samara, Russia, pp. 26–34.
- Van Brussel, H., Wyns, J., Valckenaers, P., Bongaerts, L., Peeters, P., 1998. Reference architecture for holonic manufacturing systems: PROSA. *Computers in Industry* 37, 255–274.
- Verstraete, P., Saint Germain, B., Valckenaers, P., Van Brussel, H., Van Belle, J., Karuna, H., 2008. Engineering Manufacturing Control Systems Using PROSA and Delegate MAS. *Int. J. Agent-Oriented Softw. Eng.* 2, 62–89.
- Zeulka, F., Marcon, P., Vesely, I., Sajdl, O., 2016. Industry 4.0 – An Introduction in the phenomenon. *IFAC-PapersOnLine, 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016* 49, 8–12.

COUPLAGE DE SIMULATEUR DE PROCÉDÉ ET DE MODELISATION ERTN POUR LA PLANIFICATION DE SYSTEMES ENERGETIQUES

A. LECLERCQ, G. HETREUX, R. HETREUX, P. FLOQUET

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, Toulouse, France

alexandre.leclercq@ensiacet.fr, gilles.hetreux@ensiacet.fr, raphaele.thery@ensiacet.fr, pascal.floquet@ensiacet.fr

RESUME : Dans un système industriel, la « chaleur fatale » est la part de l'énergie thermique produite mais non consommée par le procédé. Souvent présente en quantité importante, la valorisation de cette chaleur fatale issue de l'industrie constitue un potentiel important d'économie d'énergie à exploiter (ADEME, 2015). Lorsque cette valorisation ne peut se faire en interne, une valorisation en externe est envisagée pour répondre à des besoins de chaleur d'autres entreprises ou de sites urbains proches via un réseau de chaleur. Cependant, deux verrous majeurs freinent pour l'heure la généralisation de cette démarche: d'une part, la conception de la chaîne logistique mettant en relation le site producteur et le site consommateur et d'autre part, la conception d'un outil de pilotage décentralisée et autonome de ce système global. Axé sur ce dernier aspect, cet article présente le prototype d'un outil de planification court-terme pour ce type de système énergétique, en supposant l'infrastructure du réseau déjà construite. Le noyau de cet outil s'appuie sur un modèle de programmation linéaire mixte générique (PLM) déduit d'un graphe ERTN. Un simulateur de procédé en régime statique est exploité pour déterminer les paramètres nécessaires à l'instanciation de ce modèle d'optimisation. Les principes et les potentialités de l'approche sont illustrés sur un système simple.

MOTS-CLES : Chaleur fatale, Modélisation, Planification, Valorisation Energétique, Energie

1 INTRODUCTION

Force est de constater que *transition numérique* et *transition énergétique* sont intimement liées. En effet, l'essor des systèmes connectés dans l'industrie est en voie de révolutionner la manière de concevoir et de piloter les moyens de production, au point d'avoir fait émerger le concept d'« Usine du Futur ». Au moyen de ces technologies, l'ambition est de rendre les systèmes industriels plus respectueux de leur environnement grâce des modes de production moins consommateurs de ressources (matières, énergies) et moins générateurs de rejets (déchets, CO₂, ...). Rendus aussi plus intelligents, ils seront plus intégrés au cœur des territoires et connectés aux différents acteurs de leur écosystème. Dans cet objectif, l'exploitation de systèmes d'information avancés construits sur un réseau de communication peut permettre de mettre en synergie un groupe d'acteurs industriels et territoriaux afin :

- d'une part, de réduire la consommation de certaines ressources devenues critiques et de mieux appréhender les nouvelles réalités écologiques (changement climatique, GES, etc.),
- d'autre part, de gagner mutuellement en efficacité et en compétitivité dans une relation gagnant-gagnant.

La récupération de la chaleur fatale industrielle s'inscrit dans cette volonté d'une utilisation plus rationnelle de l'énergie. La *chaleur fatale* correspond aux flux de chaleur générés lors du fonctionnement d'un système industriel et qui ne sont pas récupérés. Il peut s'agir de l'énergie thermique produite grâce aux combustibles

mais qui n'est pas utilisée en totalité par le système de production (par exemple, un excédent de vapeur produit par une chaudière, un courant d'eau chaude à 95 °C qui est rejetée après utilisation à 70 °C, etc.) ou de la chaleur générée par le système de production lui-même (par exemple, une réaction fortement exothermique qui dégage un flux de chaleur qu'il faut évacuer). Aujourd'hui, la chaleur fatale perdue coûte doublement cher à l'industriel. En effet, pour des raisons techniques (traitement des fumées, etc.) ou réglementaires (seuil de température des eaux usées, etc.), son rejet nécessite souvent un refroidissement qui représente un coût non négligeable venant s'additionner aux coûts de production (coûts d'achat des combustibles, coûts d'exploitation, etc.) pour au final, une énergie qui ne sert à rien !

En France, de récentes études (ADEME, 2015) ont montré que les gisements de chaleur fatale à plus de 100°C sont très importants (51 TWh pour une consommation d'énergie globale de l'industrie de 350 TWh) et localisé dans de nombreux secteurs industriels (figure 1).

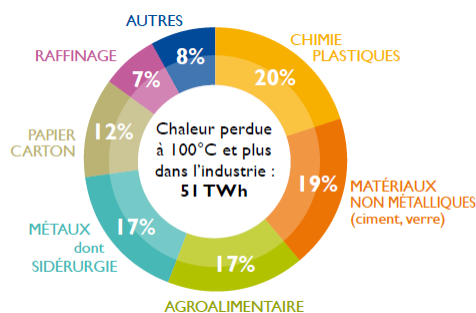


Figure 1: répartition par secteur industriel des gisements de chaleur fatale à plus de 100°C (ADEME,2015)

Les axes de valorisation de la chaleur fatale sont multiples et ont différentes motivations (Nowicki and Gosselin, 2012; Kaşka, 2014; Tigges and Link, 2017; Wahroos et al., 2018). De manière simplifiée, elle peut se faire :

- soit, sans changement de vecteur énergétique, sous forme de chaleur :
 - pour répondre à des besoins énergétiques internes à l'entreprise (séchages d'autres lignes de production, préchauffage de fluides, chauffage des locaux, etc.) lorsque les niveaux de températures le permettent. Axe généralement privilégié, on parle alors d'*intégration énergétique*,
 - pour répondre à des besoins énergétiques externes, notamment émanant d'autres entreprises voisines (notion d'*éco-parc industriel*) ou d'un bassin de population à proximité via un réseau de chaleur. Dans ce cadre, depuis le 1^{er} Janvier 2015, les *ICPE (Installations Classées pour la Protection de l'Environnement)* d'une puissance thermique totale supérieure à 20 MW ont obligation de réaliser une étude coûts-avantages en cas de rénovation substantielle ou d'installation nouvelle afin d'évaluer la rentabilité de valoriser de la chaleur fatale par un raccordement à un réseau de chaleur ou de froid.
- soit, avec changement de vecteur énergétique par la production d'électricité,
 - pour une utilisation en *autoconsommation*,
 - pour répondre à des besoins électriques collectifs externes via la mise en place de contrats spécifiques avec les fournisseurs d'électricité,

Ces deux formes de valorisation sont cohérentes et complémentaires car les niveaux de température requis sont différents. Dès lors que la chaleur récupérée atteint environ 150 - 200°C, la production d'électricité est envisageable alors que pour les réseaux de chaleur, la valorisation devient possible avec des niveaux de température inférieurs à 150°C.

Inscrit dans un projet d'*Ecologie Industrielle et Territoriale*, cette synergie économique et environnementale avec le tissu industriel représente donc une opportunité gagnante/gagnante pour tous les acteurs. Cependant, deux verrous majeurs freinent pour l'heure la généralisation de cette démarche:

- d'une part, la conception des différents maillons (conversion, transport, stockage, etc.) de la chaîne logistique mettant en relation le site producteur et le (ou les) site(s) consommateur(s). En effet, la détermination de la technologie de valorisation la plus adaptée pour exploiter une source de chaleur fatale dépend de nombreux facteurs tels que phénomènes physico-chimiques mis en jeu, niveau de température requis, variabilité et intensité des flux, distance séparant la source et l'utilisateur final, coûts d'investissement et d'exploitation, etc.
- d'autre part, la conception d'un outil de planification court-terme permettant :

- en phase de conception, d'évaluer la viabilité économique de la solution de valorisation envisagée,
- en phase d'exploitation, de piloter le système global de manière décentralisée et autonome. Dans ce cadre, l'introduction massive d'objets connectés paraît être une solution technique adaptée pour mettre en œuvre une gestion distribuée visant à coordonner les besoins et disponibilités de chaque acteur du système de la façon la plus efficiente possible.

Les travaux présentés dans cette communication sont focalisés plus spécifiquement sur la mise en place d'un modèle de planification court-terme de tels systèmes dans un objectif d'évaluation économique et opérationnelle d'une filière de valorisation. Dans (Cajot et al., 2015; Georgiou, 2016; Liu et al., 2018), la complexité et la diversité des approches de planification sont mises en évidence. Parmi les points critiques à traiter, il faut que :

- la modélisation des systèmes matériels mis en jeu soit suffisamment fidèle pour obtenir des solutions fiables,
- l'outil soit capable de prendre en compte la forte variabilité des demandes et des coûts, ainsi que le caractère souvent intermittent de la source de chaleur fatale,
- le centre de décision puisse proposer des plannings d'activité optimisés, sachant que la production et la consommation de chaleur fatale sont dépendantes de demandes sur lesquelles le système de pilotage n'a aucun contrôle.

Pour répondre à cette problématique, cet article propose une méthodologie basée sur un modèle de programmation linéaire mixte générique (PLM) instancié au moyen de paramètres issus de la simulation phénoménologique en régime stationnaire des différents composants du système. La suite de l'article s'articule de la manière suivante. La section 2 commence par présenter la méthode proposée pour construire et instancier le modèle de planification. Pour illustrer cette démarche, la section 3 décrit un exemple simple de chaîne logistique énergétique mettant en relation un site industriel et un réseau de chaleur urbain (RCU). La section 4 déroule alors la méthodologie sur ce cas d'étude et présente les premiers résultats obtenus.

2 MODELE DE PLANIFICATION D'UNE CHAÎNE LOGISTIQUE DE L'ÉNERGIE

2.1 Principes de la méthode d'élaboration du modèle de planification

Comme indiqué dans l'introduction, l'objectif des travaux présentés ici est d'établir un modèle de planification court-terme permettant d'évaluer la viabilité d'une filière de valorisation de chaleur fatale selon un critère à la fois économique et opérationnel. Pour cela, notre

approche de modélisation et de résolution s’appuie sur trois principes majeurs :

- d’une part, le modèle d’optimisation sous-jacent est formulé sous forme d’un *programme linéaire mixte (PLM)* afin de garantir l’obtention de solution dans des temps de calcul raisonnable. Ceci induit que certains phénomènes non-linéaires devront être approximatés par des fonctions linéaires continues par morceaux,
- d’autre part, le but est de pouvoir comparer différentes alternatives de chaîne logistique, autant sur le plan de la topologie du réseau constitué que des solutions technologiques de récupération, de conversion, de transport, de stockage et de restitution possibles. Par conséquent, notre approche s’est attachée à établir un *modèle d’optimisation générique* capable de traiter toutes ces alternatives sans modifier les contraintes mais par une simple instanciation de celui-ci par un jeu de paramètres.
- enfin, pour générer le jeu de paramètres spécifiques à une alternative, l’idée est de s’appuyer sur le formalisme *Extended Resource Task Network* ou *ERTN* (Fabre et al., 2011; Théry et al., 2012). Un réseau *ERTN* est un graphe orienté composé de 8 types de nœuds et de 5 types d’arcs. Il permet de modéliser de manière systématique et sans ambiguïté les processus et flux (matière, énergie) en présence mais aussi, les contraintes de précédence entre opérations, les contraintes de ressource cumulatives et disjonctives (capacité des appareils, durée opératoire, politique de stockage et de transfert, etc.). La sémantique complète de ce formalisme et les règles formelles de construction du graphe sont décrites en détail par (Hétreux, 2012). Un point fort de cette procédure est qu’à chaque élément sémantique du formalisme *ERTN* est associé un jeu de contraintes de *PLM* spécifiques paramétrées par les annotations du modèle graphique. Ceci donne ainsi la possibilité de générer automatiquement le jeu de données complet instanciant le modèle de *PLM*.

Selon ces principes, la méthode d’élaboration du modèle de planification s’effectue en deux étapes principales (cf. figure 2) :

- la première étape (*ERTN Graph Initialisation*) consiste à établir le réseau *ERTN* correspondant à la *recette générique* (selon la terminologie consacrée)

du système global à analyser. Il s’agit d’une représentation purement fonctionnelle de la chaîne logistique, indépendante de toute implantation physique et technique. A ce niveau, la séquence d’opérations associée est établie et seuls sont identifiés certains paramètres ou conditions opératoires particuliers.

- la deuxième étape (*ERTN Graph Instanciation*) consiste à construire le réseau *ERTN* correspondant à la *recette de site* du système considéré, c’est à dire à effectuer l’instanciation de la recette générique précédente pour une implantation physique et technique précise de la chaîne logistique considérée. Durant cette étape, les valeurs numériques des paramètres du modèle (tels que les durées de tâches, les délais d’obtention, les capacités min et max, les proportions de chaque flux entrants ou sortants, la température de certains flux, etc.) sont déterminés. Mettant en jeu des phénomènes physico-chimiques et thermodynamiques parfois complexes, la détermination de certains paramètres peut nécessiter de recourir à une modélisation phénoménologique détaillée reposant sur des bilans matière et énergie, des équations d’équilibres entre phases, etc. Pour cela, le simulateur de procédé en régime statique *ProSim* + est exploité pour évaluer les paramètres numériques manquants. Dans les cas où la relation entre deux grandeurs particulières n’est pas linéaire, il est nécessaire de la linéariser par morceaux, entraînant la duplication de certaines tâches du graphe *ERTN*.

Une fois ces deux étapes achevées, le jeu complet de paramètres peut être généré puis utilisé pour instancier le modèle de *PLM* (étape *ERTN Graph Planning*). Ce dernier est alors résolu au moyen du logiciel *ILOG Solver* qui intègre le solveur *CPLEX*.

2.2 Le modèle de planification court terme associé

Comme expliqué précédemment, les annotations du graphe *ERTN* modélisant la *recette de site* du système à analyser sont les paramètres du modèle de *PLM* dédié à la planification de celui-ci. Ce modèle de planification est défini sur un horizon temporel discrétisé en périodes. Les variables de décisions sont récapitulées ci-dessous :

$W_{k,t}$: variable binaire d’allocation des équipements. $W_{k,t} = 1$ si la tâche k est lancée au début de la période t et $W_{k,t} = 0$, sinon

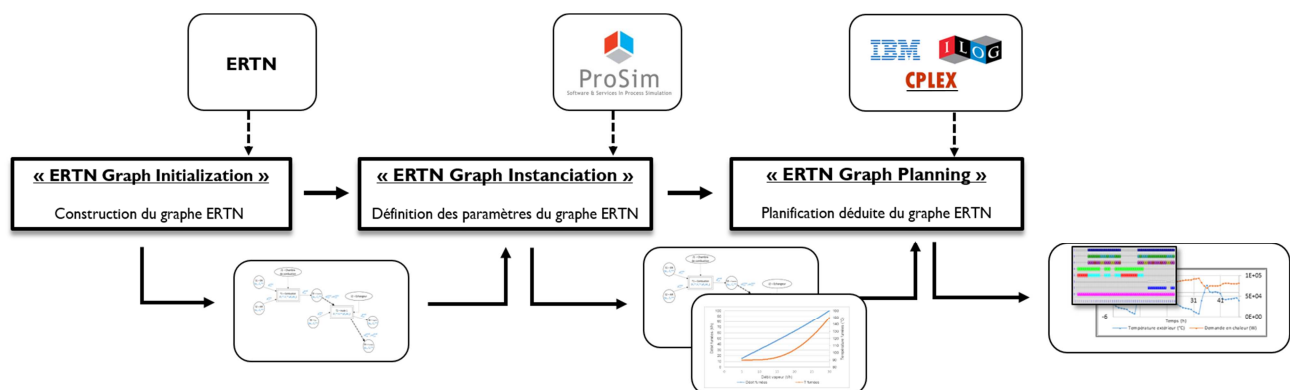


Figure 2 : Schéma de principe de la méthode

- $B_{k,t}$: quantité traitée par la tâche k lancée au début de la période t (lorsque $W_{k,t} = 1$)
- $I_{r,k,t}$: quantité de ressource cumulative r entrant dans la tâche k au début de la période t
- $O_{r,k,t}$: quantité de ressource cumulative r quittant la tâche k en période t
- $R_{r,t}$: quantité de ressource cumulative r stockée en période t
- $UO_{r,k,t}$: quantité de ressource cumulative r produite par la tâche k en période t
- $UI_{r,k,t}$: quantité de ressource cumulative r consommée par la tâche k en période t
- $In_{r,t}$: quantité de ressource cumulative r importée en période t
- $Out_{r,t}$: quantité de ressource cumulative r exportée en période t

Les contraintes du modèle de *PLM* exploité dans cet article sont indiquées ci-dessous.

$$\sum_{k \in K_r} \sum_{\substack{t'=t-p_k+1 \\ t' > 0}}^t W_{k,t'} \leq 1 \quad \forall r \in R^D, \forall t \in 1, \dots, T \quad (1)$$

$$W_{k,t} V_k^{\min} \leq B_{k,t} \leq W_{k,t} V_k^{\max} \quad \forall k \in K, \forall t \in 1, \dots, T \quad (2)$$

$$B_{k,t} = \sum_{r \in R_k^{\text{cons}}} I_{r,k,t} \quad \forall k \in K, \forall t \in 1, \dots, T \quad (3)$$

$$B_{k,t} = \sum_{r \in R_k^{\text{prod}}} O_{r,k,t} \quad \forall k \in K, \forall t \in 1, \dots, T \quad (4)$$

$$R_{r,0} = R0_r \quad \forall r \in R^C \quad (5)$$

$$0 \leq R_{r,t} \leq C_r^{\max} \quad \forall r \in R^C, \forall t \in 1, \dots, T \quad (6)$$

$$R_{r,t} = R_{r,t-1} + \sum_{k \in K} O_{r,k,t-dd_{r,k}} - \sum_{k \in K} I_{r,k,t} + \sum_{k \in K} UO_{r,k,t-dd_{r,k}} - \sum_{k \in K} UI_{r,k,t} + In_{r,t} - Out_{r,t} + \text{Approvisionnement}_{r,t} - \text{Demande}_{r,t} \quad \forall r \in R^C, \forall t \in 1, \dots, T \quad (7)$$

$$\rho_{k,r}^{\text{cons}} B_{k,t} \leq I_{r,k,t} \leq (\rho_{k,r}^{\text{cons}} + \mu_{k,r}^{\text{cons}}) B_{k,t} \quad \forall k \in K, \forall r \in R_k^{\text{cons}}, \forall t \in 1, \dots, T \quad (8)$$

$$\rho_{k,r}^{\text{prod}} B_{k,t} \leq I_{r,k,t} \leq (\rho_{k,r}^{\text{prod}} + \mu_{k,r}^{\text{prod}}) B_{k,t} \quad \forall k \in K, \forall r \in R_k^{\text{prod}}, \forall t \in 1, \dots, T \quad (9)$$

$$UI_{r,k,t} = \sum_{t'=t-p_k+1}^t (u_{f,k,r}^{\text{cons}} W_{k,t'} + u_{v,k,r}^{\text{cons}} B_{k,t'}) \quad \forall r \in R^C, \forall k \in K, \forall t \in 1, \dots, T \quad (10)$$

$$UO_{r,k,t} = \sum_{\substack{t'=t-p_k+1 \\ t' > 0}}^t (u_{f,k,r}^{\text{prod}} W_{k,t'} + u_{v,k,r}^{\text{prod}} B_{k,t'}) \quad \forall r \in R^C, \forall k \in K, \forall t \in 1, \dots, T \quad (11)$$

$$In_r^{\min} \leq In_{r,t} \leq In_r^{\max} \quad \forall r \in R^C, \forall t \in 1, \dots, T \quad (12)$$

$$Out_r^{\min} \leq Out_{r,t} \leq Out_r^{\max} \quad \forall r \in R^C, \forall t \in 1, \dots, T \quad (13)$$

où :

- K : Ensemble des tâches k
- R^D : Ensemble des ressources disjonctives
- R^C : Ensemble des ressources cumulatives r

- R_k^{cons} : Ensemble des ressources cumulatives r consommées par la tâche k ($k \in K$)
- R_k^{prod} : Ensemble des ressources cumulatives r produites par la tâche k ($k \in K$)

Les contraintes (1) assurent qu'une seule tâche k est réalisée sur la ressource disjonctive r durant la période t . Les contraintes (2) et (6) définissent les contraintes de capacité sur les flux et les états. Les contraintes (3), (4), (8) et (9) réalisent les bilans de conservation matière/énergie au niveau des tâches alors que les contraintes (5) et (7) réalisent ceux des états. Les contraintes (10) et (11) traduisent simplement une consommation/production de ressource sans bilan de conservation. Enfin, les contraintes (12) et (13) fixent les limites sur les flux importés ou exportés. Le modèle de *PLM* complet peut être trouvé dans (Hetreux, 2012).

3 PRESENTATION DU CAS D'ETUDE

3.1 Le système global

Afin d'illustrer l'approche développée dans cette contribution, un exemple simple est considéré (figure 3). Celui-ci est composé d'un site industriel dont les besoins en chaleur sont pourvus par une centrale de production de vapeur constitué d'une chaudière *UPCI*. La chaleur fatale à valoriser est portée par les fumées rejetées par cette chaudière. Dans un environnement proche, un réseau de chaleur urbain alimente un bassin de population. La chaleur est fournie grâce à deux chaudières *UPC2* et *UPC3* pouvant fonctionner en parallèle. On suppose qu'une étude de faisabilité a déjà menée et que la chaîne logistique considérée vise à valoriser la chaleur fatale en subvenant à tout ou partie des besoins du réseau de chaleur urbain.

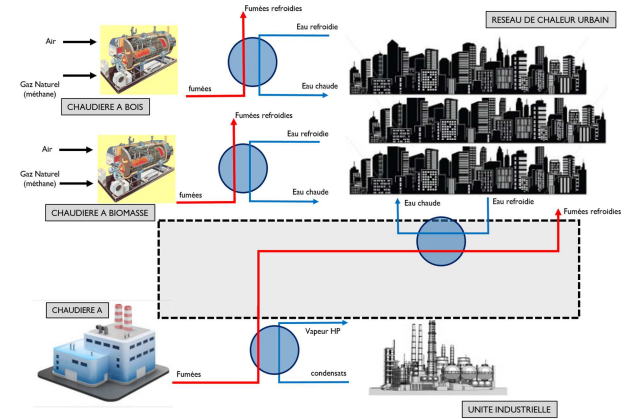


Figure 3: schéma global de l'exemple

3.2 Le site industriel

Le site industriel est composé de plusieurs unités de production pouvant fonctionner en parallèle et de manière intermittente. Par ailleurs, on suppose que ces unités ont déjà fait l'objet d'une analyse pincement

(Linnhoff 1993) afin de minimiser leurs besoins en vapeur. Dépendante des campagnes de production en cours, la demande globale en vapeur présente une grande variabilité. Une centrale d'utilité alimente les unités de production en vapeur au moyen d'une chaudière *UPCI* fonctionnant au gaz naturel. La combustion du gaz naturel avec l'air (compositions dans la Table 1) doit permettre de produire des fumées à 800°C. Le pouvoir calorifique inférieur (*PCI*) de ce gaz naturel est déterminé en utilisant les capacités calorifiques de combustion des différents constituants et est égal à $PCI = 818,4 \text{ kJ/mol}$.

	Composé	Composition molaire (%)
Gaz Naturel	Méthane (CH ₄)	0,96999
	Ethane (C ₂ H ₆)	0,022
	Propane (C ₃ H ₈)	0,003
	Isobutane (C ₄ H ₁₀)	0,001
	Azote (N ₂)	0,004
	Sulfure d'Hydrogène (H ₂ S)	0,00001
Air	Azote (N ₂)	0,7884
	Oxygène (O ₂)	0,2116

Table 1 : Composition du gaz naturel et de l'air

Ces fumées alimentent un échangeur de chaleur qui assure la production de vapeur. Les fumées récupérées à la sortie de cet échangeur sont à un niveau de température suffisant pour être valorisées. Notons toutefois que la quantité et la qualité de l'énergie apportée par ces fumées dépendent fortement de la demande en vapeur et des conditions de fonctionnement de la chaudière.

3.3 Le site urbain

Le site urbain considéré est composé de 100 habitations répondant à la norme BBC, soit une résistance thermique égale à 8 m².K/W pour le toit et 4 m².K/W pour les murs et le sol. Ces habitations sont alimentées en chaleur via un réseau de chaleur urbain. De manière générale, un réseau de chaleur est un système de distribution de chaleur produite de façon centralisée permettant de desservir plusieurs usagers (Cerema, 2011). Dans notre cas, il comprend deux chaudières, un réseau de distribution primaire dans lequel la chaleur est transportée par un fluide caloporteur et un ensemble de sous-stations d'échange, à partir desquelles les bâtiments sont desservis par un réseau de distribution secondaire. La chaudière *UPC2* est alimentée par du bois, considéré comme une ressource illimitée. La chaudière *UPC3* est alimentée par de la biomasse issue de l'agriculture, combustible supposée disponible de manière intermittente.

3.4 Couplage du site industriel et du site urbain

Dans notre exemple, on suppose que la valorisation de la chaleur fatale rejetée par le site industriel se fait via l'installation d'un échangeur de chaleur supplémentaire traversé respectivement par les fumées et le circuit primaire du réseau de chaleur urbain. Du point de vue du réseau, cet échangeur peut être considéré comme une

sorte de chaudière d'appoint. Dans ces conditions, l'objectif est de planifier le fonctionnement du système global, notamment de décider de la mise en route ou l'arrêt des chaudières dédiées à ce réseau de manière à exploiter au mieux la chaleur fatale produite par le site industriel.

4 APPLICATION

Afin de modéliser le système décrit dans la section 3, la méthode introduite en section 2.1 est appliquée.

4.1 Etape 1 : établissement du réseau *ERTN* associé à la recette générique du système

La première étape consiste à définir le système d'un point de vue fonctionnel. A titre d'exemple, la chaudière *UPCI* est décomposée en deux fonctions (figure 4) : la première correspond à la combustion du gaz naturel et la seconde, à l'échange de chaleur qui vaporise l'eau. Ces opérations sont effectuées respectivement dans la chambre de combustion (tâche T1 – ressource J1) et dans l'échangeur de chaleur entre les fumées issues de la chambre de combustion et l'eau (tache T2 – ressource J2). La tâche T1 consomme du gaz naturel (S1) et de l'air (S2) pour produire des fumées à 800°C (S3). Le paramètre $\rho_{1,1}^{cons}$ (resp. $\rho_{1,2}^{cons}$) correspond au ratio du débit de gaz naturel (resp. air) par rapport au débit total alimentant l'opération de combustion.

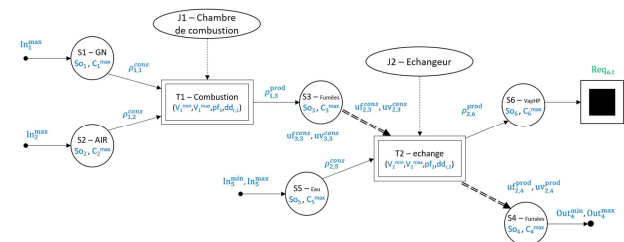


Figure 4 : Graphe *ERTN* de la chaudière *UPCI*

Ensuite, l'échangeur (J2) met en contact la fumée à 800°C (S3) et de l'eau à 90 °C (S5). Après échange, de la vapeur (S6) à 205°C et des fumées (S4) sont produits. Les paramètres $uf_{2,3}^{cons}$, $uv_{2,3}^{cons}$ correspondent aux coefficients linéaires liant les débits de fumée et les débits de vapeur. Notons que les fumées (S4) sortent à une température dépendant du régime de fonctionnement de l'échangeur. La quantité de vapeur (S6), pour sa part, est une donnée qui correspond au besoin en vapeur du site industriel. Une démarche analogue est réalisée pour chaque composant de la chaîne logistique considérée. A l'issue de cette étape, le réseau *ERTN* correspondant à la recette générique du système est établie.

4.2 Etape 2 : Etablissement du réseau *ERTN* associé à la recette de site du système

Le réseau *ERTN* obtenu à l'issue de l'étape 1 doit maintenant être instancié avec les données spécifiques du

système physique considérée afin d'établir le réseau ERTN correspondant à la *recette de site*. L'ensemble des données et paramètres de fonctionnement à fixer lors de cette étape sont récapitulés ci-dessous :

- V_k^{min} : Quantité minimale traitée par la tâche k
- V_k^{max} : Quantité maximale traitée par la tâche k
- In^{max} : Quantité maximale de ressource r importée en tout instant
- $Req_{r,t}$: Quantité requise de ressource r en période t
- $\rho_{k,r}^{cons}$: Proportion de ressource r consommée par la tâche k
- $u_{f,k,r}^{cons}$: Partie fixe de la quantité de ressource r produite par la tâche k
- $u_{v,k,r}^{cons}$: Partie variable de la quantité de ressource r produite par la tâche k

Certaines données sont fournies et correspondent, par exemple, au profil de demande en vapeur des unités de production et des besoins en chaleur du réseau urbain, estimés sur un horizon de 2 jours (soit 48 h \Rightarrow horizon de 48 périodes). Les figures 5 et 6 montrent, à titre d'exemple, les scénarios envisagés pour notre exemple.

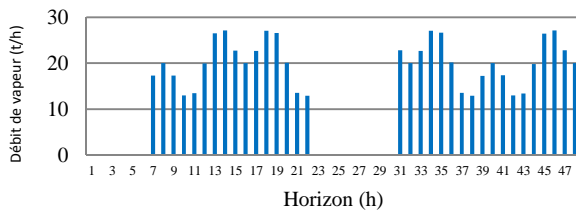


Figure 5 : Profil de demande en vapeur industrielle

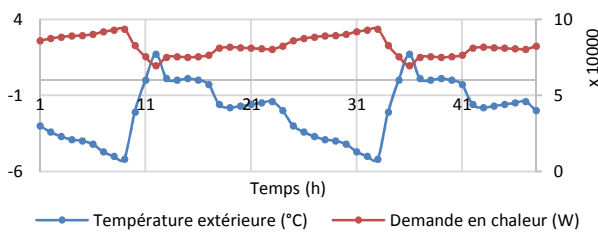


Figure 6 : Evolution de la température extérieure et du besoin en chaleur du site urbain

D'autres paramètres sont liés au fonctionnement et à la technologie des différents éléments du système. La plupart de ces paramètres sont déterminés via la simulation en régime stationnaire de modèles phénoménologiques reposant sur les lois de la thermodynamique et des bilans matière et énergie.

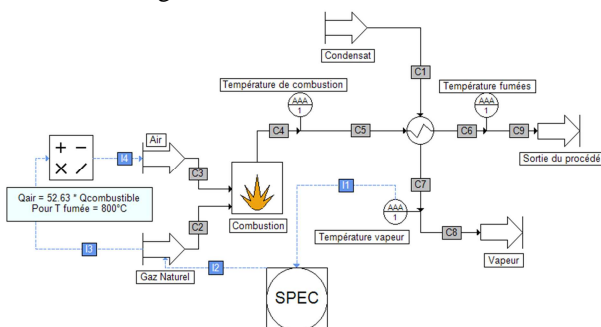


Figure 7 : Flowsheet ProSim+® de la chaudière UCPI

Pour cela, l'outil de simulation de procédés ProSim+ a été utilisé. Afin d'illustrer cette étape, la méthode est appliquée sur la chaudière UCPI (figure 7).

- Identification des paramètres d'entrée $\rho_{1,1}^{cons}$ et $\rho_{1,2}^{cons}$ de l'opération de combustion de UCPI:

Une première simulation de l'opération combustion seule est effectuée en l'alimentant avec de l'air à flux constant, et du gaz naturel à flux contrôlé par la température de sortie des fumées. Ceci permet de déterminer le ratio de gaz naturel et d'air nécessaire pour maintenir une température de sortie des fumées de 800°C. Ensuite, à ratio fixé entre les deux courants, une succession de simulations est réalisée pour différentes valeurs de débits d'entrée (figure 8).

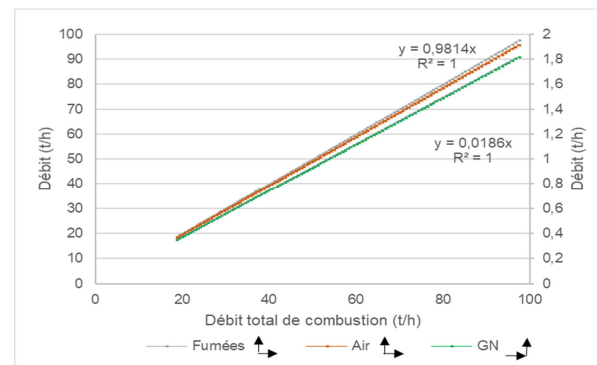


Figure 8 : Débit de fumées en fonction du débit d'air et de gaz naturel.

Mettant en évidence une relation linéaire entre débit de fumée et débit de combustible, les paramètres $\rho_{1,1}^{cons}$ et $\rho_{1,2}^{cons}$ peuvent être identifiés :

$$\rho_{Combustion, Gaz\ Naturel}^{cons} = \rho_{1,1}^{cons} = 0,019\ t/h$$

$$\rho_{Combustion, Air}^{cons} = \rho_{1,2}^{cons} = 0,981\ t/h$$

- Identification des paramètres $u_{f,2,3}^{cons}$ et $u_{v,2,3}^{cons}$ définissant le débit de fumées à 800°C en fonction du débit de vapeur à produire :

On suppose que l'échangeur associé à la chaudière a été dimensionné durant la phase de conception de la chaîne logistique. Celui-ci est caractérisé par le paramètre $UA = 88763\ W/K$, calculé à partir de son aire d'échange $A (A=472,14\ m^2)$ et son coefficient d'échange global $U (U = 188\ W/m^2/K)$, évalué via la relation suivante :

$$\frac{1}{U} = \frac{1}{h_{fumées}} + \frac{e_{inox}}{\lambda_{inox}} + \frac{1}{h_{vap}}$$

avec

$$h_{fumées} = 200\ w/m^2/K, h_{vap} = 50000\ W/m^2/K,$$

$$e_{inox} = 5\ mm, \lambda_{inox} = 16,3\ W/m/K$$

La vapeur est produite à partir d'eau à 90 °C (pression à 14 bars absolu). La mise en contact de cette eau avec les fumées à 800° C issues de la combustion l'amène à sa température d'ébullition $T_{eb} = 195,2°C$, puis est surchauffée jusqu'à 205°C avant de sortir de l'échangeur sous forme de vapeur. Les fumées sortant de l'échangeur constituent la source de chaleur fatale à valoriser.

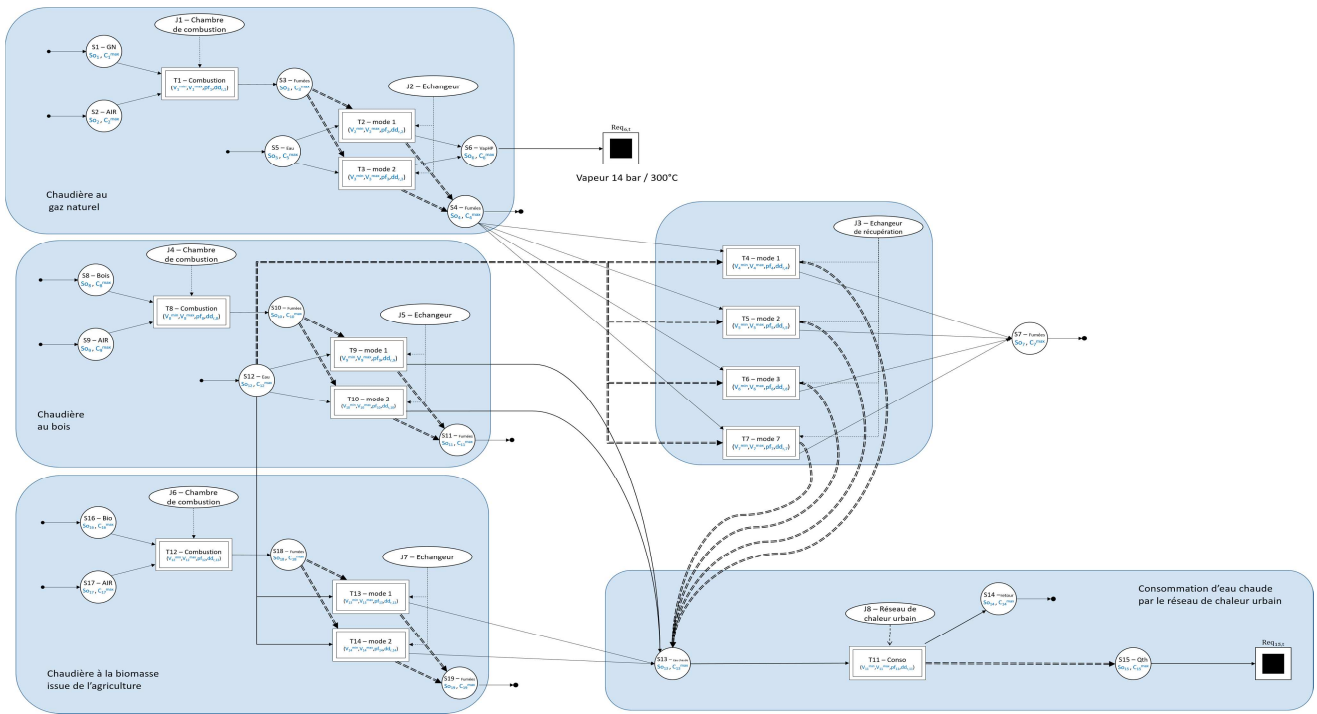


Figure 10 : Réseau ERTN complet de l'exemple correspondant à la recette de site

Sur la base de ces paramètres, la simulation du modèle ProSim[®] de la figure 7 permet de déterminer le débit et la température des fumées rejetées par la chaudière pour un débit donné de vapeur. En faisant varier le débit de vapeur entre 5t/h et 30t/h (plage usuelle de débit vapeur nécessaire aux unités de production du site) via une série de simulation, l'évolution de la température et du débit des fumées en sortie de l'échangeur est établie (figure 9). On déduit d'abord que, pour une demande en vapeur comprise entre 0 et 30 t/h, la plage de fonctionnement de l'opération de combustion varie entre 0 à 100 t/h d'où $V_1^{min} = 0$ et $V_1^{max} = 100$. Par ailleurs, cette simulation met en évidence la présence de non-linéarités.

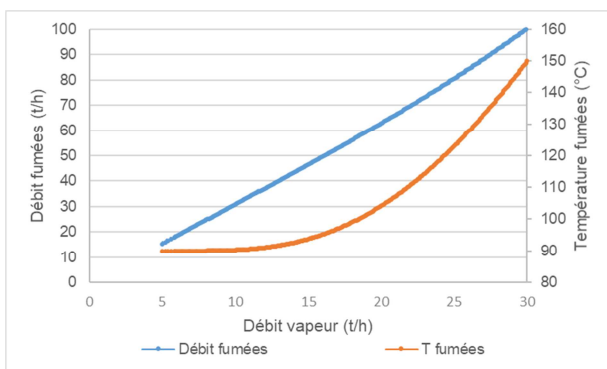


Figure 9 : Débit et température des fumées rejetées en fonction du débit de vapeur produit dans l'échangeur

Afin de tenir compte de ces non linéarités au sein du modèle de PLM de planification, cette relation est approximée au moyen d'une fonction linéaire continue par morceaux. Du point de vue du graphe ERTN, cela correspond à dupliquer l'opération d'échange en autant d'opérations, placées en exclusion mutuelle. Pour

l'exemple, la fonction est divisée en deux morceaux et l'opération d'échange est dupliquée en deux tâches T2 et T3. Les équations de ces deux droites, ainsi que leurs bornes, sont indiquées ci-dessous :

$$Q_{fumées} = 3,14 * Q_{vap} - 0,476 \text{ pour } Q_{vap} \in [5 ; 20,7] \text{ t/h}$$

$$Q_{fumées} = 3,80 * Q_{vap} - 14,3 \text{ pour } Q_{vap} \in [20,7 ; 30] \text{ t/h}$$

d'où la valeur des paramètres reportés ci-dessous :

$$V_2^{max} = V_3^{min} = 20,7$$

$$uv_{2,3}^{cons} = uv_{2,4}^{prod} = 3,14 \text{ et } uf_{2,3}^{cons} = uf_{2,4}^{prod} = -0,476$$

$$uv_{3,3}^{cons} = uv_{3,4}^{prod} = 3,8 \text{ et } uf_{3,3}^{cons} = uf_{3,4}^{prod}$$

En appliquant cette démarche pour chaque composant de l'exemple considéré, on déduit le réseau ERTN de la figure 10. Les opérations d'échange de chaque chaudière ont été dupliquées en deux, alors que celle assurant le couplage entre site industriel et site urbain a été décomposée en quatre tâches.

4.3 Etape 3 : planification du système

Le critère retenu vise à minimiser les coûts de fonctionnement, les apports de combustible (S1, S8, S16) ainsi que les rejets de fumées non valorisés (S4, S7, S11, S19) :

$$\min z = \sum_{t \in \{1, \dots, T\}} \left(\sum_{i \in \{1, 9, 12\}} 100 * W_{i,t} + \sum_{r \in \{1, 8, 16\}} 10 * In_{r,t} + \sum_{r \in \{4, 7, 11, 19\}} Out_{r,t} \right) \quad (14)$$

La figure 11 montre le plan de fonctionnement des unités sans valorisation de la chaleur fatale. Celui-ci fait apparaître les marches alternatives des différentes chaufferies et la capacité du système global à satisfaire les demandes

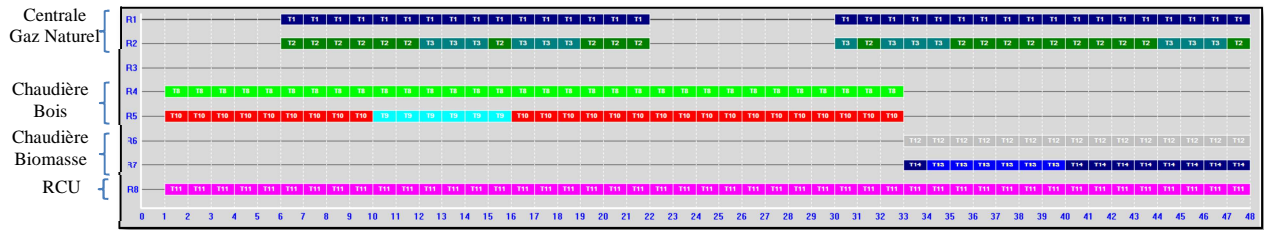


Figure 11 : Planification du cas d'étude sans valorisation

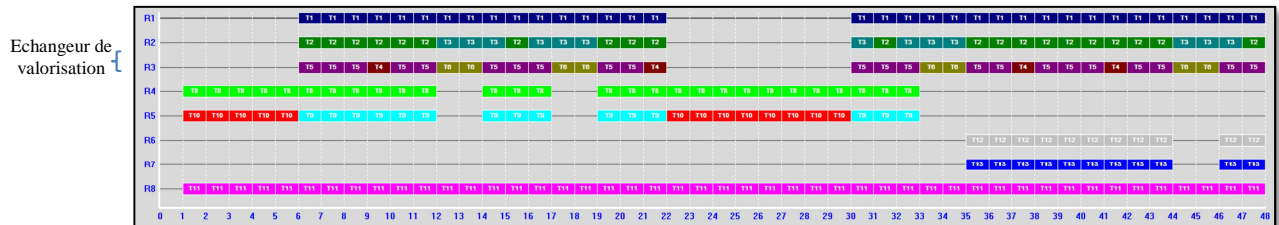


Figure 12 : Planification du cas d'étude avec valorisation

en vapeur et eau chaude du RCU tout en optimisant les coûts d'exploitation.

La figure 12 montre le plan de fonctionnement des unités avec valorisation de la chaleur fatale. On peut d'abord constater que l'échangeur de valorisation fonctionne sur toute la période de fonctionnement de la centrale de production de vapeur. Durant ces phases, on observe même des arrêts des chaudières propres au réseau de chaleur urbain (périodes 13, 14, 18, 19, 34, 35, 45 et 46). On obtient ainsi une meilleure efficacité (et donc rentabilité) de la chaudière vapeur UCPI puisque la chaleur initialement perdue par le rejet des fumées vient maintenant suppléer les chaudières du réseau urbain (amélioration confirmée par l'usage de tâches « bas régimes » par l'échangeur de la chaudière au combustible bois, R5). Cette amélioration du système global est confirmée par une réduction de la consommation de combustible (critère économique) et une diminution des rejets de CO₂ (critère environnemental).

5 CONCLUSION

Cet article a présenté une méthode d'élaboration d'un modèle de PLM dédié à la planification court-terme de chaîne logistique énergétique. Ce modèle est instancié avec les annotations du réseau ERTN associé, eux-mêmes établis, dans certain cas, par la simulation phénoménologique des composants de la chaîne logistique étudiée. Après résolution, différents indicateurs peuvent être calculés afin d'évaluer les potentialités de la solution.

Cependant, une limite de cette modélisation est que la température des états n'est pas une variable explicite du modèle, seulement une donnée (correspondant à une valeur moyenne sur une plage de débit donnée). C'est pourquoi les travaux en cours visent à introduire la température comme un potentiel dans la sémantique ERTN et comme une variable de décision du modèle. La principale conséquence est que le modèle se formule sous forme d'un programme non-linéaire mixte (formes bilinéaires), moins robuste du point de vue de la résolution.

REFERENCES

- ADEME, 2015. Chaleur fatale industrielle (La).
- Cajot, S., Peter, M., Bahu, J.-M., Koch, A., Maréchal, F., 2015. Energy Planning in the Urban Context: Challenges and Perspectives. Energy Procedia, 6th International Building Physics Conference, IBPC 2015 78, 3366–3371.
- Cerema, 2011. Constitution d'un réseau de chaleur - Réseaux de chaleur et territoires.
- Fabre, F., Hétreux, G., Le Lann, J.-M., Zaraté, P., 2011. Dynamic hybrid simulation of batch processes driven by a scheduling module. Computers & Chemical Engineering 35, 2098–2112.
- Georgiou, P.N., 2016. A bottom-up optimization model for the long-term energy planning of the Greek power supply sector integrating mainland and insular electric systems. Computers & Operations Research 66, 292–312.
- Hétreux, G., 2012. Modélisation et simulation dynamique hybride des systèmes complexes: concepts, et outils, Application en génie des procédés (Habilitation à diriger les recherches). Institut National Polytechnique de Toulouse, France.
- Kaşka, Ö., 2014. Energy and exergy analysis of an organic Rankine for power generation from waste heat recovery in steel industry. Energy Conversion and Management 77, 108–117.
- Liu, Y., Yu, S., Zhu, Y., Wang, D., Liu, J., 2018. Modeling, planning, application and management of energy systems for isolated areas: A review. Renewable and Sustainable Energy Reviews 82, 460–470.
- Nowicki, C., Gosselin, L., 2012. An Overview of Opportunities for Waste Heat Recovery and Thermal Integration in the Primary Aluminum Industry. JOM 64, 990–996.
- Théry, R., Hétreux, G., Agha, M.H., Haït, A., Lann, J.M.L., 2012. The extended resource task network: a framework for the combined scheduling of batch processes and CHP plants. International Journal of Production Research 50, 623–646.
- Tigges, K., Link, E., 2017. Simulation-based Diagnostic Concept for Waste Heat Recovery. MTZ ind 7, 44–49.
- Wahlroos, M., Pärssinen, M., Rinne, S., Syri, S., Manner, J., 2018. Future views on waste heat utilization – Case of data centers in Northern Europe. Renewable and Sustainable Energy Reviews 82, 1749–1764.

DATA VISUALIZATION IN CONCEPTUAL DESIGN: DEVELOPING A PROTOTYPE TO SUPPORT DECISION MAKING

Xin YI¹, Claude BARON¹, Rob VINGERHOEDS², Alessandro BERTONI³, Phillippe ESTEBAN¹

¹LAAS-CNRS, Université de Toulouse, CNRS, INSA, UPS, 7 av. du Colonel Roche, 31400 Toulouse, France
{xyi, claude.baron, philippe.esteban}@laas.fr

²Institut Supérieure de l'Aéronautique et de l'Espace, 10 av. Edouard Belin BP 54032, 31055 Toulouse, France
rob.vingerhoeds@isae-superaero.fr

³Blekinge institute of Technology, Karlskrona, Sweden.
alessandro.bertoni@bth.se

ABSTRACT:

In recent years, data visualization has received a lot of attention and widely applicate in many areas. Data is a treasure, a lot of valuable information hidden inside. While visualizing the hidden information concisely and intuitively is a critical task. In this paper, we explore how data visualization could support conceptual design. A specific conceptual design case for improving the fuel efficiency of wheel loader is implemented. By visualizing the complex fuel consumption data comprehensive for engineers to get insights, thus to support the decision making for a new product development. As a result, a table of classifying visual techniques by different features, and a prototype consisted of four approaches step by step to present complex data are proposed.

KEYWORDS: *complex data visualization, conceptual design, prototype, decision making*

1 INTRODUCTION

During the process of conceptual design, that takes place in the early stages of a system development project, several (often critical) decisions need to be made. A successful product is often the result of a combination of right decisions taken by engineers and managers (Bertoni et al., 2017). Uncertainties in decision-making associated with the need to shorten design cycle require increased skills and tools to support complex decision making (Platt and Huettel, 2008; Eriksson et al., 2008).

With the advent of digital technologies, important amounts of data are available, enabling engineers and project managers to access these data, explore patterns, proceed to several diagnoses and prognostics, and to collect insights for decision-making. However, effectively extracting and summarizing information so to avoid information overload and to enable its interpretation is a challenge (Keim et al., 2006). For this goal, numbers of visualization techniques have been developed to display massive data information.

The main object of this article is exploring suitable tools to visualize complex data information in a readable way in order to help engineers understand a current situation from a large collection of data before a new concept design. The paper considers a case study from product development to illustrate the method. Section 2 reviews the relative literature on conceptual design, data visualization, and the possibility as well as the challenge of applying data visualization to support conceptual design. Section 3 presents the case study and the process of design a visualization within the case, which starts with benchmarking visual techniques, followed by applying suitable tools into the case, then testing those techniques and analyzing the testing result. Section 4 presents a fi-

nal prototype. Section 5 discusses the contributions of the proposed prototype and limitations of the experiments. Section 6 draws come conclusions.

2 LITERATURE REVIEW

This chapter illustrates conceptual design, data visualization, and how data visualization can support conceptual design.

2.1 Conceptual design

Conceptual design is the first stage in the product development process. It identifies a basic solution path by the elaboration of a solution principle (Stump et al., 2004). The target of conceptual design is to study possible ideas, and after evaluation to select the best one(s) to satisfy the project goal (Stump et al., 2004).

In this stage, the comprehension of the users' needs and associated constraints is a critical activity. To support it, many computer-based tools are being developed and applied, for simulations, analyses, optimization, and visualization (Wang et al., 2002). The implementation of data visualization in concept development is a striking trend (Hasem et al., 2015).

2.2 Data visualization

(Card et al., 1999) explain data visualization as “*using computer-supported, interactive, visual representation of data to amplify cognition, where the main goal of insight is discovery, decision-making, and explanation.*” Thanks to the visualization technologies to raise availability of the abundant data, as well as the approaches to process it, data visualization will become the ‘next mass communication medium’ (Viegas and Wattenberg, 2011).

Although data visualization has existed for a long period of time, its need of serving for different purposes is growing rapidly (Donoho et al., 2000). A primary goal for data visualization is transferring information clearly and efficiently by different kinds of graphs and maps. In this research, we focus on engineering environments and explore suitable visualization tools to help engineers' visual the complex data. A variety of visualization tools are available for presenting data information; section 3.2 will later present a benchmarking table on commonly used visual tools.

2.3 Data visualization to support conceptual design

Decisions made during the conceptual design stage have a profound influence on a final product on the cost, performance, reliability, safety and environmental impact (Hsu and Liu, 2000). In addition, it has been proven that it is more challenging and costly to correct or compensate shortcomings of poor design concepts in later design stages (Hsu and Liu, 2000). It is therefore critical to support the conceptual design phase with advanced techniques so to best guide engineers.

With the advent of data collection and visualization techniques, many opportunities are available to allow flexible access to data source for value creation (Hasem et al., 2015). A clear structure of data visualization assists in knowledge production, insight gathering, relation discovery and understanding the information and underlying patterns. As such it is impacting conceptual design and decision-making (Shim et al., 2002).

(Kellen, 2005) discusses how imagery helps in making decisions. (Benn et al., 1994) indicates that imagery shows two important problem-solving and decision-making activities: distinguishing a problem from the symptoms, and deciding upon and implementing a course of action. The initial purpose of visualization is to make graphs and to support for structured thought processes (Few, 2009). Data visualization, on the one hand, gives the potential that enables people to become more engaged to select the presented information; on the other hand, it leads users to make discoveries and decisions.

2.4 The challenge of data visualization to support conceptual design

Relatively few work has been done on data visualization for the early of product design in an engineering environment (Graham et al., 2000). While data visualization tools used in these early stages has not proven to be as useful as in the final stage of product development (Anderson et al., 2004), (Mc Gown et al., 1998) concluded that many researchers have chosen to turn their attention to developing support techniques for later design stages. (Knopp, 1995) mentions that the technology push, coupled with a poor understanding of the design process, has led to the current lack of support for the conceptual designers. In this research, we explore how to visualize complex data for easy insights gathering by a concise

and easy readable way, thus to further support decision making in conceptual stage.

3 DATA VISUALIZATION DESIGN

This research follows the design research methodology proposed by (Blessing and Chackrabarti, 2009). Following this methodology, one case about wheel loader fuel consumption is applied; original data are gathered for visualizing. Following the benchmarking guidelines of (Camp, 1989), a range of visualization techniques are assembled and classified by distinguishing visual features. Subsequently, based on the needs of engineers, suitable tools are selected for first round visualization. After valuable feedbacks and suggestions are received from an experiment, empirical findings are contributing to the final visual prototype design.

3.1 Case background

Pushed by environmental protection and energy saving, construction industry desires to develop more efficient and energy-saving machines to reduce the effect on environment (Yeam et al., 2007). Improving current machine efficiency and developing new low energy consumption products are two effective solutions (Cronholm, 2013). Both solutions require a good understanding of the operation conditions of current machines (Bertino, et al., 2017); a set of operating data was collected from existing machines.

Data types in the case

The different data types guide to the different scope of selecting visualization tools in a case. In this case, 234 fuel consumptions items, from wheel loader working in different situations, are collected.

The needs

To help engineers understand the current machine' performance, some specific information is needed (Bertoni, 2017):

1. What is the general fuel consumption when operating in the different situation?
2. What is reason leading to different fuel consumption?
3. Which parameter player a greater impact on fuel consumption?

Envision data relationship

Data relations need to be revealed based on engineers' needs:

1. Showing the overall fuel consumption in one situation, which specifies tools with functions of displaying massive data and showing the relation of part to whole.
2. Displaying the different levels of fuel consumptions for pattern explore is indicating the relation of deviation.
3. Presenting the impact of one variable on the whole fuel performance.

What features of visualization tools are required?

Based on the needs and data type, the main features of visualization methods in this case are comparison, distribution, deviation, part to whole.

3.2 Benchmarking of visualization techniques

Although many novel data visualization techniques have been developed (Card et al., 1999), no visualization method is suitable to address all the problems (Grinstein, et al., 2002). A benchmarking on visualization methods was done (see Table 1 for the tools studied).

Visualization feature	Visualization techniques
Comparison	Bar chart
	Box-plot chart
	Butterfly chart
	Scatter plot
	Stacked bar chart
	Table
Distribution	Line chart
	Bubble chart
	Box plot
	Parallel sets
	Circle view
Deviation	Scatter plot
	Butterfly chart
	Waterfall chart
	Pie chart
	Parallel sets
Part to the whole	Stacked bar chart
	Treemap
	Waterfall chart
	Bar chart
	Bar chart

Table 1: Benchmarking result

3.3 Visualization techniques application

In order to allow information “hidden” in the data to become more comprehensive for engineers, the visualization is divided into three main scenes based on engineer’s needs: showing the overall fuel consumption, displaying the special cases, and comparing parameter impact on total value. The first scene aims to allow engineers obtaining an overall view of fuel consumption when wheel loader operating in different situations. Followed that, the second scene helps engineers get insights of the reason (which variables) that distinguishing the levels of fuel consumption. Then the third scenes move further to compare the impact of parameters on whole fuel performance.

Scene 1: Showing the overall fuel consumption.

In this condition, fuel consumption cases under the same operation are exhibited. Since the fuel performance varies a lot in the different operating scene, the different working scenes are designed separately. This allows for simplifying the visualization and for easier version assembling. Table 2 shows eight types of visual techniques used to exhibit the overall fuel consumption.

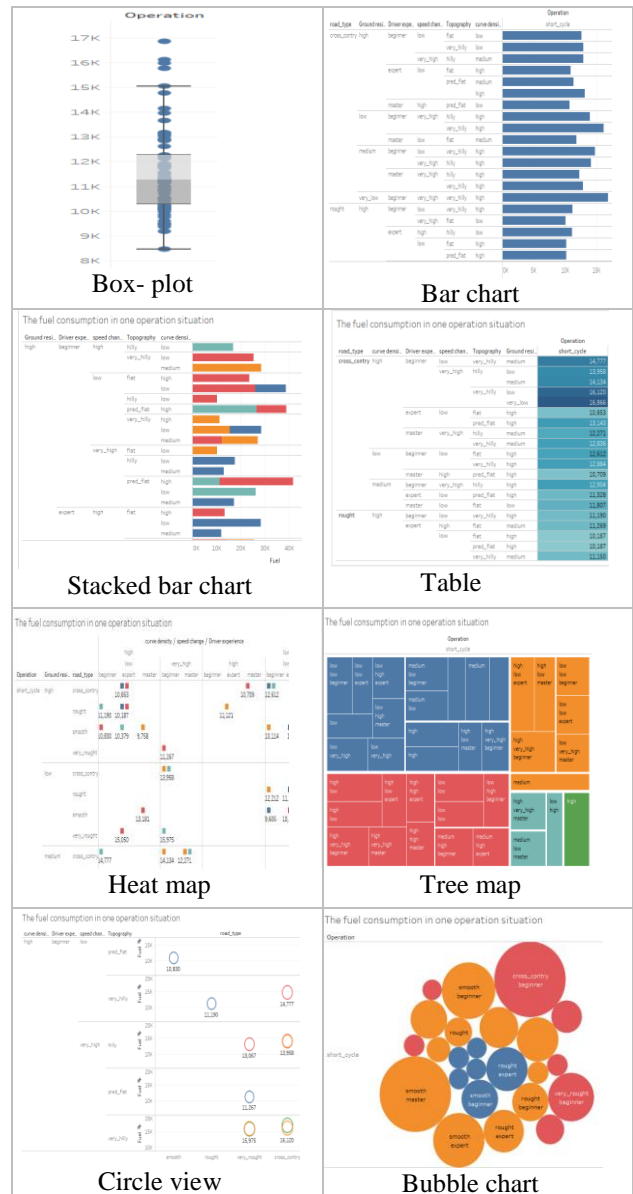
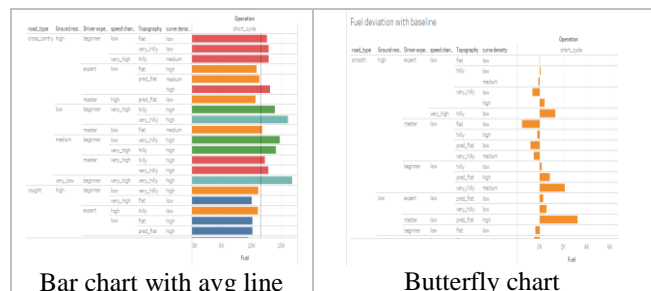


Table 2: All the visualization applied in scene 1

Scene 2: Displaying the special cases

The second visual scene is designed for indicating the special fuel performance occasions to bring convenience for engineers to figure out the reasons for fuel consumption variation. There are four tools are implemented, which are bar chart, butterfly chart, box-plot and scatter plot are presented in table 3.



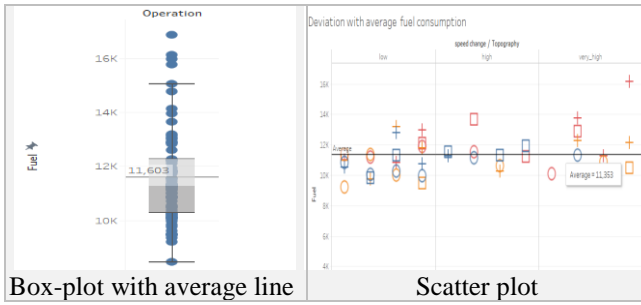


Table 3: All the visualizations applied in scene 2

Scene 3: Comparing parameter impact on total value
 The final visual situation concentrates on comparing the parameters impact on the total fuel performance. There are two modes of techniques utilized. Table 4 displays the result.

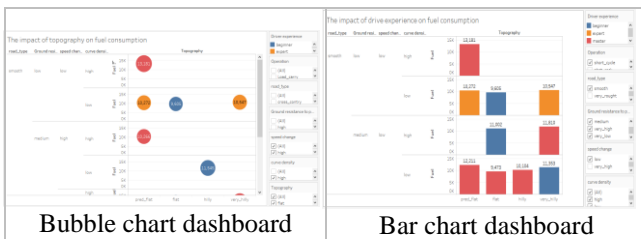


Table 4: All the visualization applied in scene 3

3.4 Experiment set-up

To test the intuitiveness of the chosen tools, an experiment was designed, and empirical data collected. The experiment involved four engineers and lasted one hour. It was started with 15 minutes introduction of task background and the visualization tools, followed by a 30 minutes session where engineers were asked to find patterns and get insights within the provided visualization techniques. After that, an additional 15 minutes session, where engineers communicated findings, sharing the suggestions for improving, and filled a questionnaire separately to score the tested tools.

3.4 Data analysis

Various techniques are applied in three visualization scenes, hence, the analysis is conducted separately under each situation to figure out which tool is more intuitive for engineers.

Table 5, 6 and 7 summarise the results on tool suitability from engineers’ perspective in each visualization scene. The technique that applied in per scene are listed in the left column, and the engineers’ labels are put in the first row. To the question “Do you think this method is clear to show the overall fuel performance?” engineers had to score all tools from zero to two (0=unintuitive, 2=intuitive). A final score for each tool is obtained by summing up the single scores.

Visual tools	EngA	EngB	EngC	EngD	Total Score
Circle view	1	0	0	0	1

Heat map	0	0	0	0	0
Table	2	1	1	0	4
Tree map	1	0	2	0	3
Staked bar chart	1	0	1	0	2
Bar chart	2	1	1	1	5
Box-plot	2	1	1	1	5
Packed bubble	0	0	1	1	2

Table 5: Analysis result in visual scene 1

In the scene 1, bar charts and box-plots were considered as the most readable tools to show the overview fuel consumption. Tables were also judged suitable. However, circle views and heat maps are clearly not applicable in this visual context.

Visual tools	EngA	EngB	EngC	EngD	Total Score
Bar chart	2	1	1	2	6
Butterfly chart	2	2	2	2	8
Box-plot	0	0	2	1	3
Scatter plot	1	2	1	1	5

Table 6: Analysis result in visual scene 2

From the analysis result, in the scene 2, all engineers agreed that butterfly charts are best suited to show the deviation. Bar charts also obtained a high score, while box-plots appear inappropriated in this visualization.

Visual tools	EngA	EngB	EngC	EngD	Total Score
Bar chart	2	0	2	0	4
Bubble chart	1	0	1	0	2

Table 7: Analysis result in visual scene 3

In the scene 3 that compares the impact of parameter to the whole fuel consumption, only two prototypes are made: a bar chart and a bubble chart. The bar chart gained more support than bubble charts in this scene.

4 PROTOTYPE

A number of visualization techniques have been applied to seek feedbacks from engineers of weighing the intuitiveness for insight gathering. Based on comments and suggestions, a final prototype consisted of several visualization techniques is put forward to deal with complex data visualization case of improved fuel efficiency from wheel loader. The visualization is divided into three scenes, including know the overall fuel consumption, patterns visualization in different clusters, and comparing the impact of parameters on total value. To corresponding each scene, four visual approaches are proposed, due to two different techniques are applied in scene two.

The logic of the designed prototype is to first present the overall fuel consumption, followed by pattern visualization in cluster. Furthermore, revealing the impact of a single variable change on the overall fuel consumption. Box-plots are used to show the overall fuel performance in one operation. See below Figure 1. This approach displays a large number of data in a concise and easy

understanding way, which had been proved on experiment. Clusters are marked with colours to divide fuel performance into five groups: low, middle, upper middle, high, and extreme. In each cluster, the number of items and average fuel level are displayed in Table 8.

Stage 1: know the overall fuel consumption

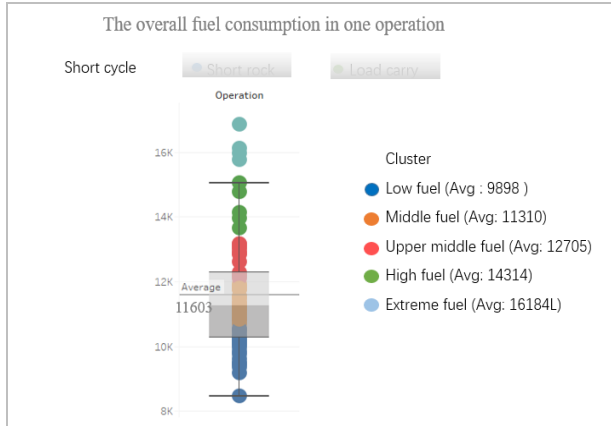


Figure 1: Approach 1 of showing the overall situation

In this visualization case, there are 22 numbers of items in cluster 1 and the average fuel consumption is 9898. Similarly, 26 numbers of elements in cluster 2 and with an average fuel consumption of 11310. Besides, there are 12, 5, 4 number of items with the average fuel consumption of 12705, 14314 and 16814 respectively in cluster 3, 4, and 5. Table 8 shows the cluster classification.

Cluster (different fuel consumption)	Number of items	Average fuel consumption
1: Low	22	9898
2: Middle	26	11301
3: Upper middle	12	12705
4: High	5	14314
5: Extreme	4	16184

Table 8: cluster classification

When clusters of fuel consumption are obtained, other questions should be addressed for further insights gathering. What is reason leading to the diverse levels of fuel consumption? What is the commonality feature of parameters in each cluster? To find answers for these questions, other visualization methods are implemented. The questions will be answered by the following visualization.

Stage 2: Patterns visualization within different cluster

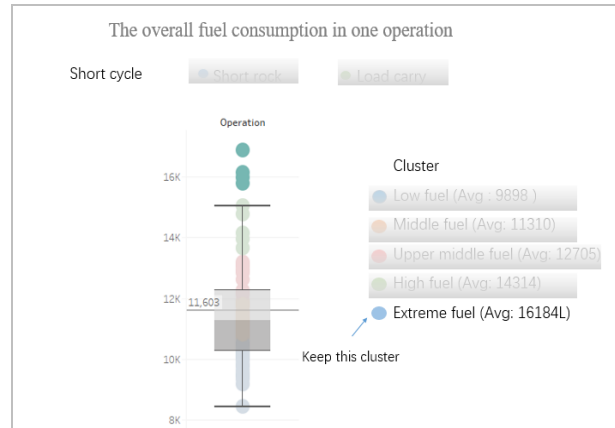


Figure 2: A cluster selection - Small amount of data sets for pattern visualization

The cluster marked in light blue (see from Figure 2), indicating the highest level of fuel performance is selected for insight observing. In this cluster, four items need to be explored further for patterns. A bar chart is used for data visualization since it with comparison function, simple for vision collection and engineers are familiar with this data visualization. Bar charts are selected to organize and visualize the deep information in the highest fuel consumption cases and identify what characteristic parameters are leading to such high fuel consumption. GIF-format information (graphic interchange format) next to each variable allows for describing different meanings of parameters. GIF combined with bar chart was retained as a solution to make visualization intelligible for engineers.

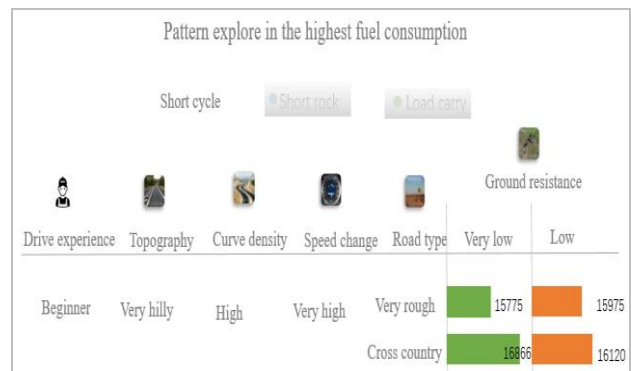


Figure 3: Approach two- Pattern explore in cluster with small number data sets

From this visualization (Figure 3), it is easy to conclude the similarity of parameters which contribute the high fuel consumption. In this case, a combination of extreme road types (very rough and cross-country) combined very hilly terrain, high curve density and very high-speed changes explained the high fuel consumption.

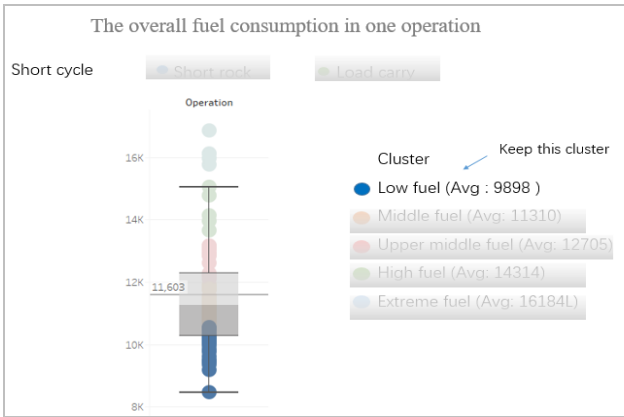


Figure 4: A cluster selection – Large amount of data sets for pattern visualization

Figure 4 displays the cluster 1, with 22 numbers of data sets, is chosen to explore commonality resulting in the lowest fuel consumption. The 22 numbers of data sets with 6 variables are not organized in the same way as in the previous visualization situation of cluster 5 that only had quintuple data sets. Consolidated the results of interviews and questionnaires, scatter plot with an added horizontal line has a higher advantage over the butterfly chart in an area of patterns exploring and parameters tracking particularly under the situation of many items and parameters. Therefore, a scatter plot substitutes the bar chart and butterfly chart to visualize patterns in a cluster with more data sets.

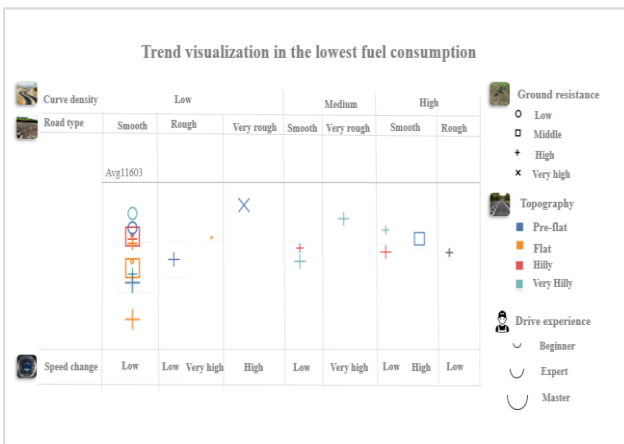


Figure 5: Approach three - Pattern explore in a cluster with more data sets

Scatter plots are applied for patterns visualization as well as with the feature of showing the fuel consumption deviation with the baseline under the situation of many data sets in one cluster. To make the visualization concise and intuitive for pattern explore, three variables (topography, ground resistance, and driver experience) are signed with different colors, symbols, and sizes in the visualization. As before, a GIF nearby each category relates to the corresponding situation of the parameter in the variables.

Many insights can be obtained from this prototype. In the visualization above, taking the pattern tracking by the size of the symbol for example, it is not hard to find

out that the symbol with big and middle sizes are account for the majority and only 3 symbols in small. It is here demonstrated that in most scenes of low fuel consumption, the machine is operated by a master and expert, as the middle size and big size are corresponding to expert and master respectively. Meanwhile, the three individual cases in small size symbols could be selected out to analyze further and figure out in which conditions that beginner also could lead to low fuel consumption. Furthermore, this visualization shows fuel consumption under the average performance – around half of the cases are gathered in the left half of the page, indicating low curve density situations with smooth road type and low speed changes.

The scatter plot provides approaches of patterns and commonality exploration in clusters. When necessary engineers can conclude of out some parameters leading to high or low fuel consumption which variables or parameters are of highest impact. The here presented visualization method allows for showing the effect of a variable on the overall fuel consumption. Bar charts provide a good method for comparative visualization. In this prototype, the highest fuel performance cases were selected to explore which parameters and variables had the highest impact on fuel consumption.

Stage 3: comparing variable’ impact on total value

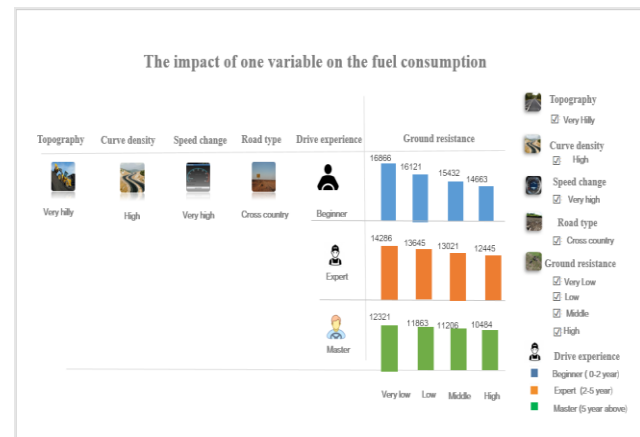


Figure 6: Approach Four - Comparing the impact of one variable to total value

In the visualization process, four variables have remained unchanged; driver experience and ground resistance are selected to make a comparison. In the right part of the dashboard, a filter is used for changing to different variables that engineers interested, and the corresponding data information would be presented in the left part of the dashboard. In this visualization, it is not hard to identify that with the ground resistance parameter changing from very low to high, the fuel consumption varies.

In comparison, fuel consumption varies more when altering drive experience from beginner to master under the same ground resistance circumstance. Therefore, it can be concluded the variable driver experience has a higher impact on total fuel consumption than ground resistance. In addition, the other insight is that fuel

consumption varies more when beginners and experts operate the machine.

5 DISCUSSION

In this section, the limitation of testing, the interesting findings from the experiment and the wider applicable scenarios of final prototype would be discussed below.

5.1 Test discussion

(Nielsen and Landauer, 1993) suggest that the best results in such research come from tests with no more than five participants and involving as many as small tests as possible. Bartlett et al. (2001) on the other hand suggest sample selection in one test to increase the reliability of data. Data visualization in design processes is a non-linear and iterative activity, which means that more tests are needed to collect feedbacks for continuing prototype improvement. So, even though in this study only four engineers were available, the by (Nielsen and Landauer, 1993) proposed approach was followed and feedback and suggestions for improvement of the tool, an experiment was carried out involving those four engineers.

5.2 Results discussion

Four different visualization approaches were investigated in series so to visualize complex data in the case of the fuel consumption improvement of a wheel loader. Even though the realized prototype is based on a special case, it is not limited only to this particular case situation. The following applicable scenarios can be seen for the prototype.

Approach one – show the overall fuel consumption

Box-plots were implemented here to show the cases of fuel performance. This approach shows to be useful only in a situation of presenting massive data from a bird view without requirements for detailed information. Color-marked clusters could be introduced to differentiate data sets so to bring convenience for further analysis (Fraley et al., 2002).

Approach two – pattern visualization in a cluster with less data sets

Similar levels of fuel consumption are assembled in one cluster to differentiate from others. Deep exploration takes place as to the reasons leading to the distinguished levels of fuel consumption in a small number of data sets. Bar charts allow for efficient data visualization of such analyses.

Approach three – pattern visualization in a cluster with more data sets

For more complex datasets, scatter plots express data in a horizontal direction bringing convenience for trend tracking. Various color, symbol and size are used to identify different variables for clear and easy pattern investigating. Scatter plots, combining the function of the commonality exploring and trend tracking, offer a

good tool for insights gathering.

Approach four – comparing the impact of variable on the whole value

Bar charts with an interactive interface allowing engineers to change parameters so to compare the impact of variable on the overall fuel consumption. GIF-pictures can offer an intuitive manner to represent different parameters in a comprehensive manner for engineers.

5.3 Discussion of the test results in relation with literature

Box-plots are considered as a useful tool to show a large number of data and highlight data distribution (Potter et al., 2006). This technique was applied and tested in two visualization situations of showing overall fuel consumption and showing the deviation with baseline. In this experiment, box-plots gained the highest scores for displaying the situation of overall fuel consumption. Low scores were obtained for showing the deviation. These results do not confirm results presented by (Potter et al., 2006) that box plots are a valuable tool to depict data as well as to spotlight data distribution. From the test, a box plot was seen not to be effective to get insights from the deviation perspective.

Scatter plots with added reference lines of average fuel consumption reached a consensus from interview results for showing deviations. They offer an interesting for box-plots and should be further investigated (see also (Kampstra, 2008)). The test also shows that scatter plots are more intuitive and easy for the pattern and trend tracking. Scatter plots with added reference lines could be taken into consideration to replace box-plots in visualization conditions in which there is the need to show deviation as well as patterns tracking of parameters

6 CONCLUSION

The aim of the research was carried out to explore how to visualize complex data comprehensively for easy insights gathering thus to support decision making in conceptual design. To reach this goal, the paper is presenting a prototype consisted of four approaches to visualize multidimensional data concisely and readably, showing “hidden” information and patterns, with as intend to support decision-making in the conceptual design phase. The research focused on complex data (multi-dimensional data) visualization and specifies an application to visualize the data of the fuel consumption of a wheel loader. The approaches in the final prototype, a single or in series, also work for other conceptual design cases when dealing with complex data visualization. The final prototype has been iteratively developed based on results of the analysis runs with engineers.

7 REFERENCES

- Anderson, E., & Lily, B. (2004). Visualization in the Design Process: Introducing two and three-dimensional sketching techniques to enhance creative thinking and communication. In International Engineering and Product Design Education Conference (pp. 2-3).
- Bartlett, J.E., Kotrlik, J.W. and Higgins, C.C., 2001. Organizational research: Determining appropriate sample size in survey research. *Information technology, learning, and performance journal*, 19(1), p.43.
- Benn, D.I., 1994. Fluted moraine formation and till genesis below a temperate valley glacier: Slettmarkbreen, Jotunheimen, southern Norway. *Sedimentology*, 41(2), pp.279-292.
- Bertoni, A.; Larsson, T.; Larsson, J.; Elfsberg, J. Mining (2017). Data to design value: a demonstrator in early design. Accepted for publication in Proceeding of the 21st International Conference on Engineering Design (ICED17), Vancouver, Canada.
- Blessing, L.T., Chakrabarti, A. (2009). "DRM: A Design Research Methodology", Springer London.
- Camp, R.C. (1989). Benchmarking: the search for industry best practices that lead to superior performance. In *Benchmarking: the search for industry best practices that lead to superior performance*. ASQC/Quality Resources.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.
- Cronholm, K., (2013). Design of Experiment based on VMEA (Variation Mode and Effect Analysis). *Procedia Engineering*, 66, pp.369-382.
- Donoho, D.L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, 1, p.32.
- Eriksson, J., Johnsson, S. and Olsson, R., (2008). Modelling decision-making in complex product development. In *DS 48: Proceedings DESIGN 2008, the 10th International Design Conference*, Dubrovnik, Croatia.
- Few, S., (2009). *Now you see it. Simple Visualization Techniques for Quantitative Analysis*, Analytics Press, Oakland, CA, pp.33-35.
- Fraley, C., & Raftery, A. E. (2002), "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, 97, 611-631
- Graham, M., Kennedy, J. and Benyon, D., (2000). Towards a methodology for developing visualizations. *International Journal of Human-Computer Studies*, 53(5), pp.789-807.
- Grinstein, G. G., Hoffman, P., Pickett, R. M., & Laskowski, S. J. (2002). Benchmark development for the evaluation of visualization for data mining. *Information visualization in data mining and knowledge discovery*, 129-176.
- Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. and Khan, S.U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47, pp.98-115.
- Hsu, W. and Liu, B., (2000). Conceptual design: issues and challenges. *Computer-Aided Design*, 32(14), pp.849-850.
- Platt, M.L. and Huettel, S.A., (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nature neuroscience*, 11(4), p.398.
- Kampstra, P. (2008). Beanplot: A boxplot alternative for visual comparison of distributions.
- Keim, D.A., Mansmann, F., Schneidewind, J. and Ziegler, H. (2006). Challenges in visual data analysis. In *Information Visualization, 2006. IV 2006. Tenth International Conference on* (pp. 9-16). IEEE.
- Kellen, V. (2005). Decision making and information visualization: research directions. Retrieved November, 11, p.2005.
- Knopp, R. and Humblet, P.A. (1995). Information capacity and power control in single-cell multiuser communications. In *Communications, 1995. ICC'95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on* (Vol. 1, pp. 331-335). IEEE.
- Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 206-213). ACM.
- Potter, K., Hagen, H., Kerren, A., & Dannenmann, P. (2006). Methods for presenting statistical information: The box plot. *Visualization of large and unstructured data sets*, 4, 97-106.
- Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization* (pp. 364-371).
- Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R. and Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision support systems*, 33(2), pp.111-126.
- Stump, G. M., Yukish, M., Simpson, T. W., & O'Hara, J. J. (2004). Trade space exploration of satellite datasets using a design by shopping paradigm. In *Aerospace Conference, 2004. Proceedings. 2004 IEEE* (Vol. 6, pp. 3885-3895). IEEE.
- Viegas, F., & Wattenberg, M. (2011). How to make data look sexy. *CNN.com*, April, 19.
- Wang, L., Shen, W., Xie, H., Neelamkavil, J. and Pardasani, A., (2002). Collaborative conceptual design—state of the art and future trends. *Computer-Aided Design*, 34(13), pp.981-996.
- Yeam I, Cavatorta JR, Ripoll DR, Kang BC, Jahn MM (2007) Functional dissection of naturally occurring amino acid substitutions in eIF4E that confers recessive potyvirus resistance in plants. *Plant Cell* 19: 2913-2928.

DEFINITION OF STATES AND MODES AS GENERAL CONCEPTS FOR SYSTEM DESIGN AND VALIDATION

R. BADUEL^{1,2} J.-M. BRUEL¹ I. OBER¹ E. DOBA²

¹IRIT / University of Toulouse
31400 Toulouse - France

²Bombardier Transport
59154 Crespin - France

{ronan.baduel,bruel,iulian.ober}@irit.fr eddy.doba@rail.bombardier.com

ABSTRACT: *This paper advocates the need of a precise definition and consistent use of the notion of states and modes in system engineering and why this is important for the early validation of system behavior. This work is part of a solution developed for Bombardier Transport in order to provide continued validation of a train behavior across all development steps. The aim is to define concepts that support and generalize the application and use of states and modes in a way that is compatible with current practice, languages and tools. The finality of our work is to provide models that specify and integrate the whole system behavior at a high level of abstraction (in both granularity and details of the information characterized), enabling its validation and then supporting traceability of the behavior and its validation across the development process. This work is conducted under the principles of Model Based Systems Engineering (MBSE) and is linked to the use of SysML models. The examples presented are taken from the train industry.*

KEYWORDS: *Early Validation, MBSE, SysML, State, Mode*

1 INTRODUCTION

Specifying a system behavior is generally done early in the development process through behavior models, such as state diagrams. If we consider the state diagram in SysML (OMG, 2017), which comes from UML (OMG, 2015), we see that what a state represents regarding the system or its behavior is not specified. It is an issue raised by Naumenko in his thesis (Naumenko, 2002) regarding UML: “*existing UML semantics are very ambiguous in presenting relations between models constructed using the language on the one hand and the subject that is being modeled on the other hand*”. While the representation and semantics of states regarding the model are specified in the standard, it is not clear what a state actually represents regarding the system to be modeled. Aside from UML, there are “state” elements or concepts used in different languages, tools and methods with different semantics, such as UPPAAL (Behrmann et al., 2006) or in Abstract State Machines (ASM) (Börger & Stärk, 2003).

There is no consensus on what a state is (Wasson, 2010 & Olver, 2014), and it is often confused with the notion of mode, which definition is also debated. When referring to the standard ISO/IEC/IEEE 24765, which references other standards, we see that there are a variety of definitions of what a state is. Since there is a confusion between states and modes but both are widely used in systems

engineering, it is necessary to define both concepts in the scope of our work. Bombardier Transport (BT) uses states and modes in its models but lacks a definition of the corresponding concepts. This leads to a difficulty in capturing and integrating the right information (such as preconditions of functions) in executable models, which are themselves state machines.

Beyond the definitions of states and modes, we are interested in their utilization. While in SysML the states are used to model a behavior, we see in the definitions proposed by Wasson that we can define different kinds of states and modes, capturing different information. Such information can describe or impact the behavior of the system without being able to fully describe, condition or model it. When modeling the system behavior in a SysML state machine, as a synthesis of the information on those conditions, the information used for the conditions has its own evolution (transitions) with specific conditions. This leads to the creation of other state diagrams or constraints on variables characterizing information about the system and not its behavior. It means that we have to identify and define each type of condition-related information in order to model the behavior.

This paper is organized as follows: in section 2 we present some issues encountered in the train industry that we address through a proper definition of states and modes. In section 3 we analyze different definitions of states and modes and propose our owns, for

which we present a concrete application in section 4. Section 5 is a conclusion on what is presented in previous sections and introduce the future applications of this work for early validation.

2 CONTEXT AND NEEDS

BT has expressed the following needs concerning the train design:

- There is a need for means to validate the behavior early and then across the whole development process. This is currently done in a test-driven approach, which does not cover unwanted and unexpected behaviors nor does it validate specifications before implementation.
- There is a need for the reuse of the functional specifications of a train. While train components and subsystems are well-mastered, what changes from one train to another is the functional specification, which is difficult to reuse, change or adapt.
- There is a need to model the information used in preconditions of the system functions and scenarios. They are currently specified only in requirements, matrix and spreadsheet files which renders them difficult to analyze and refer to in behavior specifications.

In order to validate the train behavior at each step of the development process, we need an abstract representation of the expected behavior that we can compare to the implementation. This is why we need an executable model integrating the information of the specifications. This model would be built around state machines. We aim for an approach similar to the one using the OMAG tool (Chapurlat & Daclin, 2013). Such a representation would allow to define a general and reusable functional structure, which is part of BT needs. BT is developing a solution using state machines to represent lower levels of granularity, such as functions in subsystems. Those executable models would include preconditions and information linked to them, which we have to represent. Such a representation exists for example for the train operability (the train readiness and activation status regarding the functions that can be performed), using an informal state machine. We want to formalize this process and extend it to other kinds of information.

As each language and tool have their own definition and utilization of the notion of state and/or mode, the concepts defined here will refer to common characteristics taken from several sources. They are developed so as to be useful in the scope and needs defined by BT but are not specific to the train domain and could be used for general system development.

3 STATES AND MODES

3.1 State definition

3.1.1 State of the art

We consider several definitions:

- According to the standard ISO/IEC/IEEE 24765, a state can be a characterization of the system at a given time, the value of the variables defining the system, a condition to a behavior or a function, something that determines the set of functions that can be performed, and other meanings.
- According to a system engineering handbook supported by INCOSE (Walden et al., 2015): “A system is in a state when the values assigned to its attributes remain constant or steady for a meaningful period of time”. It is otherwise specified that the handbook refers to the norm ISO/IEC/IEEE 24765 for vocabulary.
- According to (Naumenko, 2002), a state is one of two concepts at the basis of the semantic he defines, and its definition is: “an information about a thing (object) at a given time (point in continuum) inside a context (time continuum)”.
- According to (Wasson, 2010), a state is “An attribute used to characterize the current logistical employment, status, or performance-based condition of a system”.

Not all definitions available or considered are given here, only a representative group issued from formal or common sources. Some of the sources cited in this document, such as (Olver & Ryan, 2014), have already done similar work and reference different definitions, and can be consulted as additional references.

Besides the definition of states, we consider different kinds of elements used to represent them. In many finite state-machines such as UPPAAL, we have one state active at a time inside a state space without hierarchy or concurrency, at least in one diagram. In UML and SysML state machines or in Harel state-charts (Harel, 1987), we have complex state structures with concurrent states in one diagram, meaning we can have several states active at the same time. In Abstract State Machines, or ASM (Börger & Stärk, 2003), the state of the system is represented by a set of variables, called states variables, and a state is the valuation of these variable at a point in time. This can be linked to one of the definitions found in ISO/IEC/IEEE 24765. As in state diagrams, ASM allows to define transitions, conditions and constraints.

Another point is that state can be represented by discrete or continuous components, something that is highlighted in hybrid automata (Henzinger, 2000).

The state definition in UML can not be used as a definition of the concept. As per the analysis made by Naumenko, these elements are not directly associated to the system of study. In addition, UML defines states to model a behavior in what is called “behavioral state machines”, but we can define states that affect the behavior without modeling it. Taking the example of a train, we can define a state that indicate the current energy supply, be it internal or external sources. It will affect its operability and hence its behavior but does not condition capabilities directly. States can capture information used to condition capabilities. This information cannot be used to directly model the behavior, as each piece of information could be used as part of the preconditions of any capability. To represent the behavior, we have to define elements characterizing the execution of capabilities and not the information used to allow it, which seems to be the goal of the state machines in UML. This is supported by the definition found in the standard UML 2.5 (OMG, 2015): speaking of the State model element, “*a state models a situation during which some (usually implicit) invariant condition holds*”.

3.1.2 Analysis

Cross-referencing the definitions given above, we can establish a few characteristics of states:

- They characterize a thing (e.g., a system).
- They relate to a specific kind of information, knowledge domain or way of being regarding this thing (e.g., operations, readiness, energy, ...).
- They are evaluated or considered *at a given time*.

The most appropriate definition would be the one proposed by Naumenko, as it is high level, relates to the object of study that we want to model, and summarizes the main characteristics of states that are common in the definitions considered. In addition, it was thought as a core element of a semantic used to give a common meaning and understanding in the definition of models regarding the system they represent. It is one of our objectives regarding modes and states: to make it clear how to define them and what they represent regardless of which language or tool we use. It also presents the advantage of considering a temporal context in which states can be defined, which should be the system life-cycle. We will keep this notion of context as it can be applied to the other definitions.

This seems to contradict the definition given in the INCOSE engineering handbook (Walden et al., 2015),

but we argue that it is not the case. A state will always be considered at a point in time, where it will be evaluated. If we get the same values or few variations of them in regard of the information characterized during meaningful periods of time, then they can be abstracted as fixed values over time. Searching for persistent values or intervals of values can be a criterion to create state diagrams or other such models, but we consider that is it not a part of the state concept definition. We consider that a state has one fixed value at a point in time, and it can be different or not from the value at the previous or next point in time. That is why states can be modeled by variables as well as diagrams.

In SysML state machines, we can have several “states” active at the same time. By “state” we mean the modeling element rather than the concept. An “active state” is either one of the higher level “states”, or one of the “sub-states” contained in a higher level “state” or a concurrent partition. It means that there can be only one “active state” in a group of “states” or “sub-states” at a same hierarchical level and per concurrent process. We hence consider that each of these “active states” exist in their own context. It enables us to keep the definition (concept) of state as an evaluation in a time context, though defining several types of states characterizing different targets in a same diagram may not be pertinent.

The point of view from which the state is defined is also important. If we consider an unmoving train and want to evaluate its operational use, then depending on the person, it could be considered “parked”, “waiting”, “in a mission”, “reserved”, “ready for departure”, etc. When evaluating an information through the measure of physical attributes, the point of view indicates who access the information: the train control system, the pilot or an external monitoring system. Depending on how the information is accessed and on the scope of study, there can be differences in the information received.

Engineers try to define a hierarchy of states. For example, Wasson (Wasson, 2010) considers that “system states” (characterizing the use being made of the system) contain “operational states” (characterizing the mission readiness of the system). But it seems logical that a system could have a same readiness status in two different missions. Acknowledging that Wasson’s scope is oriented to the definition of scenarios and mission specifications phase by phase, we consider that in general two types of state of a same object characterize different information that may have dependencies and constraints between them but exist at a same level.

3.1.3 Definition

Using the definition proposed by Naumenko, we consider the concept of *state* as “an information about a thing at a given time inside a context”. In order to be independent of the representation of a state, we want to use informal attributes to characterize it. In order to adapt our definition to the development of complex systems, we specify two attributes impacting the information characterized: the point of view and the level of granularity. We saw that the definition of a state was different depending on who is considering the information regarding the object characterized. The level of granularity, here divided into system, subsystem and component level, enables us to define a same information at different level of details along the development process. For example, a same state of the train energy supply can be evaluated as “internal supply depleted”, “battery depleted” or a voltage measure depending on the level of granularity. This is also how we could ensure traceability of information and state along the development process, by refining states. According to this analysis, the attributes necessary to define a *type of state* are given in figure 1.

ATTRIBUTE	DEFINITION
Target	The entity characterized
Information	The type of information considered
Context	The full scope under which the state exist and can be defined
Granularity level	The level of granularity of the elements linked to the information considered
View	The point of view by which the state is expressed

Figure 1 – Attributes used to define a state type

We link our definition of *state* to concepts enabling to model it using different languages and modeling objects:

- A *type of state* is a type of information linked to the object it characterizes and the time context where both the information and the object are defined. A type of state is defined using a set of states variables.
- A *state variable* is used to express the information characterized by a type of state. A state variable has values defined over the whole time context where the type of state is defined.
- A *state space* is the set of all possible configurations of values of the state variables of a type of state.
- A *state* is the evaluation of the state variables of the type of state it belongs to. Each Element of

a state space correspond to a possible state.

We do not consider modeling concepts such as transitions, initial states or guards in our definitions. Our goal is only to know what a state represents regarding the object to model and what are the kind of elements needed to model it. The representation and evolution depend on the modeling language.

3.1.4 Example

We illustrate our definition on a simple example: checking whether a common type of door is open or closed. Our target is a door, the information is the opening status and the time context is the time where the door remains installed and in good shape. The point of view is the one of the door, the level of granularity is the door as a system. We use as a *state variable* the angle made by the plane of the door regarding the one of its frame. The *state space* is any value between 0 and 90 degrees, assuming that the door can not be opened further than at a right angle. The *state* of the door is evaluated by measuring the angle at a given time. If the angle is less than 5 degree, the door is considered closed. We can abstract the *state variable* as one having for state space only two elements, “open” and “closed”, as it is the only information we are interested in. We can then represent the *state variable* in a state diagram, with two state values. We see that a same *type of state* can be represented in several ways, and that it does not necessarily keep the same information value for extended periods of time, as we could place the door at an angle over 5 degree just for an instant before closing it again.

3.2 Mode definition

3.2.1 State of the art

Again, we consider several definitions:

- According to the standard ISO/IEC/IEEE 24765, a mode is a set of related features or functional capabilities of a product.
- According to the SMC Systems Engineering Handbook (SMC, 2005), “*The condition of a system or subsystem in a certain state when specific capabilities (or functions) are valid*”.
- According to Wasson (Wasson, 2010), a mode is “*An abstract label applied to a user (UML Actor) selectable option that enables a set of use case-based system capabilities*”.
- According to the Arcadia method (Voinin, 2017), a mode is “*a behavior expected of the system [...] in some chosen conditions*”.

Modes are linked to actions, functions or capabilities. As there are debates about what is a state and a mode, we also analyze states definitions that relate to capabilities:

- According to the standard ISO/IEC/IEEE 24765, a state can be “*something that determine the set of functions that are possible or can be performed*”.
- According to the SMC Systems Engineering Handbook (SMC, 2005), a state is “*the condition of a system or subsystem when specific modes or capabilities (or functions) are valid*”.
- According to the Arcadia method (Voirin, 2017), tooled by Capella, a state is “*a behavior undergone by the system [...] in some conditions imposed by the environment*”.
- According to Jenney (Jenney, 2011), “*States define an exact operating condition of a system, where modes define the set of capabilities or functions which are valid for the current operating condition*”.

3.2.2 Analysis

Considering those definitions and other sources, we have identified several attributes and characteristics about modes:

- They characterize the behavior of a thing (e.g., the behavior of the system under this mode).
- They express a behavior regarding a set of capabilities, functions or actions (e.g., moving forward or backward, performing flight maneuver, etc.).
- They are defined for a set of conditions (e.g., specific states of the system).

We use the following definition found in ISO/IEC/IEEE 24765 for the concept of behavior: “*the peculiar reaction of a thing under given circumstances*”. There is a distinction between the behavior characterized and the behavior expressed. The former is the whole behavior of the thing characterized, hence its reaction under any circumstances. The behavior expressed by a mode represents part or whole of the behavior of the thing, that is specific reactions under the conditions for which a mode is defined. For example, there are capabilities on a smart-phone that are valid when the phone is connected to a wifi, and that can be characterized by a “wifi mode”, but this mode does not characterize capabilities such as calls, SMS, etc. which are part

of the global behavior of the phone. “Wifi mode” could have sub-modes, such as “automatic updates”, meaning we can characterize behaviors that are not always those of the phone, but those of its main capabilities.

Behavior under specific conditions is something that we find in UML “states”. “State” elements in UML enable to call operations and are used to model a behavior. Regarding the “sub-states”, we saw they could not always be considered as state variables, making it hard to link them to the state concept. However, according to the criteria listed above, UML “states” and “sub-states” elements could correspond to modes regarding the concept. A type of mode is not always defined for the whole time context of the object it characterizes.

Individual capabilities can be characterized by more than one mode. They can have their own conditions or events needed to execute them, which can be modeled as preconditions (guards) or with others modes. Most capabilities in a train are conditioned by the train operability modes, as what can be done depends on the energy supply and the level of activation. But there are then functional modes conditioning their execution even when they are technically possible, like opening doors or going forward in a train fully activated. Modes can share the same conditions, meaning they can exist at a same time, but still be independent from each other, meaning we can not build a hierarchy around them. That may be why in the UML definition of state, which we associate to the notion of mode, the conditions are said to be often implicit.

Being in a mode or not can be an information characterized by a state. We can have a set of modes linked by transitions and defined over the whole system life-cycle. But a mode is a notion different from the one of state. A state is evaluated at a point in time, and its value(s) can possibly hold only for an instant. A mode is defined independently from time, as its conditions can hold only for an instant but characterize behaviors that last in time. As a mode is linked to conditions on state values, it may be why states are sometime defined as values holding for a period of time. We can conclude that while modes and states can be linked, they correspond to two different notions. A state characterizes a way of being of the system at a given time, a mode characterizes the way the system behaves for a particular state value. A type of state represents some knowledge about the system, a mode is part of the specification of its behavior.

Note that a mode can be abstract. A mode could for example characterize capabilities according to the way the system is used, meaning characterizing a performed behavior, which may not cover all potential behaviors. There are capabilities one wishes to in-

hibit, authorize or constrain but not as part of the system specifications. For a train, there are restrictions when you drive in a station or in an urban area, representing an information that could be included in a model of the behavior.

3.2.3 Definition

We define a mode as a characterization of a set of capabilities of a thing under a set of invariant conditions. It specifies part of its target's behavior. To characterize a mode, we need to know the source of the behavior, the kind of characterization made, the capabilities and the conditions. In order to adapt this definition to system modeling, we have to consider another attribute: the level of granularity, which characterizes the level of detail of the capabilities and whether they are attributed to the system, a subsystem or a component. The point of view is always internal, as we characterize the system capabilities. The conditions and characterization could originate from an external source, for example if we define modes of utilization, but the behavior characterized would be either the system's or the user's. According to our previous analysis, the attributes necessary to define a mode are given in figure 2.

ATTRIBUTE	DEFINITION
Source	The source of the behavior
Capabilities	The capabilities characterized by the mode
Conditions	The conditions for which the mode is defined
Characterization	The ways capabilities are characterized by the mode
Granularity level	The level of granularity linked to the capabilities characterized

Figure 2 – Attributes used to define a mode

We then link this definition to concepts enabling to model it using different languages and modeling objects:

- A *type of mode* specifies the behavior of a thing (e.g., system, function, group of functions) regarding a specific scope of study (operability). This scope is defined by sets of capabilities linked to a kind of characterization and sets of conditions.
- A *capability* is something realized by the thing characterized (e.g., actions, operations or functions)
- A *characterization* is a constraint put on a capability (e.g., enabling, conditioning, inhibiting or calling)

- A *set of conditions* is defined by conditions on a set of states variables
- A *mode space* is the set of all possible configurations of characterized capabilities and sets of conditions for a type of mode
- A *mode* is a behavior of a thing. Each element of a mode space corresponds to a mode.

3.2.4 Example

We illustrate our definition using the door example from earlier: we qualify the behavior of a door. We characterize the *capability* of a door to be opened. The *characterization* of this *capability* is enabling it. The *condition* for which the *mode* is defined is that the door is unlocked. The *level of granularity* is the one of the SOI. At the *condition* that the door is unlocked, we are in a *mode* where the *capability* “open the door” is enabled. Whether the door is locked or not could be evaluated with a *state*, which would not characterize the behavior.

4 APPLICATION

4.1 State and mode definition

An example of a state and mode definition according to our approach is given here. We start from a real train state diagram given in figure 3, which is used to specify scenarios and represent high-level conditions of the main functions of a *consist* (a consist is a vehicle forming a train. A train can have several consists). This diagram is also part of a document describing an offer to a client, meaning BT is contractually bound to respect its specifications. The guards on the transitions have been removed for more clarity. We use simple graphical representations that are not linked to a specific tool or language, as we could use several of them to represent or use the information considered.

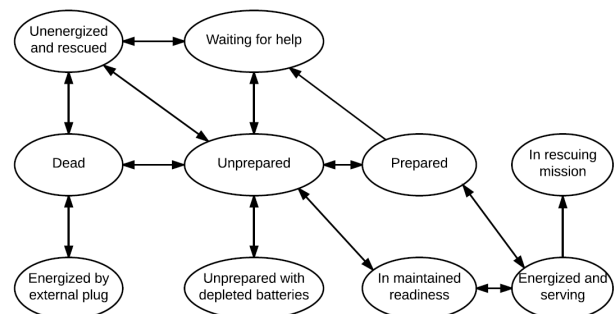


Figure 3 – Original state diagram characterizing the *consist* operability

We now evaluate the attributes that we used to characterize states in figure 4.

ATTRIBUTE	DEFINITION
Target	Consist
Information	Operability, energy supply, mission
Context	Train life-cycle
Granularity level	System or Component
View	Consist

Figure 4 – Operability state attributes

The states in this diagram provide three different kinds of information: the operability, clearly defined, the energy supply, only partially defined, and in one instance the mission (being in a rescue mission). All information are not defined over the whole context, meaning that the state space is incomplete. Mentioning batteries relate to the components granularity level, while the rest relate to system level. To settle those issues, we define two new types of states in figure 5 and figure 6, one for the operability and one for the energy supply. Dismissing the information regarding the *consist* mission, which was not in the original scope, both states now relate to the same level of granularity as the SOI. The other attributes do not change.

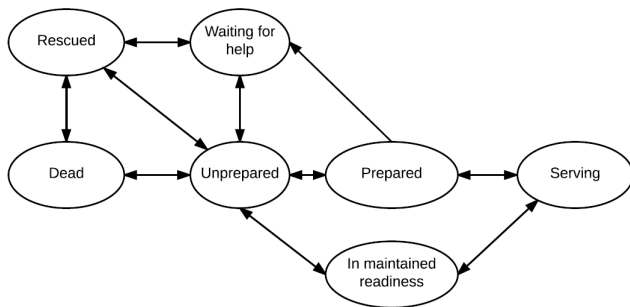


Figure 5 – Operability state of the *consist*

The dot lines in figure 6 characterize possible transitions that were not considered in the original diagram and became evident after a separate analysis. Note that the operability is only partially determined by the energy supply. The next step would be to define the constraints between the operability and the energy supply.

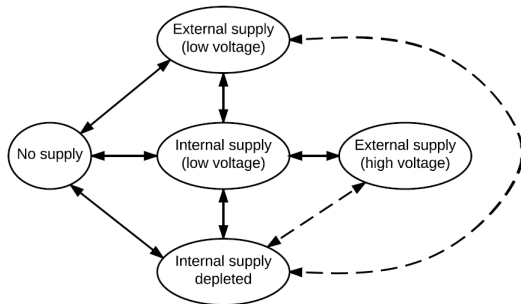


Figure 6 – Energy supply state of the *consist*

The operability state can be linked to operability modes, as we can characterize the train behavior depending on the *consist* readiness state. We define a type of mode for the operability in figure 7. For each mode of this type, we would have to specify the value(s) of the operability state as conditions as well as the detail of every functions and their characterization under this mode. For example, we could define a mode that has for condition the state “dead” and that enables the capability “power the train”.

ATTRIBUTE	DEFINITION
Source	Consist
Capabilities	All capabilities of the consist needing energy and system activation status to be performed
Conditions	Operability
Characterization	Enabling capabilities
Granularity level	System

Figure 7 – Attributes defining the operability modes

We consider another type of mode, illustrated in figure 8. This is a SysML state machine illustrating the behavior of a *TCMS* function, “Manage the TCMS HMI”, the *TCMS* being a subsystem of the *consist* and our SOI here. Another mode of the function is represented in the diagram, represented by a black circle connected to “Manage TCMS HMI”. In this mode, none of the capabilities are enabled. As said earlier, we consider the SysML “state” elements as modes. We define the attributes of the mode “Manage the TCMS HMI” in figure 9.

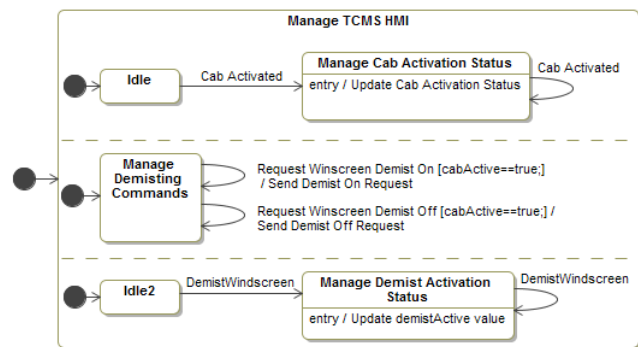


Figure 8 – Modes of a *TCMS* function

The conditions are not visible in this diagram, they are found in the documentation. This missing information could be represented by the system’s and sub-system’s states and modes, using variables. The mode characterizes the function’s behavior, not the SOI’s. Functions are abstract notions, and we prefer to not define states for them aside from modes modeling their behavior, as they are too numerous and can exist in specific time contexts. We also see that each capability is characterized by a different be-

ATTRIBUTE	DEFINITION
Source	A function managing the system interface
Capabilities	Managing the cab activation status Managing the demisting status Managing the demisting requests
Conditions	Unknown/not specified
Characterization	Enabling capabilities
Granularity level	Subsystem

Figure 9 – Attributes of “Manage the TCMS HMI”

havior with their own modes. There are information about components that could be abstracted. For a same level of granularity regarding the SOI, we can have a hierarchy of capabilities with their own behaviors, modes and time contexts. It should also be possible to model such a behavior for the *consist* itself, without referring to lower levels of granularity regarding the system. We see that beyond describing the expected behavior of the system, modes can be used to describe how it will be implemented.

4.2 Benefits

An improper definition of states and modes can have serious consequences when designing a system. There was a case where BT, as per contract, had to provide a train with troubleshooting functions executable when alimeted by batteries at low train voltage (under 63V). However, the supplied battery displayed a low level status under 50 volts, for which no functions can be executed. Due to those similar terms, BT let suppose it provided troubleshooting functions available for any voltage under 63V, even under 50V where it was no longer possible. This would have been avoided by specifying the proper target, information and granularity level for each type of state.

5 CONCLUSION

We presented in this paper a definition of states and modes as concepts that can be used in different languages and semantics, and hinted at their usefulness in specification and design of the system behavior. This is a first step in a work aiming to perform early validation in a continuous way. Indeed, states as they are defined here represent information about the system that follow its development through abstraction levels, and modes correspond to specifications of its behavior, based on conditions on state values. This should allow us to analyze and validate both the system and its behavior before even defining its functions, and then trace it to any model or product developed after. A method is currently developed with this in mind, which will be supported by a tool chain.

ACKNOWLEDGMENTS

This work is supported by Bombardier Transport SAS and the ANRT CIFRE grant #2016/0262.

REFERENCES

- Behrmann, G., David, A. & Larsen, K. G., 2006. A Tutorial on Uppaal 4.0 *Formal methods for the design of real-time systems*, p. 200-236, Springer.
- Börger, E. & Stärk, R., 2003. *Abstract state machines: a method for high-level system design and analysis*, Springer Science & Business Media.
- Chapurlat, V., Daclin, N., 2013. Proposition of a guide for investigating, modeling and analyzing system operating modes: OMAG. LIG2P, ENS Mines d’Alès
- Harel, D., 1987. Statecharts: A visual formalism for complex systems. *Science of computer programming*, 8(3): p.231-274.
- Henzinger, T., 2000. The theory of hybrid automata. *Verification of Digital and Hybrid Systems*, Springer, p.265–292.
- Jenney, J., 2011. Define Life Cycle System Modes. (See <http://themanagersguide.blogspot.com>)
- Naumenko, A., 2002. *Triune Continuum Paradigm: a paradigm for General System Modeling and its applications for UML and RM-ODP*, PhD., EPFL, Switzerland.
- Olver, A. M. & Ryan, M. J., 2014. On a Useful Taxonomy of Phases, Modes, and States in Systems Engineering. *SETE 2014*, Adelaide, Australia.
- OMG, 2015. *Unified Modeling Language*, version 2.5. (See <http://www.omg.org>)
- OMG, 2017. *Systems Modeling Language*, version 1.5. (See <http://www.omg.org>)
- SMC, 2005. *Systems Engineering Primer & Handbook*, United States Air Force SMC.
- Voirin, J., 2017. *Model-based System and Architecture Engineering with the Arcadia Method*. Elsevier.
- Walden, D. D., Roedler, G. J., Forsberg, K., Hamelin, R. D., Shortell, 2015. *Systems engineering handbook: a guide for system life cycle processes and activities*. INCOSE, Wiley.
- Wasson, C.S., 2010. System Phases, Modes, and States Solutions to Controversial Issues. *Wasson Strategic, LLC*.

DEVELOPMENT OF A FRAMEWORK FOR OPTIMAL DESIGN OF CLOSED-LOOP SUPPLY CHAINS: APPLICATION TO BATTERY RECYCLING

**M. GONZÁLEZ-RODRIGUEZ¹, C.
AZZARO-PANTEL¹, J. OCHOA ROBLES¹**

¹Laboratoire de Génie Chimique, Université de
Toulouse, CNRS, Toulouse, France,
4 allée Emile Monso, 31432 Toulouse Cedex 4
France
magnoangel.gonzalezrodriguez@ensiacet.fr,
catherine.azzaropantel@ensiacet.fr
jesus.ochoarobles@ensiacet.fr

A. AGUILAR-LASSERRE²

²Tecnológico Nacional de México, Instituto
Tecnológico de Orizaba, DEPI, Av. Instituto
Tecnológico 852 94300 Orizaba, Veracruz,
México
aaguilar@itorizaba.edu.mx

ABSTRACT: *Due to increasing environmental, economic and social constraints, the industrial sector is under a great pressure not only to develop environmentally friendly products but also to ensure the recovery after use. Moreover, customer awareness and government legislation also encourage industries to anticipate the End of Life (EoL) of their products. In this paper, a model for a reverse supply chain network has been developed in the context of the lead-acid battery manufacturing industry. This case study is particularly interesting due to the highest end-of-life recycling rate of lead among all commonly used metals. The model determines a generic framework for closed-loop supply chain (CLSC) design by connecting the forward supply chain to the point where the end-of-life products enter the reverse supply chain taking into account a multi-objective optimization approach that simultaneously considers economic and environmental criteria. The CLSC design has been formulated as an MILP problem implemented in the GAMS (General Algebraic Modelling System) environment and solved with CPLEX. The multi-objective problem is tackled by an ϵ -constraint method. The design formulation for closed loop supply chain determines the raw material level, production level, distribution and inventory level, disposal level, and recycling level at the different facilities with the objective of minimizing the total cost and the global emissions to the environment.*

KEYWORDS: *Closed loop supply chain, optimization, reverse logistic, recycling, lead acid battery.*

INTRODUCTION

The rapid economic development, population growth, and urbanization result in the generation of large amounts of wastes in cities. In most countries, insufficient facilities are available for proper management of waste materials. Additionally, the inadequate enforcement of policies and the lack of proper technologies result in indiscriminate disposal. (Periathamby et al, 2009)(Chen and Taylor, 2011). Recycling has thus become a fundamental issue because it allows reducing the amount of wastes and can also generate economic benefits. In that context, supply chain modelling has evolved towards an integrated approach considering both forward

and reverse supply chains simultaneously, i.e., the so-called closed-loop supply chain (CLSC).

The objective of this work is to develop a generic framework for the optimal design of a CLSC for the process industries taking into account various waste management strategies such as recovery, recycling and reuse options considering the multi-echelon, multi-objective and multi-period nature of the problem with multiple stakeholders. The case study of lead-acid batteries will support the analysis, that is of particular interest due to the highest end-of-life recycling rate of lead among all commonly used metals (Davidson et al, 2016). The increased use of lead-acid batteries has increased the demand for lead, so that used batteries have been identified as an important source of lead recycling. Besides, the

recycling operation has been identified as profitable, mainly because recycled lead is of high quality (Gaines 2014).

Closed-loop supply chains (CLSC) is the union of forward and reverse logistics, which focuses on taking back products from customers and recovering added value by reusing the entire product, and/or some of its modules, components, and parts. This kind of supply chain management has been identified as “the design, control, and operation of a system composed of different echelons to maximize value creation over the entire life cycle of a product with dynamic recovery of value from different types and volumes of returns over time”(Wassenhove 2002).

The CLSC supply chain has been receiving increased attention for several years. A review of some models for reverse logistics has been carried out, for example in (Govindan et al, 2015) where the lack of multi-objective approaches has been clearly emphasized despite the interest of simultaneously considering economic environmental objectives.

Two categories of works have been reported to date in the dedicated literature in particular for the case of the recycling of lead acid batteries: on the one hand, those related to the development of CLSC design models that are only based on the minimization of one economic criterion, i.e., the total supply chain cost (Kannan et al, 2010) and on the other hand, those devoted to the environmental issue that is tackled through a Life Cycle Assessment (LCA) methodology (Wang et al, 2010) (O'Donnell 2007). It must be recalled that LCA is viewed as a mature, systems-oriented and analytical tool assessing potential impacts from products or services using a life cycle perspective.

In LCA, the assessment of environment impacts is normalised by ISO 14040-44 following a four-step iterative process: goal and scope definition, Life Cycle Inventory (LCI), impact assessment (LCIA) and interpretation. In particular, the life cycle of a battery begins with the collection of the raw materials and ends with the deposit of the battery components that cannot be used anymore. The production of primary and auxiliary materials, the manufacturing, the use and the recycling of the battery are the intermediate stages in the battery's life. Transportation, involved between the several stages, cause emissions that contribute to an additional environmental damage.

The objective of the proposed work is thus to embed in one generic formulation the multi-objective formulation of CLSC design which focuses on the simultaneous optimization of economic and environmental criteria. The multi-objective optimisation step is then followed by the use of a multiple criteria decision making (MCDM) procedure that consists in finding the best alternative among a set of feasible alternatives.

PROBLEM STATEMENT

The main objective addressed in this paper is the development of a generic framework for the optimal design of a CLSC for the process industries taking into account various waste management strategies such as recovery, recycling and reuse options, involving multiple stakeholders. The forward supply chain involves the main raw materials such as lead, plastic, and sulfuric acid from different suppliers. The produced batteries are then distributed through distributors, wholesalers, retailers and then customers. After their End of Life (EoL), the used batteries (reverse logistic) are collected at the collection points and transshipped to the centralized return center where returned products are inspected for quality failure, sorted for potential repair or recycling. The useless batteries are disposed off and reusable batteries are transported to disassembly/recycling plants where the batteries are separated into different components (lead, plastic, acid etc.). Except lead the remaining components are sold to the third party for some other applications (Fig. 1). The recycled lead is transported to the battery manufacturing plants, then is mixed with the virgin lead for new battery production (Fig. 1).

The modelling of supply chains, and of closed-loop ones in particular can be formulated as a Mixed Integer Linear Programming (MILP) which a well-established method for analysing different systems (Rardin, 1998) (Taha, 1976). This approach has been adopted in this study within the framework of GAMS (General Algebraic Modelling System).

The framework considers as inputs:

- i: type of raw materials.
- s: quantity of suppliers.
- j: quantity of manufacturing plants.
- d: quantity of distributors.
- w: quantity of wholesalers.
- r: quantity of retailers.
- x: quantity of initial collection points.
- y: quantity of disposal sites.
- z: quantity of disassembly/recycling plants.
- p: quantity of products.
- t: quantity of time periods.

The goal is to determine the optimal system configuration that simultaneously minimizes the total cost of the CLSC and the global warming potential as the environmental criterion of the supply chain. By definition, LCA is a multi-criteria-oriented analysis and gives the opportunity to assess a wide range of indicators, such as Global Warming Potential (GWP), acidification, eutrophication, land-use, etc. Only GWP is considered in this work,

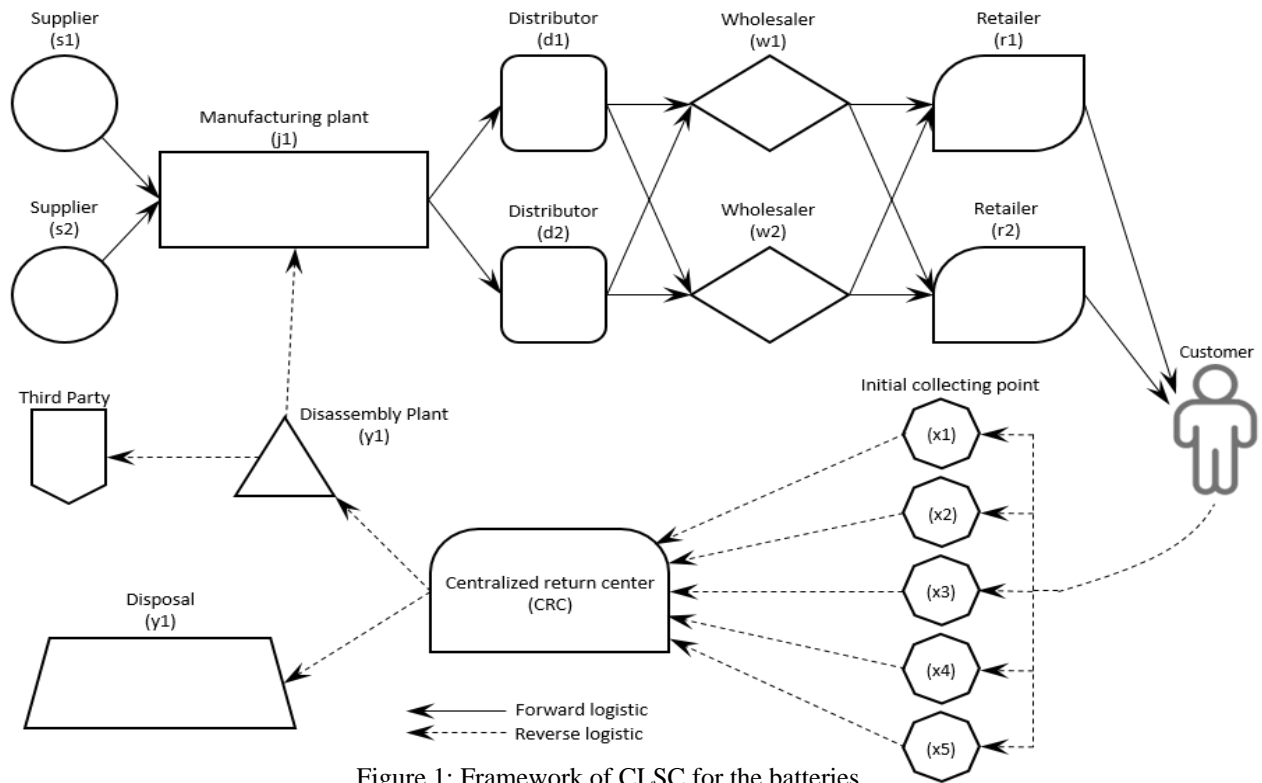


Figure 1: Framework of CLSC for the batteries.

the conceptual framework of LCA as a system-oriented environmental assessment method could yet be extended to other criteria.

MATHEMATICAL FORMULATION

3. 1 Objective functions and decision variables

Two objective functions have been considered in the mathematical formulation, the former is related to the total cost of the CLSC and the latter involves the total GWP generated through the whole CLSC.

- *Total cost objective*

For this objective the function considered in the mathematical formulation, focuses on the minimization of the total cost of the CLSC as presented in Eq. (1) (expressed in €):

$$\begin{aligned} \text{Min TotalCost} = & \text{CPu} + \text{Cpr} + \text{CTMD} + \text{CTDW} + \\ & \text{CTWR} + \text{CTCp} + \text{CTCDi} + \text{CTDissMs} + \\ & \text{CCiRm} + \text{CCiFg} + \text{CCiD} + \text{CCiW} + \text{CCp} \\ & \text{CCiCent} + \text{CDispos} + \text{CDissa} + \\ & \text{CRec3P} \end{aligned} \quad (1)$$

The following parameters and variables are involved in the expression of the objective function:

CPu: is the total cost of purchasing new raw material for the battery production (€).

Cpr: is the total processing cost involved in all the manufacturing plants (€).

The transportation costs that are involved in all the CLSC, include the transportation of final products, raw material, or recollected products, across the different phases of the CLSC:

CTMD: manufacturing plant to distributors battery (€).

CTDW: battery transportation cost from distributor to wholesaler (€).

CTWR: battery transportation cost from wholesaler to retailer (€).

CTCp: transportation cost from initial collection point to the centralized return centre (€).

CTCDiss: transportation cost from centralized return centre to disassembly/recycling plant (€).

CTDissM: transportation cost from disassembly/recycling plant to manufacturing plant (€).

The inventory costs refer to the total cost involved in handling the inventory of final product, raw material, of recollected products.

CCiRm: raw material inventory cost at the manufacturing plant (€).

CCiFg: final good inventory cost at the manufacturing plant (€).

CCiD: distributor inventory cost (€).

CCiW: wholesaler inventory cost (€).

CCp: the collection cost of the returned items is the cost involved in the recollection of returned or used batteries to the initial collection centres (€).

CCiCent: returned products inventory cost at the centralized return centre (€).

CDispo: the disposal costs are related to the cost for the discarding of raw materials from the collected batteries (€).

CDissa: disassembly/reclaiming cost at the disassembly/recycling plant involved in reclaiming material from the disassembly plant. (€).

CRec3P: recycling cost from the disassembly/recycling plant sold to the third party is the cost of the material that is recovered and sold to others companies (€).

- *Total GWP objective*

This criterion focuses on the minimization of the environmental impact based on the total Global Warming Potential (GWP) that the whole chain generates (see Eq. (2) (expressed in kg CO₂eq):

$$\begin{aligned} \text{Min GWP} = & \text{EPro} + \text{ETD} + \text{ETDW} + \text{ETWR} + \\ & \text{ECollec} + \text{EDispo} + \text{EDissa} + \\ & \text{ETRecla} \end{aligned} \quad (2)$$

The global GWP is calculated from the GWP expressed in kg CO₂eq/kg_{Battery}.

Epro: Emissions that the production of batteries generates in the CLSC (kg CO₂eq/kg_{Battery}).

ETD: emissions generated by the transport of batteries from manufacturing plant to distributor (kg CO₂eq/kg_{Battery}).

ETDW: emissions related to the transportation of batteries from distributor to wholesaler (kg CO₂eq/kg_{Battery}).

ETWR: emission generated for the transportation of a wholesaler to a retailer (kg CO₂eq/kg_{Battery}).

ECollec: emissions due to the recollection of batteries (kg CO₂eq/kg_{Battery}).

EDispo: emissions generated by the disposal of batteries (kg CO₂eq/kg_{Battery}).

EDissa: emissions from the disassembly plant (kg CO₂eq/kg_{Battery}).

ETRecla: emissions of the transportation the raw material reclaimed from the disassembly plant to the manufacturing plant (kg CO₂eq/kg_{Battery}).

The decision variables that are involved in the model are summarized in Table.1.

Decision variable	Definition
AM_{ijsjt} (amount)	Amount of raw material 'i' purchased from supplier 's' to the manufacturing plant 'j' during time period 't'
QP_{jpt} (unites)	Quantity processed of product 'p' at manufacturing plant 'j' during time period 't'
$QTJD_{jdpt}$ (unites)	Quantity transported from manufacturing plant 'j' to distributor 'd' of product 'p' at time period 't'
$QTDW_{dwpt}$ (unites)	Quantity transported from distributor 'd' to wholesaler 'w' of product 'p' at time period 't'
$QTWR_{wrpt}$ (unites)	Quantity transported from wholesaler 'w' to retailer 'r' of product 'p' at time period 't'
QRP_{ypt} (unites)	Quantity of useless returned products of 'p' transported from the centralized return centre to the disposal site 'y' at the time period 't'
$QRpP_{zpt}$ (unites)	Quantity of useful (recyclable) returned products of 'p' transported from the centralized return centre to the disassembly/recycling plant 'z' at time period 't'
RMS_{izt} (unites)	Amount of recycled raw material 'i' from the disassembly/recycling plant 'z' sold to the third party during the time period 't'
$RMRP_{izpt}$ (amount)	amount of required reclaimed raw material 'i' for new battery production transported from the disassembly/recycling plant 'z' to the manufacturing plant 'j' during the time period 't'

Table 1: Decision variables of the model

The constraints involved in this model are as follows:

- The quantity received from wholesaler to retailer must be equal to demand Eq. (3).

$$\sum_w QTWR_{wrpt} = \text{DMD}_{rpt} \quad (3)$$

- The quantity of products that the retailer receives from wholesaler should be inferior or equal to the storage of the retailer Eq. (4).

$$\sum_w \sum_p \sum_t QTWR_{wrpt} \leq \text{SR}_{rt} \quad (4)$$

- The quantity of products that the wholesaler sends to the retailer must be the quantity that the wholesaler receives from the distributor plus the wholesaler product inventory Eq. (5).

$$\begin{aligned} \sum_d \sum_w QTDW_{dwpt} + \text{Inv}W_{pt} = \\ \sum_w \sum_r QTWR_{wrpt} \end{aligned} \quad (5)$$

- The sum of the quantity that the distributor sends to wholesaler plus the wholesaler inventory of product must be less or equal to the storage of the retailer Eq. (6).

$$\sum_d \sum_p QTDW_{dwpt} + InvW_{pt} \leq SW_{wt} \quad (6)$$

- The quantity of products that the distributor sends to the wholesaler must be equal to the quantity of product coming from the manufacturing plant plus the distributor inventory of product Eq. (7).

$$\sum_j \sum_d QTJD_{jdpt} + InvD_{pt} = \sum_d \sum_w QTDW_{dwpt} \quad (7)$$

- The sum of the distributor inventory of product plus the quantity that the distributor receives from the manufacturing plant must be inferior or equal to the storage capacity of the distributor Eq.(8).

$$\sum_j \sum_p QTJD_{jdpt} + InvD_{pt} \leq SD_{dt} \quad (8)$$

- The quantity of products from manufacturing plant to distribution must be equal to the sum of the quantity of processed products plus the inventory manufacturing plant inventory of final products Eq.(9).

$$\sum_j QP_{jpt} + InvMP_{pt} = \sum_j \sum_d QTJD_{jdpt} \quad (9)$$

- The storage capacity of the manufacturing plant of final products must be inferior or equal to the sum of the quantity of processed product plus the inventory manufacturing plant inventory of final products Eq. (10).

$$\sum_p QP_{jpt} + InvMP_{pt} \leq SFG_{jt} \quad (10)$$

- The demand of raw materials for the processed products is equal to the product of the quantity of processed material by the required raw materials of each product Eq. (11).

$$\sum_p QP_{jpt} \cdot RMT_{ipt} = DemRM_{ijt} \quad (11)$$

- The demand of raw materials for the products is equal to the amount of raw materials provided by the suppliers plus the manufacturing plant raw material inventory Eq. (12).

$$\sum_s \sum_j AM_{isjt} + InvRM_{it} = \sum_j DemRM_{ijt} \quad (12)$$

- The amount of raw materials provided by the suppliers must be inferior or equal to their capacity Eq. (13).

$$\sum_i AM_{isjt} \leq SS_{ist} \quad (13)$$

- The quantity of recollected products is equal to the demand multiplied by the collection rate Eq. (14).

$$\sum_r DMD_{rpt} \cdot IndexReco_{pt} = QtyReco_{pt} \quad (14)$$

- The quantity of lost products during the collection is equal to the demand multiplied by the inverse of the collection rate Eq. (15).

$$\sum_r DMD_{rpt} \cdot (1 - IndexReco_{pt}) = QtyLost_{pt} \quad (15)$$

- The quantity of products that each collection centre have is inferior or equal to the quantity of products collected Eq. (16).

$$\sum_x AC_{xpt} \leq QtyReco_{pt} \quad (16)$$

- The quantity of returned product is equal to the product of the quantity of product collected multiplied by the product disposal rate Eq. (17).

$$\sum_x AC_{xpt} \cdot DispRate_{pt} = \sum_y QRP_{ypt} \quad (17)$$

- The quantity of returned product must be inferior or equal to the storage capacity of the disposal site Eq. (18).

$$QRP_{ypt} \leq SDP_{ypt} \quad (18)$$

- The quantity of recycled product involves the sum of the returned products inventory, plus multiplication of the quantity of collected products by the inverse of the disposal rate Eq. (19).

$$InvRP_{pt} + (\sum_x AC_{xpt} \cdot (1 - DispRate_{pt})) = \sum_z QReP_{zpt} \quad (19)$$

- The amount of recycled raw materials that is sold to the third party is equal to the product of raw materials recycled by the recycling rate Eq. (20).

$$\sum_p \text{RMREC}_{ipt} \cdot \text{ReRATE}_{izt} = \text{RMS}_{izt} \quad (20)$$

- The amount of recycled raw materials involves the quantity of recycled products and their contribution to the materials Eq. (21).

$$\sum_z \text{QReP}_{zpt} \cdot \text{PCONT}_{ip} = \text{RMREC}_{ipt} \quad (21)$$

- The amount of raw material reclaimed for a new product involves the product of the recycled raw material recycled and the recycling rate of the raw material Eq. (22).

$$\sum_p \text{RMREC}_{ipt} \cdot (1 - \text{ReRATE}_{izt}) = \sum_j \text{RMRP}_{izpt} \quad (22)$$

In addition, some assumptions have been used in the model. For instance, the manufacturing plant chooses the raw material from the recycling center first, then the raw material from other suppliers.

3.2 Solution strategy

Bi-objective optimization: In a preliminary phase, each mono-criterion problem was optimized separately to analyse how its optimal values are decreased when carrying out a bi-criteria optimization. The mathematical model is then formulated within the GAMS environment and solved using CPLEX. Then, the bi-objective optimization problem is solved by implementing the ϵ -constraint method.

Multiple choice decision making (MCDM): A modified (M-TOPSIS) (*Technique for Order Preference by Similarity to Ideal Situation*) (Ren et al. 2007) analysis is carried out on the Pareto front with the same weighting factor for the cost and environmental criteria.

4 CASE STUDY

The case study used for validating the multi-echelon, multi-objective, multi-period CLSC is based on the work presented in (Kannan, 2010) for cost evaluation. The HSC considers the following set of data:

- Number of suppliers (s) = 2
- Type of raw materials (i) = 2
- Number of manufacturing plants (j) = 1
- Number of distributors (d) = 2
- Number of wholesalers (w) = 2
- Number of retailers (r) = 2
- Number of initial collection points (x) = 5
- Number of disposal sites (y) = 1
- Number of disassembly/recycling plant (z) = 1
- Number of products (p) = 2
- Number of periods (t) = 11

Although real data have not been used, the data set reproduces the general trends that have been observed for such supply chains (Ademe, 2016).

For Life Cycle Assessment, the works of (Wang et al. 2010) and (O'Donnell 2007) have been considered for the Life Cycle Inventory of the whole process involved in the CLSC. The SimaPro software tool (version 8) has been used. Some typical results are presented in Table 2.

	GWP (kgCO _{2eq} /kgBatterie)
Production	1.51
Distribution	0.00225
Collection	0.00564
Disassembly	0.0241
Disposal	0.000281
Remanufacturing	0.0276

Table 2: GWP per Kg_{Batterie} for each section of the CLSC.

5 RESULTS AND DISCUSSION

5.1 Multi-objective optimization (ϵ -constraint method)

As a preliminary stage, two independent mono-objective optimizations were solved to obtain, the optimal value for each objective, that exhibit an antagonist behaviour, so that the ϵ -constraint method can be applied. The mono-objective optimization results are presented in Table 3.

Minimizing	Cost (€)	GWP (KgCO _{2eq} /battery)
Total Cost(€)	172892.95	520594.61
Total GWP (KgCO _{2eq})	176344.01	512220.07

Table 3: Pay off table obtained for the production by the mono-objective optimizations for all periods.

The first objective function (total cost), is chosen as the primary objective, and the second objective function, (GWP), is converted into a inferior-to-or-equal type constraint with the right-hand side specified as ϵ , and is added to the set of constraints. By selecting appropriate values for ϵ , different efficient solutions to the model are generated. The solutions presented in Table3 represent an upper bound on the total cost, and a lower bound on the pollution index, respectively.

5.2 Pareto front results

The solutions obtained lead to the Pareto front (Fig. 2), which is composed of solutions representing supply chain configurations, that are then evaluated by the TOPSIS procedure with the same weighting in cost and GWP.

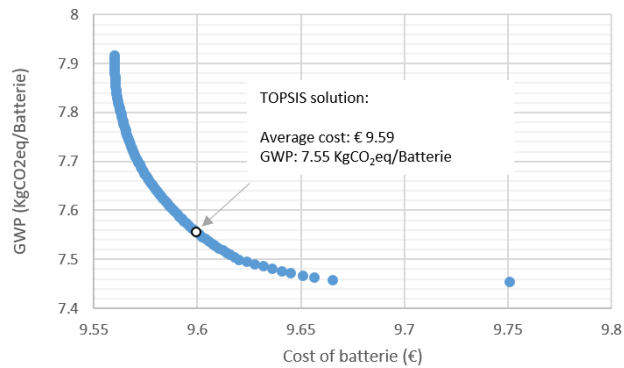


Figure 2: Pareto results for the multi-objective optimization of the CLSC.

CONCLUSIONS

This paper has presented a model for a reverse supply chain network has been developed. The case study of lead-acid batteries supports the analysis.

The model is devoted to a closed-loop supply chain design while minimizing the total cost as well as the total GWP emissions involved in all the steps of the supply chain. The model is a decision making tool for determining the levels of raw material consumption (both virgin as well as recycled), production, inventory, along the chain,

REFERENCES

- Ademe, Agence de l'environnement et de la maîtrise de L'énergie. 2017. "Piles et Accumulateurs Rapport Anuel." http://www.ademe.fr/sites/default/files/assets/documents/registre_pa_rapport_2016.pdf
- Chen, Liping, and David Taylor. 2011. "Public Awareness and Performance Relating to the Implementation of a Low-Carbon Economy in China: A Case Study from Zhengzhou." *Low Carbon Economy* 2 (2): 54–61. <https://doi.org/10.4236/lce.2011.22009>.
- Davidson, Alistair J., Steve P. Binks, and Johannes Gediga. 2016. "Lead Industry Life Cycle Studies: Environmental Impact and Life Cycle Assessment of Lead Battery and Architectural Sheet Production." *The International Journal of Life Cycle Assessment* 21 (11): 1624–36. <https://doi.org/10.1007/s11367-015-1021-5>.
- Gaines, Linda. 2014. "The Future of Automotive Lithium-Ion Battery Recycling: Charting a Sustainable Course." *Sustainable Materials and Technologies* 1–2 (December): 2–7. <https://doi.org/10.1016/j.susmat.2014.10.001>.
- Govindan, Kannan, Hamed Soleimani, and Devika Kannan. 2015. "Reverse Logistics and Closed-Loop Supply Chain: A Comprehensive Review to Explore the Future." *European Journal of Operational Research* 240 (3): 603–26. <https://doi.org/10.1016/j.ejor.2014.07.012>.
- Jolly, R., and C. Rhin. 1994. "The Recycling of Lead-Acid Batteries: Production of Lead and Polypropylene." *Resources, Conservation and Recycling* 10 (1–2): 137–43. [https://doi.org/10.1016/0921-3449\(94\)90046-9](https://doi.org/10.1016/0921-3449(94)90046-9).
- Kannan, G., P. Sasikumar, and K. Devika. 2010. "A Genetic Algorithm Approach for Solving a Closed Loop Supply Chain Model: A Case of Battery Recycling." *Applied Mathematical Modelling* 34 (3): 655–70. <https://doi.org/10.1016/j.apm.2009.06.021>.
- O'Donnell, Clayton. 2007. "Total Effect on the Environment of Electric/hybrid Electric Vehicle Batteries."
- Periathamby, Agamuthu, Fauziah Shahul Hamid, and Kahlil Khidzir. 2009. "Evolution of Solid Waste Management in Malaysia: Impacts and Implications of the Solid Waste Bill, 2007." *Journal of Material Cycles and Waste Management* 11 (2): 96–103. <https://doi.org/10.1007/s10163-008-0231-3>.
- Rardin, Ronald L. 1998. *Optimization in Operations Research*. Edited by Prentice Hall.
- Ren, L., Y. Zhang, Y. Wang, and Z. Sun. 2007. "Comparative Analysis of a Novel M-TOPSIS Method and TOPSIS." *Applied Mathematics Research eXpress*, July. <https://doi.org/10.1093/amrx/abm005>.
- Taha, Hamdy A. 1976. *Operations Research: An Introduction*. Macmillan.
- Wang, Yacan, Jing Yu, Xinfei Zhao, Tao Lu, Jie Du, and Xiaoyan Huang. 2010. "Research on the Life Cycle Analysis of the Reverse Supply Chain of the Lead Acid Batteries." In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, 1–6. IEEE. <https://doi.org/10.1109/ICBBE.2010.5518110>.
- Wassenhove, Luk N. Van, and V. Daniel R. Guide. 2002. "The Reverse Supply Chain."

l'évolution de la région comme collectivité territoriale et le transfert de l'exécutif départemental du préfet vers le président du conseil général. Les lois de 1983 précisent la répartition des compétences entre les communes, les départements, les régions et l'Etat.

L'acte II de la décentralisation est d'abord concrétisé par la loi constitutionnelle de 2003 qui considère que l'organisation de la République est décentralisée. Les collectivités territoriales (région, départements, communes) doivent avoir des ressources propres déterminantes et les transferts de compétences doivent être compensés. Cela s'est aussi traduit par le transfert de 130.000 agents de l'état en 2004.

L'acte III de la décentralisation (2014-2015) a vu la création des métropoles, la réduction du nombre régions à 12 en métropole (13 en comptant la Corse) et la facilitation de la fusion des communes. Il a donc fallu clarifier les compétences transférées : c'est l'objet de la loi NOTRe (loi NOTRe, 2015).

2.2 Déconcentration

La volonté de l'Etat français de gérer au plus près ses citoyens a toujours été un principe.

Quatre niveaux de circonscription territoriale ont été retenus :

- la zone de défense. 7 zones de défense sont définies depuis 2000 ayant chacune à sa tête un préfet de zone (un des préfets de région),
- la région. 12 préfetures de région en métropole sont définies depuis 2015. La région est l'échelon d'orientation et de coordination des politiques publiques dans tous les domaines hormis celui de la sécurité,
- le département. 96 préfetures de département existent en métropole. Le département est l'échelon pour la mise en œuvre de l'ensemble de la politique gouvernementale,
- l'arrondissement. 333 arrondissements sont placés sous la responsabilité de sous-préfets.

L'évolution des régions s'est accompagnée d'un accroissement du rôle des préfets de région en 2004 et en 2010. Le regroupement en grandes régions, s'est aussi accompagné de l'évolution des préfetures par le Plan Préfetures Nouvelle Génération (PPNG) se traduisant notamment par la dématérialisation des opérations pour les titres sécurisés (eg. cartes grises) avec quelques préfetures identifiées comme Centres d'Expertise et de Ressources des Titres (CERT). Il faut aussi noter que 8 sous-préfetures ont été supprimées entre 2015 et 2016 suite à une expérimentation en Alsace-Moselle. La création de maisons de l'Etat et des maisons de services publics regroupant plusieurs services déconcentrés remet aussi en cause l'existence de sous-préfetures (Servière, 2016). Un effort de colocalisation dans les bâtiments aujourd'hui parfois trop grands des sous-préfetures est en cours. Il faut noter un grand regroupement des directions de l'administration territoriale de l'Etat par la mise

en œuvre de la RéATE (décret 2009-1484 du 3 décembre 2009). 11 directions régionales sont fusionnées en seulement 4 directions, mais l'enseignement supérieur et la recherche n'ont pas été concernés.

2.3 Territoires

Les territoires sont le résultat des évolutions des actions de décentralisation et de déconcentration. Ces deux mouvements nécessitent parfois des adaptations.

Les ajustements issus de la décentralisation notamment par la création d'intercommunalités impliquent l'évolution des arrondissements (Savatier, 2017). Le récent rapport de la cour des comptes sur les services déconcentrés de l'état (Cour des comptes, 2017) a recensé que 229 arrondissements sur 335 ont eu leurs limites modifiées au 1er janvier 2017 afin d'être en cohérence avec les périmètres des EPCI (Établissement Public de Coopération Intercommunale).

Les périmètres des territoires et les évolutions de leurs compétences relatives ont un impact sur les relations entre les établissements d'enseignements supérieurs, l'Etat et les collectivités territoriales, mais aussi sur les relations entre les établissements d'enseignement supérieur.

3 EVOLUTION DES ETABLISSEMENTS D'ENSEIGNEMENT SUPERIEUR ET DE RECHERCHE

Si les territoires ont évolué, il en est de même pour l'enseignement supérieur et la recherche. Même l'intitulé du ministère a récemment changé pour devenir le ministère de l'enseignement supérieur, de la recherche et de l'innovation.

3.1 Evolution des missions

Les missions du service public de l'enseignement supérieur comprenaient les éléments suivants :

- la formation initiale et continue ;
- la recherche scientifique et technologique, la diffusion et la valorisation de ses résultats ;
- la diffusion de la culture et l'information scientifique et technique ;
- la coopération internationale.

Elles ont été complétées en 2007 pour intégrer :

- l'orientation et l'insertion professionnelle ;
- la participation à la construction de l'Espace européen de l'enseignement supérieur et de la recherche.

L'insertion de la partie orientation et insertion professionnelle intègre déjà la finalité professionnelle de l'université et prépare à l'adaptation des formations et de leurs effectifs suivant le marché de l'emploi.

La participation à la construction de l'espace européen de l'enseignement supérieur et de la recherche inscrit les établissements dans le vaste territoire qu'est l'Europe. Toutefois, l'enseignement supérieur français est structuré en académies qui correspondait aux anciennes régions. La formation universitaire en apprentissage est gérée par les régions. Les régions financent des projets de recherche et co-financent avec l'état des équipements lourds et l'immobilier de l'enseignement supérieur (notamment via les plans état-région). La région est donc un espace évident de structuration de l'enseignement supérieur.

3.2 Evolution des étudiants et de leurs inscriptions

L'enseignement supérieur doit faire face aux évolutions démographiques des étudiants et donc des bacheliers. S'il faut noter une très forte évolution du pourcentage d'une classe d'âge qui est diplômé du bac, on peut constater que si l'on considère uniquement les bacheliers généraux et technologiques, l'évolution en pourcentage de classe d'âge est très limitée depuis les années 1995 (cf. figure 2).

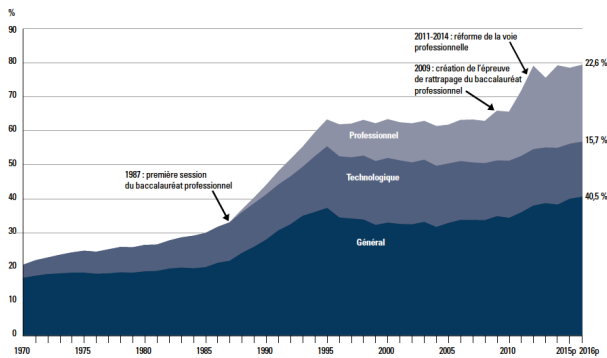


Figure 2. Evolution de la proportion de bacheliers dans une génération (DEPP, 2017).

Les effectifs des étudiants ont très fortement augmenté depuis les années 1960. En 1960-1961, il y avait presque 310.000 étudiants, contre 2,160 millions en 2000-2001 et 2,329 millions en 2010-2011. A la rentrée 2016, il y avait plus de 2,6 millions d'étudiants dans l'enseignement supérieur français (DEPP, 2017). Il faut aussi noter une très forte progression de l'enseignement privé (cf. figure 3).

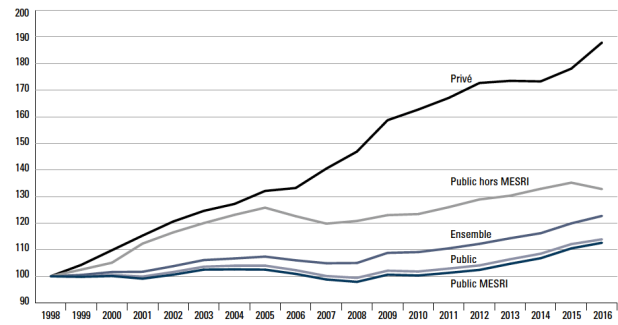


Figure 3. Evolution du nombre d'étudiants en milliers (DEPP, 2017).

3.3 Evolution du management et des ressources

Comme tous les établissements publics, les établissements d'enseignement supérieur et de recherche tentent de maîtriser leurs coûts de fonctionnement pour conserver une marge de manœuvre pour investir. La maîtrise de la masse salariale qui représente souvent plus de 50% du fonctionnement offre un levier efficace. Certains sont parfois tentés de diminuer le nombre d'emplois pour assurer la soutenabilité de l'établissement. Les promotions peuvent aussi n'être accordés que dans le cadre de la non augmentation de la masse salariale (promotions des Maîtres de Conférences en Professeurs des Universités seulement par promotion interne sans création de poste).

A titre d'illustration, voici la situation à l'université de Strasbourg. D'après les bilans sociaux, on peut par exemple représenter les évolutions des personnels (-6%) en rapport à l'évolution des étudiants (+20%) entre 2010 et 2016 (Figure 4).

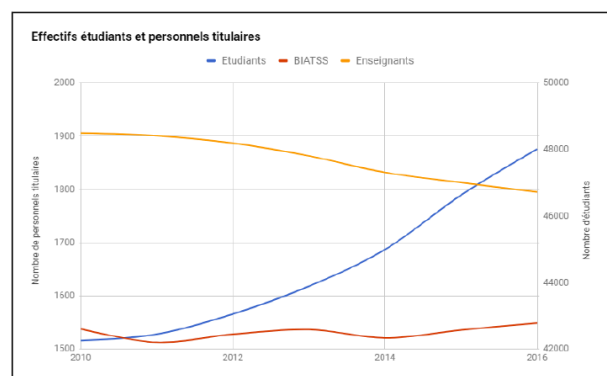


Figure 4. Evolution du nombre d'étudiants et des enseignants à l'université de Strasbourg d'après les données des bilans sociaux.

Les établissements ont développé leurs ressources propres par la création de fondations, le développement du financement de la recherche par des ressources contractuelles (création des SATT), et le développement des ressources en formation (formation continue et

apprentissage). Souvent, ces structures ont été construites soit à l'échelle de l'établissement, soit à l'échelle des anciennes académies. Par exemple, il n'existe qu'un seul CFAU (Centre de Formation par Apprentissage Universitaire) pour l'Alsace.

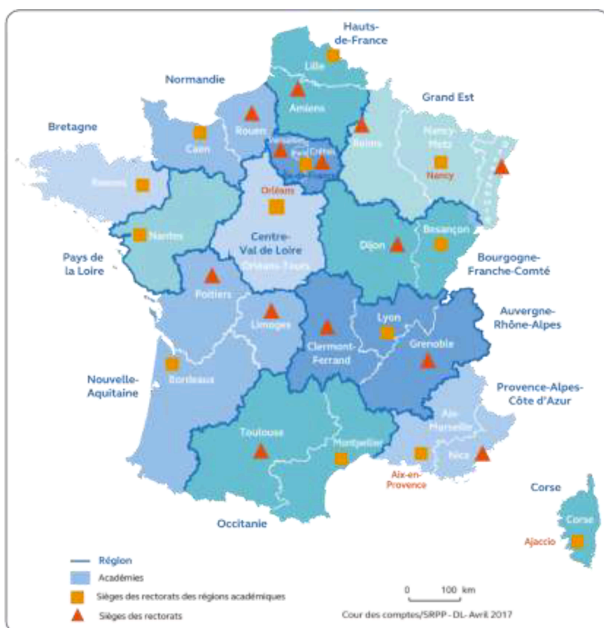
4 ADAPTATION DE L'ENSEIGNEMENT SUPERIEUR ET DE RECHERCHE AU TERRITOIRE

L'enseignement supérieur et la recherche s'est adapté (plus ou moins) aux évolutions des territoires. Comme les autres administrations territoriales de l'Etat, celles de l'enseignement supérieur et de la recherche ont évolué suivant la nouvelle organisation régionale. Les établissements se sont structurés de leur côté créant quelques incohérences.

Les influences relatives sont synthétisées.

4.1 Adaptation de l'organisation de l'enseignement supérieur (rectorats et DDRT)

La création des nouvelles régions aux contours plus larges que les académies, a conduit à création de recteurs de régions académiques (cf. figure 5). L'absence de relation hiérarchique entre le recteur de la région académique et recteurs d'académie (approche similaire aux relations préfet de région / préfets de département) ainsi que des localisations non identiques avec les préfectures des nouvelles régions sont identifiées comme des points faibles dans (Cour des comptes, 2017). Ce même rapport préconise la fermeture des rectorats de petite taille.



Source : ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche

Figure 5. Organisation des rectorats en 2017.

Une première expérience de fusions d'académies a été initiée en Normandie. Il n'y a plus qu'un recteur pour cette région académique depuis le 22 novembre 2017. Cette voie ne semble pertinente que pour les « petites » régions.

Pour les Délégations Régionales à la Recherche et à la Technologie (DRRT), la voie choisie a été de rattacher le DRRT de la nouvelle région au secrétaire général pour les affaires régionales et de transformer les DRRT des anciennes régions en DDRT-adjoints (décret n° 2015-1834 du 30 décembre 2015). Il est également conseiller du recteur de région académique. Il contribue aux "stratégies régionales de l'innovation" (S.R.I.) et aux "stratégies de spécialisations intelligentes" initiées par les conseils régionaux et élaborées à la demande de l'Union européenne dans le cadre de la mise en place des programmes opérationnels européens pour les années 2014-2020. La fusion des régions a conduit à remettre à plat les stratégies régionales en matière de recherche, d'innovation.

Liens DRRT-recteurs de région académique et DRRT-SGAR rendent encore plus nécessaire la cohérence entre les rectorats de région académique et les nouvelles régions.

4.2 Adaptation des établissements d'enseignement supérieur (COMUE, associations, fusions)

Dans la continuité des Pôles de Recherche et d'Enseignement Supérieur (PRES) de 2006, des COMMunautés d'Universités et d'Établissements (COMUE) ont été créées en 2013 pour conforter la création de pôles universitaires lisibles à l'international.

La création des COMUE en tant que nouvel établissement public complique la gestion et la bonne représentation des établissements et de leurs personnels quand de nombreux établissements intègrent la COMUE. Rappelons que les premières universités à avoir fusionné sont les universités de Strasbourg (janvier 2009) sans même passer par une phase de PRES.

Les créations, de COMUE et la fusion d'établissements ne sont pas exclusifs, ce qui complique encore le paysage.

Parfois, la confusion est entretenue entre la COMUE et la fusion de universités comme à Grenoble où la COMUE Université Grenoble Alpes et université Grenoble Alpes coexistent début 2018. Ne parlons pas des COMUE « Sorbonne Universités » et « Sorbonne Université ».

La complexité du management de cette surcouche disparaît parfois au profit d'universités fusionnées auxquelles d'autres établissements sont associées. Début 2018 marque la disparition des COMUE comme Sorbonne Universités ou Université de Champagne pour donner lieu à des associations d'établissements.

Chaque établissement tente de se placer dans ces grands regroupements en essayant de ne pas perdre son identité et ses points forts. Souvent l'enjeu pour chacun est de

conserver sa personnalité morale (et donc son budget propre et sa capacité de décision pour la gestion des ressources humaines). A ces logiques régionales, se superposent parfois des logiques de réseaux (INSA, Universités de Technologie, Ecoles Centrales, ...). Les regroupements d'établissements ont été stimulés par les appels à projets Programmes d'Investissements d'Avenir (PIA). Certains établissements sont reconnus comme IDEX (Initiatives D'Excellence). La structuration des établissements d'enseignement supérieur, initiée avant la réforme de la carte régionale, n'a pas toujours respecté les évolutions des nouvelles régions.

Chaque établissement essaie de se situer dans le nouveau paysage territorial et de conserver ses ressources issues des appels à projets régionaux en matière de recherche, d'innovation et d'apprentissage en renforçant les liens avec les autres établissements de la nouvelle région. Les regroupements universitaires (au moins de février 2016) sont représentés sur la figure 6.

Au-delà de la situation complexe de la région parisienne, la COMUE Université Bretagne Loire interpelle particulièrement.

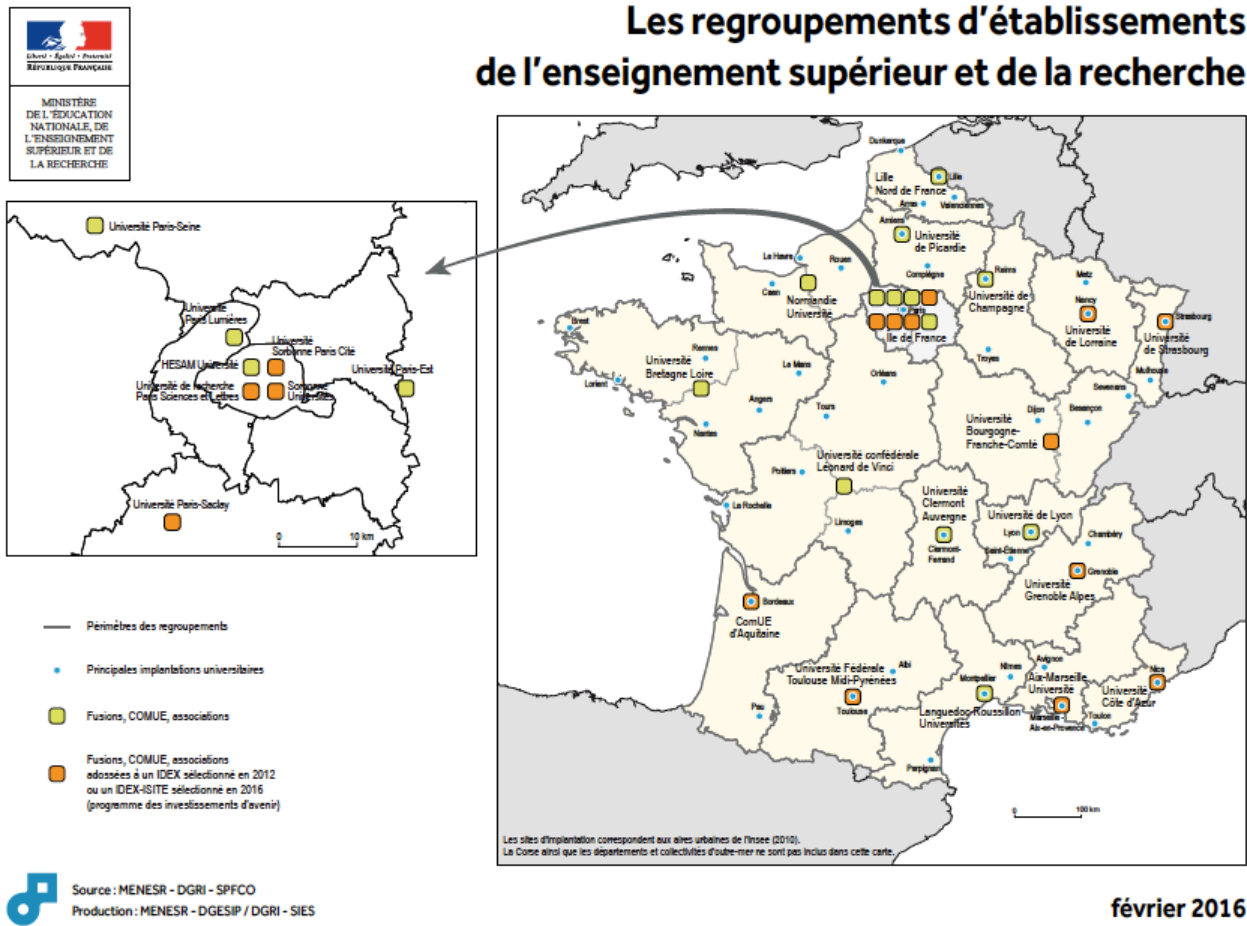


Figure 6. Regroupements universitaires en janvier 2016.

4.3 Influences entre établissements d'enseignement supérieur et territoires

Les territoires évoluent sous le double mouvement de décentralisation et de déconcentration des services de l'état. Les regroupements de communes ont conduit au redécoupage des arrondissements (sous-préfectures). Ces regroupements sont quasiment tous limités dans un département. Les régions ont connu des regroupements qui ne se sont pas traduites au niveau de l'organisation des académies (ni dans leurs contours ni dans le lieu du chef de file). Cette dernière étape sera-t'elle franchie prochainement ?

La redéfinition des compétences des collectivités territoriales semble clore la discussion. L'enseignement supérieur est de la compétence exclusive de la région. En réalité, la situation n'est pas si simple car les communes ou leurs regroupements ont clairement un

impact sur les établissements d'enseignement supérieur (par exemple l'eurométropole de Strasbourg sur l'université de Strasbourg). De même les départements ont un lien fort avec les établissements d'enseignement supérieur comme l'Université de Haute Alsace avec le Haut-Rhin).

Les gros investissements sont encore conjoints entre état (représenté par l'académie), la région et parfois le département et la commune.

L'importance du territoire pour les établissements supérieurs est telle que de plus en plus d'écoles de commerce et quelques écoles d'ingénieurs évaluent leur impact territorial (FNEGE, 2018).

En effet, pour obtenir des ressources, la démonstration de l'impact territorial est un argument de poids.

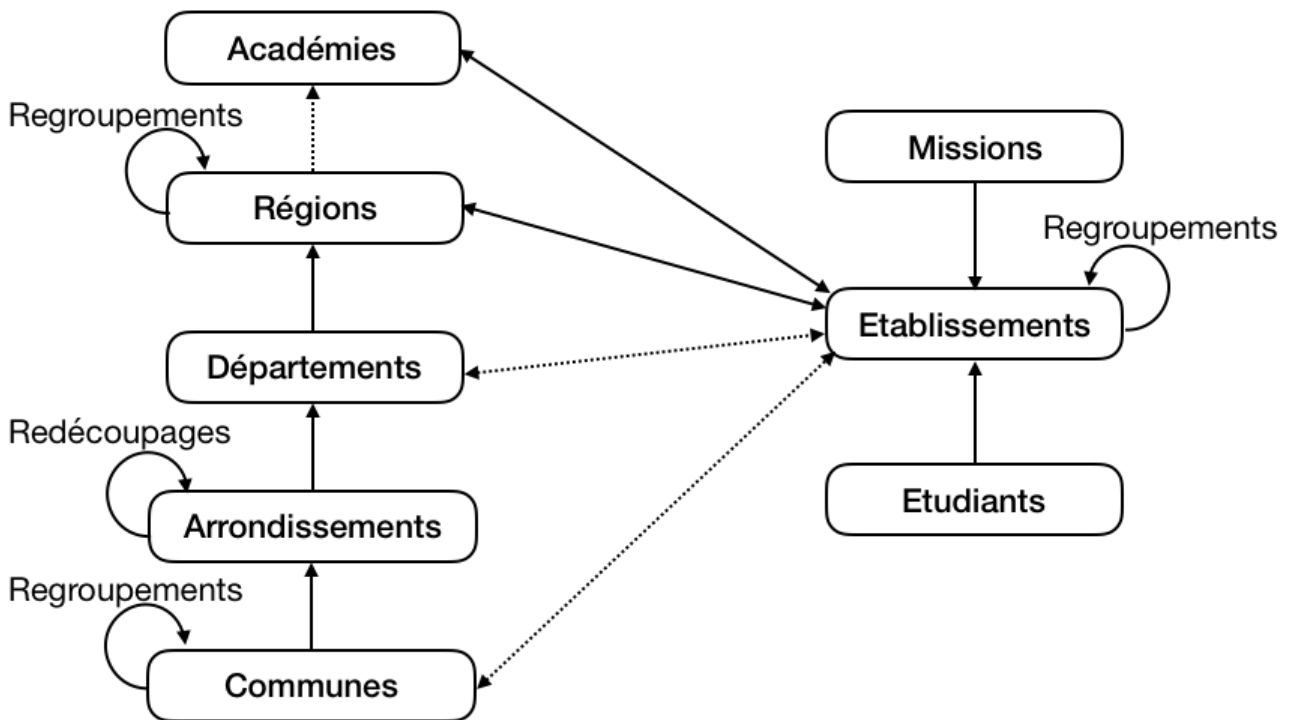


Figure 7. Influences territoires-enseignement supérieur.

5 CONCLUSIONS ET PERSPECTIVES

Deux dynamiques concourantes sont en cours : la définition des territoires et la structuration de l'enseignement supérieur et de la recherche.

Ces deux processus doivent aujourd'hui converger pour une meilleure efficacité de l'enseignement supérieur dans les territoires.

Cette première analyse devra être complétée par une modélisation suivant l'approche de modélisation des systèmes dynamiques.

Le développement du secteur privé interroge aussi. Alors que les ressources de l'Etat se concentrent sur ses fonctions régaliennes, est-il nécessaire de conserver un enseignement supérieur public ?

REMERCIEMENTS

Cette communication n'aurait jamais été écrite sans les échanges avec des collègues ces dernières années. Merci à Olivier Bistorin, Claude Pourcel, Jacques Lamothe et à Tatiana Reyes.

REFERENCES

Bistorin O., Monteiro T., Pourcel C., 2006, Modélisation du parcours apprenant dans un système de production de compétences, 6ème Conférence de Modélisation et Simulation, MOSIM 06, Rabat, Maroc.

Caillaud E., Matt M., Sonntag M., 2009. Vers une évaluation des systèmes de formation des établissements d'enseignement supérieur français. *CIGI'09*, Bagnères de Bigorre.

Chambon F., 2017. L'évolution du management public. Support de cours, Master management Public, IPAG-Université de Strasbourg.

Cour des comptes, 2017. Les services déconcentrés de l'état. Clarifier leurs missions, adapter leur organisation, leur faire confiance. Rapport public thématique.

DEPP, 2017. Repères et références statistiques. Ministère de l'enseignement supérieur, de la recherche et de l'innovation.

FNEGE, 2018. BSIS - Business School Impact System, <http://www.fnege.org/nos-programmes/bsis>, consulté le 12 janvier 2018.

Loi NOTRe, 2015. Nouvelle Organisation Territoriale de la République.

Savatier J. 2017. Rapport sur la loi de finances 2018. Annexe 3 : administration générale et territoriale de l'Etat.

Servière S-F., 2016. La discrète suppression des sous-préfectures, IFRAP, <http://www.ifrap.org/etat-et-collectivites/la-discrete-suppression-des-sous-prefectures>, consulté le 5/12/2017.

EARLY CHECKING OF SysML MODELS APPLIED TO PROTOCOLS

Pierre de SAQUI-SANNES, Rob. VINGERHOEDS

ISAE-SUPAERO – University of Toulouse
 {pdss, rob.vingerhoeds}@isae-supero.fr

Ludovic APVILLÉ

LTCl, TelecomParisTech, Université Paris Saclay
 ludovic.apvrille@telecom-paristech.fr

ABSTRACT: *The paper shares an experience in using SysML and the free, open-source software TTool for protocol modelling and communication architecture validation. A dialogue between a pilot and a control tower serves as running example to demonstrate the benefits of complementary model analysis techniques: simulation, model checking, and verification by abstraction. The proposed method may be adapted to other modelling languages and tools.*

KEYWORDS: *Modelling, SysML, Simulation, Formal Verification, TTool, Protocol.*

1 INTRODUCTION

The term “protocol engineering” (Bochmann *et al.*, 2010) was coined four decades ago to denote a set of activities that locate protocol modelling at the heart of communication software design. Protocol engineering pioneered Model-Based Systems Engineering at a time where the MBSE acronym was not yet in use.

Relying on Extended Finite State Machines and process algebra, the protocol engineering community developed its own modelling languages and called them “Formal Description Techniques”: Estelle, SDL, and LOTOS. Associated tools took formal descriptions as a reference document for early checking of design errors in the protocol’s design trajectory, for automatic generation of executable code, and for test sequences generation.

Functionally similar tools have been developed for UML 2.5. The relevance of UML to protocol engineering has been extensively discussed in (Garduno Barrera and Diaz, 2011) relying on the TAU G2 simulator. (Apvrille *et al.*, 2004) also demonstrated the possibility to apply formal verification to UML models of protocols.

The advent of SysML (OMG, 2017) now questions the applicability of SysML in the same area. To contribute to this discussion, this paper discusses a case study of protocol modelling and communication architecture validation using SysML and the use of a free, open-source SysML Editor software: TTool. TTool includes a diagram editor, a simulator, complementary verification modules (model checking, verification by abstraction), and several code generators.

The paper is organized as follows. Section 2 defines a method associated with SysML and TTool. Section 3 specifies the connection set-up procedure that serves as running example and includes assumptions into the SysML model (Saqui-Sannes and Apvrille, 2016). The requirement capture, analysis and design stages of the

life cycle are the subjects of Section 4, 5, and 6, respectively. Section 4 to 6 addresses a nominal system, whereas Section 7 discusses a degraded situation (message loss). Section 8 shows how TTool can be interfaced with the model checker UPPAAL (UPPAAL, 2017). Section 9 surveys related work. Section 10 concludes the paper.

2 METHOD

The System Modelling Language, or SysML for short (OMG, 2017), is a diagrammatic modelling language for systems engineering. The OMG standard (OMG, 2017) defines a wide spectrum language, offering tool manufacturers opportunities for customizing the SysML syntax to handle real-time or other types of systems. It is important to note that the SysML standard defines a notation, but not a way of using it. A real challenge is therefore to convince practitioners to incorporate SysML into the method in use in their development approaches and indeed in their company.

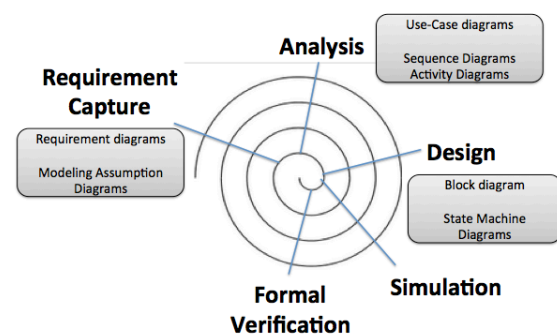


Figure 1: Method and SysML diagrams

As far as protocol engineering is concerned, the method (Figure 1) associated with SysML covers the requirements capture, analysis and design steps of traditional V life cycle. To cope with complexity, the method is further incremental, as suggested by the spiral depicted by Figure 1. Besides the three aforementioned

steps, Figure 1 underlines the importance of simulation and formal verification as two techniques enabling early checking of models against design errors. Figure 1 also indicates the names of the SysML diagrams used by each step in the method.

Requirements capture is the first stage in Figure 1. The “Modelling Assumption Diagram” is not part of the OMG-standard, but is supported by TTool and lists simplifications and other assumptions made at the time of creating the model. Requirement Diagrams, which are part of the SysML standard, define stakeholder, user, and system requirements.

The second stage in the method depicted by Figure 1 is Analysis. A use-case diagram defines the functions and services to be offered by the system and connects these functions or services to the system’s environment (actors). Use-cases are documented by scenarios (sequence diagrams) and flow-charts (activity diagrams).

The third stage in the method is design. A block instance diagram defines the architecture of the system. In each block instance lays a finite state machine diagrams defining the behaviour of the block instance.

To make the method in Figure 1 more specific to communication systems, we need to encompass concepts borrowed from the ISO Basic Reference Model (ISO, 2000): layered design, protocol, service, protocol data units and service primitives (Figure 2).

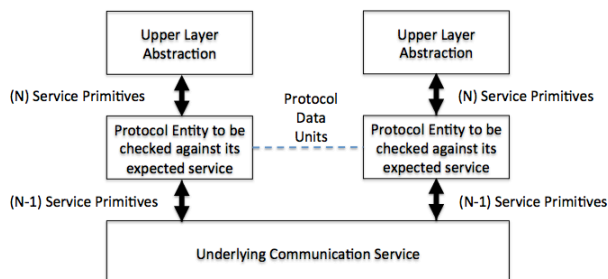


Figure 2: 3-layer architectural pattern

The 3-layer pattern of Figure 1 influences the way we structure the diagrams identified in Figure 1.

At requirement capture stage, separate modelling assumptions diagrams are created to address the protocol entities, the underlying network and the user-applications. Similarly, three families of requirement diagrams are created.

At analysis stage, two protocol entities are considered to be the system of interest, whereas the underlying network and the user-applications form the system’s environment. Use-cases are documented by means of sequence diagrams. Services scenarios are built up first, considering the system as a black box. Then, they are split up to make the protocol entities appear.

Finally, at the design stage, a 3-layer architecture is defined to comply with the pattern in Figure 2. The architecture is made up of block instances that exchange messages. Of specific interest are the service primitives. The protocol machines are precisely defined. Conversely, the upper layers and the communication service are modelled at a higher degree of abstraction.

3 MODELLING ASSUMPTIONS

This section presents the original specification and lists modelling assumptions. The original specification (Section 3.1) was extracted from the Future Air Navigation System (FANS, 2017) specification and defines a connection set up procedure initiating a dialogue between an aircraft and a control.

3.1 Original Specification

When initiated either by flight crew action or an automatic trigger, an ATS Facilities Notification Contact message is sent and Avionics Timer ATST1 is set. The aircraft AFN application then awaits an AFN Acknowledgement message from ground. Subject to prior agreement with the service provider, the supplementary address field of the ATS Facilities Notification Contact Message may contain an abbreviation of the ATC center address to which the message should be delivered (such as the corresponding IATA or ICAO code). If a successful AFN Acknowledgement message is received within the period of time of ATST1, a positive indication is given to the flight crew and timer ATST1 is cancelled. The AFN acknowledgement message will contain the full 7-character address of the ground AFN end-system. This address should be used in all subsequent AFN messages. If the aircraft AFN receives an AFN acknowledgement message with a non-zero reason code or timer ATST1 expires, an error indication is given to the flight crew.

3.2 Initial Work on the Specification

The specification in Section 3.1 is ambiguous, incomplete and partly inconsistent. Table 1 provides clarifications.

Specification/Problem	Modelling Decisions
The connection set up may be started by the crew or by an automatic trigger.	The pilot will start the connection set up procedure.
Which fields for the Facilities Notification Contact message?	The messages will have no parameter.
After issuing a connection request, the board software sets ATST1 to an undefined value.	ATST1 = 10 minutes.

Table 1 Decisions taken before modelling in SysML

3.3 Modelling Assumptions Diagram

Experience has shown that models are scarcely self-contained and need to be documented to facilitate their sharing and reuse. In particular, a model remains hard to understand for somebody who does not know about the simplifications and more generally the assumptions made by the model’s designer. For instance, a protocol developed on top of a pre-existing communication service will be modelled differently depending on whether the service is lossy or not. Consequently, TTool supports Modelling Assumptions Diagrams (MAD) to encourage the model designer to include modelling assumptions into his or her model.

Figures 3, 4 and 5 depict the MAD developed for the case study. The diagrams define tree structures with two types of nodes. The boxes with multiple fields define the assumptions and associate them with attributes that explain the origin of the assumption (e.g. the model creator), the status of the assumptions (stating whether it is applied or not in the current version of the model, and the scope indicating, e.g., that an assumption is the consequence of the modelling activity carried out by the designer and not a limitation of the verification tool. Among the relations depicted by Figure 3, let us note the use of <<impact>> to point out how modelling assumptions influences design diagrams. Another important relation is the versioning relation used in Figure 4; it contributes to make the modelling process an incremental one.

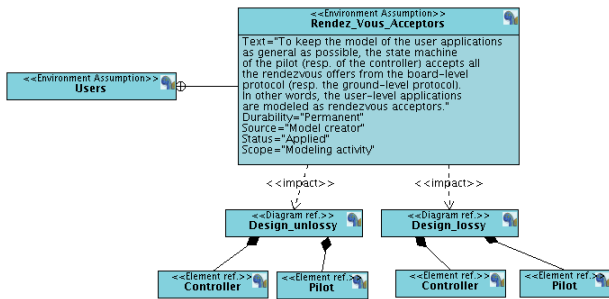


Figure 3: Modelling Assumptions Diagram

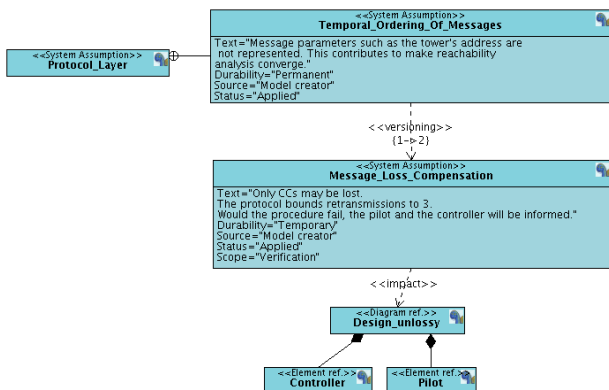


Figure 4: Modelling Assumptions Diagram

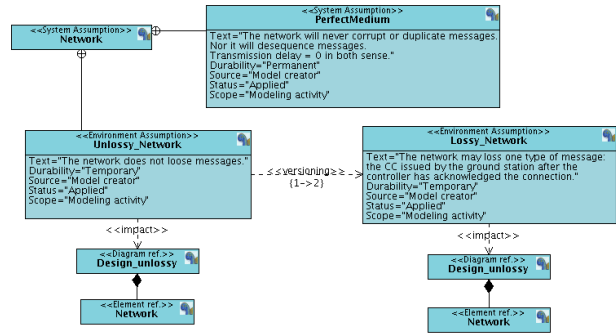


Figure 5: Modelling Assumptions Diagram

4 REQUIREMENTS CAPTURE

The goal of system architecture activities is to define a complete solution based on principles, concepts and properties logically related and consistent with each other. Such solution should have suitable characteristics and properties, matching as well as possible to the problem expressed by a set of system requirements, traceable to mission/business and stakeholder requirements, and traceable throughout life cycle phases and corresponding engineering tools (e.g., mechanical, electronics, software ...). This underlines the necessity to obtain pertinent requirements and explains why SysML supports requirement diagrams, a type of diagram not taken on board by UML. As SysML is a language and not a method, there are no constraints to the writing style of requirements. In contrast, an advantage of requirement diagrams is to oblige the system designer to structure and organize the requirements, and to show how this relates to other diagrams in the model.

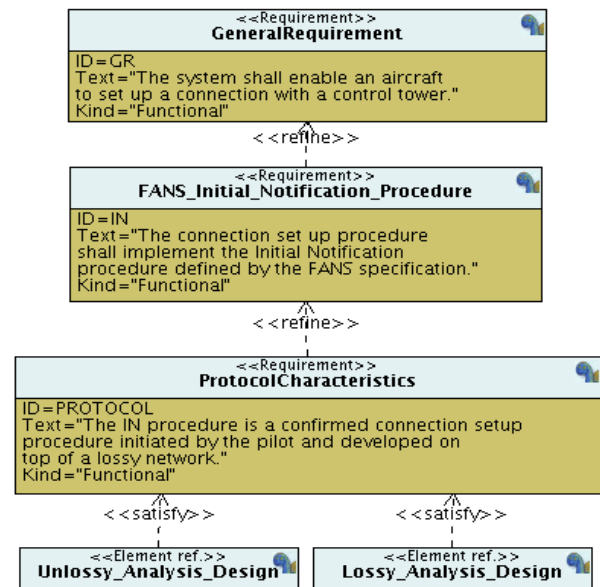


Figure 6: Requirement Diagram

Figure 6 depicts one of the requirement diagram developed for the case study. The tree structure clearly

appears with two types of nodes and two types of arrows. Each node depicted by a box contains one requirement together with its unique identifier, a text, and a categorization between functional and non-functional requirement. The <<refine>> relation from R1 to R2 allows R2 to add more precision to R1. Each <<satisfy>> relation links one design diagram element to one requirement.

5 ANALYSIS

The use-case diagram in Figure 7 defines the boundary of the communicating system. It includes one use-case named InitialNotification and connects the latter to three actors that respectively represent the Pilot, the Controller and the pre-existing network the communication system relies on.

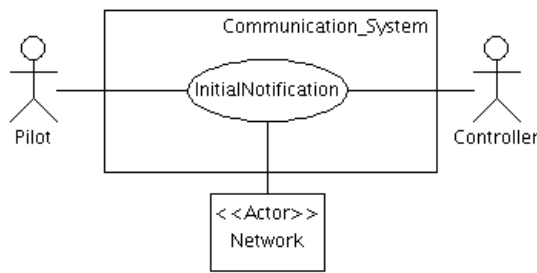


Figure 7: Use-Case Diagram

Two scenarios document the use case. Figure 8 depicts a service scenario where the communication service is a black box and the only messages are services primitives.

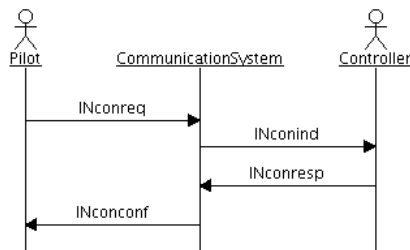


Figure 8: Nominal Case – Service scenario

The protocol scenario in Figure 9 splits up the service scenario, keeping coherence with the latter in terms of service primitives, and adds Protocol Data Units.

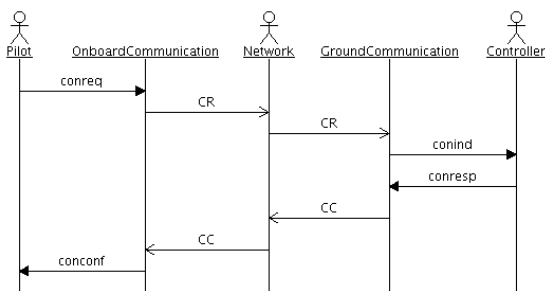


Figure 9: Nominal Case – Protocol Scenario

6 DESIGN

6.1 Architectural Design

The protocol engineering community has adopted the principle of layered design. In practice, two or more protocol entities rely on some pre-existing communication service to render their respective upper-layer user application a value-added service.

Thus, in Figure 10, OnboardCommunication and GroundCommunication both rely on a communication service modelled by a FIFO queue channel to offer a service to Pilot and Controller, respectively.

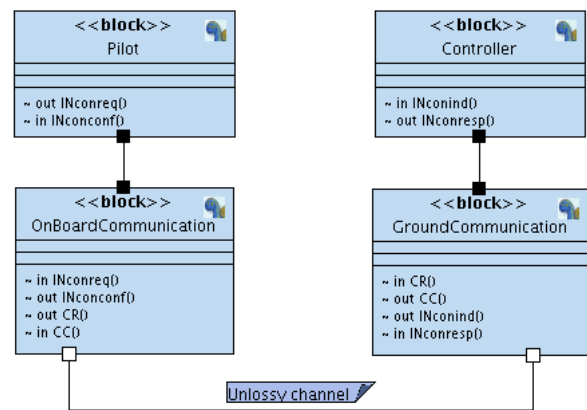


Figure 10: Architecture

Figure 10 depicts an architecture of block instances that communicate via interconnected ports. Black squares denote rendezvous ports that pairs of block instances use for synchronized communication. White square conversely denote FIFO queued communication ports.

6.2 Behavioural Design

Figures 11 to 14 depict the inner workings of the block instances of the architecture in Figure 10. The state machines use three symbols: rounded rectangles modelling states, and message reception / emission.

To keep the Pilot as abstract as possible, its state machine accepts all rendezvous from OnboardCommunication, either for initiating the procedure with INconreq or for successful completion of the connection set up procedure with INconconf.

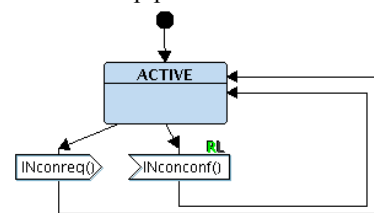


Figure 11: State Machine of the Pilot

The protocol machine of *OnboardCommunication* (Figure 12) enters the *IDLE* state and waits for the *Pilot* to issue a connection request. The protocol machine transforms the latter into *CR*, sends it via the unlossy channel and waits for the remote protocol machine to confirm. Upon reception of a *CC* conveyed by the unlossy channel, the protocol machine sends a confirmation to the *Pilot*.

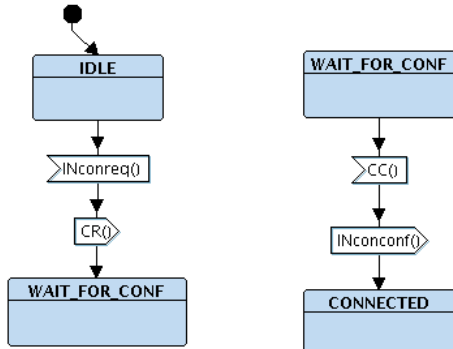


Figure 12: Protocol Machine – On Board Side

The protocol machine of *GroundCommunication* (Figure 13) enters the *IDLE* state and waits for the unlossy channel to convey a *CR* sent by the remote entity. Upon reception of *CR*, the protocol machine issues an indication towards the *Controller* and waits for the latter to respond. Upon reception of *INconresp*, the protocol machine issues *CC*, a confirmation the unlossy channel will convey towards *OnboardCommunication*.

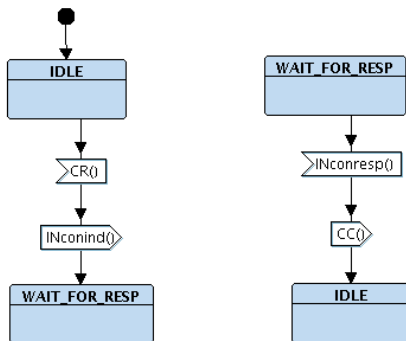


Figure 13: Protocol Machine – Ground Side

Like the state machine of *Pilot*, the one of *Controller* (Figure 14) is a rendezvous acceptor. It handles indications and responses.

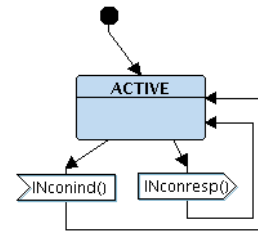


Figure 14: State Machine of the Controller

Unlike requirement capture and analysis that merely generate a set of diagrams, the design phase is not only a matter of drawing: the simulator animates design diagrams and formal verification modules explore the behaviour of the model relying on mathematics rather than chance.

6.3 Model Simulation

TTool is a free and open-source tool that supports several UML profiles, particularly SysML. Figure 15 depicts the main functions on the tool. Note that executable code generation goes beyond the scope of the paper.

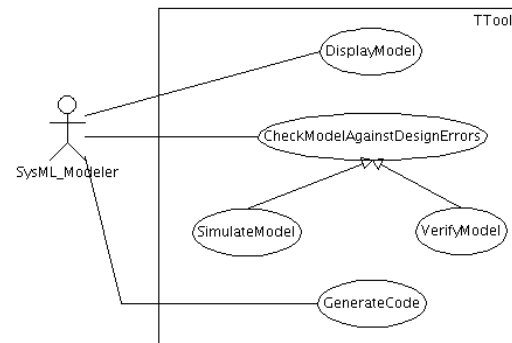


Figure 15: Overview of TTool

The Simulator of TTool (Figure 16) enables animation of state machine diagrams. It takes as input a syntactically- and type-checked SysML model and computes the model’s initial global state. Step by step firing of transitions enables early debugging of the model by joint observation of simulation traces in the form of sequence diagrams, annotations on the SysML model itself and display of the state, variables and other elements contained in the blocks the system is made up of. Random firing of transitions enables further exploration of the system’s behaviour until a deadlock situation or a termination state is encountered.

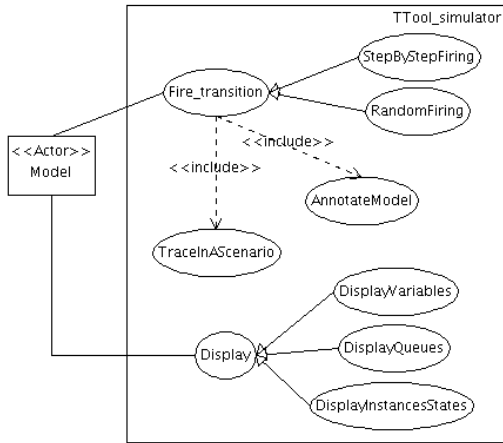


Figure 16: The Simulator of TTool

In Figure 17, the simulation traces takes the form of a sequence diagram and completely depicts the connection procedure.

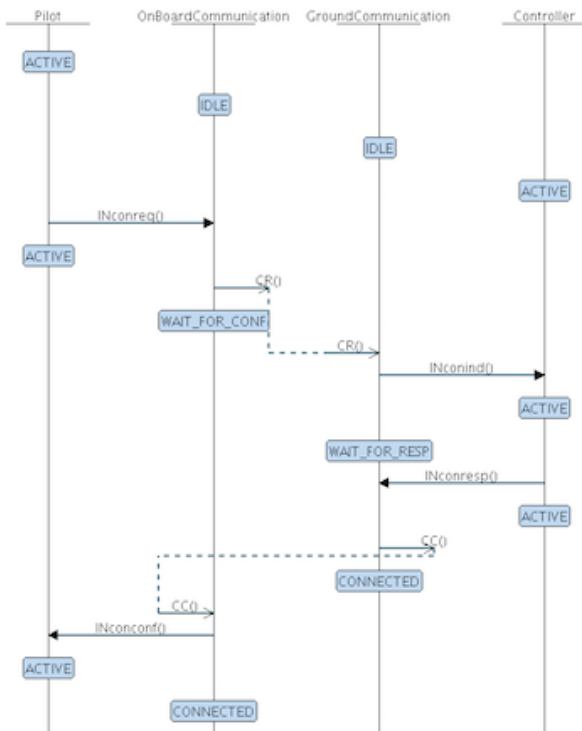


Figure 17: Simulation Trace

6.4 Model Checking

One of the most widespread verification approaches is reachability analysis. Relying on a systematic analysis of the state space of the system under design, reachability analysis may output a so-called “reachability graph” representing all the valid execution paths and states of the system starting from its initial state.

On a general principle, reachability analysis faces the well-known “state explosion problem”. Assuming the graph has been computed, the question of exploiting that

graph is asked for. An approach that checks whether one property is met or not, and provides a yes/no answer is known as “model checking” (Figure 19). An approach that processes the reachability graph as a Labelled Transition Systems and applies minimization techniques to come up with an abstract view of the system is known as “verification by abstraction” (Figure 18).

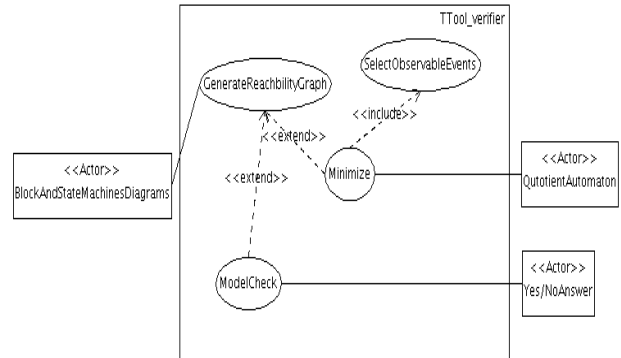


Figure 18: Verification Capabilities of TTool

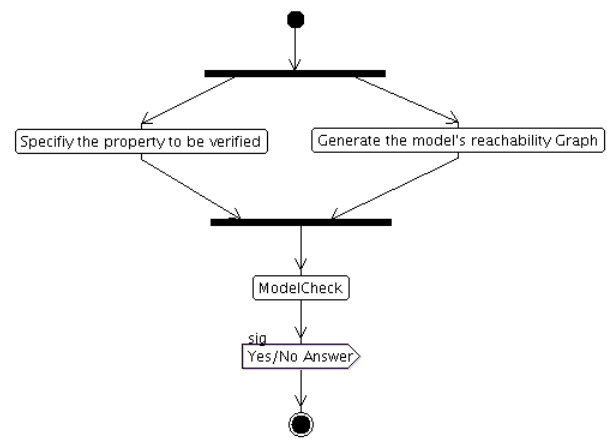


Figure 19: Principles of Model Checking

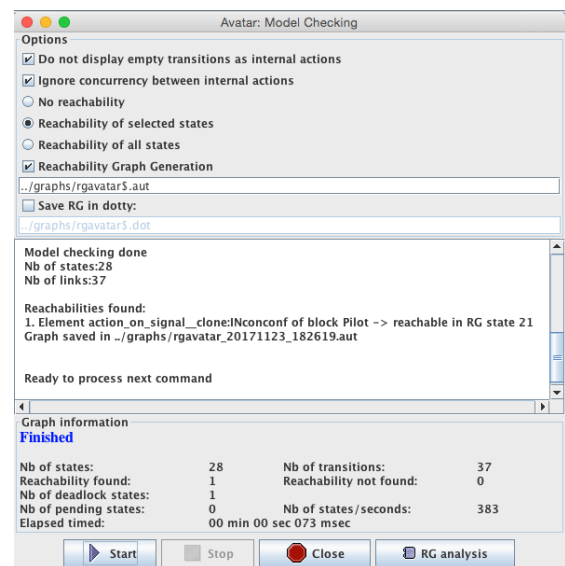


Figure 20: Model-Checker Interface

TTool natively offers a model-checker that does not require expressing a property using a formal language. Any state or event in a state machine may be checked for reachability as soon as it is marked with a RL label (see INconconf reception in Figure 11). Figure 20 shows the model-checker interface and the way it states that INconconf is reachable.

6.5 Verification by abstraction

The reachability graph (not represented here for space reasons) encounters 28 states and 37 transitions. Labelling the reachability graph with service primitives as observable events (Figure 22) enables application of minimization techniques. Minimization with respect to Milner’s observational equivalence outputs the quotient automaton in Figure 23. Thus, from a complete reachability graph that remains hard to explore by hand, verification by abstraction allows one to characterize the service provided by the protocol.

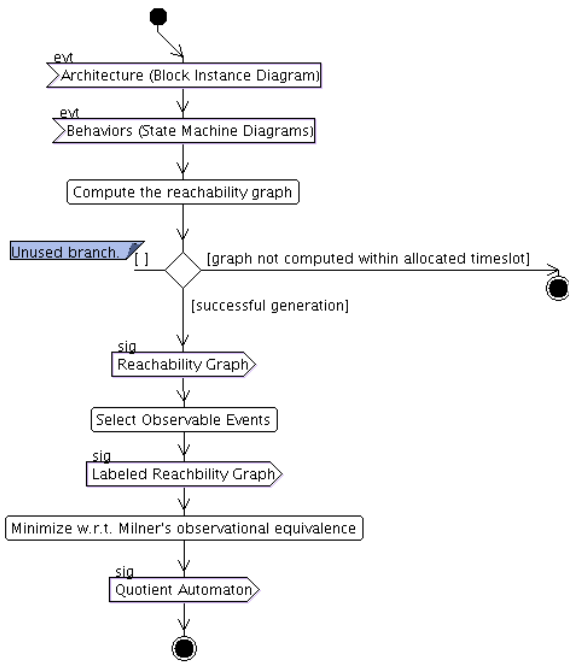


Figure 21: Principles of Verification By Abstraction

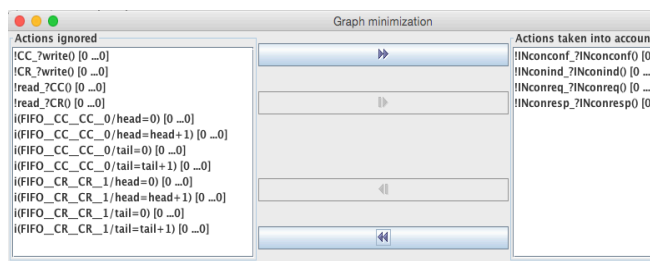


Figure 22: Service Primitives are Made Observable

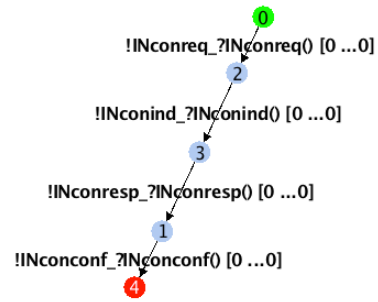


Figure 23: Quotient Automaton

7 DEGRADED CASE

Protocol design is an incremental process. The first version of the protocol machines, as well as of the service primitives list, fits with a perfect underlying communication service that neither losses nor corrupt messages. Version 2 of such model leverages initial restrictive hypotheses and approaches a more realistic version of the model where the underlying communication service may lose one type of message, namely CC.

For space reasons, the paper does not reproduce version 2 of the model entirely. Figure 24 shows how the state machine of the board protocol entity has been extended to handle a (bounded) retransmission counter and to inform the pilot in case of unsuccessful set-up procedure.

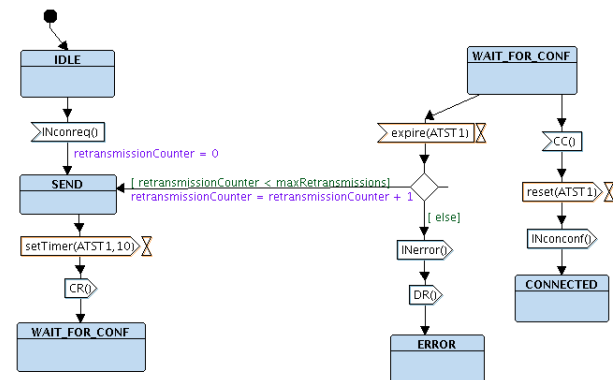


Figure 24: New Protocol Machine On the Ground Side

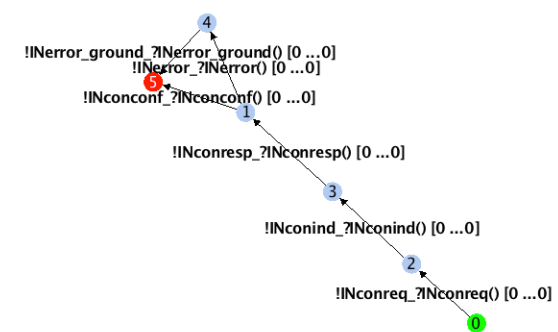


Figure 25: Quotient Automaton

Figure 25 depicts the quotient automaton obtained for the extended SysML model. The triangle distinguishes between successful completion and unset connection.

8 VERIFICATION USING TEMPORAL LOGICS

Safety pragmas can be used in design models in order to capture complex properties expressed in a reduced form of CTL. After checking the syntax of these pragmas, TTool can automatically invoke UPPAAL (UPPAAL, 2017) in order to verify these pragmas.

Pragmas must follow the following format:

- $A[] p$ means that whatever the state of the modelled system, p must be satisfied
- $A\langle\rangle p$ means that p must be satisfied in at least one state of all possible execution paths
- $E[] p$ means that p must be satisfied in all the states of at least one execution path
- $E\langle\rangle p$ means that p must be satisfied in at least one state of one execution path.
- $p \rightarrow q$ means that whenever p is satisfied in an execution path, q will eventually be satisfied in the same execution path.

Figure 26 displays two safety pragmas for the ERROR state in the `OnBoardCommunication` block. The first pragma is not satisfied, which means that the ERROR state is not reached in all execution paths. Yet, since the second pragma is satisfied, we know that there is at least one execution path that goes through the ERROR state.

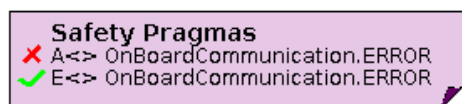


Figure 26: Safety pragmas

9 RELATED WORK

Unlike SysML tools that associate formal verification with activity diagrams (e.g., Ouchani et al., 2014) TTool applies model checking to block and state machines diagrams.

TTool also implements verification by abstraction, which, to our knowledge, is not offered by other SysML tools. Let us remark that verification by abstraction was already applied to protocol models expressed in Estelle (Courtiaat and de Saqui-Sannes, 1992).

In terms of protocol modelling and communication architecture validation, the book by Garduno and Diaz (Garduno Barrera and Diaz, 2011) discusses a complete case study of Transport-level protocol modelling using UML/SDL, the variant of UML 2 supported by the TAU G2, a tool that offers simulation capabilities similar to the one offered by TTool. Unlike TTool, TAU G2 does not offer verification capabilities.

10 CONCLUSIONS

With their capacity to abstractly model services and protocols, and to check a model against its expected service, SysML and TTool help designers in the early stages of the design trajectory of communicating systems. SysML diagrams and TTool outputs are closed to the notations protocol designers are familiar with. This paper addresses safety issues and concentrates formal verification on the temporal ordering of events. SysML-sec, another language supported by TTool (Li et al, 2017), could be used to detect security flaws. Finally, SysML models are also used to generate test sequences.

REFERENCES

- Aprville, L., Courtiaat, J.-P. Lohr, C., de Saqui-Sannes, P., 2014, TURTLE: A Real-Time UML Profile Supported by a Formal Validation Toolkit, IEEE Trans. on Software Engineering, 30 (7), p. 473-487.
- Bochmann, G.v., Rayner, D., West, C. H., 2010, Some Notes on the History of Protocol Engineering, Computer Networks, 54(128), p. 3197-3209.
- Courtiaat, J.-P., de Saqui-Sannes, P., 1992, ESTIM: An Integrated Environment for the Simulation and Verification of OSI Protocols Specified in Estelle, Comp. Networks and ISDN Systems, 25(3), p. 83-98.
- FANS 2017, Future Air Navigation Systems, https://en.wikipedia.org/wiki/Future_Air_Navigation_System.
- Garduno Barrera, D., Diaz, M., 2011, Communicating Systems with UML2: Modeling and Analysis of Network Protocols, June 2011, Wiley-ISTE, 288 p.
- ISO, 2000, ISO/IEC 7498-1:1994, Information technology -- Open Systems Interconnection -- Basic Reference Model: The Basic Model, <https://www.iso.org/standard/20269.html>.
- Li, L.W., Lugou, F., Aprville, L., 2017, Security Modeling for Embedded System Design", 4th International Workshop on Graphical Models for Security, Santa Barbara, CA, USA.
- OMG, 2017, Systems Modeling Language 1.5, May 2017, <http://www.omg.org/spec/SysML/1.5/>.
- Ouchani, S., Ait Mohamed, O., Debbabi, M., 2014, A formal verification framework for SysML activity diagrams, Expert Systems with Applications, (41) 6, pp. 2713-2728.
- Saqui-Sannes, P. de, Aprville, L., 2016, Making Modeling Assumptions an Explicit Part of Real-Time Systems Models", 8th European Congress on Embedded Real Time Software and Systems (ERTS), Toulouse, France, pp. 27-29
- TTool, 2017, <http://ttool.telecom-paristech>.
- UPPAAL, 2017, <http://www.uppaal.org/>.

EDUCATIONAL CHALLENGES FOR CYBER-PHYSICAL SYSTEMS MODELLING

Rob VINGERHOEDS, Ahlem MIFDAOUI, Pierre de SAQUI-SANNES

ISAE-SUPAERO – University of Toulouse
{rob.vingerhoeds, ahlem.mifdaoui, pdss}@isae-supero.fr

ABSTRACT: “Cyber Physical Systems” (CPS), continuously connected to the rest of the world, communicating autonomously with other systems and with a certain degree of autonomy for taking decisions, pose certain challenges to industries, but to the training of the next generation of engineers as well. The inclusion of different aspects of CPS in the educational programs is important to prepare the students for the development of such systems. The paper aims at looking at the development of CPS, identifying where and how it is important to train the students in particular. The core of the paper is dedicated to multi-paradigm modelling as a key enabler for mastering the complexity of CPS and focuses on educational challenges and relies on the authors’ experience with the SysML tool TTool and the WoPANets tool that implements network calculus theory.

KEYWORDS: Cyber-Physical Systems, Systems Models, Computer-Aided System Design, Education.

1 INTRODUCTION

“Cyber Physical Systems”, or CPS for short, denote a family of systems that are continuously connected to the rest of the world, communicating autonomously with other systems and having a certain degree of autonomy for taking decisions. One can think of a fleet of drones that jointly perform tasks and that therefore need to exchange information, search for weather information, etc. Another example concerns autonomously driving vehicles (e.g. Google Cars), exchanging information with each other, search for the shortest routes, getting up-to-date weather information, etc. One may also think of the next generation of production systems, already sometimes referred to as Cyber-Physical Production Systems (Monostori, 2014).

Such systems with a high degree of complexity pose challenges to industries, but also to educational institutes such as technical universities, for the training of the next generation of engineers to prepare them for the development of these critical systems. The topic is not new, but now that such systems are gaining more and more attention in industrial environments the needs for improved training become clearer.

Whereas frequently engineers still today continue to reason over solutions rather than over requirements, this approach may induce problems in CPS development due to non-deterministic environment in which CPS operate and to the embedded intelligence of the CPS itself. A structured approach, addressing the overall picture, looking at all of the requirements, gradually zooming in on designs and eventually on solutions is essential. It is therefore important that students can appreciate the com-

plexity of CPS in the design phase, understand the difference between the behaviour of one system and the behaviour of several systems working together, and being able to analyse the impact of different systems working together.

Mastering complexity is an important issue in CPS design and development, and engineer-training programs need to address this. This is where modelling languages, tools and methods, successfully experienced with real-time systems, have a role to play. Among numerous mature tools, the paper addresses the TTool (Apvrille et al., 2016) and the WoPANets (WoPANets) tools for they have been used in both industry projects and engineering training sessions. TTool edits and formally analyses models expressed in SysML (OMG, 2017), a wide-spectrum systems modelling language with a high potential for CPS design. WoPANets enables proving properties on communication links, a function that usually does not exist in SysML tools.

When training students on the development of CPS, a real hands-on experience allows students to really “feel” the complexity as well as the possibilities that such systems offer.

In the paper, joint use of TTool and WoPANets is part of a more general discussion on the development of CPS and on the training of the next generation of engineers that will need to develop such CPS. To achieve that goal, the paper is organized as follows. Section 2 first gets back to the existing situation with embedded systems and highlights potential pitfalls. Section 3 continues on the development of CPS as a logical evolution from embedded systems, and derives the needs for improving the education of the next generation of engineers. Sec-

tion 4 presents the TTool and WoPANets tools respectively. Section 5 proposes the way in which those two tools together can help to address a part of the highlighted educational challenges. A roadmap on how such things could be integrated in today's curriculum is presented. Section 6 illustrates the use of TTool and WoPANets. Section 7 concludes the paper.

2 DEVELOPING EMBEDDED SYSTEMS

With ever-faster evolutions in technology, ever more challenging objectives and even stronger constraints imposed by certifying authorities, designing embedded systems has changed dramatically over the last 15 years. This has a strong impact on all involved stakeholders in embedded systems design, e.g. integrators (OEM's) or suppliers.

It is fair to say that the overall complexity of embedded systems is increasing rapidly, leading to a strong need for support during the development cycle. As a result, it is no longer possible to predict with certainty the total system behaviour just by examining the individual components or the sum of them. Each may have a certain defined behaviour, but bringing several components together and having the components interact may bring unexpected results. A strong coherent design approach is

needed, allowing bringing the individual blocks together, and analysing the overall system behaviour.

The structured use of a systems engineering approach in this respect is logical. The main phases are shown in Figure 1 and are discussed here. Starting from feasibility studies, a concept of operations, expressed user requirements, and several other information sources, the system requirements are obtained. In successive steps the designs are refined step-by-step passing via higher-level designs into detailed designs.

For certain targeted functionalities, different realisation technologies can be chosen, e.g. an electronic component vs. a software implementation. These choices lead to a set of activities that can be done in each of the realisation technologies, e.g. electrical and electronics, mechanical, software. For each of the retained technologies, separate development cycles start for the detailed design phases, implementation and unit tests. Each of those development cycles has its own characteristics; for example the time that is needed for evolutions is inherently different for each of the technologies (software evolutions can be made significantly faster available for testing than for example a new layout for electronics).

At the system integration stage the complete system is being built. Verification and validation can now continue

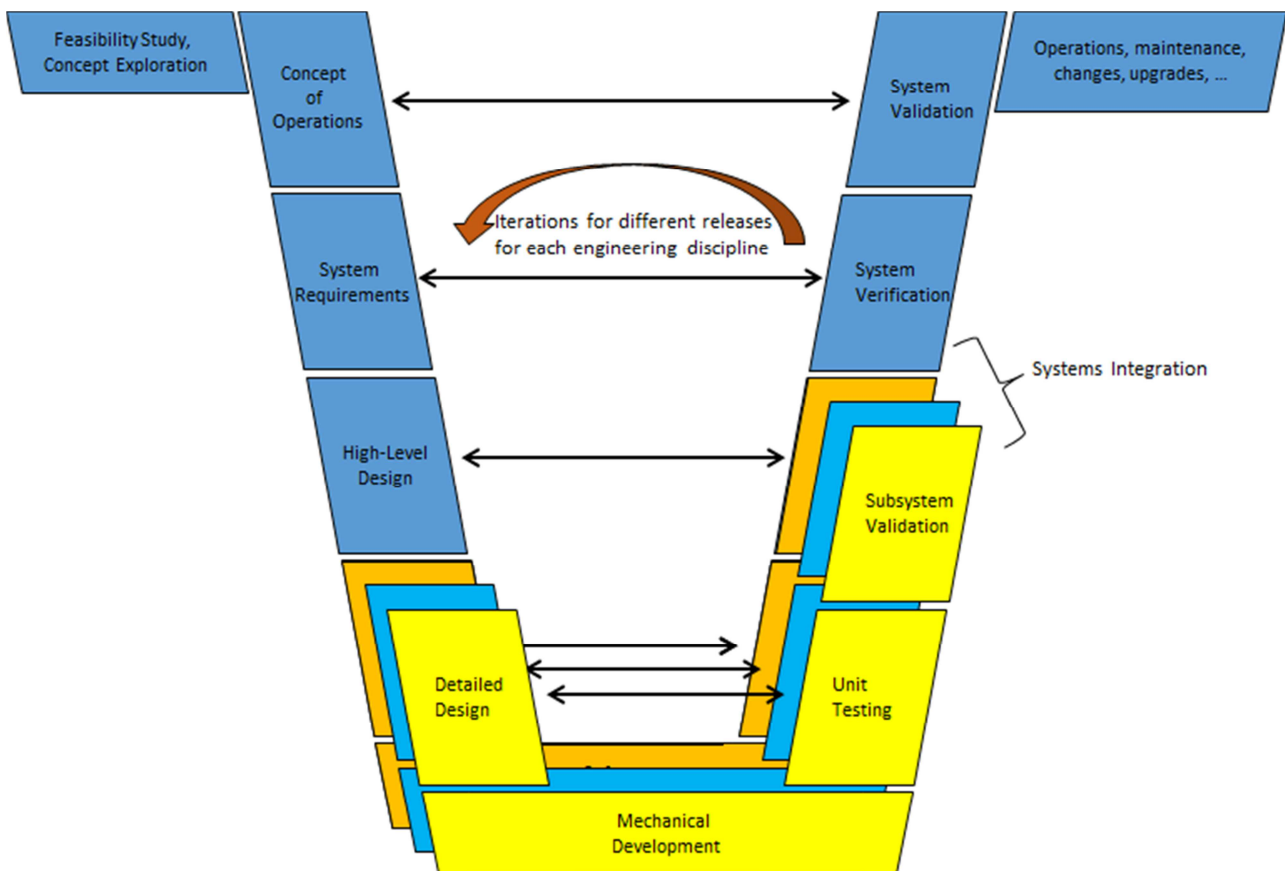


Figure 1: Typical development cycle following the V-Cycle model

on the basis of the complete system and if final validation and acceptance by the customer take place, the preparation for instance for production and operations can take place.

An iterative approach with feedback loops allows for getting back to earlier phases to change design options, or correct issues leading to an inherent recursiveness in the development process. A major challenge is to ensure that students fully appreciate the necessity of such structured approaches and that they live them accordingly.

Model-based approaches are used to better support and guide the system designers. In the first phases of the development process, the Systems Modelling Language (OMG, 2017) can be used to specify the system's requirements and constraints. SysML allows for a strong inter-disciplinary cooperation. This language can be used with tool chains such as TopCased (Vernadat, 2006) and TTool (Apvrille and de Saqui-Sannes, 2013), proposing different model checking tools to the system designer. At higher design levels these models are refined, enriched and are often linked to other models to give more insight in specific areas. Co-simulation of phenomena covering different systems, or different physical aspects of the system at hand, allows for the system designer to better understand the different impacts (Fitzgerald et al., 2015). Finally, at detailed design level discipline-specific models are used for the different realisation technologies allowing for a full expression of the required characteristics for the specific technology. A major challenge for students is to see the contribution of each of such models and to realise that it is vital to keep all of the different models coherent between them; any change made at a certain design level needs to be cascaded to the impacted models in a controlled manner.

3 DEVELOPING CYBER-PHYSICAL SYSTEMS

As a logical evolutionary step, CPS have seen the light of day, but important differences exist with embedded systems, which differences have an essential influence on the system's behaviour. Embedded systems are characterised by timed communication and information exchange in a controlled manner with a limited number of other systems; whereas CPS can observe time-delayed communication and interaction with an unknown number of (largely) unknown other systems (Jeschke, 2013). These characteristics make the inter-connectivity a vital part in the operations of the system and as such on its behaviour. The problem is that it is virtually impossible to model exhaustively this behaviour and therefore difficult to take it into account during the development cycle.

An example of such behaviour can be seen with autonomous self-driving cars. Such a car searches for weather forecasts and traffic information and will privilege a certain route over another one depending on the received information. When a person crosses the road in front of

the car, the latter will hopefully capture this presence and take corrective actions. The weather information, as well as the environment around the car detected by sensors, will play a vital role in the way the car will react. Whereas the sensors that detect the environment instantaneously will be able to give an accurate image to the car, the same cannot be said for the weather information and/or traffic information. Still this information will determine to a large extent the way the car will react. The behaviour of autonomous vehicles is induced by the not-mastered information (weather information and circulation of various sources, ...), meaning that the developer of the vehicle or the underlying sub-systems has no control on the information that the vehicle will receive, but which information is used in the reasoning process on board of these vehicles. This can lead to unpredictable (emergent) behaviour (Motus et al. 2005a and 2005b). Since such behaviour cannot be foreseen, it is not possible to manage it during the development (i.e. eliminate unwanted behaviour or reinforce required behaviour). However, it is possible to take measures to facilitate early detection of any emergent behaviour during operation, allowing addressing the impact.

To support engineers during the development, innovative cost-effective methods and tools are needed to design, analyse and verify the hardware and software architecture of such systems components, to guarantee predictability and reliability requirements.

The modelling part of CPS development needs particular attention (Broy, 2013). The real-time aspects of the system(s) and the communication network, often misunderstood, is a crucial topic for CPS. The system designer needs to ensure that the system itself can provide a suitable, acceptable response within the necessary time limit (Motus et al. 2005a and 2005b).

Complex systems are systems that display behaviour that is unexpected, emerging, and/or unpredictable. Characteristics of a complex system, as described by Flood (1990) and Poel (2009) typically include:

- a large scale of application,
- many parts, and many dependencies between parts and with the environment,
- the involvement of many stakeholders with different, sometimes conflicting, interests and goals, requiring extensive cooperation and coordination throughout the system's lifecycle,
- decision-making by stakeholders on the basis of uncertain, incomplete, inconsistent, or ambiguous information,
- in case of a networked system: autonomous sub-systems with different norms and values, rules of engagement and agreement, communication architectures, and requirements for trust,
- continuous change that may span many years, and
- not one unique objective measure to determine the quality of a design.

CPS are part of this class of systems and the main challenge is to master this complexity.

The challenges for training the next generation engineers, focussing on the final years of technical universities and/or specialised Masters Programs, can be formulated as follows:

1. Allowing students to better appreciate the complexity of CPS in the design phase,
2. Allowing students to see immediately the impact of changes in one system affecting the behaviour of another,
3. Allowing students to better appreciate how using the different realisation technologies also brings a more complexity to the development cycle itself, and
4. Allowing students to “touch” the challenge on how to satisfy two contradictory requirements, namely “guaranteeing the system predictability,” and “improving its adaptability” while meeting the certification requirements.

These objectives are to be seen in addition to a normal engineering curriculum that includes systems engineering, control and automation, and computer science.

4 TTOOL AND WOPANETS

This section presents the tools TTool and WoPANets, which are used within the framework of CPS education.

4.1 TTOOL

Users of the free software TTool may develop a system model by following a tree-step and iterative design process that mostly reuses the syntax of the OMG-based SysML (OMG, 2017) with some extensions and one major restriction (no continuous flow).

1) *Requirement capture.* A Requirement Diagram hierarchically organizes safety and security requirements with explicit notion of refinement and derivation of technical requirements from logical ones. TTool further makes modelling assumptions as an explicit part of the model in the form of Modelling Assumptions Diagrams.

2) *Analysis.* One or several Use-Case Diagrams identify the main functions and services to be offered by the system. Use cases need to be documented by scenarios (Sequence Diagrams) and flow charts (Activity Diagrams).

3) *Design.* A Block Instance Diagram defines the architecture of the system. Each block has an internal behaviour defined in the form of a State Machine Diagram.

The Block Instance and State Machine diagrams constitute an executable design model. They may be step-by-step debugged and randomly run using the simulator. Further, a model-checker enables formal checking of design models against logical and temporal properties. Note: TTool implements a press-button approach that does not request writing or reading any piece of formal code.

4.2 WOPANETS

To cope with the real-time issues, in particular the predictability of the overall timing behaviour, the system designer will need to select the right network parameters, in the face of the complexity described above. Without the support of dedicated tools, choosing the right network parameters respecting the system constraints becomes a difficult task for the designer. Different methods and tools exist to perform analysis for the communication network part within the framework of an overall systems development approach. The WoPANets tool (Worst-case Performance Analysis of embedded Networks) is one of them, which:

- performs a system level performance evaluation based on Network Calculus (Le Boudec 2001)
- includes an optimization process for design space exploration, which allows to enhance the design process time and costs,
- is capable of using input designs based on existing design approaches (UML, AADL),
- integrates high-level models of the most common real-time networks,
- is easy to use by any designer without any specific knowledge of the used analytical formalism due to an ergonomic GUI.

Within the framework of enhancing educational programs for CPS, the WoPANets tool offers diverse possibilities.

5 PROPOSAL

Whereas the necessity for special focus on CPS development in the training programmes is clear, different methods can be distinguished to realise this (Atkins and Bradley, 2013) (National Research Council, 2015), such as:

- enhancing existing control and automation courses with topics related to communication, real-time systems, and systems engineering,
- reinforcing computer science courses in the final years of engineering studies to master the CPS complexity,
- reinforcing courses on formal methods to promote formal verification and validation,
- introducing special and new specialisations aiming at CPS.

As CPS require integral, multidisciplinary design, any new specialisation needs to propose contributions from the concerned domains, e.g. control and automation, systems engineering, computer engineering, communication, signal processing, and mathematical modelling. Additionally contributions can be expected on emerging CPS applications, such as multi-core and many-core architectures, artificial intelligence, internet of things, cloud computing and Big Data.

The approach proposed in this paper focuses on the development and the design of CPS though enhancing existing control and automation courses for graduate students and extending them with hands-on experience.

The proposed approach can be applied to three different programs: the standard engineering program, specialised Master's programs, and continuous education.

- For the standard engineering program, the specialisation track builds on a strong common scientific program, a strong contribution on systems engineering, and specialisation courses in real-time and autonomous systems, as well as communication networks.
- When looking at a specialised Master's type of diploma on CPS, contrary to the example mentioned above, the idea is not to have a program building uniquely on computer and electrical engineering. There is also a need for mechanical engineering, control and signal processing. The proposal therefore aims at allowing persons with a Bachelor's degree in mechanical engineering, aerospace engineering, electrical and electronic engineering or computer science to start this Master's program. There is a need for a semester adapted to the students' background to master the CPS basics, followed by a common course program to master advanced CPS features.
- A similar approach can be proposed for more continuous-education type of training targeting engineers already working in industry looking for gaining understanding in the design of CPS.

The key element of the proposed approach concerns the hands-on experience. In such complex development situations, the best way for students to really appreciate the inherent complexity is to allow them to play with such scenarios. The selected tools need to allow the students to draw the necessary conclusions out of their experiments. The hands-on experience includes modelling of complex systems and analysing their behaviour based on both simulation and formal analyses.

Modelling complex systems requires a structured, step-by-step, top-down approach. This entails amongst others modelling the system's behaviour, the interfaces to other systems, and the behaviour of those interfaces. Hands-on

experience is a suitable way to teach this need for structured top-down modelling approaches, while postponing any solution-specific issues until later. Based on a unique expertise on modelling techniques for system and software engineering, the proposal is to build upon the Systems Modelling Language (SysML), adapted to better meet the expectations of real-time system designers. A SysML model conveys a point of view of some complex systems, and the tools developed for SysML model analysis (simulation / verification) are key enablers for early detection of design errors in the design trajectory of real-time systems. The experience gained with real-time systems is an asset for addressing CPS challenges.

The rationale behind the use of SysML is to encompass the various facets of a CPS in a single model and to apply the later complementary analysis techniques for checking the system design against design errors; the latter include inappropriate ordering of interactions between the CPS components and time mismatches. Modelling and analysing real-time behaviour and situation awareness, as well as handling of uncertainty in design parameters, will bring in an additional topic for which potential extensions to the modelling language; such extensions are subject of a separate research activity. Following the particular challenges related to emergent behaviour, the focus will be on how to identify such emergent behaviour early enough to allow taking the appropriate actions.

Using the modern techniques as in CPS to bring new features, enhance efficiency and improve productivity comes at the price of security risks due to the Internet openness, such as intentional intrusion into the system. This may provoke serious problems where an incorrect sequence, message insertion or duplication may bring the CPS to behave differently than what was anticipated. Even worse, particular attention needs to be assigned to eliminate the possibilities of intentional take-over of a CPS at distance, which could have dramatic consequences. These security problems need to be mitigated to satisfy the security requirements of CPS.

6 ILLUSTRATION

The proposed approach for hands-on experience for learning CPS development is built around TTool and WoPANets (see Figure 2).

The students need to develop their CPS in TTool, with all the features necessary to gain a full understanding of its requirements, the environment in which the CPS will operate, the other systems it will work with, the communication networks, etc. Such models should in a first step already allow the students to see more clearly as to the complete picture. Eventually other simulation models can be used so to enrich this picture, better understand some parts of it. TTool allows for executing a variety of analysis methods, based on simulation and formal meth-

ods. The idea is that the complete system is as such developed within TTool. Students can perform checking of their models using for example the approach explained in (Saqui-Sannes et al., 2018).

Let's take as example autonomous vehicles driving on roads. The starting point of the design is for the students to realise the mission of such vehicles. Brazier et al. (2018) define the mission of autonomous vehicles as

“To provide safe, comfortable, environmentally friendly, and timely autonomous road transportation of human passengers and baggage”.

This includes not only transportation, but also collision avoidance (safety), determining and adapting the route to take based on actualised real-time traffic information, weather information, user preferences, potentially communication with other vehicles driving on the road, and all of this energy efficient.

In the modelling phase, the functional requirement “can determine and adapt route on the basis of real-time information on traffic, weather, user preferences, services on route” requires specific attention. Indeed, it consists of communication tasks and interpretation tasks, for the moment to be modelled without having identified the specific sub-systems that will ultimately perform these functions. The students are expected to explore current technologies to see if and how such functionalities can be realised and in successive steps model to propose solutions for such sub-systems. The same applies for the collision avoidance and driving control sub-systems. These too need to be modelled in this stage, including all the components that may contribute to the realisation of this functionality.

When modelling such systems, one needs to model not only the vehicle and its subsystems (including the embedded intelligence in the vehicle, responsible for the

effectuation of the different reasoning tasks), but also the “receiving and sending” parts of the systems with which the vehicles communication (weather and traffic information, other vehicles, other road services ...).

This modelling may bring to the light for example inconsistencies or situations that need to be covered. For example, a question that merits specific attention is how to take on board specific and very local weather information. Hail storms for example have a short life-time and take time before they are included in the “overall” weather forecast, by which time it may be too late.

Once the students feel that all they functions and services related to autonomous vehicles are properly modelled in SysML in the TTool environment, attention needs to be turned to the vehicles and their communication networks. The WoPANets tool imports the design as from an early stage. The tool allows for performing worst-case analysis for the network performance, which will allow the student to quickly gain insight in how the overall system performs and where eventual adaptations will need to be made. To be able to perform such a network performance analysis, a priori knowledge on the communication traffic is essential. However, it is not always possible for adaptive dynamic systems to know sufficiently in advance how intense this traffic may be. A first estimation needs to be done based on the available information, and in iterations such assumptions may need to be adjusted.

In the case of very local weather phenomena for autonomous vehicles, some high-level designs are made in TTool using SysML, gradually adding more detail to the design, searching for solutions such as car-to-car communication or car-to-and-from-road-services communication. In this example, the road services receive from all cars within a specific range up to date weather information and make it in turn available to other vehicles driving in the vicinity. Note that if such a solution is

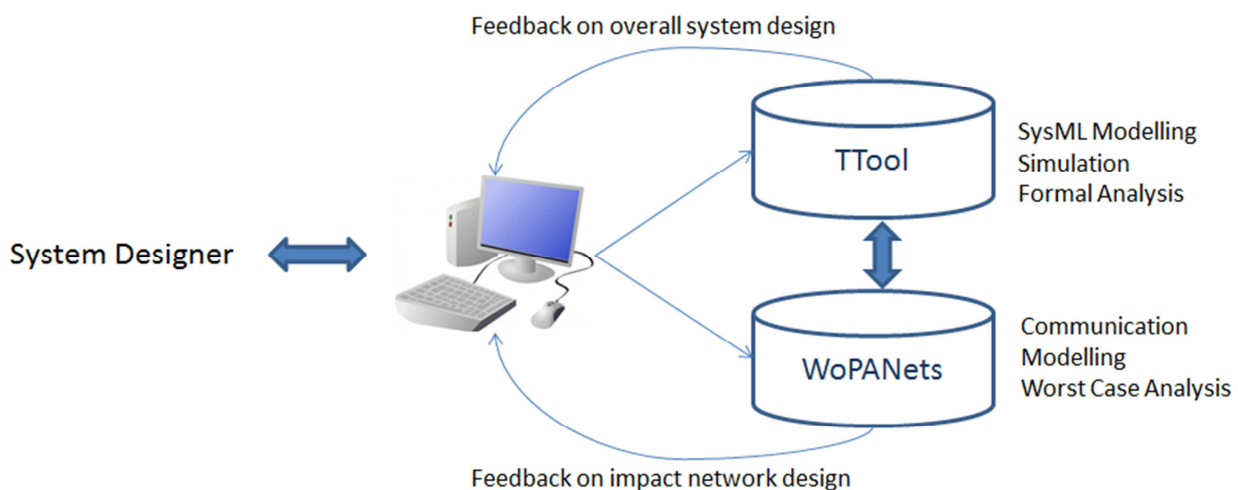


Figure 2: Set-Up for Hands-On Experience

retained, that in the SysML model of the vehicle there also will need to be a weather information module that records the local weather and sends it to the road services.

As soon as sufficient information is available on the communication between the autonomous vehicles and the road services, the necessary parts of the design are imported into WoPANets and using the first estimations on the traffic intensity first analyses can be made. This will give a first impression on the feasibility of such a solution and may yield useful feedback as specific additional technical requirements to the sub-systems both on vehicle and road services side.

In successive steps the normal development cycle as shown in Figure 1 is followed until the implementation stage is reached.

Looking at the autonomous vehicles example this means that the different control systems and their associated communication to road services, weather and traffic information are designed to a sufficient level of detail that the realisation can be started accordingly. For the students, this means having realised a complete design, including the necessary sub-systems and links to other systems, and having gained insight in the associated complexity.

The goal of the proposed approach is to confront in an interactive manner the student with the design choices he/she makes. Both TTool and WoPANets show a decisive role in this approach, each providing a part of the complete picture. Such hands-on experience will allow the student to see more clearly in the impacts of the different sub-systems, and will appreciate the way they all work together

7 CONCLUSION

The paper discusses the complexity for designing embedded systems and cyber-physical systems. The development process, following a systems engineering approach, is presented, and the points where specific attention is needed are highlighted. The paper continues with highlighting the inherent complexity that arrives with cyber-physical systems, and that require additional attention during the design phase. This complexity not only concerns the systems themselves, but also the communication aspects play a major role.

The paper discusses the use of SysML and the free software TTool, as well as the WoPANets tool throughout the design trajectory of cyber-physical systems. The chosen approach aims at confronting final year students with the complexity of cyber-physical systems, letting them have a real hands-on experience and therewith gaining awareness, learning to seek for pitfalls, and following a well-structured approach to CPS design.

Finally, the paper proposes a rough outline for inclusion of CPS in specialisation tracks for technical universities, for specialised Master's programs, and for continuous education programs.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Prof. Ludovic Apvrille who has developed TTool.

REFERENCES

- Apvrille, L., Saqui-Sannes, P. de, (2013), Requirements Analysis, chapter in *Embedded Systems: Analysis and Modeling with SysML, UML and AADL*, Edited by F. Kordon, J. Hugues, A. Canals and A. Dohet, Ed. ISTE / Wiley.
- Apvrille, L., Li, L., Roudier, Y. (2016) *Model-Driven Engineering for Designing Safe and Secure Embedded Systems*", IEEE workshop on Architecture Centric Virtual Integration, Venice, Italy..
- Atkins, E.M., Bradley, J.M. (2013), *Aerospace Cyber-Physical Systems Education, Guidance, Navigation and Control*, AIAA, August 19-22, Boston, USA.
- Brazier, F., Langen, P. van, Lukosch, S., and Vingerhoeds, R. (2018), *Design, Engineering and Governance of Complex Systems*, in: «Projects and People – Mastering Success», NAP Foundation Press, accepted for publication.
- Broy, M. (2013), *Modeling Cyber-physical systems, towards a holistic integrated approach*, Summer School CPS, Grenoble.
- Fitzgerald, J., Gamble, C., Gorm Larsen, P., Pierce, K., Woodcock, J. (2015), *Cyber-physical systems design: foundations, methods and integrated tool chains*, IEEE/ACM 3rd FME Workshop on Formal Methods in Software Engineering (FormaliSE), Florence, Italy, pp. 40-46.
- Flood, R. L., (1990). *Liberating Systems Theory*. In: *Liberating Systems Theory*. Contemporary Systems Thinking. Springer, Boston, MA.
- Jeschke, S. (2013), *Cyber-physical systems – History presence and future*, Industrial Advisory Board, RWTH Aachen University.
- Le Boudec, J.-Y, and Thiran, P. (2001). *Network calculus: a theory of deterministic queuing systems for the internet*. Springer Science & Business Media.

- Monostori, L. (2014), Cyber-physical production systems: roots, expectations, and R&D challenges, *Procedia CIRP*, Vol. 17, pp. 9-13.
- Motus, L, Vingerhoeds, R.A., Meriste, M. (2005a), Challenges for Real-Time Systems Engineering, Part 1: State-of-the-Art, *Proceedings of the Estonian Academy of Sciences and Engineering*, Vol. 11, Nr. 1, pp. 3-17.
- Motus, L, Vingerhoeds, R.A., Meriste, M. (2005b), Challenges for Real-Time Systems Engineering, Part 2: Towards Time-Aware Technology, *Proceedings of the Estonian Academy of Sciences and Engineering*, Vol. 11, Nr. 1, pp. 18-30.
- National Research Council (2015), *Interim Report on 21st Century Cyber-Physical Systems Education*, Report, The National Academic Press.
- OMG, 2017, *Systems Modeling Language 1.5*, May 2017, <http://www.omg.org/spec/SysML/1.5/>.
- Poel I. (2009). Values in engineering design. In: Meijers A, editor. *Handbook of the Philosophy of Science. Volume 9: Philosophy of technology and engineering sciences*. Amsterdam: Elsevier, pp. 973–1006.
- Saqui-Sannes, P. de, Vingerhoeds, R., Apvrille, L. (2018), *Early Checking of SysML Models Applied to Protocols*, MOSIM'18, accepted for publication.
- Vernadat, F., Percebois, Ch., Farail, P., Vingerhoeds, R., Rossignol, A., Talpin, J.P., Chemouil, D. (2006), *The TOPCASED Project - A Toolkit in Open-source for Critical Applications and System Development*. In: *Data Systems In Aerospace (DASIA 2006)*, Berlin, Germany, 22/05/2006-25/05/2006, European Space Agency Publications.
- WoPANets: Worst-Case Performance Analysis of embedded Networks, <http://websites.isae.fr/wopanets>.

ELABORATION D'OFFRES EN RELATION CLIENT-FOURNISSEUR : VERS UNE EXPLOITATION DES CONNAISSANCES POUR L'INGENIERIE DES RISQUES

R. AYACHI^{1,2}, D. GUILLON^{1,3}, F. MARMIER¹, E. VAREILLES^{1,4}, M. ALDANONDO^{1,4}, T. COUDERT²,
L. GENESTE², Y. BEAUREGARD⁴

¹CGI, Université de Toulouse, IMT Mines Albi - France

²LGP, Université de Toulouse, INP-ENIT - France

¹prénom.nom@mines-albi.fr ²prénom.nom@enit.fr

³ESTIA Recherche - Bidart - France

⁴ETS - Montréal - Canada

³p.nom@estia.fr ⁴prénom.nom@etsmtl.ca

RESUME : Cette communication s'intéresse à l'ingénierie des risques lors de la définition des processus de réalisation d'offres en relation client-fournisseur. Nous nous situons dans le cas où lors de la remise de l'offre au client, celle-ci n'est pas basée sur une conception détaillée du produit/système à fournir mais seulement sur une connaissance des principes à mettre en œuvre. Il y a prise de risque car lorsque le client accepte l'offre, le fournisseur doit finaliser (i) la conception du produit/système objet de l'offre ainsi que (ii) son processus de mise à disposition. L'ingénierie de ce type de risque est le plus souvent effectuée manuellement par des experts. L'objectif de cette communication est de définir et discuter des éléments d'un système à base de connaissances permettant d'assister cette ingénierie. La problématique est d'abord détaillée, un modèle de connaissances proposé et des possibilités d'exploitation discutées.

MOTS-CLES : relation client-fournisseur, processus de réalisation, ingénierie des risques, connaissances

1 CADRE DE TRAVAIL ET OBJECTIF

Cette première partie définit progressivement l'objectif de la contribution, l'existant scientifique et présente l'organisation de la communication.

1.1 Offres et relation client-fournisseur

Nos travaux s'intéressent à la mise au point d'outils à base de connaissances pour aider l'élaboration d'offres en relation client-fournisseur. Dans cet article, nous nous intéresserons uniquement aux offres concernant des produits ou systèmes principalement d'origine manufacturière, que nous nommons par la suite systèmes.

Nous considérons que ces systèmes entrent dans un processus de développement de type ETO (Engineer To Order) ou ingénierie à la commande (Rivest et al 2010). C'est-à-dire que la mise à disposition d'un système pour un client nécessitera une phase de conception ou d'ingénierie plus ou moins conséquente. Par contre, nous considérons que cette phase de conception est routinière (Chandrasekaran 1986). Ceci signifie que la conception du système composant l'offre exploite au maximum des solutions ou principes de solutions techniques déjà éprouvés, en les adaptant ou en les complétant si besoin.

Pour élaborer son offre, le fournisseur doit définir (i) la composition physique du système, qui permet d'obtenir un coût système et une définition technique de la solution proposée, mais aussi (ii) le processus de mise à disposition, incluant la production qui est indispensable pour déterminer un délai de mise à disposition et compléter le coût système avec un coût processus (Guillon et al 2017).

1.2 Prise de risque en élaboration d'offre

La mise en concurrence de plus en plus systématique fait que les appels d'offres sont de plus en plus nombreux (Sylla et al 2017a) et mettent les fournisseurs sous pression. Deux comportements sont rencontrés en entreprise :

- (1) Le fournisseur étudie en détail la demande et fait un travail de conception détaillée, aussi bien pour la nomenclature du système que pour son processus de mise à disposition. Il vérifie ainsi que son offre satisfait les besoins techniques (fonctionnalités et performances) et établit précisément les coûts et délais de mise à disposition. Si le client accepte l'offre, il aura une grande confiance dans la bonne réalisation du processus de mise à disposition.
- (2) Le fournisseur étudie succinctement la demande et fait un travail de conception préliminaire en identifiant uniquement les principes de solutions et les choix clés à mettre en œuvre. Ne pas étudier en détail son offre (c'est-à-dire le système à fournir et son processus de mise à disposition) entraîne une prise de risques. En effet, si l'offre est acceptée, le processus de mise à disposition comprendra alors une étape de finalisation de conception qui peut s'avérer délicate et conduire à ne pas arriver à réaliser l'offre en respectant les besoins techniques, les coûts et délais (Sylla et al 2017b).

Le premier comportement requiert une charge de travail conséquente, ce qui est regrettable si le client ne donne pas suite mais permet une relation client-fournisseur sereine minimisant les risques de ne pas arriver à satis-

faire le client une fois l'offre acceptée. Le second comportement est beaucoup moins lourd en termes de charge de travail mais occasionne une prise de risque conséquente de ne pas arriver à réaliser l'offre proposée et acceptée par le client (en termes de fonctionnalités, coût, délais...).

Nos travaux se situent dans le cadre de ce second comportement et notre contribution propose une ingénierie des risques à base de connaissances pour minimiser le risque fournisseur de ne pas arriver à réaliser l'offre proposée une fois que celle-ci est acceptée par le client.

1.3 Ingénierie des risques en élaboration d'offre

Nous faisons l'hypothèse que l'offre résulte d'un processus d'ingénierie produit/processus supporté par des outils d'aide à base de connaissances de manière similaire aux travaux de : O.Sullivan (2002), Yannou et al (2003), ou encore Yvars (2008). Nous situons notre contribution en aval de ce processus, au sein d'un second processus appelé Ingénierie des Risques processus (IR), qui peut exploiter toutes les informations décrivant le système objet de l'offre et son processus de mise à disposition (Figure 1). Cette Ingénierie des Risques s'appuie sur les travaux de Marmier et al (2013) et sur un progiciel nommé ProRisk en découlant.

L'information principale alimentant cette ingénierie des risques est le processus de mise à disposition de l'offre, qui comprend par exemple des tâches de : finalisation de la conception, approvisionnement, fabrication, assemblage, test ou livraison. Chaque tâche de ce processus est alors analysée par un expert risque qui identifie pour chaque risque :

- (1) l'évènement négatif correspondant au risque avec sa probabilité d'occurrence,
- (2) les impacts de l'occurrence de cet évènement, sous forme d'augmentation de coût et/ou de durée de certaines tâches,
- (3) des actions correctives et/ou préventives pour faire face aux risques,
- (4) des réductions d'impact et/ou de probabilité d'occurrence du risque issues de ces actions correctives et ou préventives.

Ceci étant fait, l'approche préconisée dans (Marmier et al 2013) exploite un simulateur à événements discrets qui simule toutes les combinaisons possibles d'occurrences de risques et toutes les combinaisons d'actions correctives et préventives (appelées *stratégies risque*). L'expert risque dispose alors pour chaque stra-

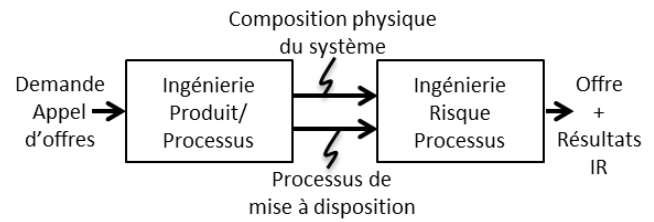


FIGURE 1 – Ingénierie produit/processus et risques

tégie risque d'un ensemble de scénarios de déroulement de la mise à disposition de l'offre caractérisé par un délai, un coût et une probabilité de réalisation. L'expert peut alors choisir la stratégie risque (c'est-à-dire quelles actions correctives et préventives) à mettre en œuvre.

1.4 Modélisation et exploitation de connaissances en ingénierie des risques pour l'élaboration d'offre

Actuellement, l'expert fournit toute la connaissance risque (partie gauche de la Figure 2). L'objectif de cette communication est de proposer les premiers éléments de modélisation pour élaborer une base de connaissances risque et un système permettant de l'exploiter (partie droite de la Figure 2). Ceci permettrait de diminuer le niveau d'expertise requis pour faire l'ingénierie des risques conventionnels caractérisés (expert risque junior par exemple) et de permettre à l'expert confirmé de se concentrer sur les risques non conventionnels (nouveaux risques ou risques critiques par exemple).

La formalisation et l'exploitation des connaissances risque se feront suivant deux approches :

- (1) d'une part à l'aide du raisonnement par analogie en recherchant des "données risques" dans une base de cas regroupant les processus de mise à disposition qui se sont déroulés dans le passé complétés par des informations sur les risques,
- (2) et d'autre part, à l'aide de déductions ou plus exactement de propagation de contraintes sur des modèles d'ingénierie de risque formalisés comme des problèmes de satisfaction de contraintes (CSP).

1.5 Travaux en rapport et organisation de la communication

En ce qui concerne les risques en relation client-fournisseur, la plupart des travaux se réclament d'approches marketing, de problématiques de logistique ou de sélection de fournisseurs (Ryals et al 2005), (Thun

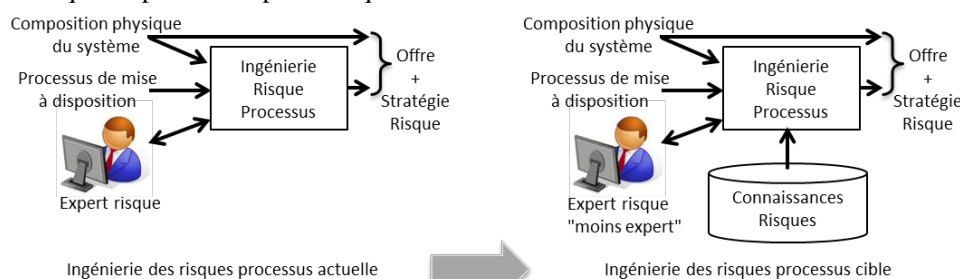


FIGURE 2 – Ingénierie des risques actuelle et avec exploitation de connaissances (cible)

et al 2011), (Hallikas et al 2005), (Lee 2009). Nous retenons les travaux de Hallikas et al (2005) car : (i) ils proposent une classification des risques suivant le type de relation client-fournisseur, (ii) ils différencient clairement les risques "acheteur" et "vendeur" et (iii) ils insistent sur la nécessité de prendre en compte le point de vue du fournisseur. Nos travaux s'inscrivent pleinement dans ce dernier point.

En ce qui concerne les risques en élaboration d'offres, nous n'avons pas trouvé de travaux abordant le problème comme nous l'avons formulé. Il existe par contre des travaux plus nombreux et des éléments normatifs concernant les risques en gestion de projet (ISO 2009), (PMBok 2009), (Tixier et al 2003), (Ward et al 2003). Nous nous inscrivons dans ce flux de travail et notamment dans la continuité des approches proposées dans (Marmier et al 2013) et (Fang et al 2012), où une notion de stratégie de traitement de risque est proposée.

En ce qui concerne la modélisation et l'exploitation de connaissance risque pour aider l'élaboration d'offres en relation client-fournisseur, les travaux sont beaucoup plus rares. Il est possible de citer en génie civil (Tah et al 2001) et (Yildiz et al 2014) ou en projet de système d'information (Alhawaria et al 2012). A notre connaissance, seuls (Yildiz et al 2014) proposent des éléments de modélisation des connaissances pour les risques. Nous nous inscrivons dans ce type de contribution et proposerons des éléments de modèles de connaissances.

La communication est décomposée de la manière suivante. Dans une seconde section, une analyse de l'ingénierie des risques « manuelle » (c'est-à-dire conduite par un expert risque) permet d'identifier formellement les connaissances apportées par l'expert. Il en découle en section 3 un premier modèle de connaissances pour l'ingénierie des risques dans notre contexte d'élaboration d'offre. La quatrième section analyse dans quelles mesures (i) le raisonnement par analogie peut exploiter le modèle de connaissance proposé, et (ii) l'ingénierie des risques visée peut être considérée comme un problème de satisfaction de contraintes et exploitée de même le modèle proposé.

2 IDENTIFICATION DES CONNAISSANCES DE L'INGENIERIE DES RISQUES

Dans le but de proposer des modèles considérant la connaissance "risque" sous la forme de cas et sous la forme de problème de satisfaction de contraintes, il est nécessaire d'identifier ces connaissances "risque". Pour ce faire nous allons analyser le travail de l'expert lors de l'ingénierie des risques et noter les connaissances qu'il apporte. Un premier modèle de connaissance pour l'ingénierie des risques découlera de cette analyse.

2.1 Données d'entrées

2.1.1 Données Processus

Le processus de mise à disposition constitue logiquement une première donnée d'entrée de l'ingénierie des

risques, car les risques et leurs traitements vont être définis sur les tâches de ce processus. Nous décrivons ce processus de la manière suivante :

- Un processus de mise à disposition est associé à un système objet d'une offre. Un même système peut être associé à plusieurs processus de mise à disposition, afin de pouvoir comparer des variantes de processus.
- Un processus est décomposé en tâches reliées par des contraintes de précédence (Figure 3). Chaque tâche i est réalisée par une ressource clé (ressource nécessaire pour la réaliser) et est caractérisée par un coût c_i et une durée d_i (d'autres métriques pourraient être définies).

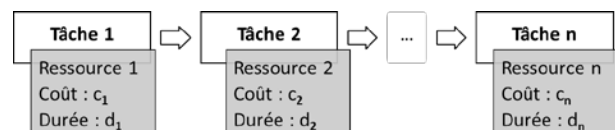


FIGURE 3 – Données Processus

Ces données ne sont pas suffisantes pour réaliser l'ingénierie des risques et nous proposons de les compléter avec des données décrivant (i) le système objet de l'offre, (ii) le contexte de l'offre (Guillon et al 2017).

2.1.2 Autres données d'entrée

- (1) Les données décrivant le système correspondent à des caractéristiques du système impactant l'ingénierie des risques, par exemple : la complexité ou la taille du système, la maturité des technologies employées, la fiabilité des composants utilisés, etc.
- (2) Les données décrivant le contexte de l'offre correspondent à des caractéristiques de l'offre impactant l'ingénierie des risques, par exemple : l'importance du client, la récurrence de la demande, la charge de travail de l'entreprise, la période de l'année...

Ces caractéristiques peuvent impacter le bon déroulement du processus de mise à disposition, dont (i) la probabilité d'occurrence des risques, (ii) leurs impacts, mais également (iii) les actions correctives ou préventives à mettre en œuvre et (iv) les réductions d'impacts associées.

L'expert risque du fournisseur a la connaissance de ces caractéristiques et de leurs impacts sur les tâches du processus de mise à disposition. Nous proposons de décrire ces caractéristiques à l'aide du triplet :

- (1) Élément conceptuel, par exemple : système grue, composant moteur, offre, client, etc.
- (2) Attribut décrivant le concept, par exemple : complexité de la grue, maturité du moteur, récurrence de l'offre, importance du client, etc.
- (3) Valeur de l'attribut décrivant le concept, par exemple, très fort / fort / faible / très faible, haut/moyen/bas ...

L'ingénierie des risques dispose donc de deux types d'entrées : le processus de mise à disposition et les caractéristiques du système et de l'offre impactant les risques. Ces éléments étant définis, nous allons mainte-

nant identifier l'apport de l'expert durant l'ingénierie des risques en suivant le déroulé du processus de traitement des risques.

$$Ic = \sum_i (ac_i + c \times mc_i)$$

$$Id = \sum_i (ad_i + d \times md_i)$$

2.2 Apports de l'expert et connaissances risque

Nous avons vu en section 1.3 que l'expert documente tâche par tâche : (1) la probabilité des risques, (2) les impacts, (3) les actions correctives et préventives et (4) les réductions d'impacts et de probabilité d'occurrence des risques. Nous proposons dans cette section d'identifier les connaissances risques apportées par l'expert durant cette ingénierie.

2.2.1 Risque et probabilité d'occurrence

Nous considérons le risque comme un couple (tâche, évènement). C'est-à-dire que l'évènement qui se produit durant cette tâche correspond au risque. Ceci est discutable mais permet de dissocier l'analyse des conséquences d'un même évènement sur des tâches différentes. Par exemple, cela permet d'analyser différemment l'évènement « Chute de neige et route bloquée », suivant qu'il se produit durant la tâche « Approvisionnement de composant » ou durant la tâche « Livraison du système au client ».

Un risque est caractérisé par la probabilité d'occurrence de l'évènement associé, sur la tâche identifiée.

Une même tâche peut être la source de plusieurs risques. Par exemple, la tâche « Finalisation de la conception » peut être l'objet de deux risques « Tâche plus difficile que prévue » et « Ressource clé indisponible ».

2.2.2 Impacts d'un risque

Un risque est associé à un ensemble d'impacts. Par exemple, le risque « Tâche finalisation de la conception plus difficile que prévue » peut avoir deux impacts, un sur la tâche "Finalisation de la conception" et l'autre sur la tâche "Assemblage et test". Un impact est défini pour un seul risque. Un impact est défini par :

- (1) la tâche impactée,
- (2) la nature de l'impact : coût ou durée (ou autre),
- (3) le mode de calcul de l'impact : forfaitaire/additif ou proportionnel/multiplicatif,
- (4) la valeur de l'impact.

Une tâche peut être l'objet de plusieurs impacts issus de risques R_i différents. Elle subira alors un impact global coût (Ic) et un impact global durée (Id), calculés de la manière suivante :

- chaque impact proportionnel est calculé individuellement par rapport aux caractéristiques de la tâche,
- tous les impacts proportionnels et fixes sont ensuite additionnés.

Ainsi, si l'on considère :

- ac_i : Impact coût additif du risque R_i ,
- mc_i : Impact coût multiplicatif du risque R_i ,
- ad_i : Impact délai additif du risque R_i ,
- md_i : Impact délai multiplicatif du risque R_i ,

les deux impacts globaux sont les suivants :

2.2.3 Les stratégies locales du risque

Pour faire face à ces impacts, l'expert risque définit des stratégies locales de traitement. Une stratégie locale regroupe : (i) un ensemble d'actions correctives et/ou préventives, (ii) un ensemble de réductions d'impact et éventuellement (iii) une réduction de probabilité d'occurrence du risque. Une stratégie locale est définie pour un seul risque, et un risque peut avoir plusieurs stratégies locales. Par exemple, si le traitement d'un risque peut être l'objet d'une action corrective AC et d'une action préventive AP, quatre stratégies locales (sl_i) sont envisageables :

- sl_1 = ne rien faire,
- sl_2 = AC,
- sl_3 = AP,
- sl_4 = AC et AP.

Chaque action corrective ou préventive est une tâche reliée aux tâches du processus de mise à disposition par des contraintes de précédence. Une tâche corrective ou préventive peut être associée à plusieurs stratégies locales d'un même risque. Chaque action corrective ou préventive est caractérisée par une durée et un coût. Par exemple, pour le risque « Tâche finalisation de la conception plus difficile que prévue », il est envisagé une action préventive AP, visant à identifier un expert pour la tâche « Finalisation de la conception » et une action corrective AC, visant à revoir le plan de test du système assemblé pour la tâche « Assemblage et test ».

Chaque réduction d'impact est définie pour une stratégie locale et pour une tâche impactée par le risque. Une réduction est définie par :

- (1) la tâche impactée objet de la réduction d'impact,
- (2) la nature de l'impact objet de la réduction : coût ou durée,
- (3) le mode de calcul de la réduction d'impact qui est toujours proportionnel,
- (4) la valeur de la réduction d'impact.

Par exemple, l'ensemble AP et AC (i) réduit l'impact en durée de la tâche "finalisation de la conception" mais augmente celui de son coût (réduction d'impact négative) et (ii) réduit les deux impacts en coût et durée de la tâche "assemblage et test".

Une tâche peut être l'objet de plusieurs réductions d'impacts. Nous faisons l'hypothèse que chaque réduction d'impact s'effectue de manière homogène sur tous les impacts additifs et multiplicatifs. Les réductions d'impact globales coût (RIc) et durée ($RIId$) sont calculés de la manière suivante :

- chaque réduction d'impact est calculée individuellement par rapport à la valeur de l'impact,
- les réductions d'impacts sont ensuite additionnées.

Ainsi, si l'on considère :

- $redc_i$: Réduction d'impact coût stratégie risque R_i
 - $redd_i$: Réduction d'impact durée stratégie risque R_i
- les deux réductions d'impacts sont les suivantes :

$$RIc = \sum_i [redc_i \times (ac_i + c \times mc_i)]$$

$$RI d = \sum_i [redd_i \times (ad_i + d \times md_i)]$$

Une stratégie mettant en œuvre une action préventive peut être la source d'une réduction de probabilité du risque pour lequel cette stratégie est définie. Le mode de calcul de la réduction de probabilité est toujours proportionnel.

2.2.4 Synthèse des entités ingénierie des risques

Durant l'ingénierie des risques, l'expert définit donc :

- pour chaque tâche :
 - o les risques associés à la tâche : couple tâche-événement et probabilité d'occurrence,
- pour chaque risque :
 - o les impacts du risque,
 - o les stratégies locales associées au risque,
- pour chaque stratégie :
 - o un ensemble de réductions d'impact,
 - o une réduction de probabilité d'occurrence du risque (optionnel),
 - o un ensemble d'actions correctives et préventives

L'expert doit ensuite décider quelles actions correctives et préventives mettre en œuvre, c'est-à-dire pour chaque risque quelle stratégie locale choisir. Une stratégie globale de traitement des risques d'un processus de mise à disposition lors de l'élaboration d'une offre se définit alors comme un ensemble de couples (risques, stratégie locale). Il est expliqué dans (Marmier et al 2013) comment ces stratégies globales peuvent être simulées et évaluées et comment le choix de stratégie globale peut être assisté.

3 MODELE DE CONNAISSANCE RESULTANT

Le but de notre communication est de proposer un modèle de connaissances et des outils d'aide pour assister l'expert dans ce travail de définition des items synthétisés en section 2.2.4. En conséquence, il est nécessaire au minimum de pouvoir stocker ces items et d'identifier des connaissances complémentaires. Dans ce but, nous proposons le modèle de classe de la Figure 4 dont nous commentons les entités clés et les connaissances complémentaires.

3.1 Tâches, événements et risques

Les tâches du processus de mise à disposition, données d'entrée de l'ingénierie des risques, correspondent à la classe *Task OE* (Offer Elaboration). Les tâches correctives et préventives proposées par l'expert sont décrites dans la classe *Task RE* (Risk Engineering). Ces deux types de tâches sont généralisés dans la classe *Task*. L'évènement est défini dans une classe *Event*.

Un risque est défini avec la classe *Risk*, pour une seule tâche et un seul évènement (lien 1).

3.2 Concepts

Le raisonnement humain ou le système d'aide exploitant la connaissance a besoin de typer de manière abstraite ces trois entités (tâche, évènement et risque). En effet, lorsque l'expert décide d'associer un risque à une tâche, il fait appel à ses connaissances et associe mentalement, par exemple, un type d'évènement à un type de tâche et en déduit un risque. Cette connaissance est modélisée par la notion de concept, que nous définissons dans la classe *Concept*.

Entité	Exemples de concepts
Tâche du processus de mise à disposition	conception, approvisionnement, production, assemblage, test, livraison, mise en service...
Evènement	tâche plus difficile que prévue, ressource indisponible, erreur de référence composant, modification des entrées de la tâche...

Pour les risques, les concepts correspondent des couplages de concepts tâche et évènement, par exemple « Conception plus difficile que prévue » ou « Ressource indisponible en livraison ». Ces concepts abstraits permettent d'une part de définir et stocker à l'avance des risques probables (par le biais des couplages tâche-évènement) et d'autre part, de pouvoir rechercher des données risque dans la base de connaissances en fonction de ces concepts.

3.3 Impacts, stratégie locale, réduction d'impacts.

L'impact est représenté par la classe *Impact*. Un impact ne pointe que sur un seul risque et sur une seule tâche impactée (lien 1).

La stratégie locale est représentée par la classe *Local Strategy*. Une stratégie locale ne pointe que sur un seul risque, et sur un seul concept parmi : ne rien faire, corrective, préventive ou mixte (lien 1). Elle comprend des tâches correctives et/ou préventives et des réductions d'impact (lien 0.*) et éventuellement une seule réduction de probabilité (lien 0.1).

La réduction d'impact est représentée par la classe *Impact Reduction*. La réduction pointe sur un seul impact et sur une seule stratégie locale (lien 1).

3.4 Modulation des entités risques

En complément des concepts, l'expert risque dispose également d'une connaissance de l'influence des caractéristiques de l'offre sur toutes les entités de l'ingénierie des risques. Nous avons évoqué dans la section 2.1.2 des caractéristiques systèmes et des caractéristiques offres qui peuvent impacter des entités de l'ingénierie des risques. Par exemple, un risque « Conception plus difficile que prévue » aura une probabilité plus forte si le système est « complexe » et de « grande taille », ou si le

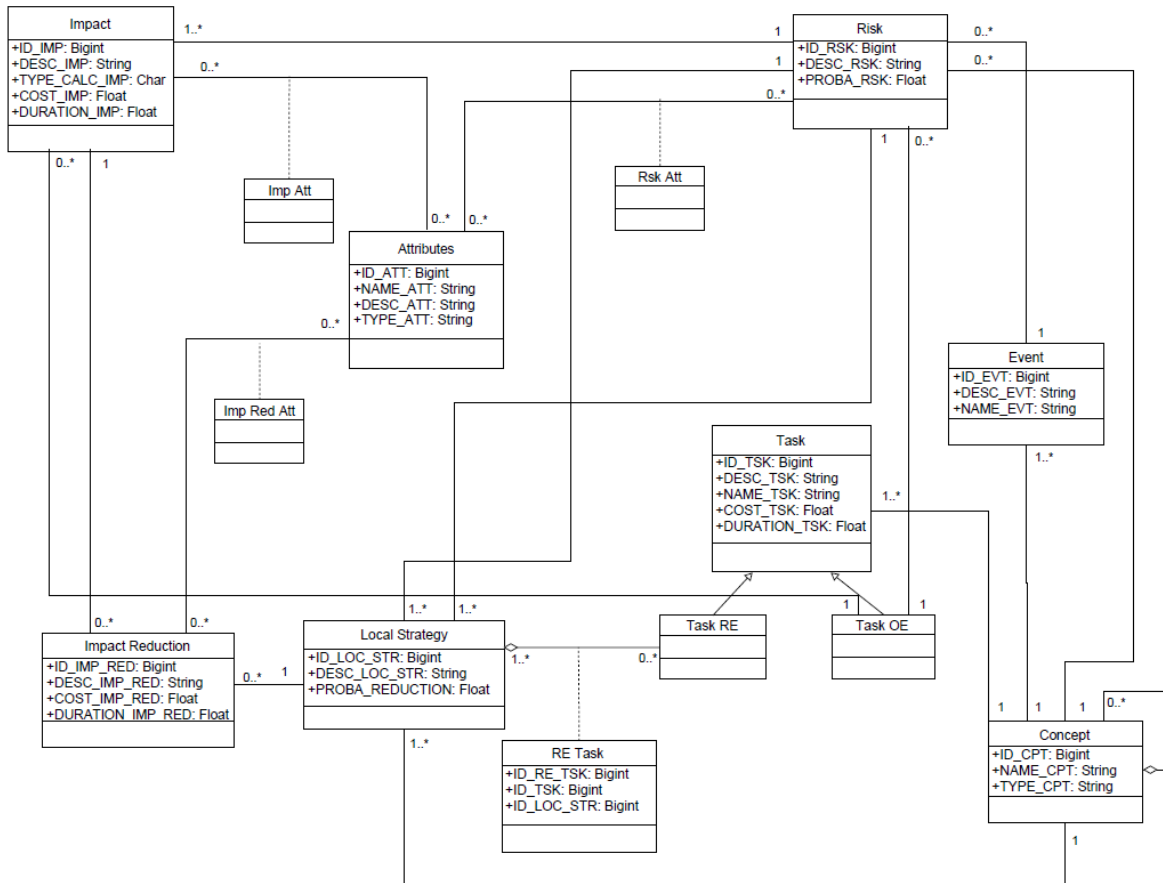


FIGURE 4 – Modèles de classes des connaissances de l'ingénierie des risques du processus de mise à disposition

client est « nouveau ». Il en va de même pour les impacts et les réductions d'impacts qui peuvent, par exemple, être respectivement modulés par une « charge de travail entreprise importante » ou une « importance client stratégique ». Comme pour les concepts, cette connaissance est stockée par la classe *Attribute* et ses associations avec les classes *Risk*, *Impact* et *Impact Réduction*.

4 MODELE DE CONNAISSANCE ET EXPLOITATION DES CONNAISSANCES

Les entités de l'ingénierie des risques et les connaissances associées étant modélisées, l'objet de cette section est d'évaluer comment il est possible d'exploiter les connaissances ainsi modélisées pour aider l'ingénierie des risques en élaboration d'offre.

4.1 Ingénierie des risques et analogie

En nous basant sur les principes du raisonnement par analogie ou raisonnement à partir de cas (Aamodt et al 1994), notre idée est de stocker les informations décrivant la réalisation effective des processus de mise à disposition passés (ces processus correspondent à des offres acceptées par le client, sinon le processus ne serait pas réalisé). En présence d'une nouvelle offre, l'objectif est de rechercher dans ce stockage des offres similaires et d'extraire les informations risques nécessaires à l'ingénierie des risques.

Ce principe a priori simple qui a fait ses preuves dans de nombreuses activités industrielles ne se met pas simple-

ment en place pour notre problématique d'aide à l'ingénierie des risques.

Considérons le processus de mise à disposition ci-contre (Figure 5). Les notations utilisées sont celles de la section 2.2.2.

L'ingénierie des risques a identifié que la tâche 1 était sujette à un risque R1 ayant un impact coût sur la tâche 3. Une action corrective AC1 qui réduit l'impact de moitié est définie. La tâche 2 est sujette à un risque R2 ayant aussi uniquement un impact coût sur la tâche 3. De même, une action corrective AC2 qui réduit l'impact d'un quart est définie.

Lors du suivi de la réalisation effective du processus de mise à disposition, les seules informations facilement accessibles et stockées pour réutilisation sont :

- pour les toutes les tâches : durées et coûts, prévisionnels et effectifs,
- pour chaque risque : occurrence effective ou pas de l'évènement associé.

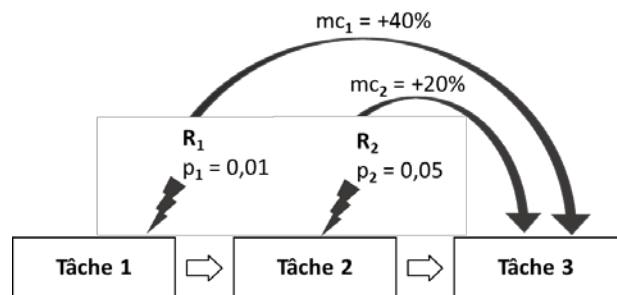


FIGURE 5 – Cumul d'impacts sur une tâche

Si les événements associés aux deux risques se produisent et que les actions correctives AC1 et AC2 sont effectuées, le coût de la tâche 3 est impacté de manière prévisionnelle à hauteur de : $c_3 + 0.4 \times c_3 + 0.2 \times c_3 = 1.6 \times c_3$, puis réduit à hauteur de $c_3 + 0.5 \times 0.4 \times c_3 + 0.75 \times 0.2 \times c_3 = 1.35 \times c_3$. Ce calcul exploite quatre informations risque fournies par l'expert : deux impacts et deux réductions d'impact. Les informations stockées concernant ce seul cas ne permettent que d'obtenir l'écart entre le prévisionnel et l'effectif (c'est-à-dire 1.35) mais en aucun cas la valeur des impacts et des réductions d'impact. Seule une analyse de données exploitant plusieurs occurrences de ce processus avec différentes combinaisons d'occurrences de risques et d'actions correctives pourrait extraire les valeurs des informations risques, mais ceci n'entre pas dans le cadre du raisonnement par analogie.

En conséquence, pour aider l'expert, il est indispensable d'associer dans la base de cas :

- les informations décrivant la réalisation effective des processus de mise à disposition,
- les informations prévisionnelles du modèle de données proposées en section précédente.

Disposant de ces éléments, l'expert peut alors requêter la base de cas sur (1) les différents concepts de tâches, événements et risques, et (2) les caractéristiques du système et de l'offre, pour identifier des offres passées et en extraire les informations risques prévisionnelles détaillées.

4.2 Ingénierie des risques et satisfaction de contraintes

Un problème de satisfaction de contraintes, ou CSP, (Montanari 1974) se définit comme un triplet regroupant : un ensemble de variables pour décrire le problème, un domaine de définition de chaque variable, et un ensemble de contraintes autorisant certaines combinaisons de valeurs de variable. La résolution interactive d'un CSP consiste à choisir une valeur pour chaque variable séquentiellement. Entre chaque choix, la propagation de contrainte supprime les valeurs incompatibles et guide ainsi la résolution vers des solutions faisables (cependant sans garantie car les propagations couramment employées ne peuvent pas garantir la cohérence). De très nombreux travaux ont exploité avec succès ce formalisme pour modéliser et traiter les problèmes de configuration de produits (Felfernig 2014). Notre but ici est d'évaluer dans quelle mesure il est possible de faire de même pour l'ingénierie des risques.

L'emploi de la propagation de contraintes en configuration et en conception n'est possible et intéressante que s'il est possible de définir une famille de produit, ou si la conception peut être caractérisée de routinière. En ingénierie des risques, cela nécessite que les risques soient parfaitement identifiés et connus. Les échanges avec les entreprises montrent qu'une partie des risques sont iden-

tifiables ainsi que leurs impacts et stratégie de traitements.

Nos travaux ont convergé vers un modèle de contraintes paramétriques pour aider l'ingénierie des risques. Les variables définissant : le risque, ses impacts, ses actions correctives et préventives ainsi que ses réductions d'impact, sont définies de manière rigide non modifiable. Par contre, les valeurs : de la probabilité du risque, des impacts du risque, des durées et coût des actions correctives et préventives, et des réductions d'impacts, sont modulables suivant les caractéristiques du système et de l'offre. C'est-à-dire que des contraintes peuvent relier les variables de l'ingénierie des risques aux variables décrivant les caractéristiques du système et de l'offre. La Figure 6 donne des exemples de telles contraintes. La première table montre comment la probabilité d'un risque est modulée par une « Complexité système » et une « Récurrence client » et la seconde comment cette même complexité module des impacts en coût et durée.

Complexité Système	Réurrence Client	Probabilité Risque
Très faible	Habituel	0,001
Très faible	Nouveau	0,005
Faible	Habituel	0,01
Faible	Nouveau	0,02
Fort	Habituel	0,05
Fort	Nouveau	0,08
Très fort	Habituel	0,1
Très fort	Nouveau	0,2

Complexité Système	Type Impact coût	Impact value €	Type Impact durée	Impact value day
Très faible	add fix	1000	add fix	1
Faible	add fix	1500	add fix	2
Fort	add fix	2500	add fix	3
Très fort	add fix	5000	add fix	4

FIGURE 6 – Contraintes modulant les données risques

Un premier modèle de contraintes est en cours de codage et de test sur ces principes. Il est également prévu que l'utilisateur puisse déroger aux valeurs préconisées.

5 CONCLUSIONS

L'objectif de cette communication était de poser les bases d'un système à base de connaissances pour aider l'ingénierie des risques du processus de mise à disposition de l'offre lors d'une réponse à appel d'offres. Ces travaux entrent dans le cadre du projet ANR OPERA¹.

L'étude des travaux existant a montré que s'il existait des travaux concernant les risques en relations client-fournisseur dans les communautés marketing et gestion de projet, ils sont beaucoup plus rares dans le domaine de la modélisation et de l'exploitation des connaissances.

Nous avons ensuite identifié et proposé au sein d'un premier modèle de connaissances un ensemble d'entités "risque" ainsi que la connaissance nécessaire à l'ingénierie des risques. Puis nous avons analysé comment ces connaissances pouvaient être intégrées et exploitées dans un système d'aide à l'ingénierie des risques.

Les premiers éléments rapportés montrent qu'il est possible, pour les risques conventionnels, de mettre en place un système d'aide à base de connaissances aidant l'acteur en charge de l'ingénierie des risques. Cette assistance permet alors de diminuer le niveau d'expertise requise pour faire l'ingénierie des risques et permet à l'expert confirmé de se concentrer sur les risques non conventionnels (nouveaux risques ou risques critiques par exemple).

REFERENCES

- Alhawaria S., Karadshehb L., Talet A.N., Mansoura E., 2012, Knowledge-Based Risk Management framework for Information Technology project, *International Journal of Information Management* vol 32 p.50 - 65
- Aamodt, A., & Plaza, E. 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications* Vol 7, issue 1, p.39–52.
- Chandrasekaran B. 1986, Generic Tasks in Knowledge Based Reasoning: High Level Building Blocks for Expert System Design," *IEEE Expert*, Vol. 1, issue 3, p. 23–30,
- Fang C., Marle F. 2012. A simulation-based risk network model for decision support in project risk management, *Decision Support Systems* Vol 52, p. 635–644.
- Felfernig A., Hotz L., Baglay C., Tiihonen J., Knowledge-based configuration From Research to Business Cases. 2014.
- Guillon D., Sylla A., Vareilles E., Aldanondo M., Villeneuve E., Merlo C., Coudert T., Geneste L. 2017, Configuration and response to calls for tenders: an open bid configuration model, Proc. of 19th International Configuration Workshop p 46-53
- Hallikas J, Puumalainen K., Vesterinen T., Virolainen V.M. 2005. Risk-based classification of supplier relationships. *Journal of Purchasing and Supply Management* Vol 11, Issues 2–3, p. 72-82
- ISO 31000, 2009. International Standards for Business, Risk Management – Principles and Guidelines.
- Lee A.H.I. 2009. A fuzzy supplier selection model with the consideration of benefits, opportunities, costs and risks *Expert Systems with Applications*. Vol 36, Issue 2, Part 2, p. 2879-2893
- Marmier F., Gourc D., Laarz F. 2013, A risk oriented model to assess strategic decisions in new product development projects, *Decision Support Systems* Vol 56 p. 74–82
- Montanari U. 1974, "Network of Constraints: Fundamental Properties and Applications to Picture Processing," *Inf. Sci.*, vol. 7, p. 97–132.
- O'Sullivan B., 2002, Interactive constraint-aided conceptual design, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol 16, p. 303–328.
- PMBok 2009 A Guide to the Project Management Body of Knowledge, 4th edition Project Management Institute.
- Rivest L., Desrochers A., Brie A. 2010, Adaptive generic product structure modelling for design reuse in engineer-to-order products, *Computer in Industry*, Vol. 61, p. 53–65.
- Ryals L.J., Knox S. 2005. Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value, *European Journal of Marketing*, Vol. 39 Issue 5/6, p.456-472
- Sylla A, Vareilles E., Aladanondo M., Coudert T., Geneste L., Kirytopoulos K. 2017a. Customer/Supplier Relationship: reducing Uncertainties in Commercial Offers thanks to Readiness, Risk and Confidence Considerations. *Advances on Mechanics, Design Engineering and Manufacturing. Lecture Notes in Mechanical Engineering* p. 1115-1122.
- Sylla A., Vareilles E., Coudert T., Kirytopoulos K., Aldanondo M., Geneste L. 2017b. Readiness, feasibility and confidence: how to help bidders to better develop and assess their offers. *International Journal of Production Research*, Vol. 55, Issue 23, p. 7204–7222,
- Tah J.H.M., Carr V. 2001. Knowledge-Based Approach to Construction Project Risk Management, *Journal of Computing in Civil Engineering* Vol 15 Issue 3
- Thun J.H., Hoening D. 2011. An empirical analysis of supply chain risk management in the German automotive industry *International Journal of Production Economics*. Vol 131, Issue 1, p 242-249
- Tixier J., Dusserre G., Salvi O., Gaston D. 2002. Review of 62 risk analysis methodologies of industrial plants, *Journal of Loss Prevention in the Process Industries* Vol 15, issue 4, p. 291–303.
- Ward S., Chapman C. 2003. Transforming project risk management into project uncertainty management, *International Journal of Project Management* Vol 21, issue 2, p. 97–105.
- Yannou B., Harmel G., (2005), Use of Constraint Programming for Design, *Advances in Design, Springer-Verlag: London*, Chapter 12, p. 145-155.
- Yildiz A.E., Dikmena I., Birgonul M.T., Ercoskunb K., Alten S., 2014, A knowledge-based risk mapping tool for cost estimation of international construction projects, *Automation in Construction* Vol 43, p.144–155
- Yvars, PA. 2008. Using constraint satisfaction for designing mechanical systems. *Int J Interact Des Manuf* vol 2, issue 3, p. 161-167.

ⁱ <https://reaseach-gi.mines-albi.fr/display/OPERA>

ENHANCING OPERATIONAL FAULT DIAGNOSIS BY ASSESSING MULTIPLE OPERATIONAL MODES

Juan José MONTERO JIMENEZ

TEC-Costa Rica Institute of Technology
juan.montero@itcr.ac.cr
ISAE-SUPAERO – University of Toulouse
juan-jose.montero-jimenez@student.isae-supaero.fr

Rob VINGERHOEDS

ISAE-SUPAERO – University of Toulouse
rob.vingerhoeds@isae-supaero.fr

ABSTRACT: *Operating a modern technical system, such as a train or aircraft, calls for good organised engineering, operation and maintenance to keep the system in an optimal operational condition. Predictive maintenance is being studied and has as aim to identify errors early enough to still be able to propose a suitable solution before a real incident occurs. After all, technical problems in service may lead to delays or even interruptions of service due to extensive repair actions, such as the replacement of components. Often, predictive maintenance aims at recognising patterns in time series of monitored data and classifying these patterns as known conditions (faulty or correct). As such it provides a vital source of information for maintaining a healthy operational status. However, these approaches are still in their early phases and rely still heavily on skill and experience from the expert. In this paper, the use of self-organising maps for predictive maintenance is being discussed, applied to data of a jet engine. The aim of the study was to assess the usability of such approaches to real-life situations, assessing the learning and validation phases.*

KEYWORDS: *Maintenance, condition monitoring, fault diagnosis, neural networks.*

1 INTRODUCTION

Safe and efficient operation of modern technical systems involves several online and off-line diagnostic tasks as well as the incorporation of the diagnostic results in timely maintenance. Off-line diagnostic tasks involve the handling of recorded data, analysis and interpretation, and organised maintenance. The information gathered during operation, such as that from crew reports and data recordings form vital sources of input for such off-line diagnostic tasks.

Maintenance is vital during a systems life cycle to ensure their functionality. Several types of maintenance can be distinguished; corrective, preventive and predictive maintenance. A good combination of the three is vital to make a system's operation reliable. Even though corrective and preventive maintenance remain today important research topics, a lot of attention is currently being given towards predictive maintenance that aims for defining the best possible moment to trigger maintenance actions. Triggering too late may lead to failure occurrence, causing financial losses, sometimes image damage and may even lead to casualties and/or losses of human lives. Triggering too early may lead to replacing components that are not faulty through costly interventions. Predictive maintenance aims at proposing a solution by monitoring the health status of the systems and components, identifying incipient faults and forecasting the exact moment of failure.

In this paper, the use of Self-Organising Maps (SOM) is discussed to detect degradation patterns on aircraft jet engine data as part of a fault diagnosis approach within the framework of predictive maintenance.

The paper is organised as follows. Section 2 describes the diagnostic problem as studied within the framework of this project. Section 3 describes the case study, an aircraft jet engine. Sections 4 and 5 then explain the use of Self-Organising Maps for this case study, first the more regular case as for SOM for a given condition and then the case for bringing together several operational modes into one SOM. Section 6 draws conclusions and indicates some paths for future research on this topic.

2 DIAGNOSTICS FOR PREDICTIVE MAINTENANCE

Diagnostics approaches address the data acquisition, processing and fault identification using explicit methods and implicit methods. Prognostics approaches are oriented to determine the Remaining Useful Life (RUL) of system components using reliability and availability criteria. Both approaches intervene and contribute to the overall idea to better support the decision-making process for operations engineers when recommending or not a maintenance intervention.

Explicit methods group the techniques in which the diagnostic of faults is made "explicitly" from data by experts linking the symptoms to the failures and faults and pointing to the root cause. Here one can think of

techniques such as Rule-Based Reasoning (RBR), Case-Based-Reasoning (CBR), decision trees, fuzzy logics, etc. Implicit methods, on the other hand, involve the use of algorithms to transform complex data sets into useful data that can be interpreted afterwards by experts in form of reports or alarms. One can think of Neural Networks (NN) and other learning algorithms.

From a top-level point of view, one can see that predictive maintenance is mainly aimed to trigger maintenance actions at the suitable time, defining an analysis strategy to address complex data, by means of specialized methods and tools in order to identify symptoms related to known failures that will lead to determine the actual state (diagnosis approach) and the remaining useful life (prognosis approach) of the component or system (see Figure 1).

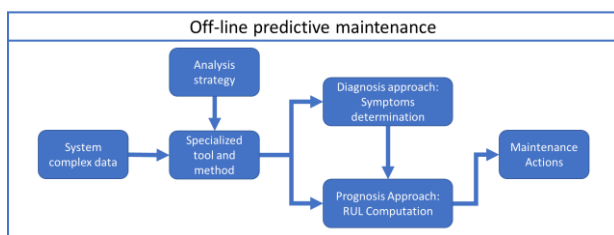


Figure 1: Predictive maintenance model

3 CASE-STUDY ON JET ENGINE

Keeping aircraft engines in an optimal operational condition is a vital aid for operating modern airlines. Engine Condition Monitoring (ECM) is used to have a regular overview of the proper functioning, a good health, of the aircraft engines, by analysing in-flight measured aircraft and engine parameters. Modern ECM is built around performance trend analysis approaches but at the same time also relies heavily on human expertise. The aircraft and engine parameters are visualised and the evolution over time (trends) of the most important parameters is analysed. Specific parameter trend evolutions often precede the occurrence of incidents and allow for early detection of deterioration of engine modules, failures, and/or malfunctions. The application of modern artificial intelligence so to enhance trend analysis to anticipate incidents by determining the need for maintenance and detecting the components or systems where maintenance is required has been part of several studies (e.g. Vingerhoeds, 1995). Trend analysis can offer a good support for performance engineers in airlines engineering departments in detecting problems early. In this case-study, a neural network approach (in this case a self-organising map) will be used to assess the engines health state. (Vingerhoeds, 1995) gives more information on the general conditions for such predictive maintenance approaches, and in particular the operational conditions for diagnosing jet engines.

Performance engineers use a variety of different techniques to assess the engine's health state. In this study, measured in-flight behaviour is used for assessing the engine's health state. It provides early warning on ongoing or imminent problems prior to serious malfunction. Effects of for example bird strikes can be seen and performance engineers can use witnessed changes in the engine's prime parameters to determine the eventual need for corrective maintenance. Other sources of information (flight trouble reports, etc.) offer complementary information that can be used to enhance the overall diagnostic approach.

The complete view for a performance engineer is complex, many parameters come into play and eventual wrong decisions on corrective maintenance (if in the end there was no need) would have a strong financial impact. The overall goal of this work is to improve condition monitoring to a point where performance engineers can rely on these techniques for their decisions.

In this paper, use is made of the database made available for the PHM08 benchmark for the International Conference on Prognostics and Health Management in 2008, as described in (Saxena, 2008). The data was created using a commercial software tool, C-MAPSS, that allows for simulating a realistic large commercial turbofan engine.

The data set consists of multivariable cycle series, divided into training, test and validation subsets with multiple jet-engines of the same type and under comparable nominal operational conditions. Each engine has some unknown initial wear and manufacturing variability that should not be confused with faults. The engines degrade along the operational cycles and upon reaching a specific limit are to be taken out of service. The definition of these limits is not provided and the predictive maintenance approach needs to be able to identify such limits.

The data set contains three input variables representing the engine control settings: the flying altitude, the Mach number and the throttle resolver angle. 6 different combinations of these control settings are used, defining the operational modes in the data sets. The rest of the data correspond to 21 out of 58 different outputs available on the C-MAPSS tool (see Table 1). These 21 variables correspond to sensor information from different engine parts (see a simplified engine diagram in Figure 2).

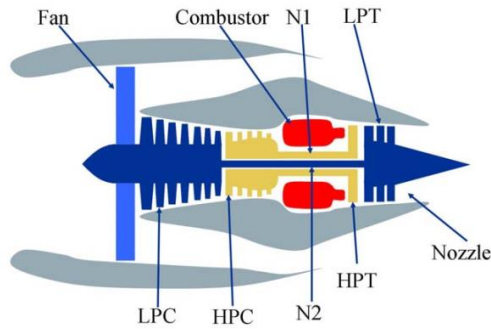


Figure 2: Simplified diagram of engine simulated in C-MAPSS (Saxena 2008)

- HPT: High pressure turbine.
- LPT: Low pressure turbine.
- HPC: High pressure compressor.
- LPC: Low pressure compressor.
- N1: Outer shaft.
- N2: Inner shaft.

Table 1. Sensors list present in the PHM08 Challenge (Saxena 2008).

Number of sensor	Symbol	Description	Units	Column in data set
1	T2	Total temperature at fan inlet	° R	6
2	T24	Total temperature at LPC outlet	° R	7
3	T30	Total temperature at HPC outlet	° R	8
4	T50	Total temperature at LPT outlet	° R	9
5	P2	Pressure at fan inlet	psia	10
6	P15	Total pressure in bypass-duct	psia	11
7	P30	Total Pressure at HPC outlet	psia	12
8	Nf	Physical fan speed	rpm	13
9	Nc	Physical core speed	rpm	14
10	epr	Engine pressure ratio (P50/P2)	**	15
11	Ps30	Static pressure ratio at HPC outlet	psia	16
12	phi	Ratio of fuel flow to Ps30	pps/psi	17
13	NRf	Corrected fan speed	rpm	18
14	Nrc	Corrected core speed	rpm	19
15	BPR	Bypass Ratio	**	20
16	farB	Burner fuel-air ratio	**	21
17	htBleed	Bleed Enthalpy	**	22
18	Nf_dmd	Demanded fan speed	rpm	23
19	PCNfR_dmd	Demanded corrected fan speed	rpm	24
20	W31	HPT coolant bleed	lbm/s	25
21	W32	LPT coolant bleed	lbm/s	26

Analysis on the data showed that only 7 out of the 21 data deliver significant input. Others are either constant or binary, or redundant information with other information. This discrimination process is important for identifying which parameters really come into play for diagnostics. At the same time, as the computation time for the SOM approach is a function of the data dimension, a reduction of the number of output data will allow for a reduction in this computation time.

4 SELF-ORGANISING MAPS FOR DIAGNOSTICS

Self-Organising Maps (SOM) are Neural Networks with unsupervised learning inspired by the way the human brain functions (Kohonen, 1990). The development of a SOM consists of two main phases; the training phase and the labelling phase. Those phases are strongly linked and both must be carefully developed to obtain a reliable tool. The training phase is oriented to teach the algorithm with the dataset in such a way that the neural network clusters similar data features on specific regions on the map. The labelling phase, once the training phase has been finished, links these regions to identified features of the dataset (labelling). This allows the algorithm to identify similar data patterns in complex data. When similar data patterns are presented to the SOM, the right region on the map is identified, hinting, for the current study, a degradation status of the engine.

Usually a two-dimensional, rectangular architecture is used for the SOM's. Other architectures include 2D hexagonal maps, but also 1D and 3D architectures. In this paper a rectangular map is selected for the health monitoring (see Figure 3).

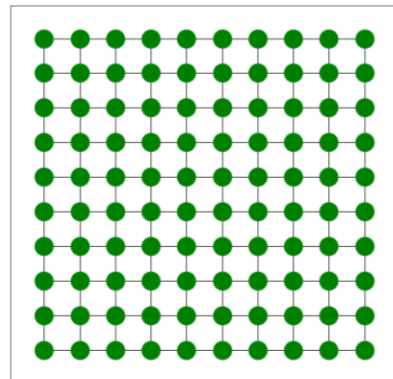


Figure 3: Rectangular SOM architecture.

Taking a closer look at the SOM training phase shows that this phase consists two different main steps; competitive and convergence. During the competitive step, a set of input vectors x_p is presented to the 2D map (p denotes the dimension of the input and output data, for the jet-engines database it is 7). The map is constituted of " n " neuron with coordinates i and j on the map and which have a weight w_{ij} . The weights are vectors with the same dimension as input vectors. At the start of the training process the weights values are set randomly in the intervals $[0,1]$. A normalisation of the input vectors is performed prior to the training phase so to avoid algorithmic problems. Such normalization allows to bring variable values of very different orders of magnitude to a similar range. This requires expert knowledge on the application domain, on the viable range of variable values so to identify the actual limits.

During this competitive process, the Euclidean distance is computed between one input and each neuron. When the lowest distance among all of them is found, the neuron related to this distance is declared as the winner neuron (see equation (1)), or in more formal words, the Best Matching Unit (BMU). The BMU is surrounded by its neighbourhood constituted by neurons within a certain radius from the BMU.

Now, the Euclidean distance is computed between the coordinates of the BMU and the coordinates of all the rest of the map, if the result is lower than the threshold, the neuron become part of the neighbourhood. The radius determining the neighbourhood starts large at the beginning of the training phase (around the half the size of the map) and will decrease gradually using decay function (2) along the learning process. When the neighbourhood is constituted only by the BMU it-self (equations (3)-(4)), the training phase ends.

$$BMU = \arg \min_{ij} \{ \|x_p(t) - w_{ij}(t)\| \} \quad (1)$$

$$decay(t) = \exp \frac{-t}{\lambda} \quad (2)$$

with t being the current iteration and λ a time constant.

$$\lambda = \frac{\text{number of iterations}}{\log r_0} \quad (3)$$

with r_0 being the initial neighbourhood radius (half of the map size), then

$$r = r_0 \times decay(t) \quad (4)$$

After finding the BMU, the weights of the winner neuron and the neighbourhood are updated using the following equation:

$$w(t+1) = \begin{cases} w(t) + \alpha(t)\beta(t,r)(x(t) - w(t)) & \text{if } n_{ij} \in \Lambda_{BMU}(t) \\ w(t) & \text{if } n_{ij} \notin \Lambda_{BMU}(t) \end{cases} \quad (5)$$

where:

$\alpha(t)$ is the learning rate depending on the iteration.

$\beta(t,r)$ the influence rate depending on the iteration and the radius from the BMU.

n_{ij} a neuron with its i and j coordinates on the map.

$\Lambda_{BMU}(t)$ Neighbourhood of the BMU for the current iteration.

The weights updating is strongly linked with the influence rate $\beta(t,r)$. It defines the magnitude of the change when updating the weights of the BMU and its neighbours, depending on the distance from the winner neuron. The further away the neighbour is, the lesser the influence. There are some influence functions which

affect negatively the furthest neighbours, this means, there is an inhibitory effect on them (e.g. the ‘‘Mexican hat’’ function). In this study the well-known Gaussian Function, is the one used.

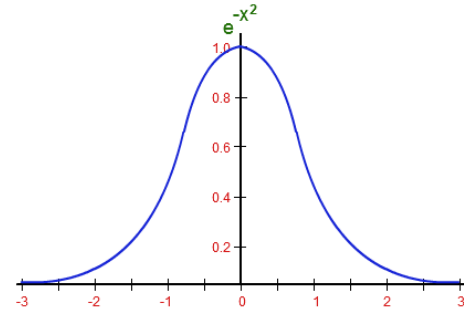


Figure 4: Gaussian function as influence rate.

Once the weights are updated, a second input is shown to the map, the BMU is found and the weight are updated again. This process is repeated with all the inputs and several times (iterations) to ensure a good learning process. In this study, the maximum number of iterations was set to 500 times the number of nodes in the map, described in literature as a reasonable estimate for good convergence (Kohonen, 1995).

Another important factor in the training phase is the learning rate, the speed with which the neural network learns. At the beginning of the learning process, the learning rate is rather large (in this study 0.9) and it will decrease exponentially along the learning process. When the learning rate is high, neurons with similar features are rapidly on a map region, this is known as the topological step in the training phase. As the learning rate is decreasing, the convergence step begins when the weights of the neurons are slightly refined. Equation (6) has been used to determine the learning rate along the training process.

$$\alpha(t) = \alpha(t)_0 \times decay(t) \quad (6)$$

To summarize the SOM algorithm, the following steps are proposed:

1. Define the map architecture
2. Initialize the neuron weights randomly.
3. Show the first input vector to the SOM.
4. Find the BMU.
5. Define the neighbourhood.
6. Update the weights of the winner neuron and its neighbourhood
7. Repeat steps 3 to 6 with all the inputs vectors.
8. Show all the input vectors several times to the neural network to facilitate the SOM convergence.

If the algorithm converges properly, some defined regions, or trends, will appear on the map. These regions

concern areas where comparable inputs are recognised. This allows to see what data is considered, by the algorithm, to be similar. It is the task of the human expert to assess these results, so to ensure that the data concerns indeed similar patterns. Knowing now these regions on the map, allows for labelling these regions with clear identification on what the cluster is about. It concerns therefore a manual identification of the different regions / clusters by SOM algorithm. A name, numeric value or any other identification character will be assigned to the map regions or neurons so that new data could be assessed and identified automatically by the trained SOM.

5 APPLICATION OF SOM'S FOR DIAGNOSTICS

SOM have been applied in several approaches for diagnosis and prognosis in predictive maintenance precisely due to its identification abilities in complex data. For example, (Germen et al. 2014) presents a SOM application to determine different types of faults in induction motors using sound based data. (Côme et al. 2010) use self-organising maps to monitor different parameters of aircraft engines for health diagnosis purposes. (Hu et al. 2013) determines the dynamic degradation on bearings using self-organising algorithms from incipient faulty conditions to the failure state.

This paper proposes a different approach of analysis with SOM on the jet-engine health diagnosis, assessing the engine degradation trend on different operational modes scenarios at the same time. The degradation trend is quantified by a degradation index obtained from the trained SOM for each operational mode, allowing to determine the engine health status in a more integrated way, assessing multiple operational modes at the same time to avoid false positives (suggesting a fault where it is still good), or false negatives (suggesting no faults while a fault has occurred). The rationale behind this approach is to be able to compare diagnostics from different operational modes so to build up “evidence” for a suspected fault as time goes along and different control setting are encountered.

The study case is built on the database described in paragraph 3. For each engine cycle in the database, only one operational mode is used to record discrete data. For each operational mode, the data patterns are very different in terms of range so that an unambiguous determination of a degradation trend is not possible over all operational modes at the same time. A first approach therefore has been to develop an SOM for each operational condition, leading to 6 different SOMs, one for each operational mode. This allows to follow the degradation evolution through the lifecycle on each single scenario for the given operational condition.

The results on the 6 SOMs were similar. On each map two main regions were clustered: a white region representing the good operational conditions of the engine and a black region representing the faulty conditions. Besides these regions, a grey scale witnessing the transition over engine lifetime between both regions appeared on the rest of the map, therefore representing the engine degradation in the intermediate states. Figure 5 represent one of these single operational mode SOMs. The red circles within some neurons connected by arrows, represent the lifecycle of one extra engine tested with the trained SOM. It is possible to observe the degradation evolution on the tested engine from the white to the black region, going through the grey neurons along its lifecycle.

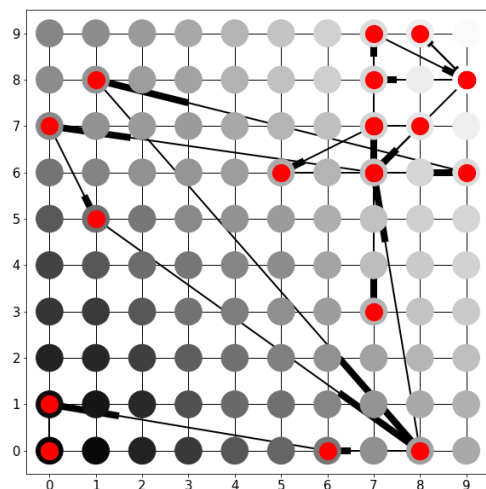


Figure 5: Engine degradation trend observed on a SOM trained with one control setting data.

Even when consistent degradation trends were obtained on individual operational modes, showing the evolution from the good to the faulty operational conditions, it is possible to obtain incomplete degradation trends on these individual SOM. Considering the data format, only one operational mode is encountered for each engine cycle (changing randomly from one cycle to another among six different options) and this could yield a lack of data for a specific operational mode at some stages of the engine lifecycle. This condition is more critical at the end of the lifecycle, where a lack of information on a specific operational mode might hide a faulty engine condition.

Besides, in real life, an engine passes via multiple operational conditions during one single flight. As such, information from these different operational modes could be gathered and used to improve the confidence in the diagnostic result. If information is used only from one operational condition, many external factors, like bad weather conditions, avoidance manoeuvres or gusts might induce a false result in the degradation indicator and this

could lead to trigger maintenance actions in the wrong moment.

To address this situation, a degradation assessment on several different operational modes is proposed to enhance the decision-making process. A new SOM was developed, this time using the information from the 6 operational modes available on the dataset. The SOM size for this case was determined following the method suggested by (Delgado et al. 2017), resulting a 15x15 map the best option. Verification has shown that no significant improvements were obtained using SOMs with a higher dimension maps. The result of the trained SOM with the 6 operational modes is shown in Figure 6. Each bounded region on the trained SOM represents an operational mode that can be easily labelled using the known features of each mode in the dataset.

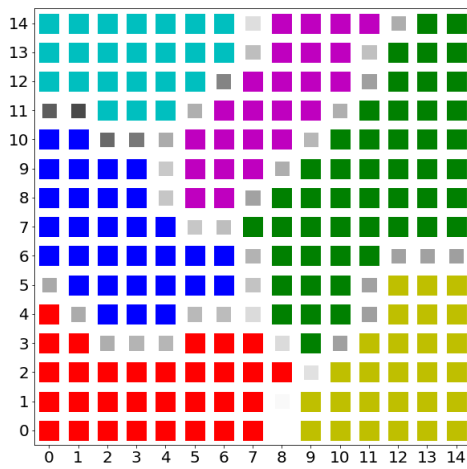


Figure 6: 15x15 SOM trained with full jet-engine dataset.

Once the regions were identified a test was carried out to prove SOM accuracy at clustering the engines cycles to the corresponding operational mode. All engines available in the data set were shown to the SOM and it succeeded allocating each cycle to the correct map region. This shows that each operational mode is well bounded to a single map region allowing analysing the degradation trend on each operational mode at the same time for a single engine. Figure 7 shows a single engine tested with the 15x15 SOM. The black point within the coloured neurons represent the exited neurons of each region and arrows represent the degradation evolution on each operational mode.

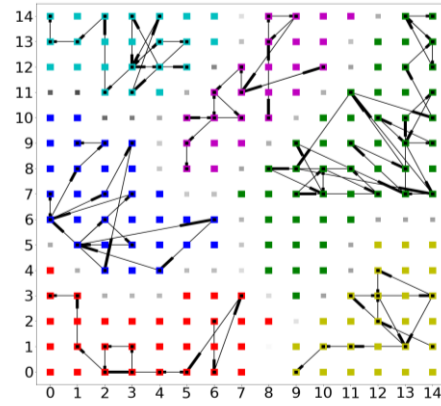


Figure 7: Testing engine sample from training subset on the 15x15 SOM.

Since the difference between each operational mode is large, the small differences within each region are not easily observable as it was on SOM trained with one single operational mode. This poses a serious problem to follow the engine degradation trends. Yet, even when these variations between the neurons on a map region are too small to be observed by the colours of the map, variation on the neurons weights (numerical values) exist and this information can be used to describe the engine degradation process.

The norm of each neuron weight vector is computed, this value is normalized between 0 and 1 for each operational mode. Let be OP_m a set of neurons that constitute an operational mode 'm', then:

$$\forall n_{ij} \in OP_m$$

$$d_m = \frac{\|W_{ij}\|_2 - \min(\|W_{ij}\|_2)}{\max(\|W_{ij}\|_2) - \min(\|W_{ij}\|_2)} \quad (7)$$

with d_m being the degradation index on each neuron n_{ij} and, w_{ij} being each neuron weight within the same operational mode.

The obtained values are used as labels for each neuron and once the labelling task was finished, a clear evolution from 0 to 1 is easily observed on each map region. This value has been called "degradation index" and its use is key for diagnosis strategy. Figure 8 shows the degradation index (d_m) as labels on the 15x15 SOM, being the 0 the best operational conditions, the 1 the faulty conditions and the values between them the degradation process on the engine lifecycle.

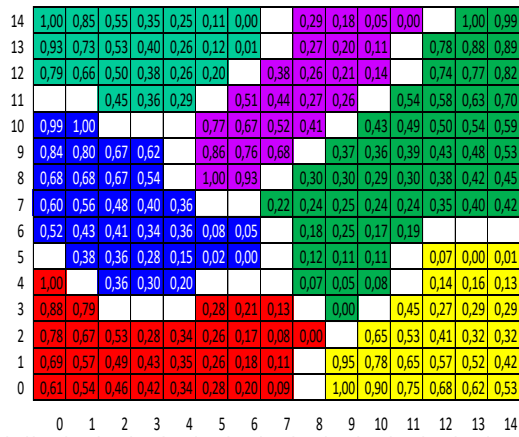


Figure 8: 15x15 SOM labelled with degradation values on each region.

As the goal is to follow the current health state of the engines, the approach is to determine whether the engines condition is faulty or nominal. First, individual degradation indexes d_m are obtained for all operational modes in one single flight. Once the individual degradation indexes d_m are obtained it is possible to combine them in a consolidated degradation index D_C . This consolidated value allows to observe a general degradation evolution considering as many operational modes as possible (equation 8).

$$D_C = \sum_{m=1}^l \frac{d_m}{l} \tag{8}$$

D_C : consolidated degradation index at any flight.

d_m : individual degradation indexes for each m operational mode.

l : number of operational modes registered during the flight.

To compute D_C , the ideal case is when all possible operational modes are registered in the same flight (six operational modes for the current study). However, the engine might not register one or several operational modes during a flight, if this happens, d_m might be obtained from the previous flights, as long the data is still considered as valid. A first experimental approach in the current study established the previous 5 flights as valid to take missing d_m for the latest flight D_C computation.

The consolidated degradation index D_C encounters different perspectives of the degradation evolution using a single value for each engine flight. Plotting this value versus the number of flights (discrete cycles for the simulated data), an increasing evolution of the degradation is observed until a certain threshold, close to value 1. Figure 9 shows the degradation evolution along the lifecycle of one engine, using D_C approach.

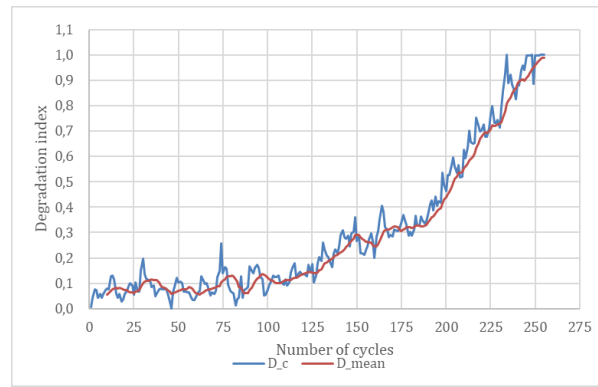


Figure 9: Degradation evolution of one cycle

The fluctuations of curve D_C pose a problem to study the degradation trends and determine an experimental degradation index threshold. To soften D_C curve, the arithmetic moving average is applied considering the measurements of the last 10 flights. These 10 flights represented the best experimental compromise between softening the curve and keeping the variation of the degradation index as wide as possible. This produce the curve D_{mean} versus the number of flights (cycles), which facilitate the degradation trends study. Figure 9 also shows the D_{mean} curve for the current engine under study.

Obtaining D_{mean} curve for multiple failed engines and plotting all of them in the same graphic, it is possible to determine an experimental threshold for the proposed method of degradation index. For the current jet-engines analysis a threshold of 0.93 for D_{mean} is proposed. Engines below of such threshold are considered in nominal conditions and those overcoming the threshold are considered as failed. In real life applications, such thresholds must be supported by knowledge of experts, regulations and manufacturers specifications.

Once the threshold is known other engines were evaluated with the proposed method to confirm the current health status of the engine. Table 2 summarize the data for the last measured flight of some of the assessed engines. Columns 2 to 7 show the individual degradation indexes for the last measurement of the engines Columns 8 and 9 present D_C and D_{mean} correspondingly and last column shows the confirmed health status of the engine.

Table 2. Engine condition diagnosis from the degradation index.

Engine	d_1	d_2	d_3	d_4	d_5	d_6	D_C	D_{mean}	Condition
Engine1	0,99			1,00		1,00	1,00	0,99	Failed
Engine2	1,00	1,00			1,00	1,00	1,00	0,98	Failed
Engine3			0,99	1,00		0,93	0,97	0,93	Failed
Engine4		0,78	0,78		1,00	0,73	0,82	0,77	Nominal
Engine5	0,41	0,42	0,35	0,44	0,14		0,35	0,38	Nominal
Engine6	0,38		0,24	0,18		0,38	0,30	0,31	Nominal

As explained in section 3, there are two subsets in the data base; the training subset and the testing subset. Engines from 1 to 3 in table 2 were taken randomly from the training subset, in which all engines reached the faulty threshold. For such engines, the degradation index value confirms the reaching of a faulty condition since the value is greater than the experimental threshold. Engines from 4 to 6 were selected randomly from the testing subset and in this case, it concerns engines that remained in good operational conditions (nominal condition). The degradation index confirms this, as the value remains under the degradation threshold.

6 CONCLUSION

The correct diagnosis of health state allows to improve the decision-making process in maintenance departments, bringing benefits to the companies. When a system faces a normal lifecycle, it is submitted to degradation thought the time and sometimes is not easy to identify the degradation evolution from complex data so that specialized methods such as SOM should be applied to facilitate the data analysis.

Using SOM is has been possible to identify the degradation trend on simulated aircraft jet-engines, considering first, a single operational mode of the engines at the time. This yielded a SOM in which was possible to observe the evolution of the degradation by means of a colours transitions on the map neurons, from a white region (optimal conditions) to a black region (faulty conditions) on the trained map.

Since an Aircraft faces several operational modes during a flight, a diagnosis based on one single operational mode is not enough, since the data might be altered for external reasons. A multiple operational mode approach is proposed, training another SOM with all operational modes available on the jet-engines dataset. To facilitate the analysis, a degradation index has been computed from each neuron weight vector and normalized from 0 to 1 on each operational mode. These degradation indexes allowed to determine if the tested engines are in the faulty condition or not.

Future work will be dedicated to the computation of Remaining Useful Life of the jet-engines considering the degradation indexes obtained from the trained SOM, this includes further analysis among the degradation index threshold. Finally, based on the obtained results, different SOM architectures will be studied so to assess their suitability to the problem at hand, and looking at the applicability in terms of computation time for training, convergence, etc.

REFERENCES

- Côme, E., Cottrell, M., Verleysen, M., and Lacaille, J., 2010, Self-Organising Star (SOS) for health monitoring. Michel Verleysen Ed. European conference on artificial neural networks, pp. 99-104.
- Côme, E., Cottrell, M., Verleysen, M., and Lacaille, J., 2010, Aircraft engine health monitoring using Self-Organising Maps, Petra Perner Ed. 10th Industrial Conference, ICDM 2010, Berlin Springer 6171, pp. 405-417.
- Delgado, S., Higuera, C., Calle-Espinoza, J., Morán, F., and Montero, F., 2017, A SOM prototype-based cluster analysis methodology, *Expert Systems with Applications*, vol. 88, pp. 14-28.
- Germen, E., Basaran, M., & Fidan, M. 2014. Sound based induction motor fault diagnosis using Kohonen self-organising map. *Mechanical Systems and Signal Processing*, 46, 45-58.
- Hu, J., Zhang, L., and Liang, W., 2013, Dynamic degradation observer for bearing fault by MTS-SOM system, *Mechanical Systems and Signal Processing*, vol. 36, pp. 385-400.
- Jardine, A.K.S., Lin, D., and Banjevic, D., 2006, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20, pp. 1483-1510.
- Kohonen, T., 1990, The self-organising map, *Proceedings of the IEEE*, vol. 78, n° 9, pp. 1464 -1480.
- Kohonen, T., 1995, Self-organising map. Second Edition. Springer Series in Information Sciences. Ed. T.S. Huang, T. Kohonen and M.R Schroeder. Berlin Heidelberg: Springer-Verlag.
- Nguyen, K.-A., Do, P., and Grall, A., 2015, « Multilevel predictive maintenance for multi-component system », *Reliability Engineering and System Safety* 144, pp. 83-94,.
- Saxena, A., Goebel, K., Simon, D., and Eklund, N., 2008, Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation, NASA, Integrated Vehicle Health Management.
- Vingerhoeds, R.A., Janssens, P., Netten, B.D., and Aznar Fernández-Montesinos, M., 1995, Enhancing off-line and on-line condition monitoring and fault diagnosis, *Control Engineering Practice*, Vol.3, pp. 1515-1528.

ENJEUX FINANCIERS DE DDMRP : UNE APPROCHE SIMULATOIRE

S. BAYARD,

F. GRIMAUD

Ecole Nationale des Ingénieurs de Saint Etienne –
58 rue Jean Parot, 42000 Saint Etienne
stephanie.bayard@enise.fr,

Mines de Saint-Etienne, LIMOS UMR CNRS 6158
158 cours Fauriel, 42023 Saint-Etienne Cedex 2
frederic.grimaud@mines-stetienne.fr

RESUME : *Cet article porte sur l'évolution des modes de pilotage industriels et l'impact du choix sur les performances, notamment financière. La plupart des entreprises adoptent un mode de planification basé sur les principes du MRP2 qui datent de la fin des années 70'. Force est de constater que les conditions de l'environnement économique ne sont plus les mêmes. Les entreprises ont donc mis en place des systèmes qui visent à donner plus de réactivité au MRP2 (Kanban, Lean, Conwip). Mais face à la variabilité, les performances ne sont pas toujours au rendez-vous et la multiplication des sécurités (stock ou délai notamment) conduit à l'inertie du système et à des contre-performances en termes financiers et de service client. Face à ce constat la théorie du Demand Driven semble offrir une alternative intéressante. L'objectif de cet article est de montrer, via la simulation à événements discrets, quel impact a ce système sur le BFR dans un contexte de demande stochastique et si l'on peut envisager de diminuer les coefficients de dimensionnement pour améliorer la performance financière.*

MOTS-CLES : *pilotage industriel, DDMRP, simulation à événements discrets, stock, performance financière*

1 CONTEXTE

Depuis quelques années, l'approche DDMRP (Demand Driven Material Planning) a été lancée par Ptak et Smith (2011) promettant à la fois de faciliter la planification de la production et son exécution mais annonçant également de meilleures performances en terme de service clients et de résultats financiers. Cette approche séduit de plus en plus d'industriels qui mettent en place ce système de pilotage comme ils ont adopté d'autres systèmes par le passé pour faire face aux différentes évolutions du marché et pour répondre au mieux à ses exigences, mais également en vue de préserver leurs performances financières.

Il existe, en effet, des liens forts entre les méthodes d'organisation industrielle et la performance financière. Une meilleure productivité permet par exemple de mieux répartir les coûts et donc de dégager un résultat final supérieur. Une réduction des délais et du niveau de stock permettent d'abaisser sensiblement le niveau du besoin en fonds de roulement (BFR). Ce dernier représente le besoin de financement de l'activité courante de l'entreprise et est un bon indicateur de la performance financière. Réduire le BFR libère des ressources financières pour l'entreprise et lui permet d'afficher une meilleure situation patrimoniale.

Si les objectifs opérationnels affichés des différents systèmes de gestion de production sont souvent loin de ces considérations, ce n'est pas le cas de la réflexion stratégique au sein de laquelle les équilibres financiers sont une préoccupation majeure visant à soutenir la croissance de l'entreprise et assurer sa pérennité. Comme nous le verrons par la suite, ce constat explique en partie

le parallèle entre les évolutions des systèmes de production et la recherche de la performance financière.

2 PROBLEMATIQUE :

Dans le contexte actuel de reprise économique annoncée, l'enjeu du financement de la croissance des entreprises est capital. Il nous paraît donc opportun de faire le lien entre cette nouvelle méthode de pilotage industriel - le DDMRP - et ses implications financières notamment en termes de BFR, d'autant plus que c'est l'un des arguments affichés de ces auteurs. L'indicateur financier qu'ils utilisent est celui du Working Capital défini comme la valeur du stock moyen qui a un impact majeur dans le volume du BFR.

Dans ce travail préliminaire nous souhaitons, dans un premier temps, vérifier l'impact sur le BFR de la méthode DD MRP dans un contexte de demande stochastique par rapport à une approche traditionnelle basée sur le MRPII (Manufacturing Resources Planning), et dans un deuxième temps, vérifier si les coefficients de dimensionnement peuvent être réduits pour améliorer le BFR.

Nous avons choisi une approche simulatoire basée sur une modélisation du système de gestion de production via un modèle de simulation générique sous Arena couplé à une base de données (tableur). Une application sur un cas d'étude simple, permettant d'effectuer des expérimentations sur les coefficients de dimensionnement de la méthode DDMRP, a été menée.

Dans cet article, nous commençons par retracer l'évolution des systèmes de production et leur impact financier, nous présentons ensuite le modèle DDMRP,

puis la démarche simulateur et enfin nous concluons sur les résultats et les perspectives ouvertes par ce travail

3 EVOLUTION DES SYSTEMES DE PRODUCTION ET DE LEUR SYSTEME DE GESTION

3.1 Depuis le début du 20^e

Le travail schéma ci-dessous établi par Y Koren (Koren 2010) met en évidence ce lien temporel entre les business-modèles et les enjeux présents au moment de leur émergence.

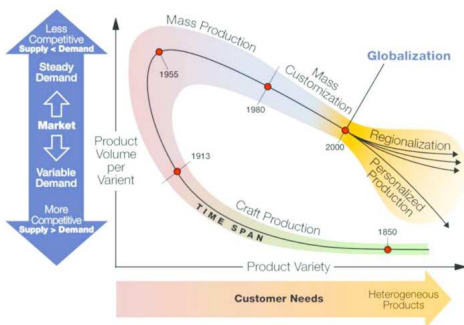


Figure 1 : The drivers to new paradigms are market and society needs Y Koren (2010)

Nous pouvons également ajouter un parallèle entre ces business-modèles et les modes productifs eux-mêmes. Ainsi, la production de masse portée par Ford répondait à la nécessité de combler une demande croissante et peu exigeante en termes de variété. En termes de gestion de production, l'enjeu essentiel était le fonctionnement des lignes en continue donc notamment la maintenance des équipements et la sécurisation des approvisionnements. Ce n'est pas la planification qui est au centre des préoccupations mais plutôt la productivité de l'appareil productif.

La mass customization, qui répond à une demande forte mais davantage diversifiée, s'appuie, quant à elle, sur les principes de la théorie du MRP développé par Orlicky (1975) dans les années 70'. Cette méthode permet de gérer les besoins en fonction de la demande et des prévisions en limitant le surstockage. Elle permet une certaine visibilité des besoins de production et d'approvisionnement en fonction de la demande réelle et des prévisions. Cette méthode s'est généralisée dans les années 80 grâce à l'informatisation des calculs. Le MRP présente donc l'avantage de mieux maîtriser les stocks de composants et les approvisionnements. Ce n'est donc pas un hasard qu'il soit largement adopté par les entreprises occidentales au début des années 80 qui font alors face à un ralentissement économique durable qui met en péril leur équilibre financier.

Cette méthode de calcul est encore au cœur des systèmes de la planification des entreprises car elle est la base des systèmes ERP (Entreprise Resources Planning) et des

systèmes de GPAO (Gestion de Production Assistée par Ordinateur) dont disposent les entreprises.

D'après Koren (2010), les années 2000 marquent l'accroissement de la demande de personnalisation mais aussi et surtout d'une variabilité exacerbée. Nombre d'entreprises occidentales se tournent alors vers les méthodes issues du lean manufacturing (Womack et Jones 1990) et de la méthode OPT (Optimized Production Technology). La première repose sur les principes développés dès les années 50 chez Toyota au Japon : production en fonction de la demande, fractionnement des lots de fabrication, élimination de toutes les sources de gaspillages, qualité totale, standardisation, Kanban (Ohno and Mito 1990)... Elle permet de répondre rapidement à une demande hétérogène grâce à une réactivité et une flexibilité accrue. Elle permet également d'améliorer sensiblement la situation financière des entreprises en abaissant les coûts et en réduisant les stocks. La seconde est la méthode OPT, développée par Goldratt et Cox (1993) à partir de 1984. L'idée centrale est d'ordonner la production à partir de la machine la moins productive (goulot) et de ne produire que ce qui est demandé. L'objectif financier est clairement affiché par cette théorie qui mesure les résultats grâce 3 indicateurs :

- « throughput » : indicateur mesurant « le rythme auquel le système génère de l'argent par les ventes
- « operations expenses » : les dépenses de fonctionnement
- L'inventary qui représente l'ensemble de l'actif immobilisé ainsi que les stocks

Il est à noter que les grandes entreprises, notamment du secteur automobile, avaient adopté ce type de démarche dès les années 90. Pour les grands groupes occidentaux, ces années sont marquées par de grandes réorganisations industrielles avec des mouvements fusion-acquisition importants et un recentrage sur le cœur de métier. L'objectif est de concentrer la valeur ajoutée sur leur cœur de métier. De ce fait, les méthodes japonaises et OPT leur permettent de dégager des ressources financières notamment en réduisant les stocks.

3.2 La situation récente

Depuis plus de 15 ans ces théories deviennent la règle dans les entreprises occidentales et se développent sous la nouvelle appellation « excellence opérationnelle » dans laquelle l'amélioration continue a une place centrale. Mais elles induisent également une plus grande nervosité du système MRP qui continue à alimenter le système de prévisions des entreprises. Ce phénomène a été décrit par Forrester (1958) qui constate que la nervosité de demande se répercute avec un effet d'amplification, « effet coup de fouet » au minimum le long de la nomenclature mais aussi tout le long de la supply chain. Or les années 2000 marquent l'ère de la globalisation avec des entreprises impliquées dans plusieurs supply chains et souvent à l'échelle mondiale.

Cette nervosité a donc des conséquences néfastes sur le fonctionnement des chaînes logistiques et plus largement sur la performance de l'ensemble des maillons des SCM (à des degrés différents en fonction de la position dans la chaîne et du pouvoir exercé par chaque maillon). Ainsi les donneurs d'ordre sont moins impactés que leurs fournisseurs car ils reportent leurs contraintes sur eux par le biais contractuels.

Pour faire face à cette nervosité et continuer à répondre à la demande toujours plus volatile, les entreprises sont contraintes d'augmenter les stocks et les délais de sécurité. Ce type de réaction est d'autant plus commun que la variabilité de toute nature augmente et que la fenêtre de prévision est courte.

Si ces sécurités devraient permettre de conserver des taux de performance acceptables en termes d'OTD (On-Time Delivery), ce n'est pas toujours le cas d'un point de vue financier. L'augmentation des stocks et des délais pénalisent le besoin en fonds de roulement et immobilisent les ressources de l'entreprise. Paradoxalement, malgré ces contre-mesures, le respect du délai client n'est pas toujours atteint et des ruptures de stock apparaissent. Ce phénomène est notamment constaté dans la filière aéronautique qui face à des montées en cadence forte liées aux succès commerciaux et qui ne parvient pas toujours à assurer les livraisons alors que la demande finale est pourtant connue plusieurs années à l'avance.

Force est de constater qu'actuellement les méthodes lean cohabitent avec la méthode MRP alors même que leurs principes de fonctionnement divergent, notamment du point de vue de la planification. De nombreuses études, notamment synthétisées dans une étude taxonomique de littérature de Powell et al (2013), ont été menées montrant la réalité du phénomène. Actuellement, on peut donc parler d'une situation hybride où les entreprises tentent d'adapter des méthodes qui ne sont plus en phase avec les contraintes de leur environnement. Elles jouent notamment sur les délais et stocks de sécurité ce qui conduit à des contreperformances. Dans cette démarche, le positionnement et le dimensionnement des stocks reprennent tout leur sens alors que les théories appuyées sur les principes du Juste à Temps prônaient leur réduction au minimum. D'après Ptak et Smith (2011, 2016), la question fondamentale concernant la gestion de stock devient « Où » et non plus « Combien et Quand ». Mais on peut, à juste titre, s'interroger sur le dimensionnement qui est une question fondamentale de la méthode et qui a un impact fort sur le BFR.

3.3 Le DDMRP

3.3.1 Définitions

D'après Ptak et Smith (2011), la méthode DD MRP « is a formal multi-echelon planning and execution method to protect and promote the flow of relevant information through the establishment and management of strategically placed decoupling point stock *buffers*.» [...]

Selon APICS Dictionary (2015), il s'agit – A method for planning material needs that enables a company to build more closely to actual market requirements.

D'après ces définitions, il s'agit bien d'une méthode hybride à l'instar de ce qui se pratique déjà en entreprise mais l'objectif est de proposer une réelle nouvelle approche formalisée réconciliant les méthodes lean, OPT et MRP.

3.3.2 Objectifs

Cette nouvelle approche est fondée sur l'idée que des stocks de découplage permettraient de réduire la nervosité du système et d'améliorer les indicateurs de performance que sont l'OTD et le Working capital.

Mais au-delà des objectifs de performance globale, la méthode ambitionne aussi de simplifier la planification et la gestion des alertes sous-MRP qui est devenue beaucoup trop complexe pour les personnes en charge de cette tâche.

3.3.3 Les principes

Selon les auteurs, le DDMRP repose sur 5 composants comme le montre la figure ci-dessous.

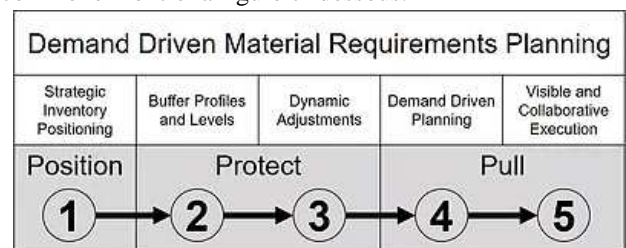


Figure 2 : The five components of DDMRP (Ptak et Smith, 2016)

Dans ce travail nous nous concentrons sur les 2 premiers points qui sont ceux qui entrent en jeu dans le paramétrage de base.

Les auteurs ont également introduit de nouvelles notions qui devraient permettre de simplifier la planification et son exécution. Pour mener à bien notre étude, nous avons eu recours à plusieurs de ces nouvelles notions :

- Délai découplé : c'est-à-dire la plus longue séquence de la nomenclature non protégée par un *buffer*
- demande qualifiée : elle intègre la demande résultant uniquement de commandes fermes ainsi que les commandes en retard,
- l'équation de flux : le stock physique augmenté des ordres d'approvisionnement en cours et diminué de la demande qualifiée.

3.3.4 Le positionnement stratégique

Selon les auteurs, il repose sur des heuristiques avec pour objectif principal que le plus long délai découplé soit inférieur au délai client. Le processus est itératif et doit permettre réduire au maximum l'impact de la variabilité en découplant la nomenclature.

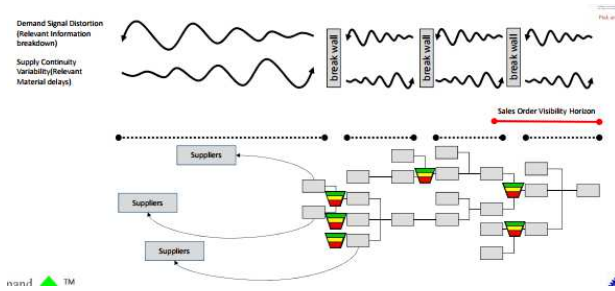


Figure 3 :A system with decoupling point buffer (Ptak et Smith 2016)

3.3.5 Les profils de buffer

Le *buffer* est le stock de découplage et est divisé en trois zones distinctes Red, Yellow et Green qui symbolisent les différentes plages au sein desquelles le niveau de stock peut évoluer.

Les différents niveaux sont définis comme suit :

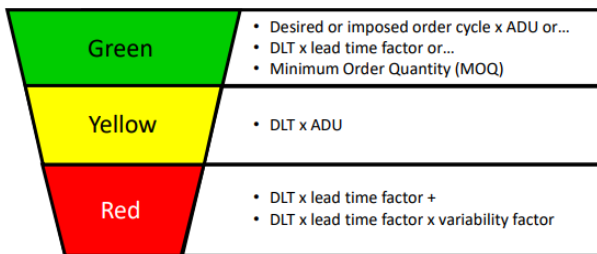


Figure 4 :The *buffer* equation summary (Ptak et Smith, 2016)

Les coefficients, Lead Time Factor et Variability Factor, utilisés comme base pour les calculs de la zone rouge reposent sur 3 critères

- Le type de pièce :
 - o article parent c'est-à-dire vendu par l'entreprise
 - o sous-ensemble fabriqué par l'entreprise
 - o article acheté
- la longueur du délai divisé en trois catégories (long, moyen ou court) dont les seuils qui varient d'un type de pièce à l'autre
- la variabilité divisé elle aussi en 3 niveaux

Grâce à ces critères, on obtient 36 catégories d'articles et donc des profils de *buffers* différents pour lesquels les auteurs recommandent une fourchette de variation des facteurs de variabilité et de délais.

		Part Type				
		Purchased	Manufactured	Distributed	Intermediate	
Lead Time Category	Short	PSL	MSL	DSL	ISL	Low
		PSM	MSM	DSM	ISM	Medium
		PSH	MSH	DSH	ISH	High
Medium	PML	MLM	DML	IML	Low	
	PMM	MLM	DMM	IMM	Medium	
	PMH	MLH	DMH	IMH	High	
Long	PLL	MLL	DLL	ILL	Low	
	PLM	MLM	DLM	ILM	Medium	
	PLH	MLH	DLH	ILH	High	

Figure 5 :*Buffer* profiles Ptak et Smith (2016)

Variability Factor Range	
High Variability	61-100%+ of Safety Base
Medium Variability	41-60% of Safety Base
Low Variability	0-40% of Safety Base

Lead Time Factor Range	
Long Lead Time	20-40% Average Daily Usage (ADU) x Decoupled Lead Time (DLT)
Medium Lead Time	41-60% ADU x DLT
Short Lead Time	61-100% ADU x DLT

Figure 5 : proposition de fourchette pour les coefficients de dimensionnement (Ptak et Smith, 2016)

Nous nous interrogeons sur l'influence de ces coefficients utilisés pour le dimensionnement des buffers sur le BFR. Une première étude menée par Miclo et Al (Miclo et al 2016) a montré que ces paramètres pouvaient être ajustés par simulation permettant une amélioration du BFR sans dégradation de l'OTD. Les résultats étaient probants menant à un ajustement de ces coefficients référence par référence et dans le cadre d'une demande variable mais égale à la demande prévisionnelle du scénario de base. Ces conclusions nous conduisent à vouloir vérifier dans quelle mesure ils peuvent être abaissés dans les calculs de base dans un contexte de demande stochastique.

4 LA MODELISATION

Notre étude est une première approche portant sur un modèle simple qui devrait permettre de montrer l'impact de la méthode DDMRP par rapport à un modèle basé sur MRP en contexte de demande stochastique et vérifier s'il est possible de diminuer les coefficients de base pour permettre d'améliorer le BFR.

Le modèle est un modèle générique construit sur le logiciel Arena, un logiciel de simulation à événements discrets, couplé à une base de données permettant d'intégrer le système de gestion et son paramétrage. Il existe une version pour chaque méthode.

4.1 La base de données

Les données sont globalement structurées de la même manière dans les 2 cas (MRP et DDMRP) mais les modes de calculs sont basés sur les hypothèses liées au mode de pilotage.

Elles sont regroupées en 4 tables :

- Data : regroupe l'ensemble des paramétrages de base
- CBN ERP : permettant le calcul des besoins
- CBN : synthèse des besoins par période
- BOM : liens de nomenclature
- Dem PF : demande Client

Dans le cas MRP, les données de base utilisées dans la table DATA sont :

- le code article,
- le stock initial : déterminé de façon à couvrir la demande pendant le délai de réapprovisionnement

ment lors de la première commande augmentée du stock de sécurité

- la taille de lot
- le délai de fabrication ou d'approvisionnement
- le cout unitaire,
- la valeur totale du stock
- le stock de sécurité.

Les autres tables ont été construites sur la base des paramètres de base en respectant la méthode MRP.

Concernant le fichier DDMRP, il inclut en plus des paramètres initiaux :

- les calculs des différents niveaux de buffers
- le niveau du « Targeted on Hand » qui sert de base pour dimensionner le stock initial des articles *bufferisés*

La méthode du calcul des besoins est également adaptée en cohérence avec la méthode DDMRP.

Dans les 2 cas, les tables CBN et DemPF, permettent la lecture de la demande par article dans Arena.

Des liens ont été construits vers Arena qui permettent :

- de configurer les paramètres initiaux
- De rechercher le besoin, dans le cas DDMRP pour chaque référence en fonction des liens de nomenclature et de la demande qualifiée
- De mettre à jour le niveau de stock de chaque composant et la valorisation

4.2 Le modèle Arena

Nous avons choisi de travailler à partir d'un modèle générique dont l'évolutivité repose sur la gestion des attributs et des variables et le couplage avec le tableur.

Le modèle s'articule autour de deux sous modèles :

- Le premier permet de gérer la demande de produits finis
- Le second permet de gérer les mouvements de stock de chaque référence en fonction des besoins

Il est géré à partir d'un set de stations dont le nombre est dépendant du nombre de références présentes dans le tableur.

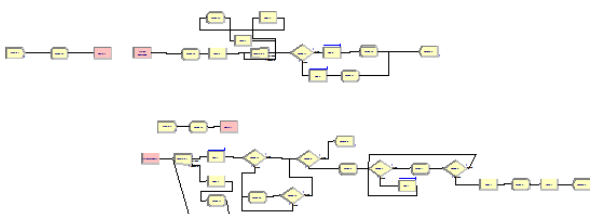


Figure 5 :Modèle générique MRP développé sous Arena

Il fonctionne grâce à 6 variables générales dont 3 sont en lien avec le tableur ce qui permet de lire des paramètres utiles au fonctionnement du modèle.

4.3 Les indicateurs

Les indicateurs recueillis dans le modèle Arena sont les stocks moyens par référence et leur valorisation.

Ils reposent sur les variables qui correspondent au stock initial et de la valeur du stock par référence.

Leur agrégation ultérieure permettra le calcul du BFR.

Nous recueillons également un OTD rapportant pour les produits finis, pour chaque période, le nombre de commandes livrées complètes et à temps, sur le nombre total de commandes. Cet indicateur n'a que valeur de vérification dans les cas de demande stable.

5 EXPERIMENTATION

Pour expérimenter notre modèle, nous avons décidé d'exploiter le modèle COMPANY ABC utilisé par les auteurs de la méthode DDMRP, C. Ptak et C. Smith, dans leur livre de référence « DD MRP ; Demand driven Material Requirement Planning ».

L'entreprise ABC fabrique 3 types de produits finis ayant les nomenclatures suivantes.

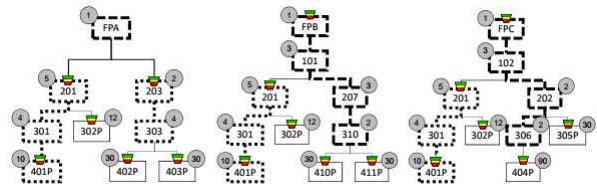


Figure 6 : nomenclature des produits de la Company ABC (Patck et Smtih 2016)

5.1 Les hypothèses du cas d'étude

Les coefficients de nomenclature sont de 1 pour chaque lien mais les délais sont variables d'une référence à l'autre. Les emplacements de buffer retenus sont ceux de la dernière itération présentée par les auteurs

La demande de produits finis dans le scénario de base est stable avec une consommation quotidienne comme indiqué ci-dessous en nombre d'unités

Référence	Consommation
FPA	250
FPB	100
FPC	300

Tableau 1 : Consommation moyenne quotidienne

Dans le cadre de la méthode DDMRP, les coefficients retenus pour le dimensionnement en fonction des profils de buffers sont les suivants

Parent Make Items (M)	Lead Time Factor	Variability Category	Variability Factor	
Long Lead Time	8+ days	.25	High	.75
Medium Lead Time	3-8 days	.4	Medium	.5
Short Lead Time	1-2 days	.7	Low	.25
Intermediate Make Items (I)				
Long Lead Time	8+ days	.25	High	.7
Medium Lead Time	3-8 days	.4	Medium	.5
Short Lead Time	1-2 days	.7	Low	.2
Purchased Items (P)				
Long Lead Time	20+ days	.3	High	.75
Medium Lead Time	11-19 days	.5	Medium	.5
Short Lead Time	1-10 days	.7	Low	.25

Tableau 2 : coefficients en fonction des profils de buffer

L'utilisation du tableur sous MRP traditionnel basé sur un stock de sécurité calculé avec $SS=LT*D$ conduit aux paramétrages de base suivants avec un stock initial de 23M€. Les différentes hypothèses de dimensionnement seront testées dans le plan d'expérience.

nom article	item type	stock initial	délai	taille de lot	coutfab/ac hat	valeur stock
		1	2	3	9	10
FPA	1	500	1	250	450	225 000
FPB	2	200	1	100	500	100 000
FPC	3	600	1	300	445	267 000
101	4	600	3	100	500	300 000
102	5	1800	3	300	445	801 000
201	6	6500	5	650	210	1 365 000
202	7	1200	2	300	250	300 000
203	8	1000	2	250	290	290 000
207	9	600	3	100	235	141 000
301	10	5200	4	650	110	572 000
303	11	2000	4	250	290	580 000
306	12	1200	2	300	100	120 000
310	13	400	2	100	110	44 000
302P	14	15600	12	650	250	3 900 000
305P	15	18000	30	300	125	2 250 000
401P	16	13000	10	650	110	1 430 000
402P	17	15000	30	250	110	1 650 000
403P	18	15000	30	250	140	2 100 000
404P	19	54000	90	300	90	4 860 000
410P	20	6000	30	100	200	1 200 000
411P	21	6000	30	100	125	750 000
						23 245 010,00

Tableau 3 : Paramétrage initiaux MRP

Les paramètres de base avec un système DDMRP sont les suivants et nous conduisent à une valeur initiale de stock de 21,7 M€ répartie comme il suit :

nom article	item type	stock initial	délai	taille de lot	coutfab/ac hat	valeur stock
		1	2	3	9	10
FPA	1	250	1	250	450	112 500
FPB	2	1151	1	100	500	575 500
FPC	3	3451	1	300	445	1 535 695
101	4	2688	3	100	500	1 344 000
102	5	8062	3	300	445	3 587 590
201	6	3650	5	650	210	766 500
202	7	0	2	300	250	-
203	8	1275	2	250	290	369 750
207	9	0	3	100	235	-
301	10	9305	4	650	110	1 023 550
303	11	3075	4	250	290	891 750
306	12	1800	2	300	100	180 000
310	13	600	2	100	110	66 000
302P	14	15935	12	650	250	3 983 750
305P	15	9000	30	300	125	1 125 000
401P	16	7963	10	650	110	875 930
402P	17	8000	30	250	110	880 000
403P	18	8000	30	250	140	1 120 000
404P	19	25200	90	300	90	2 268 000
410P	20	3000	30	100	200	600 000
411P	21	3000	30	100	125	375 000
						21 680 525.00

Tableau 4 : Paramètres initiaux DD MRP

Cette valeur est inférieure de 1.5M€ à celle du système MRP. Cela étant le but de ce travail est de vérifier comment ce comporte cet indicateur dans le cadre de la simulation et comment les hypothèses de dimensionnement l'impactent.

Les buffers ont été dimensionnés sur la base des paramètres suivants

nom article	green V1	green V2	green V3	green zone	yellow zone	red base	red safety	TOR	TOY	TOG
FPA	0	0	0	0	0	0	0	0	0	0
FPB	575	250	300	575	2 300	575	288	863	3 163	3 738
FPC	1 725	250	900	1 725	6 900	1 725	863	2 588	9 488	11 213
101	0	0	0	0	0	0	0	-	-	-
102	0	0	0	0	0	0	0	-	-	-
201	1 463	250	1 950	1 950	5 850	1 463	293	1 755	7 605	9 555
202	0	0	0	0	0	0	0	-	-	-
203	750	250	750	750	1 500	750	150	900	2 400	3 150
207	0	0	0	0	0	0	0	-	-	-
301	0	0	0	0	0	0	0	-	-	-
303	0	0	0	0	0	0	0	-	-	-
306	0	0	0	0	0	0	0	-	-	-
310	0	0	0	0	0	0	0	-	-	-
302P	3 900	650	1 950	3 900	7 800	3 900	975	4 875	12 675	16 575
305P	2 700	300	900	2 700	9 000	2 700	675	3 375	12 375	15 075
401P	4 550	250	1 950	4 550	6 500	4 550	1 138	5 688	12 188	16 738
402P	2 250	250	750	2 250	7 500	2 250	563	2 813	10 313	12 563
403P	2 250	250	750	2 250	7 500	2 250	563	2 813	10 313	12 563
404P	8 100	300	900	8 100	27 000	8 100	2 025	10 125	37 125	45 225
410P	900	100	300	900	3 000	900	225	1 125	4 125	5 025
411P	900	100	300	900	3 000	900	225	1 125	4 125	5 025

Tableau 5 : Résultat du paramétrage des zones de buffers

Les articles non bufferisés sont gérés en mode MRP sans stock de sécurité puisque, à priori, ils sont protégés par les buffers.

5.2 Les scenarii

Afin de démontrer l'impact des coefficients choisis pour la base de calcul dans DDMRP, nous avons d'abord testé

trois scenarii de base sur MRP sans variabilité de la demande et en faisant varier le stock de sécurité (scenarii 1,2, 3). Le 1^{er} est basé sur l'absence de stock de sécurité. Le second repose sur une méthode empirique utilisant la consommation moyenne durant le délai d'approvisionnement moyen. Le troisième utilise un Stock de Sécurité = Ecart type de la Demande X Coefficient de sécurité X racine carré du Délai de réapprovisionnement avec un taux de service à 90%.

Nous avons ensuite induit, dans ces 3 scenarii, de la variabilité dans la demande en utilisant une loi normale avec pour moyenne la consommation sur la période avec un écart-type de 10% pour vérifier son impact sur le BFR et sur le taux de service (scenarii 4,5, 6)

Ces six scenarii permettront la comparaison avec les résultats obtenus avec les scenarii basés sur la méthode DDMRP avec demande quotidienne stable (scenario 7) avec une variabilité induite au niveau de la demande finale (scenario 8). Dans les 2 cas, nous utilisons les paramètres de dimensionnement utilisés par les auteurs puis nous testerons la possibilité de réduire par sauts successifs de -0.05 points par rapport aux coefficients de base (scenarii 9,10, 11, 12).

	Scenario	Hypothèse de variabilité	Stock sécurité	Coefficient DD MRP
1	MRPV1	Aucune	aucun	
2	MRPV2	Aucune	Délai*conso moyenne	
3	MRPV3	Aucune		
4	MRPV4	Demande : loi normale (conso moyenne, σ 10%)	Ecart-type	
5	MRPV5	IDEM	Ecart-type	
6	MRPV6	IDEM	Ecart-type	
7	DD MRP V1	Aucune		Coefficient de base
8	DD MRP V2	demande10%		Coefficient de base
9	DD MRP V3	demande10%		-0.05 pt
10	DD MRP V4	demande10%		-0.1 pt
11	DD MRP V5	Idem		-0.15 pts
12	DD MRP V6	Idem		-0.2 pts

Tableau 7 : Tableau des scenarii testés

6 RESULTAT ET PERSPECTIVES

6.1 Les résultats en avenir certain

En comparant les résultats de nos simulations en avenir certain, nous constatons que la méthode DDMRP (7) conduit à un BFR plus élevé quelle que soit la méthode de dimensionnement du stock de sécurité retenue.

N°	Scenario	VARIABILITE	Stock sécurité /coefficients	BFR en K€	OTD
1	MRPV1	0%	0	2,2	100%
2	MRPV2	0%	approche empirique	13,2	100%
3	MRPV3	0%	approche probabiliste	2,6	100%
7	DD MRP V1	0%	COEFF DE BASE	14,0	100%

Tableau 8 : Résultats en avenir certain

Elle se rapproche des résultats d'un MRP dont le stock de sécurité est dimensionné en fonction d'une approche empirique (2). L'écart est de l'ordre de 6%. En termes de

performance client, la demande étant connue, l'OTD est nécessairement maximum.

Ces résultats ne peuvent constituer une conclusion et nécessitent d'être confrontés à la variabilité de la demande.

6.2 Les résultats en avenir incertain

Les différents scenarii faisant appel à la variabilité (4 à 6 et 8 à 12) de la demande ont été testés sur 10 réplifications de 100 jours afin de fiabiliser les résultats qui suivent :

N°	Scenario	VARIABILITE	Stock sécurité /coefficients	BFR en K€	OTD	σ
4	MRPV4	10%	0	2,9	85%	3,5
5	MRPV5	10%	approche empirique	13,2	92%	7,1
6	MRPV6	10%	approche probabiliste	2,9	93%	2,5
8	DD MRP V2	10%	coefficient de base	14,3	98,2%	1,2
9	DD MRP V3	10%	0,05	13,2	98,1%	0,4
10	DD MRP V4	10%	0,1	12,0	98,4%	0,7
11	DD MRP V5	10%	0,15	10,1	98,2%	0,4
12	DD MRP V6	10%	0,2	10,7	98,1%	0,5

Tableau 9 : Résultats avec demande stochastique

Concernant la méthode MRP, on constate une nette dégradation de l'OTD. Le meilleur résultat est obtenu par la méthode probabiliste du dimensionnement du stock de sécurité (6) tout en conservant un très bon niveau d'OTD.

Quel que soit le scenario considéré, les résultats obtenus avec la méthode DD MRP montrent des performances client nettement supérieures quelle que soit les coefficients envisagés. Tous présentent, en effet, des taux de service supérieur à 98%.

Le scenario (8) avec les coefficients de dimensionnement proposé par Ptak et Smith, met en évidence un meilleur OTD et mais un BFR largement supérieur à celui obtenu par le scenario de référence MRP (respectivement 98% soit +5 pts) et 4 fois plus élevé). Face à la dégradation du taux de service, les entreprises auraient donc intérêt à adopter la méthode DDMRP. Mais le problème du financement du BFR se pose.

De ce fait, les autres scenarii DDMRP (9 à 12) montrent que la diminution des coefficients permet d'améliorer sensiblement le montant du BFR tout en conservant des performances supérieures au scenario de référence de la méthode MRP (scenario 5) en termes d'OTD. L'abaissement de 0.15 (scenario 11) montre des résultats proches avec un OTD de 98,1% et un BFR de 10,1 K€ qui représentent une amélioration de l'ordre de 29% par rapport au scenario DD MRP de référence (8). Par contre, il semble que la poursuite de la diminution (scenarii 9 et 10) soit contreproductive puisque le BFR et l'OTD se dégrade. De fait, une réflexion autour du seuil des coefficients semble pertinente.

7 CONCLUSION ET PERSPECTIVES

Cette expérimentation était un travail préliminaire visant à prouver le bien-fondé d'une réflexion autour des l'abaissement des coefficients de départ dans la méthode DDMRP et les conséquences sur l'OTD et le BFR. Les systèmes de pilotage de production ayant un lien indéniable avec la performance financière, il nous paraît op-

portun dans le contexte économique actuel de tester ce nouveau mode pilotage sur cet aspect.

Suite à ce travail, il apparaît que, face à une demande stochastique, si la priorité est donnée à l'OTD, la méthode DDMRP semble effectivement plus performante en termes de BFR et d'OTD. Par ailleurs, l'expérimentation montre qu'un abaissement empirique des coefficients de dimensionnement des *buffers* améliore effectivement la performance financière par rapport aux coefficients utilisés dans le cas présenté par Ptak et Smith. Cette amélioration atteint 29% de réduction du BFR. Dès lors, il apparaît opportun de poursuivre les investigations pour affiner ces coefficients.

De ce fait, les travaux à venir devront s'attacher à parfaire le modèle de simulation. En effet, il n'est pas tenu compte de la valeur ajoutée par le travail de l'entreprise dans la valeur des stocks ce qui conduit à des approximations dans le calcul du BFR. Par ailleurs, nous n'avons pas intégré la capacité finie de l'entreprise ni d'autres facteurs de variabilité tels que la fiabilité de l'outil productif ou encore celle des délais que ce soit d'approvisionnement ou de fabrication. De prochains travaux appuyés sur le même type d'approche simulateur et intégrant ces différents éléments nous paraissent indispensables pour pouvoir tirer des conclusions plus significatives.

REFERENCES

- APICS Dictionary, 2016 15th Edition,
- Banks, J., J.S. Carson II, B.L. Nelson and D.M. Nicol, 2014. *Discrete-Event System Simulation*, Pearson New international Edition, Pearson Education Limited,
- Forrester, J.W., 1958; A major breakthrough for decision makers", *Harvard Business Review*, n° 36, p. 37-66
- Giard, V. and G. Mendy, 2007, Production à flux tirés dans une chaîne logistique, *Revue Française de Génie Industriel*, vol. 26, n° 1, p. 87-110
- Goldratt, E. and J. Cox, 1993. *Le but, un processus de progress permanent*, AFNOR
- Kelton, W.D., R.P Sadowski, and N.B. Zupick, 2015. *Simulation with Arena, 6th edition*, Mac Graw Hill,
- Koren, Y., 2010, *The Global Manufacturing Revolution: Product-Process-Business Integration and Reconfigurable Systems*, Wiley.
- Miclo, R., F. Fontanili, M. Lauras, J. Lamothe, and B. Milian, 2016. *An empirical study of Demand-Driven MRP*, Information Systems Logistics and Supply Chain (ILS 2016), Bordeaux France
- Miclo, R., 2016, *Challenging the "Demand Driven MRP" Promises : a Discrete Event Simulation Approach. Modeling and Simulation*. Thèse de Doctorat, Ecole des Mines d'Albi-Carmaux, France
- Observatoire des PME, 2011, *La défaillance des entreprises, Etude sur données françaises entre 2000 et 2010*, OSEO
- Ohno, T. and S. Mito, 1993, *Présent et avenir du Toyotisme*, Masson,
- Orlicky, J., 1975, *Material Requirements Planning*, McGraw-Hill, New York,
- Powell, D., J. Bas and I. Alfnes, 2013, Integrating Lean and MRP: A Taxonomy of the Literature , *Advances in Production Management Systems*. (APMS 2013): State College, PA, USA, September 9-12, 2013, Proceedings, Part II, Springer Berlin Heidelberg, 2013 , 485-492
- Ptak, C., and C. Smith, 2011, *Orlicky's Material Requirements Planning*. McGraw Hill Professional.
- Ptak, C., and C. Smith, 2016, *DD MRP ; Demand driven Material Requirement Planning*; Industrial Press Inc., U.S.
- Womack, J.P., D.T. Jones, and D. Ross, 1990, *The Machine that Changed the World*, Rawson Associates

ENTREPOT DU FUTUR : FORMALISATION DES CARACTERISTIQUES CLEFS

Dmitry KUCHARAVY, David DAMAND

EM Strasbourg Business School,
Université de Strasbourg
HuManiS (EA 7308)
61 avenue de la Forêt-Noire, F-67085 Strasbourg Cedex
dzmitry.kucharavy@etu.unistra.fr,
damand@em-strasbourg.eu

Marc BARTH*, Ridha DERROUCHE

*INSA Strasbourg /
EM Strasbourg Business School,
Université de Strasbourg
HuManiS (EA 7308)
61 avenue de la Forêt-Noire, F-67085 Strasbourg Cedex
marc.barth@insa-strasbourg.fr, derrouiche@unistra.fr

RESUME : *Le contexte général de ce papier est la planification stratégique des entrepôts de type plateformes logistique. Il s'agit de prévoir l'évolution de l'entrepôt en fonction d'objectifs et d'enjeux liés généralement à l'évolution du marché (par ex. : commerce électronique, etc.) et de l'environnement (par ex. : logistique du dernier kilomètre, etc.). La planification passe par la maîtrise des conséquences des choix effectués mais également leurs interactions. Pour motiver et justifier des choix technologiques prévisionnels, il est crucial d'identifier et de comprendre les règles de conception souvent contradictoires. Il y a problème lorsqu'il y a contradiction ! Pour soutenir la planification stratégique, l'aide à la décision proposée dans ce papier consiste à extraire et formaliser l'ensemble des problèmes sous forme de règles de conception contradictoires.*

MOTS-CLES : *Systèmes d'entreposage, Planification stratégique, Prévisions technologiques, Evolution des entrepôts, Cartographie de contradictions.*

1 INTRODUCTION

Les coûts logistiques des entrepôts sont dans une large mesure déterminés pendant la phase de conception (Roveillo *et al.*, 2012). En règle générale, à partir d'une description fonctionnelle, la phase de conception de l'entrepôt consiste à choisir une implantation et les technologies associées des quatre activités de base (la réception, le stockage, la préparation de commandes et l'expédition), et, un mode de planification des opérations liées à ces activités (Gu *et al.*, 2007 ; Rouwenhorst *et al.*, 2000). L'évolution de cette description fonctionnelle est largement corrélée à l'évolution rapide du marché (De Koster *et al.*, 2017). L'introduction de nouveaux modes de consommation comme par exemple le commerce électronique est susceptible d'imposer l'introduction de nouvelles technologies de stockage dédiées, représentant de forts investissements et de surcroît de forts risques financiers.

Par conséquent, il est vital de planifier et contrôler l'évolution stratégique des entrepôts logistiques. Les décisions de planification stratégiques, généralement caractérisées par un horizon de 10 à 15 ans, concernent la détermination des politiques générales et des plans d'utilisation des ressources du futur entrepôt (Harrison, 1995).

Dans certains cas, il est possible de répondre rapidement à ces évolutions par des décisions programmées ; les

activités associées à la décision sont des activités d'exécution (Roy, 1985). En revanche, quand les décisions afférentes à la nouvelle situation sont indisponibles, une activité de résolution de problème est requise. Ces décisions non-programmées sont plus difficiles à étudier car elles impliquent un ensemble de connaissances très larges (endogènes et exogènes à l'entrepôt). Cet ensemble de connaissances est souvent contradictoire.

Les travaux de recherche entrepris en conception d'entrepôt portent principalement leur attention sur les paramètres endogènes d'entreposage (Chen, Paulraj, 2004), (Schnetzler, *et al.*, 2007). Peu de travaux prennent en compte simultanément les paramètres endogènes et exogènes. Le futur entrepôt dépend de l'évolution de l'environnement. Les demandes, les limites et les tendances du marché induisent des innovations et des changements opérationnels, technologiques, organisationnels (De Koster *et al.*, 2017). Dans ce contexte, deux questions de recherche sont posées. Première question : quels sont les caractéristiques clefs pour la prise de décision stratégique en conception d'entrepôt ? Deuxième question : comment définir une méthode systématique d'identification des paramètres clefs en conception d'entrepôt ?

Dans ce papier, une contribution est proposée dans le cadre de la première question. La contribution attendue est la formalisation d'un réseau de règles de conception génériques (moyen/effet) en utilisant un langage adapté

(Chen, Paulraj, 2004). Le caractère des règles de conception attendu est la contradiction.

Le papier est structuré comme suit. Le chapitre 2 présente le cadre théorique. Le chapitre 3 présente le cadre méthodologique pour identifier et extraire des règles de conception génériques. Le chapitre 4 présente une application industrielle effectuée chez un prestataire logistique de type PL3. Enfin le chapitre 5 conclut ce papier et énonce les perspectives de recherche.

2 REVUE DE LA LITTÉRATURE – CADRE THEORIQUE

Un entrepôt peut être représenté comme un réservoir de stockage (Fender, Pimor, 2013). Il a une fonction d'amortisseur compte tenu de la variabilité et de l'incertitude inhérente à la supply chain. Son cycle de vie inclut trois grandes étapes : conception et construction ; exploitation, aménagement et réorganisation ; déclassement ou retraitement. Pour envisager un design conceptuel sur un horizon de 5 et 10 années aux niveaux de l'architecture organisationnelle et des processus, le besoin d'une vision fiable sur les demandes, les limites et les tendances du futur entrepôt est nécessaire (De Koster *et al.*, 2017). La décision stratégique liée à l'évolution de l'entrepôt doit considérer les changements dynamiques des activités et de fonctions aussi bien que des particularités du cycle de vie du système (Andersson, Wemner, 2008).

La prise de décisions stratégique en conception d'entrepôt est associée aux questions génériques suivantes :

- Quels sont les niveaux de performance atteignables et non atteignables ?
- Est-il nécessaire de modifier l'état du système actuel ?
- Quel est l'effet des critères de décisions dans la nouvelle situation ?
- Quels sont les critères de décisions les plus significatifs sur la performance du système ?
- Dans le cas de la non-possibilité de modifier les critères de décisions simultanément, quel est l'ordre de modification des critères dans le temps ?
- Etc..

L'approche traditionnelle en phase de conception consiste à inventer, créer les alternatives possibles entre lesquelles il faudra choisir. Le modèle décisionnel rationnel (Simon, 1955), (Dawes, 1988), comporte 6 étapes (Robbins *et al.*, 2014) : 1. définir le problème ; 2. identifier les critères de décision ; 3. pondérer les critères ; 4. élaborer les alternatives ; 5. évaluer les alternatives (règle si critère=... alors alternative=...); 6. choisir la meilleure alternative. Ce modèle représente une structuration séquentielle et logique de l'information à analyser (Harrison, 1995), (Robbins, 2003). Les approches multicritères, généralement usités (Roy, 1985), peuvent être résumées par l'approche du critère unique de synthèse et

l'approche du système relationnel de synthèse (Damand, 2000). Dans le cas où les alternatives ne sont pas compatibles avec les valeurs des critères de décisions alors il faut prendre position vis-à-vis des contradictions.

Dans ce papier, nous posons l'hypothèse que le processus de résolution de problème en phase de conception passe par l'identification d'une contradiction ou des contradictions occasionnées par le nouveau contexte de décision (par ex. évolution du marché). Les contradictions sont alors l'expression d'un problème par des conséquences des choix effectués, mais il faudra évidemment qu'il y ait au moins deux conséquences antagonistes pour qu'il y ait problème. En demandant à exprimer les relations de cause à effet entre les moyens et les performances, le processus d'identification des contradictions révèle les règles utilisées. On explicite ainsi de façon incidente des zones de conflit, mais également et surtout les raisons des choix technologiques prévisionnels.

Pour faire tendre le processus de décision vers un choix de solution technologique, la littérature propose plusieurs méthodes de prévisions technologiques. Les premiers travaux significatifs datent du début du XX siècle (Ayres, 1969), (Linston, Turoff, 2011), (Martino, 1972). La fonction principale de ces méthodes est de fournir une vision consensuelle de la future situation technologique (Kostoff, Schaller, 2001). La veille des revues scientifiques, des rapports institutionnels et des livres publiés (Porter *et al.*, 2011), (Armstrong, 2002), (Grübler, 2003) montre l'avancement méthodologique pour la prévision technologique pour les niveaux socio-économiques et technologiques, plutôt que pour le niveau technologique lui-même. Les principales améliorations enregistrées par les méthodes de prévision des changements technologiques sont liées à la combinaison des techniques et des modèles existants et l'affinement des méthodes existantes (Glen, Gordon, 2003), (Miranda, Lima, 2010). La croissance exponentielle des publications, des institutions et des chercheurs impliqués dans la prévision et le développement des méthodes de prévision technologique lors des cinq dernières décennies a produit une multitude de techniques et de méthodologies (Slupinski, 2013).

Après une revue de la pratique actuelle des méthodes de prévision technologique à long terme les méthodes de prévision peuvent être classées en quatre groupes (Cascini *et al.*, 2015) :

- (1) les modèles phénoménologiques (par exemple, des extrapolations de données de séries chronologiques, des régressions),
- (2) les modèles intuitifs (par exemple, méthode Delphi, interviews structurées et non structurées),
- (3) la surveillance et la cartographie (par exemple, revues de la littérature et des sources publiées, les scénarios, la cartographie de l'information existante),
- (4) les modèles causaux (par exemple, l'analyse de l'analogie, l'analyse morphologique, les lois et les modèles d'évolution de système).

Dans le chapitre suivant la méthode sélectionnée basée sur les modèles causaux est décrite.

3 METHODOLOGIE DE RECHERCHE

La méthode de recherche intervention choisie s'appuie sur la méthode « Researching Future » (méthode de prévision technologique) (Kucharavy, De Guio, 2008), (Kucharavy, 2013). Cette méthode combine des cartes de contradictions (Kucharavy *et al.*, 2007) avec des S-courbes des fonctions logistiques (Modis, 1999, 2013). La méthode de recherche effectue le principe de triangulation méthodologique. Dans les sciences gestion, la triangulation est souvent utilisée. Selon Miles et Huberman (2003), Rothbauer et Paulette (2008), «la triangulation est censée confirmer un résultat en montrant que les mesures indépendantes qu'on en a faites vont dans le même sens, ou tout au moins ne le contredisent pas». Denzin (1978) a mis en évidence trois types de triangulation : l'utilisation de sources de données différentes pour le temps, l'espace et les personnes ; l'utilisation de chercheurs différents ; l'utilisation de méthodes différentes en même temps.

La méthode Researching Future utilise les principes des méthodes suivantes :

- le modèle « Opérateur du Système » issu de la Théorie de la Résolution de Problèmes Inventif (Altshuler, 1984),
- le modèle « Contradiction » issu de TRIZ et développé pour la modélisation des réseaux des problèmes au sein de OTSM-TRIZ (Khomeenko, 2010),
- les S-courbes (Modis, 1999, 2013),
- les modèles de substitution technologique on utilise pour mesurer le temps et la capacité des processus d'évolution et de substitution technologique (Marchetti, Nakicenovic, 1979), (Modis, 2013),
- les schémas d'interprétation des connaissances issus du modèle DITEK (Grundstein, 2011).

Le langage de représentation de la contradiction est illustré par la Figure 1a.

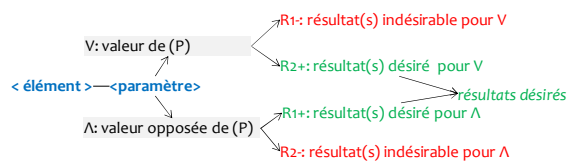


Figure 1a : Modèle pour décrire une contradiction

La sémantique utilisée est la suivante :

- *Élément* – chacune des choses (matériel ou immatériel) qui entre dans une combinaison pour former un tout ou un ensemble,
- *Système* – est un groupe d'éléments interagissant, liés et interdépendants effectuant une certaine fonction,

- *Super-système* – est un groupe d'éléments interagissant où le système défini est un composant ; c'est-à-dire que le super-système ne peut pas fonctionner sans système,
- *Tendance* – est un changement de caractéristique d'un système sur une période longue,
- *Force motrice* – est une ressource qui favorise une tendance,
- *Barrière* – est une limitation de ressources (par ex., l'espace, temps, énergie),
- *Questionnement* – formulation de problème basé sur « système + tendance + barrière » avant la formulation des contradictions ; c'est une première expression d'une contradiction,
- *Caractéristique* – est une propriété qui constitue un caractère spécifique, distinctif de quelque chose, ce qui sert à caractériser quelque chose (par ex., performances d'un appareil, d'une machine),
- *Paramètre* – est une caractéristique d'un élément sur laquelle on peut agir,
- *Indicateur* – est une conséquence d'action due à la valeur d'un paramètre (par ex. résultat désiré) ; chaque indicateur a une valeur mesurable,
- *Unité de mesure* – est un étalon nécessaire pour la mesure d'un indicateur,
- *Résultat* – est la valeur de l'indicateur,
- *Famille d'indicateurs* – est un rassemblement d'indicateurs dans un groupe généralisé.

La syntaxe utilisée est la suivante :

Contradiction – est une modèle pour décrire un problème via la description d'un conflit d'intérêts. Le modèle de contradiction inclut un élément, un paramètre, une valeur d'un paramètre et une valeur opposée d'un paramètre, les résultats désirés (R1+, R2+) et les résultats indésirables (R1-, R2-). La contradiction apparaît dans le cas suivant : l'évolution de la valeur d'un paramètre dans le sens de R2+ désiré, conduit à R1- indésirable, et, quand l'évolution de la valeur opposée d'un paramètre dans le sens de R1+ désiré, conduit à R2- indésirable.

Un exemple de contradiction est décrit Figure 1b et correspondant au questionnement suivant :

Dans le *système* <centre de distribution> la *tendance* <hétérogénéité de flux importante> se heurte à une *barrière* <rentabilité de stockage>.

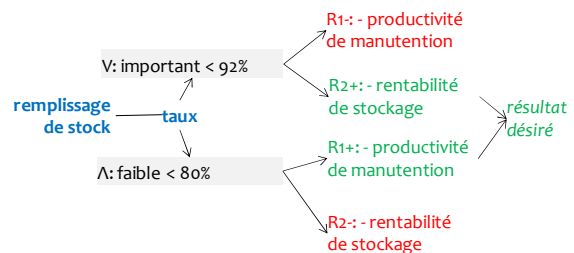


Figure 1b : Un exemple de contradiction

4 CAS D'ÉTUDE

Une recherche intervention est conduite chez un 3PL français. Son chiffre d'affaires en 2016 est de 2,045 M d'euros. Le périmètre d'étude est constitué des 25 entrepôts. Les clients sont principalement issus des secteurs de l'agro-alimentaire, de la grande distribution et de la santé.

Un groupe de travail a été créé. Le groupe est constitué de 6 personnes, respectivement 3 experts 3PL et 3 chercheurs. Le groupe a mené 26 réunions de 4h de travail réparties sur 12 mois. La méthode de gestion de projet est initiée par les étapes génériques :

- identification des besoins,
- spécification du périmètre d'étude,
- identification des acteurs,
- consolidation des questions expertes principales,
- définition de la précision de l'analyse,
- consolidation de l'expression des résultats attendus.

Pour définir le périmètre du système d'entrepasage (SE) la fonction principale est formulée et précisée : « Mettre à disposition des clients des produits en quantité et délai souhaité ».

Afin de saisir les changements importants dans la conception et le processus de l'entrepasage, nous avons collecté des données chronologiques sur la surface des entrepôts (m²) du PL. La surface construite ainsi que les valeurs prévues sont des informations confidentielles pour le PL, par conséquent, aucun chiffre réel ne sera présenté. Les résultats partiels d'étude sont présentés sur les figures 2 et 3. Les S-courbes obtenues positionnent le SE actuel sur son cycle d'évolution. Les S-courbes sont développées sur la base de 85 SE réparties sur 13 pays.

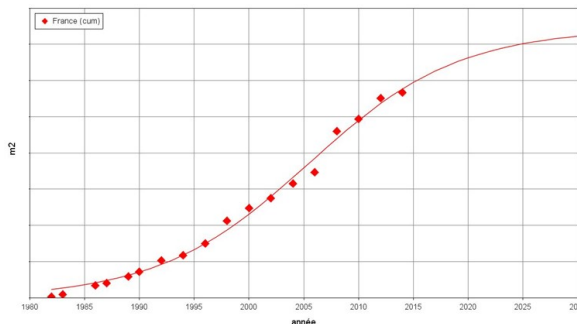


Figure 2 : Exemple d'évolution de la surface des SE construite en France en m²
($T_m=2005.5$; $\Delta t=30.1$; $R_{square}=0.993$)

L'évolution de la surface des SE existant en France arrive en fin de croissance en 2020 (figure 2). Au niveau mondial la saturation est estimée à 2022. Les nouvelles technologies d'entrepasage doivent permettre la croissance du chiffre d'affaires sans augmentation de la surface de SE.

Les caractéristiques principales de ces nouvelles technologies sont justifiées par le réseau de contradiction du SE. Chaque contradiction est élaborée sur la base d'un questionnement, tendance et d'une barrière.

Lors de la formulation des contradictions, certains questionnements ont été exclus de l'étude car caractérisés comme non-pertinents. Deux nouvelles contradictions ont été identifiées. Plusieurs questionnements ont été regroupés et représentés par des mêmes contradictions.

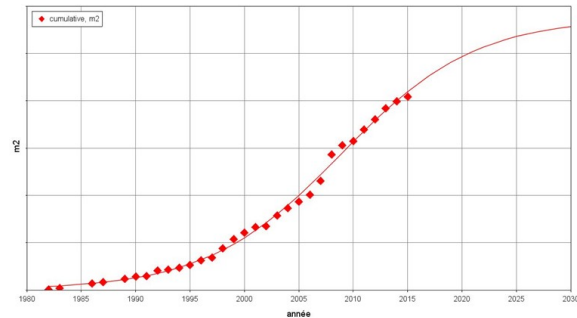


Figure 3 : Evolution de la surface des SE construite dans le monde en m²
($T_m=2008.9$; $\Delta t=27.2$; $R_{square}=0.996$)

Après consolidation, le résultat est caractérisé par : 21 tendances, 48 forces motrices, 49 barrières et 281 résultats désirés. Pour faciliter la connexion des contradictions, les 281 résultats sont regroupés dans six familles d'indicateurs (Tableau 1). La figure 4 présente un fragment d'une cartographie de 58 contradictions résultants.

	Famille	nombre d'indicateurs de performance / famille	nombre de résultats désirés / famille
1	Délai	6	37
2	Coût logistique	35	115
3	Investissements	15	43
4	Méthode d'acheminement	8	33
5	Productivité	16	34
6	Flexibilité	8	19
	TOTAL :	88	281

Tableau 1 : Distribution des indicateurs de performance et des résultats désirés par famille

Pour tendre vers une description définitive des contradictions les plus importantes (du point de vue de l'évolution de SE), il est nécessaire d'étudier les liaisons directes entre les contradictions formulées. Cette activité est en cours d'étude.

La cartographie de contradictions représente les conflits d'intérêts entre sept acteurs principaux (Figure 5). Le nombre de contradictions les plus importantes concerne le prestataire logistique et le producteur industriel qui

sont par conséquent les acteurs principaux de l'évolution.

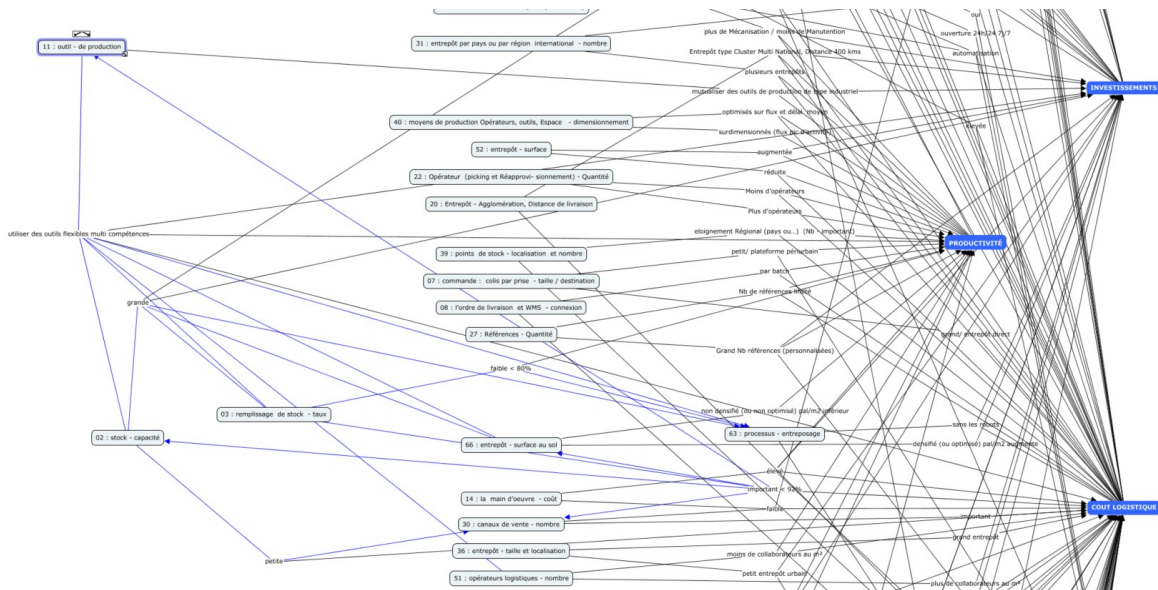


Figure 4 : Fragment du réseau des 58 contradictions avec quelques liaisons directes entre les contradictions

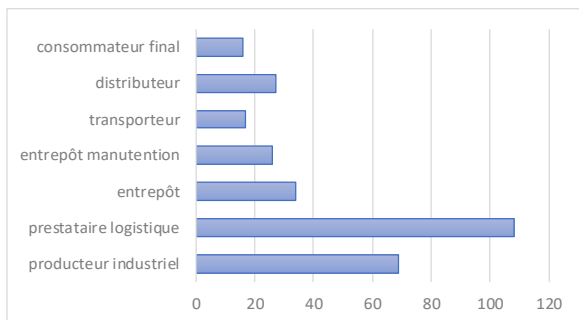


Figure 5 : Le nombre de contradictions classé par acteur

5 DISCUSSIONS ET PERSPECTIVES

Dans ce papier la question traitée est la suivante : « *quels sont les caractéristiques clés pour la prise de décision stratégique en conception d'entrepôt ?* ». Le résultat est la définition d'un réseau de problèmes d'évolutions endogènes et exogènes des systèmes d'entrepôt. Les problèmes sont représentés sous forme de contradictions permettant d'identifier les caractéristiques principales de SE du future. Le résultat obtenu est décrit par une cartographie de 58 contradictions. Ces contradictions sont obtenues en reliant 281 résultats désirés (par ex., localisation de l'entrepôt), 88 indicateurs de performance (par ex. heures de livraison, etc.) classifiés en 6 familles (par ex. coût logistique, etc.). Ces contradictions concernent 7 acteurs (par ex. consommateur final, distributeur, etc.). Le concept de contradiction est pertinent pour initier les problématiques d'évolution stratégiques des SE.

Les résultats obtenus ne permettent qu'une généralisation partielle. En effet le champ d'application est caractérisé par 25 SE des secteurs de l'agro-alimentaire, de la grande distribution et de la santé. Les perspectives de cette étude sont d'une part, la hiérarchisation des contradictions et, d'autre part l'élaboration d'une méthode systématique d'identification des contradictions.

REMERCIEMENTS

Cette étude a été financée par FM Logistic Corporate SAS. Nous remercions les membres de FM Logistic qui ont généreusement consacré leur temps à cette étude.

REFERENCES

Altshuller, G.S. & Williams, A., 1984. *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*, Gordon and Breach Science Publishers.

Andersson, J. & Wemner, T., 2008. *A Strategic Decision-Making Model for Supply Chain*. Lund University.

Armstrong, J.S. ed., 2002. *Principles of Forecasting: A Handbook for Researchers and Practitioner*. 1st ed., Boston / Dordrecht / London: Kluwer Academic Publishers.

- Ayres, R.U., 1969. *Technological Forecasting and Long-Range Planning*. McGraw-Hill book Company.
- Bartholdi III, J.J. & Hackman, S.T., 2016. *Warehouse & Distribution Science*, Atlanta, GA, USA: Georgia Institute of Technology. www.warehouse-science.com
- Cascini, G. et al., 2015. FORMAT - Building an original methodology for Technology Forecasting through researchers exchanges between industry and academia. *Procedia Engineering*, 131, pp.1084–1093.
- Chandler, A., 1962. *Strategy and structure: chapters in the history of the enterprise*, MIT Press
- Chen, I. & Paulraj A., 2004. Towards a theory of supply chain management: the constructs and measurements. *Journal of Operations Management*, Vol. 22, No. 2, pp. 119-150
- Damad, D. & Barth, M., 2016. Graphe de problèmes-solutions de référence pour la conception des activités d'entrepôt. *Logistique & Management*, 24(2), pp.121–133.
- Damad, D., 2000. *Aide au Pilotage de la Performance en Gestion de Production*. l'Université Louis Pasteur de Strasbourg.
- Dawes R., 1988. *Rational choice in an uncertain world*, USA, Harcourt Brace & Company
- De Koster, R.B.M., Johnson, A.L. & Roy, D., 2017. Warehouse design and management. *International Journal of Production Research*, 55(21), pp.6327–6330.
- Denzin, Norman K. (1978): *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: McGraw-Hill
- Fender, M. & Pimor, Y., 2013. *Logistique & Supply Chain*. 6e édition., Paris, France: Dunod.
- Glenn, J.C. & Gordon, T.J., 2003. *Futures Research Methodology*. Version 2.0. , p.700.
- Grant, R. M., 2005. *Contemporary strategy analysis*. (5th Ed.), Singapore, Blackwell Publishing Ltd.
- Grübler, A., 2003. *Technology and Global Change*. Cambridge: International Institute of Applied System Analysis.
- Grundstein, M., 2011. Three Postulates That Change Knowledge Management Paradigm. In H.-T. Hou, ed. *New research on knowledge management models and methods*. InTech, pp. 1–22.
- Harrison, F.E., 1995. *The Managerial Decision-Making Process*. (4th Ed.), USA, Houghton Mifflin Company
- Jantsch, E., 1967. *Technological Forecasting in Perspective*. Paris: OECD.
- Khomenko, N. & Yoon, H., 2010. OTSM-TRIZ as a Response to the Request from the Specialized and Inter-disciplinary Problem Situations. In *Korea TRIZCON 2010*. p. 16.
- Kostoff, R.N., Schaller, R.R., 2001. Science and Technology Roadmaps. *IEEE Transactions on Engineering Management*. 48(2), pp.132-143.
- Kostoff, R.N., 2012. Literature-related discovery and innovation — update. *Technological Forecasting and Social Change*. 79(4), pp.789-800.
- Kucharavy, D., De Guio, R., Gautier, L., Marrony, M., 2007. Problem Mapping for the Assessment of Technological Barriers in the Frame-work of Innovative Design. In *16th International Conference on Engineering Design, ICED'07*. Paris, France: Ecole Centrale Paris.
- Kucharavy, D. & De Guio, R., 2008. Technological Forecasting and Assessment of Barriers for Emerging Technologies. *IAMOT 2008*, p.20.
- Kucharavy, D., 2013. Combination of Contradictions Based approach and Logistic Curves models for Strategic Technological Forecasting. In *Global TRIZ Conference 2013 in Korea*. Seoul, Korea.
- Lee, H., 2004. The triple-A Supply Chain. *Harvard Business Review*, Vol. 82, No. 10, pp. 102-112.
- Linstone, H.A. & Turoff, M., 2011. Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*. 78(9), pp.1712-1719.
- Mantel, S.P., Tatikonda, M. V. & Liao, Y., 2006. A behavioral study of supply manager decision-making: Factors influencing make versus buy evaluation. *Journal of Operations Management*, 24(6), pp.822–838.
- Marchetti, C. & Nakicenovic, N., 1979. *The Dynamics of Energy Systems and the Logistic Substitution Model*. , (RR-79-13), p.73.

- Martino, J.P., 1972. *Technological Forecasting for Decision Making*. 1st ed., New York: Elsevier Publishing Company.
- Mentzer, J.T., DeWitt, W., Keebler, J.S., Min, S., 2001, Defining supply chain management. *Journal of Business Logistics*. Vol. 22, No. 2, pp. 1-26
- Miranda, L.C.M. & Lima, C.A.S., 2010. A new methodology for the logistic analysis of evolutionary S-shaped processes: Application to historical time series and forecasting. *Technological Forecasting and Social Change*. 77(2), pp.175-192
- Modis, T., 1999. A Second Lease on Life for Technological Forecasting. *Technological Forecasting and Social Change*. 62(1), pp.29-32.
- Modis, T., 2013. *Natural Laws in the Service of the Decision Maker: How to Use Science-Based Methodologies to See More Clearly further into the Future, Growth Dynamics*.
- Le Moigne, R., 2017. *Supply Chain Management : Achat, Production, Logistique, Transport, Vente*. 2e ed., Malakoff: Dunod.
- Porter, A.L. et al., 2011. *Forecasting and Management of Technology*. 2nd Revise., John Wiley & Sons Ltd.
- Robbins, Stephen P. (2003), *Organizational Behavior*, (10th Ed.), New Jersey, Prentice Hall
- Robbins S. & Judge T., 2014. *Comportements organisationnels*. (16e édition), Pearson France
- Rothbauer & Paulette, 2008. Triangulation. In Given, Lisa (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods*. Sage Publications. pp. 892-894.
- Roy, B., 1985. *Méthodologie Multicritère d'Aide à la Décision*, Paris, France: Economica.
- Rusjan, B., 2005. Model for manufacturing strategic decision making. *International Journal of Operations & Production Management*, 25(8), pp.740-761.
- Schnetzler, M.J., Sennheiser, A, & Schönsleben, P., 2007. A decomposition-based approach for the development of a supply chain strategy. *International Journal of Production Economics*, Vol. 105, No. 1, pp. 21-42
- Sfez, L., 1988. *La décision*. 2e édition., Paris, France: Presses Universitaires de France.
- Simon H. A., 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, Vol. 69, No. 1, pp.99-118.
- Slupinski, M., 2013. *Technology Forecasting – State of the Art Update / Deliverable 2.3 FORMAT Project*. Milan, Italy. Available at: <http://www.format-project.eu/deliverables/public-reports-and-white-papers/deliverable-2.3/view>
- Stank T., Davis B. & Fugate B., 2005. A strategic framework for supply chain oriented logistics. *Journal of Business Logistics*, Vol. 26, No. 2, pp. 27-46.
- Tseng, S.-C. & Hung, S.-W., 2014. A strategic decision-making model considering the social costs of carbon dioxide emissions for sustainable supply chain management. *Journal of Environmental Management*, 133, pp.315-322.
- Whittington, R., 2001. *What is strategy – and does it matter?* (2nd Ed.), London, UK, Thomson Learning
- Wu, Z. & Pagell, M., 2011. Balancing priorities: Decision-making in sustainable supply chain management. *Journal of Operations Management*, 29(6), pp.577-590.

GESTION DE LA DEMANDE EN PUISSANCE EN MILIEU INDUSTRIEL : ÉTAT DE L'ART ET STRATÉGIES POUR LA PRODUCTION DE BOIS D'OEUVRE

**Laurence B. DUMONT, Nadia LEHOUX, Louis
GOSELIN**

Département de génie mécanique, Université Laval,
Québec, Québec, Canada

Laurence.B-Dumont.1@ulaval.ca,

Nadia.Lehoux@gmc.ulaval.ca,

Louis.Gosselin@gmc.ulaval.ca

Hugues FORTIN

Institut de recherche d'Hydro-Québec (IREQ)

Shawinigan, Québec, Canada

Fortin.Hugues@ireq.ca

RÉSUMÉ : *Une utilisation responsable et pensée de l'énergie permet d'atténuer les impacts environnementaux associés à sa production. Hydro-Québec, société d'État et principal fournisseur d'électricité de la province de Québec, Canada, fait face à un problème récurrent à chaque période de grand froid : une hausse de la demande en puissance. En ce sens, Hydro-Québec a mis en place un programme de gestion de la demande en puissance visant à minimiser l'appel en puissance durant les périodes de pointes. L'industrie du bois d'œuvre est ciblée pour souscrire au programme. Or, avant d'adhérer au programme, les entreprises doivent revoir leurs pratiques pour évaluer si la réduction d'appel de puissance durant des périodes ciblées est possible, sans trop d'impacts négatifs. Dans cet article, une revue de la littérature a permis d'identifier les bonnes pratiques en matière de gestion de l'énergie et les façons de faire pour réduire la demande en puissance au sein des organisations. Diverses stratégies sont adaptées à la production de bois d'œuvre et leurs effets sur certains indicateurs de performance clés du milieu, mesurés. L'une des stratégies considérées repose sur la révision de l'ordonnancement des opérations afin de déplacer la fabrication des produits les plus énergivores hors pointe.*

MOTS-CLES : *Gestion de la demande en puissance, gestion de l'énergie, planification opérationnelle, industrie forestière, industrie du bois d'œuvre.*

1 INTRODUCTION

Le secteur industriel se veut l'un des plus grands consommateurs d'énergie au monde et près de la moitié de l'énergie mondiale est consommée par celui-ci (Liu *et al.*, 2013). Cette consommation significative a un impact direct sur l'environnement puisque les sources d'énergie utilisées ne sont pas toujours renouvelables, comme c'est le cas du charbon ou du pétrole (Andruleit, *et al.*, 2013). C'est pourquoi il devient judicieux de mettre en place divers programmes de réduction de la consommation énergétique afin de réduire l'empreinte écologique.

Au Québec, la consommation énergétique totale du secteur industriel s'élève à 650 PJ (Transition énergétique Québec, 2017), dont 46% de celle-ci est électrique. Cette électricité est produite en grande partie par Hydro-Québec, société d'État québécoise et principal producteur et distributeur d'électricité de la province. Celle-ci fournit à sa clientèle une source d'énergie propre composée à 99 % d'énergie renouvelable grâce, entre autres, à une capacité de production hydroélectrique de 36,9 GW (Hydro Québec, 2016). Néanmoins, les périodes de froid extrême (inférieur à ~ -20°C) auxquelles fait face la province entre les mois de décembre et de mars influencent fortement la consommation des usagers, en particulier celle du secteur résidentiel. Dans le but d'assurer un approvisionnement

sans interruption aux utilisateurs de son réseau, Hydro Québec, a développé un nouveau programme de gestion de la demande en puissance (GDP) offert à la clientèle de moyenne puissance commerciale-institutionnelle et industrielle (CI&I) (Hydro Québec, 2017). Les adhérents à ce programme s'engagent à mettre en place des stratégies permettant de réduire leur appel de puissance durant les périodes de pointes hivernales (décembre à mars), c'est-à-dire lorsque la demande en électricité sur le réseau électrique est exceptionnellement élevée. Les compagnies y adhérant reçoivent un appui financier de 70 Can\$/kW-année proportionnel à la réduction de puissance moyenne atteinte pendant ces périodes (Hydro Québec, 2017). Ainsi, tout au long de l'hiver, Hydro-Québec peut demander aux entreprises faisant partie du programme de diminuer leur appel en puissance, le matin entre 6h et 9h et/ou le soir entre 16h et 20h. Pour les organisations, une telle diminution d'appel en puissance peut souvent impliquer de cesser temporairement la réalisation de certaines activités ou la production de certains produits, ce qui peut alors entraîner un retard de commandes. Il devient donc judicieux d'identifier des stratégies internes permettant de diminuer l'appel en puissance durant ces périodes ciblées, puis d'évaluer l'impact de leur mise en œuvre avant d'adhérer à un tel programme. L'intégration des contraintes énergétiques à l'intérieur même des modèles de planification des opérations est toutefois apparue comme un élément souvent négligé par les auteurs dans la littérature.

Au Québec, il y a plus de 194 usines qui opèrent dans la fabrication du bois d'œuvre (Gouvernement du Québec, 2017). Cependant, peu d'écrits dans la littérature se sont intéressés à une utilisation plus responsable de l'énergie dans le secteur de fabrication de bois d'œuvre, malgré le fait que cette industrie est un secteur économique important au Québec (Transition énergétique Québec, 2017).

Cet article s'intéresse à la gestion de l'énergie et plus particulièrement aux bonnes pratiques à mettre de l'avant dans le secteur de la production de bois d'œuvre de façon à utiliser l'énergie de façon durable. Pour ce faire, une revue de la littérature est d'abord présentée à la section 2, afin de dresser un portrait des mécanismes de gestion de l'énergie dans le milieu industriel de même que des stratégies de réduction de l'appel en puissance associées aux activités industrielles. La méthodologie de recherche et le contexte à l'étude sont par la suite présentés à la section 3. Cette section décrit plus précisément le contexte à l'étude, soit le secteur de la production de bois d'œuvre, ainsi que les étapes suivies permettant l'adaptation des stratégies de réduction de la consommation énergétique explicitées à la section 4. Finalement, une brève conclusion du travail effectué est présentée à la section 5.

2 REVUE DE LITTÉRATURE

La revue de littérature présentée dans cette section porte principalement sur la gestion de l'énergie dans le secteur industriel. Pour y arriver, les recherches ont été principalement effectuées à partir de la base de données *Engineering Village* à l'aide de mots clés liés au bois d'œuvre, au secteur industriel, à la gestion de l'énergie dans ces deux secteurs, aux stratégies de réduction de la consommation énergétique et aux méthodes pour recueillir les données d'énergie. Puisque le domaine de la gestion de l'énergie a fait l'objet de plusieurs avancées dans les dernières années, la recherche a ciblé des articles publiés de 2010 à 2017. Toutefois, les articles pertinents cités par ceux dégagés de la recherche et datant d'années précédentes ont également été pris en considération. *Engineering Village* a de plus été la base de données retenue, puisqu'elle permet l'accès à de la recherche scientifique de haut niveau dans le domaine de l'ingénierie.

2.1 La gestion de l'énergie dans le secteur industriel

Comme le secteur industriel s'avère l'un des plus importants consommateurs mondiaux d'énergie (Liu *et al.*, 2013), il semble alors une cible de réduction de choix dans un processus de gestion durable de la ressource. La norme ISO50001, qui concerne les systèmes de gestion de l'énergie, propose aux compagnies d'élaborer une politique d'utilisation plus efficace de l'énergie, d'établir des cibles et des objectifs pour mettre en œuvre la politique, de regrouper toutes les données nécessaires afin de bien connaître l'usage et la consommation énergétique du système de production considéré, d'évaluer les résultats obtenus, d'examiner l'efficacité de la politique, puis d'améliorer la gestion de l'énergie en continu. Afin de mettre en œuvre

un système de gestion de l'énergie, cette même norme offre aux entreprises un cadre qui guide les diverses actions à mettre de l'avant. Aux États-Unis, les entreprises ayant franchi le pas ont réalisé jusqu'à 30% d'économies sur une période de trois ans (U.S. Department of energy, 2017).

Dans le même ordre d'idées, (BizEE Software Limited, 2017) propose quatre étapes fondamentales à suivre pour toute entreprise désirant mieux gérer sa consommation énergétique. La Figure 1 illustre ces étapes et montre comment le processus s'effectue de manière continue. De plus, chacune de ces étapes est présentée en détail dans cette section.



Figure 1: Étapes pour améliorer la gestion de la consommation énergétique

2.1.1 Mesurer la consommation énergétique et collecter les données énergétiques

La mesure et la collecte de données concernant la consommation énergétique d'une entreprise manufacturière s'avère le premier pas à effectuer pour mettre en place une gestion responsable de l'énergie. (Thiede *et al.*, 2012) proposent ainsi quatre étapes à suivre afin d'estimer la consommation énergétique d'une organisation. Les deux premières étapes consistent à générer une liste de tous les équipements contenus dans l'organisation (étape 1) et de déterminer pour chacun leur puissance nominale (étape 2). Il est possible de trouver cette donnée dans tous les livres techniques fournis par le fournisseur de l'équipement considéré. Cette valeur est plus élevée que la puissance consommée, mais elle permet de servir de base lorsqu'il est impossible de faire appel à des appareils de mesure plus précis. L'étape 3 nécessite de déterminer le temps de fonctionnement de chaque équipement tout dépendant du produit fabriqué. Une fois ces données recueillies, la consommation énergétique (W) de chacun des équipements pour chacun des produits fabriqués peut être évaluée grâce à l'équation (1) :

$$W = FC * P * t \quad (1)$$

où FC est un facteur de charge, P est la puissance nominale de la machine (kW) et t est le temps de fonctionnement de la machine (h). L'étape 4 repose sur la création d'un portfolio permettant de classer chaque équipement de l'entreprise selon sa demande en puissance et son temps d'opération. Cette action permet de distinguer quels équipements ont le plus d'influence sur la consommation énergétique et de porter une attention plus particulière sur

ceux-ci. À noter qu’il faut constamment tenir à jour la liste des équipements afin d’augmenter la fiabilité des données acquises. Le Tableau 1 montre les étapes en question.

Étape	Description
1	Faire une liste complète des machines
2	Déterminer la puissance nominale de chaque machine
3	Déterminer le temps d’utilisation des machines pour la production de chacun des produits permettant de déterminer la consommation énergétique.
4	Création d’un portefeuille

Tableau 1: Étapes à suivre pour déterminer la puissance requise par chacune des machines de l’entreprise selon (Thiede *et al.*, 2012)

(Swat *et al.*, 2013) proposent quant à eux un processus de collecte de l’information en trois phases. La première phase fait référence au choix de la méthode de mesure. Cela implique de déterminer quels appareils de mesure doivent être utilisés en tenant compte de leur coût, des interfaces de communication disponibles ainsi que de leur précision. Les données d’utilisation de l’énergie sont normalement récoltées par des dispositifs de mesure, tels que des compteurs d’énergie, sous la forme de lectures étendues de la consommation moyenne par unité de temps pour une période donnée. (Marzouk *et al.*, 2015) proposent l’utilisation du dispositif de mesure d’énergie *Power-Visa* qui fournit des données de consommation moyennes, seconde par seconde. (De Carvalho et De Oliveira Gomes, 2015) proposent quant à eux l’utilisation d’un analyseur de puissance portatif qui permet également la prise de mesure en puissance des équipements. La deuxième phase consiste à déterminer ce qui doit être mesuré, où et quand. Elle consiste également à déterminer la durée de la période de mesure et le nombre de répétitions requises. Finalement, la phase d’acquisition de données est réalisée et elle implique de noter l’état de fonctionnement de l’équipement, le numéro de l’équipement, le produit fabriqué durant la mesure, l’ensemble des paramètres choisis ainsi que l’horodatage. Suite à l’acquisition de données, les auteurs proposent une stratégie d’évaluation. Celle-ci fournit des instructions sur la manière dont les données capturées doivent être traitées. L’idée consiste à tracer un profil de consommation selon les intervalles de données recueillies, à déterminer quels indicateurs de performance peuvent être calculés à partir des profils de consommation établis, puis à faire le suivi des indicateurs choisis au fil du temps. Le Tableau 2 résume la démarche globale à suivre d’après les auteurs.

Phase	Description
1	Déterminer la méthode de mesure
2	Déterminer la procédure de mesure
3	Procéder à l’acquisition et à l’évaluation des données

Tableau 2: Phases pour l’acquisition de données (puissance requise, énergie consommée, etc.) proposées par (Swat *et al.*, 2013)

Ces deux approches, différentes en soi, permettent de collecter des données reliées à l’utilisation de l’énergie en s’adaptant aux conditions spécifiques de l’entreprise (aspect monétaire, ressources disponibles, etc.). Une fois les données recueillies, il est possible dans les deux cas de calculer différents indicateurs de performance qui seront utiles afin de comparer la situation actuelle à celle « améliorée ».

2.1.2 Trouver des solutions pour mieux utiliser l’énergie et estimer la quantité d’énergie que chaque solution pourrait permettre d’économiser

La deuxième étape processus de gestion de l’énergie consiste à trouver des opportunités permettant de minimiser la consommation énergétique. D’après (Thiede *et al.*, 2012), il existe trois façons principales d’y parvenir. La Figure 2 montre sur quoi reposent ces trois avenues.

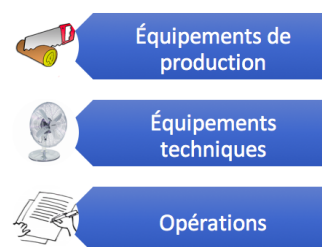


Figure 2: Éléments sur lesquels reposent les principales actions à effectuer afin de minimiser la consommation d’énergie d’un système de production

La première façon consiste à établir un inventaire de tous les équipements de production et de déterminer leur consommation énergétique respective. Suite à cette étape, il est possible de déterminer quels équipements s’avèrent les plus énergivores dans le processus. L’entreprise peut donc décider de modifier la façon de configurer les paramètres de l’équipement afin de voir si un changement peut améliorer la situation actuelle, changer l’équipement ou tout simplement assumer les frais d’inefficacité encourus par l’utilisation de cette ressource.

La deuxième façon consiste à revoir les équipements techniques qui se retrouvent à l’intérieur des bâtiments et qui sont nécessaires, à un moment ou un autre, à la production. On parle ici de systèmes d’éclairage, de systèmes de climatisation et de chauffage, etc. (Partnership, Renewable Energy and Energy Efficiency et UNIDO, 2008) proposent par exemple l’utilisation d’ampoules écoénergétiques, la suppression de postes d’éclairage là où ceux-ci ne sont pas nécessaires au travail ou le changement de commutateurs de courant afin de fermer l’éclairage dans certaines zones de travail tout en conservant l’éclairage dans les autres secteurs où il y a de l’activité.

Finalement, un dernier élément consiste à revoir la planification et le contrôle des opérations de production. (Rackow *et al.*, 2015) proposent par exemple de décaler le début de certaines tâches, de revoir l’ordonnancement des opérations, de décaler les périodes de pause ou encore

d'ajuster les temps de travail. De telles façons de faire s'avèrent particulièrement profitables lorsque le prix de l'énergie varie en fonction de la période de la journée. Dans le cas d'Hydro-Québec, la structure tarifaire ne varie pas en fonction du temps ou de la période.

2.1.3 Mettre de l'avant les solutions choisies

Une fois les opportunités trouvées, il est primordial de mettre de l'avant les solutions retenues afin de générer des économies d'énergie. De telles solutions prennent la plupart du temps la forme d'un changement dans la manière de pratiquer l'action, d'un changement de paramètres pour un ou des équipements ou d'un remplacement d'équipement. Il est important dans cette partie du travail de sensibiliser les gens aux avantages de la nouvelle pratique ou du nouvel équipement. Cette étape nécessite ainsi un investissement en temps et en argent qui peut devenir substantiel (BizEE Software Limited, 2017).

2.1.4 Suivre les progrès en analysant les données de mesure afin de voir à quel point les efforts d'économie d'énergie ont fonctionné

Finalement, la dernière étape consiste à mesurer la performance des solutions implantées. Il est donc important de continuer à collecter des données et de faire le suivi des indicateurs choisis afin de voir si les changements apportés se sont avérés bénéfiques. Suite à cette étape, le processus de gestion de l'énergie peut recommencer et de nouvelles opportunités peuvent être implantées dans l'entreprise afin de minimiser progressivement la consommation énergétique (BizEE Software Limited, 2017).

2.2 La gestion de l'énergie dans le secteur de la fabrication du bois d'œuvre

Peu de travaux touchant à la gestion énergétique du secteur du bois d'œuvre ont été répertoriés dans la littérature. En Europe toutefois, un regroupement d'institutions de recherche ainsi que d'industriels, (Norwegian Institute of Wood Technology *et al.*, 2015), a réalisé un projet portant justement sur le potentiel d'économie d'énergie dans le milieu du sciage du bois d'œuvre. Le projet en question s'est déroulé sur une période de trois ans et visait l'implantation d'un système de gestion de l'énergie permettant de diminuer la consommation énergétique annuelle des scieries de 1 TWh sur une période de trois ans. Cependant, à la fin du projet, une économie estimée entre 190 GWh et 490 GWh a été remarquée. Le projet visait par ailleurs à développer une méthodologie pour l'installation de compteurs de puissance et de chaleur afin de faciliter le suivi la consommation d'énergie de l'industrie du bois d'œuvre tout en mettant en lumière les meilleures pratiques en matière d'efficacité énergétique au niveau des technologies de production et des processus.

Une tournée des 54 scieries européennes participant au projet a d'abord été effectuée. Plusieurs pratiques intéressantes en sont ressorties. Au niveau des équipements de production, l'utilisation d'entraînements à vitesse variable, d'échangeurs de chaleur sur les compresseurs d'air

et d'unités de récupération de chaleur au niveau de procédés de séchage se sont révélées des technologies à utiliser dans l'optique d'une production plus écoénergétique. Au niveau des équipements techniques, l'utilisation d'ampoules particulières et d'un entraînement à vitesse variable pour l'unité d'aspiration des débris produits lors du sciage sont apparues comme deux moyens de favoriser une utilisation moindre de l'énergie. Finalement, au niveau de la planification et du contrôle des opérations, un suivi mensuel de la consommation énergétique, l'embauche d'un employé dédié à l'énergie et l'arrêt de certains équipements aux heures de pointe en hiver sont quelques exemples de mesures proposées.

La revue de la littérature a contribué à mettre en relief une démarche structurée à suivre afin de mettre en place une gestion durable de l'énergie par le milieu industriel ainsi que des éléments sur lesquels intervenir pour générer des retombées positives. Les prochaines sections porteront plus spécifiquement sur l'adaptation de la pratique concernant la planification et le contrôle des opérations au secteur de production de bois d'œuvre.

3 MÉTHODOLOGIE DE RECHERCHE

Dans cette section, une description des grandes étapes suivies pour mener la recherche à bien, de même qu'une description du contexte à l'étude sont proposées.

3.1 Déroulement de la recherche

La première étape de la recherche a été de réaliser une revue de la littérature (résumée dans la section précédente) portant sur la gestion de l'énergie dans le milieu industriel de même que dans l'industrie du bois d'œuvre (Figure 3).

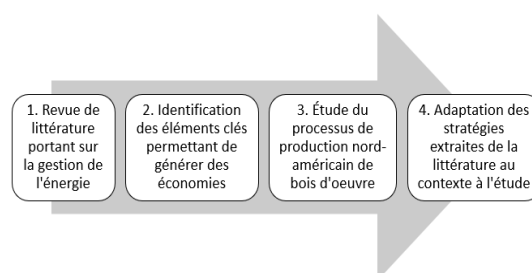


Figure 3: Déroulement de la recherche

Suite à l'élaboration de cette revue, il a été possible d'identifier trois éléments clés sur lesquels agir afin d'économiser de l'énergie dans les usines : les équipements de production, les équipements techniques et la gestion des opérations. De plus, comme les travaux s'intéressent plus particulièrement à l'industrie du bois d'œuvre, certaines recherches ont été effectuées afin de mieux connaître les processus de fabrication et les modèles de planification existants au sein de cette industrie. Au Québec, il existe divers programmes incitant les compagnies à acquérir des équipements de production à haute efficacité énergétique. Cette solution est souvent la plus populaire chez les industriels qui désirent réaliser des économies. Par contre, une fois les modifications apportées

aux équipements, des opportunités au niveau de la gestion des opérations demeurent. C’est pourquoi la recherche s’est principalement intéressée à ce sujet La prochaine sous-section montre une synthèse du contexte étudié.

3.2 Contexte à l’étude

Le processus de production de bois d’œuvre est principalement constitué de trois grandes étapes, soit le sciage, le séchage et la finition. La Figure 4 illustre ces trois étapes.

La première étape consiste à scier les billots arrivant de la forêt selon différentes grandeurs tout en fonction des produits en demande et de leur valeur sur le marché. Le bois scié se dirige ensuite vers le séchage afin de réduire sa teneur en humidité (Marier *et al.*, 2014a). Le séchage est effectué par lots dans des grands séchoirs et le temps de séchage varie généralement entre 16 et 150 heures suivant l’essence séchée, la technologie employée et les périodes de l’année où l’activité est réalisée (Marier *et al.*, 2014b). Les paquets d’un même lot à sécher peuvent être de différentes longueurs, mais ils doivent en général être de même essence et de mêmes dimensions. Une fois le bois séché, celui-ci doit passer par la raboteuse afin d’obtenir l’épaisseur souhaitée et la finition de surface demandée. De plus, la longueur et la largeur du bois peuvent être ajustées durant cette opération (Gaudreault *et al.*, 2009). Le temps de cycle de cette opération peut être déterminé par la vitesse de la raboteuse. Toutefois, ce paramètre joue un grand rôle sur la qualité du bois obtenue. Il arrive parfois que cette opération n’ait pas lieu à la scierie même puisque la capacité de la raboteuse demeure souvent limitée (Marier *et al.*, 2014b). En 2013, la consommation énergétique lors du processus de fabrication se divisait de façon suivante : 55% de bois (sous forme de biomasse pour alimenter en vapeur les séchoirs). 22,6% d’électricité, 17,2% de gaz naturels et 5,1% autres (Gouvernement du Canada, 2017).

3.2.1 Modèles de planification des opérations existants

Le processus de production de bois d’œuvre étant qualifié de divergent, c’est-à-dire qu’un billot mènera systématiquement à la fabrication de plusieurs produits différents, sa planification demeure complexe. C’est pourquoi différents auteurs se sont penchés sur cette problématique au cours des dernières années afin de proposer des modèles mathématiques facilitant la prise de décisions lors de l’élaboration des plans de production tactiques et opérationnels. C’est le cas de (Marier *et al.*, 2014a) qui ont développé un modèle de planification tactique permettant d’aborder les règles d’allocation qui définissent quelle unité ou quel groupe d’unités d’une même compagnie est

responsable de l’exécution des différentes activités du réseau et quelles ressources ou quels groupes de ressources seront utilisés. Il précise également les contraintes à respecter en termes de délais de production et de distribution, de dimensionnement des lots et de politiques d’inventaire, tout en tenant compte de la fluctuation des prix de vente du bois d’œuvre sur le marché afin de capturer des opportunités de revenu. Toute entreprise forestière ayant recours à un tel modèle peut donc être en mesure de planifier globalement ses opérations sur une période d’un an tout en ayant une bonne idée de ce qu’il sera possible de réaliser.

Suite à l’étude du processus de fabrication du bois d’œuvre et du modèle de planification tactique développé par (Marier *et al.*, 2014a) pour ce secteur, il est alors devenu possible d’adapter les stratégies identifiées dans la revue de littérature à la situation québécoise à l’étude. De telles stratégies pourront par la suite être incluses dans un modèle de planification tel que celui développé par (Marier *et al.*, 2014a) afin de mieux gérer la consommation énergétique derrière le processus de production de bois d’œuvre, tout en mesurant les effets d’une réduction de l’appel en puissance durant certaines périodes données sur le système.

4 STRATÉGIES DE GESTION DE L’ÉNERGIE POUR LA PRODUCTION DE BOIS D’OEUVRE ET EFFET SUR LES MESURES DE PERFORMANCE

La littérature démontre la pertinence d’apporter des changements au niveau de la planification et du contrôle des opérations afin de mieux gérer la consommation énergétique. Cette pratique de gestion peut engendrer des économies non négligeables lorsque jumelée à la participation à des programmes de GDP tel que celui proposé par Hydro-Québec. Dans cette section, des stratégies proposées dans la revue de littérature sont adaptées au contexte du bois d’œuvre afin de faciliter la gestion de l’énergie dans ce secteur d’activités.

4.1 Stratégies de réduction de la consommation énergétique et de l’appel en puissance lors d’un événement de GDP

Une première stratégie est de revoir l’ordonnancement des opérations. Lorsqu’on effectue la planification de la

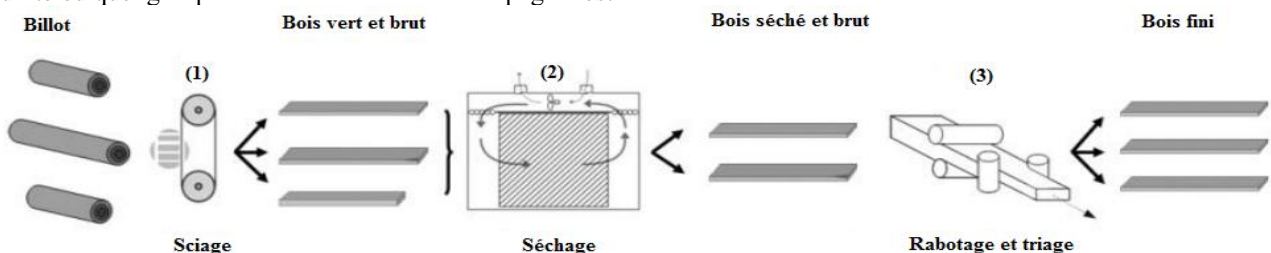


Figure 4: Étapes du processus de transformation du bois traduit de (Marier *et al.*, 2014a)

production de bois d'œuvre au niveau tactique, l'ordonnement des tâches se traduit principalement par le choix du type de produit à fabriquer selon les périodes de l'année. Comme il existe une multitude de produits se différenciant par leur essence et leur grade, il est pertinent de regarder quels sont les produits qui entraînent une plus grande consommation d'énergie et lesquels ont un appel en puissance plus élevé lors de leur production. On cherchera à obtenir un ordonnancement minimisant l'appel en puissance en déplaçant par exemple à plus tard la production d'un certain panier de produits si son appel en puissance pour une période donnée s'avère trop élevé. Un tel travail nécessite de recueillir préalablement des données concernant la consommation énergétique des processus selon les méthodes de (Thiede *et al.*, 2012) ou de (Swat *et al.*, 2013). Cette stratégie pourrait conduire à retirer certains produits de la liste de choix possible à affecter aux périodes de production.

La deuxième stratégie concerne le décalage des intervalles de pause de production. On s'intéresse principalement aux moments où la production de la scierie sera en arrêt complet (à des fins de maintenance, en raison des horaires de travail, durant les vacances, etc.) ainsi qu'aux moments où il y aura des pauses durant les journées (pauses des employés ou maintenance ciblée). Les arrêts complets seraient optimalement planifiés durant les périodes de l'année les plus énergivores. Au niveau des pauses de production durant la journée, elles peuvent par exemple être tenues en alternance par les employés afin d'éviter un arrêt complet de la ligne de production, ce qui causerait une surconsommation énergétique lors de la reprise des opérations. Cette stratégie aurait un effet sur les temps de production disponibles par période.

La troisième stratégie concerne le réarrangement de l'horaire de travail. En se fiant aux conditions météorologiques des années passées, il est possible de voir que les périodes les plus froides de l'année, au Québec, se situent entre les mois de décembre et de mars. Durant ces intervalles, il devient alors possible d'ajuster les horaires de travail de manière à ce que la production soit minimale ou arrêtée le matin et le soir, soit les périodes convoitées par le programme de gestion de la demande d'Hydro Québec. Dans le même ordre d'idées, il devient judicieux d'arrêter seulement une partie de la production, soit au sciage et/ou au rabotage. Une telle stratégie aurait encore une fois un effet sur les temps de production disponibles par période pour chaque activité de production.

Un dernier groupe de stratégies (Fortin & Richard, 2017) montre un potentiel intéressant sans toutefois avoir un impact direct sur la planification de la production. Celui-ci concerne l'arrêt d'équipements non essentiels en périodes de grand froid, l'utilisation d'une source d'énergie de substitution ainsi qu'une stratégie de préchauffage des espaces de production avant la tenue d'événements de GDP.

Les stratégies proposées dans cette section peuvent être appliquées directement, sans avoir recours à un modèle

d'optimisation, mais il demeure judicieux de mesurer préalablement leurs effets à l'aide d'un modèle dans le but de minimiser leur impact sur la production et la satisfaction de la demande.

4.2 L'influence des stratégies sur quelques indicateurs de performance typiques d'une usine de fabrication du bois d'œuvre

Lorsqu'une stratégie de gestion de l'énergie est implantée dans une organisation, il demeure primordial d'en faire le suivi de manière à bien mesurer ses effets. Un bon moyen d'y arriver consiste à regarder les effets qu'elle peut avoir sur les différents indicateurs de performance de l'organisation. Un indicateur de performance est un outil de mesure qui permet de suivre la réalisation d'un objectif, la mise en place de stratégies ou l'accomplissement d'un projet (Gouvernement du Québec, 2012). Chaque scierie peut donc avoir ses propres indicateurs de performance qui sont souvent en lien avec des aspects critiques de son système de production.

Il devient donc intéressant d'évaluer quels pourraient être les effets de l'implantation des stratégies de gestion de l'énergie proposées précédemment sur quelques indicateurs de performance utilisés dans le milieu du bois d'œuvre. Un tel travail est présenté dans le Tableau 3. Quatre indicateurs de performance ont été sélectionnés parmi ceux utilisés dans les scieries, soit le taux de production, le taux de qualité, le budget de production et la satisfaction des clients. Le taux de production concerne le nombre de produits fabriqués par jour. Le taux de qualité concerne le pourcentage de produits fabriqués respectant les normes de qualité sur le nombre de produits totaux fabriqués. Le budget de production se traduit par la valeur, selon la période et le marché, que vaudrait la production de la journée si elle était vendue immédiatement. Finalement, la satisfaction du client concerne la qualité des produits reçus, les commandes remises en temps, etc. Le Tableau 3 présente l'impact des trois principales stratégies proposées à la section précédente sur ces indicateurs de performance, la première consistant à revoir l'ordonnement des opérations, la seconde à décaler les intervalles de pause et la dernière à réarranger les horaires de travail.

	Ordonnement	Décalage des pauses	Horaires de travail
Satisfaction du client	↓	↓	↓
Taux de qualité		↑	
Taux de production	↑	↓	↓
Budget de production	↑	↓	↓

Tableau 3: Impacts significatifs des stratégies sur des indicateurs de performance typiques
Si on examine tout d'abord la première stratégie, on peut anticiper un impact sur le budget de production et sur le

taux de production puisque l'on enlève ou remet à plus tard la possibilité de fabriquer certains produits. Ces impacts peuvent s'avérer positifs ou négatifs tout dépendant du panier de produits concernés. Elle peut également avoir un effet négatif sur la satisfaction des clients dans l'éventualité où la scierie décide de ne pas fabriquer certains produits même s'ils apparaissent dans le carnet de commandes.

La deuxième stratégie peut quant à elle diminuer le taux de production puisqu'à certains moments durant la journée, le nombre d'employés peut être diminué (intervalles de pause alternée entre groupes). Dans le même ordre d'idée, on peut anticiper un impact négatif sur le budget de production et sur la satisfaction du client puisque la production en question sera ralentie et que cela pourrait entraîner des retards de commande. La stratégie peut également influencer positivement le taux de qualité puisque des arrêts de production fréquents ont tendance à augmenter les non-conformités (Ecoinflow SawBenchmark, 2014-2015).

Finalement, la troisième stratégie peut avoir un impact négatif sur le budget de production et sur le taux de production puisque certaines activités seront ralenties voire même arrêtées dans certains secteurs de l'usine. De plus, elle peut avoir une influence négative sur la satisfaction des clients puisque l'arrêt ou la réduction de la production pourrait amener des retards de commande.

Ainsi, il est possible de constater que l'implantation de stratégies de gestion de l'énergie peut avoir une influence non négligeable sur le système de production, et par le fait même dans le calcul des indicateurs de performance des scieries. Il apparaît toutefois pertinent, pour les entreprises, d'ajouter une série d'indicateurs de performance qui permettrait de comparer les stratégies entre elles au niveau énergétique. L'intégration de telles façons de faire à l'intérieur d'un modèle de planification devient donc particulièrement intéressante pour chiffrer les économies effectuées, d'un point de vue financier et d'un point de vue énergétique, et mieux guider le choix d'adhérer ou non à un programme de gestion de GDP.

5 CONCLUSION

Les organisations sont aujourd'hui appelées à utiliser l'énergie de façon plus durable dans une optique de réduction de l'empreinte écologique. La connaissance de leurs différentes charges électriques leur permet la mise en place de stratégies de gestion énergétique. Avec ces informations, ceux-ci peuvent plus facilement évaluer les opportunités qu'offre le programme de gestion de la demande en puissance proposé par la société d'État Hydro-Québec. Les travaux présentés dans cet article ont donc cherché à mettre en relief les bonnes pratiques à adopter par le milieu industriel afin de mieux gérer l'énergie. Une revue de la littérature a été menée, de même qu'une analyse plus approfondie du processus de fabrication de bois d'œuvre. Des stratégies de réduction de la consommation

énergétique et de l'appel en puissance adaptées au contexte de production de bois d'œuvre ont ensuite été proposées, reposant sur le choix des produits à fabriquer durant certaines périodes, les horaires de pauses et d'arrêts de production à mettre de l'avant de même que la possibilité de revoir les horaires de travail afin de minimiser ou d'arrêter la production lors de périodes de forte demande en énergie et en puissance sur le réseau.

Lorsqu'une stratégie est implantée, il est primordial de faire le suivi de ses effets sur le système. Différentes hypothèses ont mis en lumière l'influence que pourrait avoir l'implantation de telles stratégies sur quelques indicateurs de performance typiques d'une scierie, soit le budget de production, la satisfaction du client, le taux de qualité et le taux de production. Ainsi, la révision de l'ordonnement des opérations semble avoir une influence plus marquée sur le budget et le taux de production et la satisfaction du client. Le décalage des pauses de production semble, quant à lui, influencer plus fortement le taux de production et la qualité. Finalement, le réarrangement des horaires de travail semble avoir une influence sur la satisfaction des clients, le budget et le taux de production.

Pour les industriels, l'implantation de la facette énergétique dans leurs processus décisionnels est une avenue à considérer afin de bénéficier d'économies considérables. La littérature a en effet prouvé qu'il serait préférable pour eux d'envisager un meilleur système de gestion de l'énergie à l'intérieur de leurs activités. Pour tout type de scierie confondue, cette avenue peut prendre la forme d'un changement au niveau des équipements de production, de l'enveloppe du bâtiment ou encore de la gestion de opérations. Si cette dernière avenue est retenue, les industriels auront alors à définir, tel que proposé, un ensemble de stratégies.

Pour la suite, une analyse complète faite à partir d'un modèle de planification tactique de la production comme celui proposé par (Marier *et al.* 2014a) sera nécessaire afin de confirmer les hypothèses posées, c'est-à-dire d'implanter les stratégies proposées et d'en ressortir leurs impacts précis sur les différents indicateurs de performance disponibles. Il sera également pertinent de mesurer concrètement l'impact global qu'un événement de gestion de la demande en puissance peut avoir sur le système de production.

RÉFÉRENCES

- Andruleit, H., Bahr, A., Babies, H. G., Franke, D., MeBner, J., Pierau, R., . . . Weihmann, S. (2013). *Energy Study 2013: Reserves, Resources and availability of energy resources*. Récupéré sur https://www.bgr.bund.de/EN/Themen/Energie/Downloads/energiestudie_2013_en.pdf?__blob=publicationFile&v=2
- BizEE Software Limited. (2017). *The What, Why, and How of Energy Management*. Récupéré sur <https://www.biz-ee.com/energy-management>

- <http://www.energylens.com/articles/energy-management>
- De Carvalho, H. M., & De Oliveira Gomes, J. (2015). Energy Efficiency Evaluation for Machining Process in Flexible Manufacturing Systems — A Case Study. *The 22nd CIRP Conference on Life Cycle Engineering*.
- Ecoinflow SawBenchmark. (2014-2015). *Gestion de l'énergie*.
- Fortin, H., & Richard, M.-A. (2017). Gestion de la demande en puissance : Enjeux et opportunités pour la petite et moyenne industrie au Québec. *Présentation au Séminaire conjoint du CIRRELT et FORAC*. Québec.
- Gaudreault, J., Forget, P., Frayret, J.-M., Rousseau, A., & D'amours, S. (2009). *Distributed Operations Planning in the Lumber Supply Chain : Models and Coordination*. CIRRELT.
- Gouvernement du Québec. (2017). *Chiffres-clés du Québec forestier*. Récupéré sur <https://www.mffp.gouv.qc.ca/publications/forests/chiffres-cles.pdf>
- Gouvernement du Canada. (2017). *Tableau 128-0006 : La consommation énergétique de combustibles pour les industries manufacturières, en gigajoules, selon le système de classification des industries de l'Amérique du Nord (SCIAN)*. Récupéré sur Statistique Canada: <http://www5.statcan.gc.ca/cansim/a47>
- Gouvernement du Québec. (2012). *Office québécoise de la langue française*. Récupéré sur Fiche terminologique: http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=504208
- Hydro Québec. (2016). *RAPPORT SUR LE DÉVELOPPEMENT DURABLE 2016*. Récupéré sur <http://www.hydroquebec.com/data/developpement-durable/pdf/rapport-developpement-durable-2016.pdf?v=20170529>
- Hydro Québec. (2017). *Programe Gestion de la demande en puissance*. Récupéré sur hydroquebec: <http://www.hydroquebec.com/pdf/fr/aff-guide-gestion-demande-puissance-mai2017.pdf>
- ISO, 2016. (2016). *ISO 50001 Système de management de l'énergie*. Consulté le décembre 04, 2017
- Liu, Y., Dong, H., Lohse, N., Petrovic, S., & Gindy, N. (2013). An investigation into minimising total energy consumption and total weighted tardiness in job shops. *Journal of Cleaner Production*.
- Marier, P., Bolduc, S., Ben Ali, M., & Gaudreault, J. (2014a). *S&OP Network Model for Commodity Lumber Products*. CIRRELT.
- Marier, P., Gaudreault, J., & Noguer, T. (2015). Planification opérationnelle multi-périodes du séchage du bois d'œuvre avec composition dynamique des patrons de chargement. *11e CONGRES INTERNATIONAL DE GENIE INDUSTRIEL – CIGI2015*. Québec.
- Marier, P., Gaudreault, J., & Robichaud, B. (2014b). Implementing a MIP model to plan and schedule wood finishing operations in a sawmill : lessons learned. *10ème Conférence Francophone de Modélisation, Optimisation et Simulation. MOSIM 2014*. Nancy.
- Marzouk, A. M., ElMaraghy, H. A., & ElMaraghy, W. H. (2015). Effect of changing operating policies on energy use consumption. *48th CIRP Conference on MANUFACTURING SYSTEMS - CIRP CMS 2015*.
- Ministère de l'Énergie et des Ressources naturelles du Québec. (2016). *Consommation*. Récupéré sur Politique énergétique du Gouvernement du Québec: <http://politiqueenergetique.gouv.qc.ca/energie-au-quebec/consommation/>
- Norwegian Institute of Wood Technology; SP; FCBA; Thünen Institute; InnovaWood. (2015). *Energy Management in the European Sawmill Industry*. Ecoinflow.
- Partnership, Renewable Energy and Energy Efficiency; UNIDO. (2008). *africa-toolkit*. Récupéré sur Module 14 - Demand-side management: <http://africa-toolkit.reeep.org/modules/Module14.pdf>
- Rackow, T., Kohl, J., Canzaniello, A., Schuderer, P., & Franke, J. (2015). Energy flexible production : Saving electricity expenditures bu adjusting the production plan. *12th Global Conference on Sustainable Manufacturing*, (pp. 235-240).
- Swat, M., Stock, T., Bähre, D., & Seliger, G. (2013). Monitoring production systems for energy-aware planning and design of process chains. *GCSM*.
- Thiede, S., Bogdanski, G., & Herrmann, C. (2012). A systematic method for increasing the energy and resource efficiency in manufacturing companies. *1st CIRP Global Web Conference: Interdisciplinary Research in Production Engineering*.
- Transition énergétique Québec. (2017). *Fiche diagnostique/enjeux Industrie*. Récupéré sur <http://transitionenergetique.gouv.qc.ca/fileadmin/medias/pdf/consultation/Fiche-diagnostic-Consultation-TEQ-Industrie.pdf>
- Transition énergétique Québec. (2017). *Industrie : État de la situation*. Récupéré sur Transition énergétique Québec: <https://consultation.teq.gouv.qc.ca/project/industries/step/etat-de-situation-6>
- U.S. Department of energy. (2017). *Superior Energy Performance*. Récupéré sur OFFICE of ENERGY EFFICIENCY & RENEWABLE ENERGY: <https://energy.gov/eere/amo/superior-energy-performance>

GENERATION D'UN MODELE DE SITUATION EN TEMPS REEL : APPLICATION A LA GESTION DE CRISE

Audrey FERTIER¹, Anne-Marie BARTHE-DELANOË²,
Aurélie MONTARNAL¹, Sébastien TRUPTIL¹, Frédérick BÉNABEN¹

¹Centre Génie Industriel, Université de Toulouse,
Albi, France
audrey.fertier@mines-albi.fr,
aurelie.montarnal@mines-albi.fr,
sebastien.truptil@mines-albi.fr,
frederick.benaben@mines-albi.fr

²Laboratoire de Génie Chimique, Université de
Toulouse, CNRS, INPT, UPS,
Toulouse, France
annemarie.barthe@ensiacet.fr

RESUME : *Chaque jour, de plus en plus de sources de données sont rendues accessibles, et ce à l'échelle mondiale. Ceci constitue une opportunité sans pareil pour les organisations d'améliorer leur connaissance de la situation (« situational awareness ») dans un environnement de plus en plus instable et imprévisible. Cette connaissance sert de base à la prise de meilleures décisions et ce dans un laps de temps plus court qu'auparavant. Le projet GéNéPi a vu le jour afin d'aider cette prise de décision, dans un contexte multi-acteurs. Il vise à faciliter la collaboration inter-organisationnelle, en temps réel, quelque soit la complexité ou l'instabilité de l'environnement collaboratif. Afin de définir et tester la solution proposée par GéNéPi, un cas d'étude de gestion de crise a été développé en partenariat avec des institutionnels. Cet article présente une nouvelle approche pour automatiser la création et la mise à jour d'un modèle représentatif de la situation, en temps réel, afin de : (i) fournir aux décideurs une vue opérationnelle commune, (ii) établir un cadre de définition et de maintien des processus collaboratifs au cours du temps afin de veiller à la bonne poursuite des objectifs du réseau collaboratif.*

MOTS-CLES : *Modélisation, Big Data, Connaissance de la situation, Gestion de crise, Vue opérationnelle commune.*

1 INTRODUCTION

Durant une situation de crise, l'interconnexion des réseaux contraint fortement les différents acteurs en charge de la réponse à la crise (services de secours, institutions, opérateurs de réseau d'électricité, d'eau, de télécommunications, etc.). Afin de rendre la collaboration possible entre ces différents acteurs, des cellules de crise sont établies. Elles opèrent au niveau opérationnel, comme aux niveaux décisionnels. Le but principal de ces cellules inclut le partage d'information et la prise de décision (centralisée). Ceci nous amène aux constats suivants. D'une part, la gestion des multiples informations entrantes et sortantes de la cellule de crise monopolise une grande partie du temps des acteurs de la cellule. Par ailleurs, la nature instable et imprévisible de la situation complique la tâche des décideurs : bien que physiquement éloignés du terrain, ils doivent avoir la meilleure connaissance possible de la situation, en temps réel.

L'accroissement des sources de données disponibles durant la dernière décennie représente une opportunité sans pareil d'obtenir rapidement des informations sur la situation courante. Malheureusement, si les cellules de crise ne veulent pas être submergées par une avalanche de données brutes et d'informations, il leur faut limiter leurs recherches aux capteurs et appareils déjà identifiés

et mis en place au préalable sur le terrain, ou aux rapports internes.

Dans ce contexte, afin de venir en aide aux différentes parties prenantes présentes sur le terrain de la collaboration et afin d'appuyer les décideurs dans leur travail, le projet GéNéPi ambitionne de concevoir, orchestrer et surveiller en temps réel la réponse à une crise (vue comme un ensemble de processus collaboratifs). Un des points innovants présenté par le projet est la génération automatique d'un modèle de la situation collaborative, alimenté via les données brutes émises par le terrain. Ce modèle est créé et mis à jour en deux étapes : (i) un modèle issu de la vérité terrain décrit les faits provenant directement de sources de données connues et fiables ; (ii) de nouvelles informations sont déduites sur la base de cette vérité terrain, en s'appuyant sur une base de connaissances et à l'aide de règles de déduction. Durant ces deux phases, le système instancie un méta-modèle unique, en tant que « spécification explicite d'une abstraction » (Bézivin and Gerbé 2001). Dans le cas du projet GéNéPi, un méta-modèle est utilisé pour décrire la collaboration des nombreuses parties-prenantes pris dans la réponse à une crise. Ceci est un cas complexe de situation collaborative en milieu instable et imprévisible, avec une contrainte temporelle forte.

Ce papier présente une nouvelle méthodologie pour aider les décideurs dans la gestion des données et de l'information, au sein des cellules de crise françaises. Le résultat consiste en un modèle de situation, qui sert d'entrée à : (i) la déduction et la surveillance des processus collaboratifs ; (ii) la génération d'une vue opérationnelle commune, ou COP (Common Operational Picture), mise à disposition pour chaque cellule de crise.

La seconde section de l'article s'intéressera aux opportunités et aux menaces que représente l'avalanche de données disponibles lors d'une situation de crise. La troisième section montrera comment en tirer profit compte tenu des caractéristiques particulières à une situation de crise. La quatrième section s'intéressera aux événements qui seront à prendre en compte durant la réponse à la crise. Enfin, la dernière section présentera deux approches permettant de répondre à la question détaillée dans la troisième section, dont une nouvelle méthodologie, adaptée aux caractéristiques des événements décrits dans la quatrième section.

2 UNE QUANTITE EXCESSIVE DE DONNEES DISPONIBLES A TRAITER EN TEMPS REEL

D'après Ackoff (Ackoff 1989), la donnée représente « les propriétés des objets et des événements », l'information est « contenue dans les descriptions » (Qui ? Quoi ? Où ? Combien ?) et la connaissance est « transmise par les instructions » (Comment ?). Cette section explique d'où proviennent les données disponibles, pourquoi il est nécessaire de les traiter dans leur ensemble, et à quelle vitesse elles doivent être traitées par les cellules de crise.

2.1 Les besoins des décideurs

Les réseaux, qu'ils soient sociaux, de capteurs ou d'organisations, vont croissants et génèrent de plus en plus de données. Ces sources de données sont nombreuses et hétérogènes. Elles introduisent une complexité dans la gestion de crise à plusieurs niveaux.

Si les décideurs veulent tirer parti de cette nouvelle situation, ils devront : localiser les sources de données ; les identifier ; déduire de nouvelles informations décrivant la situation de crise ; résumer et partager leurs informations et leurs connaissances à l'intérieur de la cellule de crise ; etc. Leur objectif est d'obtenir la vision la plus fiable et la plus précise possible de la situation de crise, en tant qu'image opérationnelle commune (i.e. COP, Common Operational Picture), afin de pouvoir prendre des décisions en toute connaissance de cause (Dautun and Lacroix 2013).

2.2 La contrainte temporelle

Une connaissance partagée de la situation (que l'on retrouve nommée « shared situational awareness » dans la littérature) peut être définie comme « the degree to which team members have the same SA on shared SA

requirement » (Endsley and Jones 2001). Cette connaissance (et son partage) doit être mise à jour en temps réel pour s'adapter à l'instabilité intrinsèque de la situation de crise. Pour la générer, les données et les informations doivent être collectées auprès des parties prenantes impliquées dans la réponse à la crise. Malheureusement, il faut un temps important et précieux pour obtenir et diffuser les bonnes données (ou informations) après de la bonne personne –ou du bon appareil, du bon capteur– au bon moment. C'est la raison pour laquelle les décideurs choisissent volontairement de limiter le nombre de leurs sources de données.

2.3 L'équilibre entre rapidité et qualité

Ces limites volontairement posées permettent aux cellules de crise de gagner du temps dans leur processus de prise de décision, mais elles soulèvent néanmoins plusieurs problèmes :

- **Une perception datée de la situation.** En situation de crise, l'instabilité est la norme, ce qui rend les choses presque impossibles à prévoir. Il est donc essentiel d'être informé à tout moment d'une quelconque évolution de la situation. Les parties prenantes impliquées dans une crise s'attendent à ce que les systèmes respectent cette exigence de temps réel et jugent sévèrement les échecs. Par exemple, l'application pour smartphones SAIP (Système d'Alerte et d'Information des Populations) (Ministère de l'Intérieur 2016a) s'est avérée inutile durant l'attentat terroriste qui a frappé la ville de Nice au soir du 14 juillet 2016. En effet, l'application a relayé l'alerte avec un retard de trois heures et n'a été publiée qu'à 1h34 dans la nuit du 14 au 15 juillet (Ministère de l'Intérieur 2016b).
- **L'erreur de jugement.** Les plans d'urgence des cellules de crise françaises comprennent des réunions planifiées régulièrement. L'un des objectifs de ces réunions est le partage de l'information pour initier les prises de décisions régulières. Néanmoins, l'ordre d'arrivée des données peut impacter le processus de prise de décision. Par exemple, les observations de : (a) un mouvement de foule ; (b) quelques personnes blessées ; (c) des bruits soudains et forts ; peuvent être interprétés de plusieurs façons suivant la séquence de ces observations. Dans la séquence (a, b, c), la situation peut être comprise comme un mouvement de foule classique. Dans le cas de la séquence (c, b, a), la situation peut être interprétée comme un mouvement de foule conséquent à des coups de feu.
- **La perte d'information.** Certaines données, tels que les signaux faibles, n'attirent pas l'attention par nature. D'autres données quant à elles ne correspondront pas (en terme de format et ou de sémantique) au système d'information des cellules de crise. De fait, toutes ces données sont ignorées et définitivement perdues.

En résumé, le nombre croissant de données disponible représente une opportunité unique d'améliorer la connaissance de la situation en temps réel, au sein des cel-

lules de crise, tout en facilitant la prise de décision à travers une COP toujours plus précise. Dans le même temps, les acteurs pris dans la réponse à la crise doivent faire face aux problèmes liés aux données elles-mêmes, comme les délais, les erreurs de jugement ou les pertes. Pour les éviter, les systèmes d'informations dédiés à la gestion de crise sont nécessaires pour aider les décideurs en temps réel (Finn et al. 2015) et générer le meilleur modèle possible de la situation afin d'obtenir la COP attendue.

Comment générer automatiquement un modèle de situation décrivant la situation collaborative en cours et comment le mettre à jour en temps réel pour répondre à l'instabilité inhérente à une situation de crise ?

3 LA GENERATION D'UN MODELE DE SITUATION EN TEMPS REEL

L'exploration de données (data mining) peut être considérée comme une réponse pertinente au problème de la génération automatique d'un modèle de situation. Les approches d'exploration de données sont plutôt dédiées aux entreprises qu'aux situations de crise, et il existe de grandes différences entre ces deux champs d'application. Cette section vise à mettre en lumière ces différences et à définir comment l'exploration de données pourrait être utilisée pour gérer la grande quantité de données de crise disponibles, compte tenu des difficultés soulignées dans la section 2.

3.1 Les particularités du traitement de données dans l'industrie

Les techniques de data mining sont fréquemment utilisées par les entreprises pour surveiller les relations clients (CRM), ou pour identifier les meilleures opportunités de marché (BI), avec pour objectif final une aide à la prise de décision. H. Wang et S. Wang (Wang and Wang 2008) soulignent qu'une approche répandue consiste à utiliser le data mining pour relier (i) la collecte de données, l'accès aux données et l'analyse des données et (ii) la création et l'exploitation de la connaissance. Dans ce contexte, trois hypothèses sont faites :

- **Les sources de données sont connues au préalable.** Les experts métier, en connexion avec les data miners, planifient les extractions de données en fonction de leurs objectifs. Ceux-ci étant connus en avance, il est finalement plus aisé de choisir les sources les plus intéressantes et d'analyser les flux de données avec des règles précises.
- **Une portée réduite des décisions.** Dans la plupart des cas, les entreprises ont décomposé leurs fonctions de façon précise. Par conséquent, la portée des actions qui peuvent être prises pour résoudre des problèmes ou atteindre un objectif est restreinte à un périmètre étroit.
- **Une grande fenêtre temporelle.** L'extraction des données peut être effectuée sur une période de temps

importante afin d'améliorer l'exactitude et la pertinence des résultats.

3.2 Les particularités du traitement de données pendant la réponse à la crise

Les techniques de data mining dépendent de leur champ d'application. Par exemple, elles sont dirigées par les objectifs dans l'industrie, alors que dans le cadre d'une gestion de crise, elles sont dirigées selon les sources de données disponibles : le nombre et la diversité des sources étudiées tendent à diminuer les manques, les pertes d'information, les malentendus ou les jugements sans discernement. Cela représente un énorme volume qui doit être traité en temps réel, quelle que soit la variété des données, qui peuvent arriver sous forme d'images satellites, de photographies, de rapports, de tweets, d'articles, d'emplacements GPS ou de données capteurs. Si le traitement de ces données était automatisé, alors il serait possible d'inférer de précieuses connaissances pour compléter la COP représentant la situation de crise, c'est-à-dire une carte localisant les personnes, les acteurs de la réponse, les risques, les faits (Norheim-Hagtun and Meier 2010), les sentiments (Caragea et al. 2014) ou les activités en cours de réalisation.

Les situations de crise sont assez éloignées des cas industriels et de leurs hypothèses de travail :

- **Des sources de données floues et inconnues.** Les data miners ne peuvent pas prédire les résultats de leur travail, ni l'emplacement des données selon leur niveau de pertinence ou de précision. Ils ne peuvent pas non plus accéder à des données qui ne sont pas connues au préalable, tout en ayant pour objectif d'en étudier le plus grand nombre (afin de fiabiliser le résultat).
- **Des utilisations imprévisibles.** Une donnée peut faire référence à plusieurs types d'informations, liées ou non aux objectifs du décideur. Par conséquent, les règles de classification doivent être à la fois permissives, pour traiter le plus de données possibles, et précises, pour répondre aux attentes des décideurs.
- **Une évaluation de la performance difficile à prévoir.** Le nombre d'activités disponibles pour répondre à un risque ou pour traiter un effet néfaste augmente avec le nombre d'acteurs dans la cellule de crise (tous niveaux opérationnels et stratégiques confondus). Par ailleurs, il n'est pas possible d'établir des indicateurs de performance en avance de phase, et les résultats du processus de réponse à la crise sont difficilement évaluables sur le court terme.
- **Une fenêtre temporelle étroite.** Un seul événement est suffisant pour déclencher une crise en très peu de temps ou en modifier les caractéristiques majeures. Dans ces conditions, le système d'information doit être capable de détecter de tels événements en quelques minutes.

3.3 Un système informatique à trois niveaux servant de base à la gestion des données en situation de crise

Les particularités de l'industrie et de la gestion de crise révèlent deux différences importantes portant sur: (i) le volume et la qualité des données qui peuvent être extraites en temps réel; (ii) le nombre des actions possibles à entreprendre. C'est pourquoi il est intéressant d'utiliser ces grands flux de données, non seulement pour maintenir un modèle de la situation, mais aussi pour définir et surveiller les actions à entreprendre. Dans cette perspective, un système d'information à trois niveaux est nécessaire:

- **Le niveau « données ».** Les données brutes sont directement extraites de sources de données hétérogènes et variées (telles que Systèmes d'Information Géographiques, capteurs, réseaux sociaux ...). Les données extraites sont hétérogènes, nombreuses, plus ou moins fiables et datées.
- **Le niveau « information ».** Ce niveau correspond à la contextualisation des données précédentes. Concrètement, le niveau d'information vise à fournir un état de la situation de crise en temps réel. La COP ainsi obtenue consiste en une carte où les instances importantes et pertinentes pour la cellule de crise sont localisées. Une fois mis en œuvre, ce niveau pourrait soutenir la prise de décision à l'intérieur de chaque cellule de crise. Un méta-modèle est utilisé pour standardiser et faciliter cette modélisation.
- **Le niveau « connaissance ».** Il s'agit des activités à mener compte tenu de la situation de crise caractérisée par le niveau « information ». Les actions font par exemple partie du processus collaboratif qui est (i) déduit par les partenaires de GÉNéPi et (ii) exécuté par les parties prenantes. Le processus de sélection des activités offre un large éventail de possibilités.

Les sections suivantes se focalisent sur les solutions pour traiter en temps réel les données émises par le théâtre de la crise et obtenir les informations les plus pertinentes afin d'alimenter le modèle de situation.

4 PARTICULARITES DES EVENEMENTS EMIS PAR UNE CRISE

L'événement peut être compris comme "la chose qui est arrivée" (comme défini par (Chandy and Schulte 2009; Etzion and Niblett 2010; Luckham and Schulte 2011)). Par exemple, il peut s'agir d'un tsunami, d'un niveau d'eau atteignant un seuil, de la réception d'un courriel, de la nouvelle version d'une page Web, d'un commentaire sur un statut Facebook, etc. Lorsqu'un événement se produit sur le terrain ou est sur le point de se produire, son abstraction est créée dans un système d'information (Luckham and Schulte 2011) sous la forme d'un objet, d'un message ou d'un tuple. Dans cet article, l'événement sera considéré comme étant l'événement « informatisé » et il contiendra au moins:

- Un identifiant unique,

- L'heure de l'événement terrain (i.e. dans le monde physique),
- Une description de l'événement terrain (données ou informations sur le terrain).

Pour générer automatiquement le modèle de situation représentant la collaboration, le système doit décrire le contexte, les partenaires et la crise elle-même. Le contexte et les partenaires peuvent tous deux être représentés pendant la phase de préparation. C'est pourquoi le système pourrait, dans un premier temps, se concentrer sur la modélisation des risques et des effets de la crise.

4.1 Description générale des événements terrain à observer

Pour pouvoir éviter les erreurs ou les retards dans la génération de la COP à partir du modèle de situation, les événements terrain doivent être observés via l'Internet des Evénements (IoE). Il consiste en quatre classes telles que définies par Van Der Aalst (Van der Aalst 2016) :

L'Internet des Objets. Cette classe intègre tout objet physique connecté au réseau via une connexion Internet ou grâce à des étiquettes d'identification uniques telles que des codes à barres ou des étiquettes RFID. Pendant une crise, les systèmes d'identification, comme le «système d'information Internet sans fil pour la réponse médicale aux catastrophes», sont utilisés pour signaler la position des victimes et le dossier médical pendant le triage (Koenig and Schultz 2010).

L'Internet des Personnes. Ces données sont générées à partir des interactions sociales entre les personnes. Il peut s'agir d'e-mails, de messages envoyés via les réseaux sociaux tels que Facebook, Twitter ou Instagram. Pendant une crise, les gens ont tendance à se concentrer sur les médias sociaux pour communiquer ou pour être informés sur la crise en cours, en raison du principe « 1-to-n »: poster un seul message sur Twitter et n abonnés seront avertis en même temps. Ce crowdsourcing émergent est particulièrement intéressant dans le contexte de la gestion de crise pour soutenir la prise de conscience du contexte (Meier 2015).

L'Internet des Lieux. Cela inclut toutes les données avec une dimension spatiale. Par exemple, les tweets, les publications sur Facebook et même les images peuvent intégrer des attributs de géolocalisation fournis par les smartphones et les appareils photo modernes. Lors d'une crise, un pourcentage significatif (49% selon Kouadio et Douvinet (Kouadio and Douvinet 2015)) des victimes potentielles dispose d'un smartphone et peut être localisé sur une carte. Cette représentation de crise peut donc être utilisée comme une COP pour soutenir la prise de décision locale au sein des cellules de crise.

L'Internet du Contenu. Il représente les informations créées par les humains, comme les pages Web, Wikipedia, YouTube, les flux d'actualités. Pendant une crise, les recommandations officielles pour aider les victimes

pourraient être transmises directement aux personnes sur place qui veulent aider, etc.

Comme l'a souligné van der Aalst (Van der Aalst 2014), les données sur les événements constituent la principale source d'information. Tous ces événements disponibles sont nombreux et les données et informations qu'ils contiennent sont plus ou moins fiables, proviennent de sources variées, de différents types et formats, et sont datées.

4.2 Problèmes dus au volume, à la variété, à la vélocité et à la véracité des données terrain

Les événements disponibles décrivent des données et des informations provenant du champ de la crise, des cellules de crise de niveau supérieur et inférieur ou d'Internet. Cette section décrit les caractéristiques de ces entrées hétérogènes, également appelées *4V du Big Data* :

Le volume. Il se réfère à la quantité de données générées en continu par unité de temps. Le volume est proportionnel à la taille des données (Demchenko et al. 2013) et donc au type de données (Krishnan 2013). Un équilibre doit être trouvé entre l'élargissement de la zone d'influence et l'énergie dépensée à traiter les données à temps. Pourtant, plus il y aura de données, plus la découverte d'informations et de comportements cachés sera importante pour les décideurs (Hashem et al. 2015). À mesure que le temps passe, le volume de données disponibles continue d'augmenter tandis que la capacité disponible pour les traiter reste constante (Fan and Bifet 2013).

La variété. Elle fait référence à la diversité des types de données, comme les vidéos ou les journaux de données (Chalmers et al. 2013; Fan and Bifet 2013; Hashem et al. 2015; Krishnan 2013; Ohlhorst 2012; Power 2014), et les formats de données, en tant qu'ensembles structurés ou non structurés (Demchenko et al. 2013; Hashem et al. 2015). Les problèmes de contournement liés à la variété sont possibles grâce aux métadonnées identifiant ce qui est contenu dans les données réelles (Krishnan 2013), même si beaucoup de détails sont perdus dans ce processus.

La vélocité fait référence à la vitesse nécessaire pour récupérer et traiter les données dans le temps (Krishnan 2013; Power 2014). Pour éviter de perdre des données, qui pourraient être essentielles pour inférer des informations critiques aux décideurs, il est crucial d'améliorer la rapidité de toutes les étapes de l'analyse (Kaisler et al. 2013). La vélocité fait également référence à la vitesse à laquelle les données sont générées (Demchenko et al. 2013; Power 2014).

La véracité se réfère à la fiabilité, l'exactitude, la cohérence et la sécurité des données (Demchenko et al. 2013; Gantz and Reinsel 2011). Lukoianova et Rubin (Lukoianova and Rubin 2014) identifient également l'objectivité, la véracité et la crédibilité comme trois dimensions com-

plémentaires de la véracité. (Leskovec et al. 2014) vont plus loin en affirmant que, compte tenu de l'évolution permanente des situations étudiées tout domaine considéré, le processus d'agrégation de données hétérogènes doit être réalisé en temps réel afin de pouvoir fournir une vue pertinente de la situation.

5 APPROCHES DE GENERATION DE MODELE DE SITUATION A PARTIR DE DONNEES BRUTES

D'après les caractéristiques énoncées ci-dessus, un bon outil d'analyse doit accéder à autant d'événements que possible (volume, vitesse), à partir de sources hétérogènes (variété). Il doit également calculer la véracité et la valeur de toutes les données à l'intérieur du système. Des méthodologies couvrant au moins un de ces points existent déjà. Cette section en présente deux.

5.1 Les approches générales traitant a minima un des 4V

KDD (Knowledge discovery in databases) par (Fayyad et al. 1996). Le processus de découverte de connaissances dans des bases de données structurées a été mis en place pour cartographier des données de bas niveau, trop volumineuses pour être facilement agrégées. Le processus utilise des modèles pour lier des données ensemble. Ces modèles doivent être validés (véracité) et potentiellement utiles (valeur). Ce processus tient en cinq étapes : sélection ; pré-traitement ; transformation ; data mining ; schémas d'interprétation et évaluation.

SA (Situational Awareness) par (Endsley 1995). La conscience de la situation d'un système est mesurée par sa capacité à percevoir des éléments dans un environnement particulier, à les comprendre et à les projeter dans le futur. L'ensemble du processus est composé de trois étapes principales (Fayyad et al. 1996):

- Perception des éléments: observer et récupérer le statut, les attributs et la dynamique des éléments pertinents sur le terrain.
- Compréhension de la situation actuelle: recueillir des éléments récupérés et déduire de nouveaux éléments qui véhiculent de plus en plus de significations pour les utilisateurs et leurs objectifs.
- Projection du statut futur: prévoir l'évolution du statut des éléments pour projeter sa future dynamique dans l'environnement.

L'objectif est de fournir les connaissances requises par les décideurs sous contrainte de temps (vélocité), sans perdre de vue de leurs objectifs (valeur).

Ces deux méthodologies couvrent partiellement les caractéristiques des données présentées précédemment (4 V du Big Data).

5.2 Une nouvelle approche prenant en compte les 4 V du Big Data

Le projet GÉNéPi intègre nativement la gestion des données massives au système d'aide à la collaboration. Les événements sont directement analysés pour générer des modèles à jour, durant l'exécution des processus collaboratifs. Dans le cas d'une situation de crise, ils peuvent faciliter la déduction et l'orchestration de la réponse à la crise mais aussi fournir une COP à chaque cellule de crise.

Du point de vue de l'intelligence artificielle, le système GÉNéPi peut être vu comme un agent (Russell and Norvig 2009): il peut percevoir son environnement, à

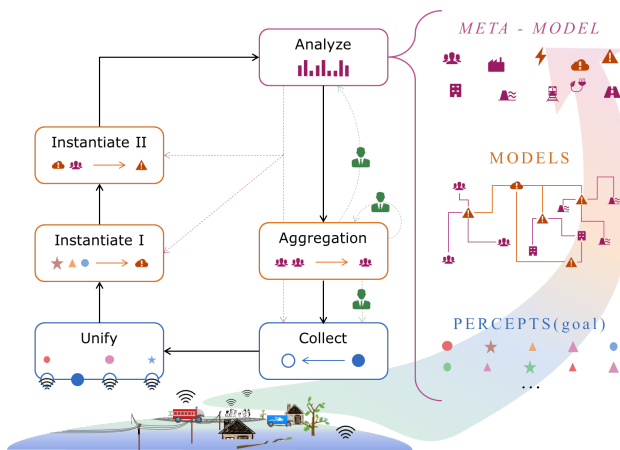


Figure 1. Les modules composant le système d'information à trois niveaux de GÉNéPi.

travers la réception des percepts, et agir sur l'environnement, à travers l'orchestration d'actions coordonnées. Dans le cadre du projet GÉNéPi, l'Interpreter GÉNéPi est la brique logicielle responsable de la prise de conscience de la situation.

La structuration des informations. L'Interpreter GÉNéPi illustre la situation de crise en instanciant un méta-modèle (commun à tous les modèles de situation). Le méta-modèle (Bénaben et al. 2016) est composé de deux types de concepts:

- La première partie du méta-modèle peut être instanciée principalement avant la crise, lors de la phase de préparation. Cela concerne les capacités des parties prenantes, désireuses d'être impliquées ou prévues dans une procédure ; et les enjeux dans le domaine. Plus tard, lors d'une réponse à une crise, des partenaires inconnus peuvent émerger et de nouveaux enjeux peuvent être identifiés et le modèle peut être complété en temps réel ;
- La deuxième partie du méta-modèle concerne la crise elle-même : le danger, les risques et les événements affectant le territoire.

L'objectif de l'Interpreter GÉNéPi. Le module Collect (voir Figure 1) détermine régulièrement l'objectif de

l'Interpreter. Il indique quels aspects de l'environnement doivent être considérés. Les percepts - décrivant l'environnement de crise et perçus par le système - sont ainsi toujours adaptés aux priorités actuelles. Par exemple, au début de la réponse, les cellules de crise les plus basses dans la hiérarchie essaient d'abord de détecter les conséquences et les risques liés à la crise. Grâce à elles, les cellules de crise de niveau hiérarchique supérieur pourront évaluer l'ampleur de la crise. Dans ce cas, le module Collect doit prioriser la collecte de données pouvant être utilisées pour instancier la seconde partie du méta-modèle. Cette étape de *perception* est critique (Endsley 1995) compte tenu de la quantité de données à traiter dans le temps.

L'interopérabilité de tous les composants de l'Interpreter. Les données et informations sont reçues par le module Unify. Les informations reçues de sources fiables et connues peuvent être automatiquement ajoutées au modèle de situation tandis que les autres sont redirigées vers le module Instanciation (voir Figure 1). Au sein de l'Interpreter, des problèmes se posent rapidement en raison de la variété des données. C'est l'un des plus grands obstacles à la compréhension d'un grand volume de données (Kaisler et al. 2013). Pour contrer cela, le module Unify transforme automatiquement les données d'entrée en documents XML. Ces nouvelles données sont des événements informatiques. Ils comprennent (* obligatoirement): * une date de création ; * un identifiant ; * un sujet ; une géolocalisation ; et enfin une description de ou des événements observés sur le terrain et rapportée par la donnée ou l'information reçue par le module Unify.

Des événements bruts aux modèles de situation. Le module Instanciate vise à comprendre la signification des événements informatiques: il relie des données disparates pour développer (Endsley 1995) les connaissances que le système a de la situation courante (« *Situation awareness* »). Ce module est divisé en deux parties. Instancier I, reçoit les événements provenant du module Unify et les transforme en instances du méta-modèle. Instanciate II déduit de nouvelles instances directement du modèle de situation. Par exemple, lorsqu'Instanciate I déduit un nouveau danger d'inondation à partir de l'évolution prédite de la rivière, Instanciate II ajoute un nouveau risque pour chaque bâtiment sensible menacé dans la zone de danger.

∀ Enjeu,

$$\text{Risque} = \text{Danger} * \text{Vulnérabilité}(\text{Danger})$$

Un danger menace tous les enjeux qui lui sont vulnérables dans son périmètre. Par exemple, un risque de victime apparaîtra pour chaque foyer situé en zone inondable et un risque de coupure électrique apparaîtra pour chaque transformateur en zone inondable.

L'étape de compréhension. L'un des objectifs du module Analyze est de prévoir le futur, sous la forme d'un nouveau modèle, dérivé du modèle de situation, appelé

modèle projeté. Il est produit en combinant l'avancée du processus de réponse en cours, les progressions naturelles (prévisibles) de la crise et le modèle de situation courant. Par exemple, il peut prendre en compte l'évolution du niveau d'eau et du débit d'eau, prévue par le Service de Prévision des Crues. L'objectif est d'anticiper les divergences entre la situation en cours, représentée par la COP, et la situation qui devrait résulter de la réponse en cours, pour pouvoir adapter au plus tôt la réponse des autorités. Cette étape de *compréhension* permet au système d'atteindre le dernier niveau de conscience (« *Situation awareness* ») décrit par (Endsley 1995).

L'autre objectif du module Analyze est de surveiller la qualité (Wang and Strong 1996) des percepts, les modèles et le méta-modèle manipulés par l'Interpreteur GénÉPi. Pour réussir, le module Analyze cherche à confirmer les résultats, à combler les informations manquantes ou à améliorer la véracité des modèles. Ces itérations permettent également d'alerter l'administrateur lorsqu'une source de données ou une règle d'exploitation doit être remise en question.

La visualisation des informations. La COP illustre à la fois la situation et les modèles projetés. Le but (Endsley 2015) est de communiquer efficacement sur «ce qui est entrain de se passer» et sur «ce qui risque de se produire ensuite», tout en réduisant de manière significative les incertitudes et le stress liés aux prises de décision (Dautun and Lacroix 2013).

Le point de vue utilisateur. Les modèles contiennent tellement d'informations que l'utilisateur doit pouvoir accéder à une granularité et un horizon géographique précis. Par exemple, au lieu du nom, de la fonction et de la géolocalisation de chaque victime potentielle, la cellule de crise de la zone de défense préférerait connaître le nombre total de victimes sur sa zone. C'est pourquoi, le module Aggregate a été configuré pour ajuster, en temps réel, la COP à chaque cellule de crise. Typiquement, pour la crue de la Loire, cela représente neuf points de vue différents (Fertier et al. 2016). Pour réussir, la COP est affichée dans une forme familière aux parties prenantes: une interface de SIG (Système d'Information Géographique) (voir Figure 2).

Le module Aggregate gère également les droits d'accès et les demandes d'informations complémentaires. Par exemple, si une information a été classée au niveau national, une autorisation formelle sera requise de tout maire, préfet ou préfet de zone qui voudrait y accéder.

6 CONCLUSION

Les décideurs impliqués dans la réponse à une crise n'ont pas le temps de prendre en compte de nouvelles sources de données, alors que ce serait l'occasion d'avoir une vue plus complète de la situation courante. Pour les aider, le projet GénÉPi propose à la fois de soutenir la collaboration et la gestion des données dans les cellules de crise. La solution pourrait être tirée du monde de l'industrie, si le monde de la gestion de crise n'avait pas ses propres

hypothèses de travail: sources de données inconnues, utilisation imprévisibles des données, nombre d'actions possibles infini, temps passé à la collecte et l'analyse de données compté. Dans ce contexte, deux approches ont été étudiées et évaluées en considérant le nouveau volume de données à gérer au sein des cellules de crise. La combinaison de ces deux approches a permis la mise en œuvre d'un nouvel outil, ancré : (i) dans la suite d'outil existante IO Suite (Bénaben et al. 2016) qui aide déjà à la collaboration de toutes les parties-prenantes à la réponse à la crise ; et (ii) dans le prototype GénÉPi concurrent à la solution IO Suite. Les différentes règles (métier, transformation de modèle) ont été rédigées avec l'aide du CEREMA, en sa qualité d'expert métier.

A l'avenir, un tel outil, permettant la génération automatique d'un modèle de situation en temps réel, quelle que soit la complexité et l'instabilité du contexte de décision, bénéficiera non seulement aux décideurs pris dans une réponse à la crise, mais aussi :

- Aux industriels désireux de s'adapter rapidement aux variabilités de l'approvisionnement et/ou de la demande, tout en accédant aux informations cachées dans les données qui leurs sont maintenant accessibles en grand nombre (Agostinho et al. 2016) ;
- A toutes les personnes en prise avec des tableaux de bords de plus en plus complexes qui représentent directement la couche « données », et non la couche « information » comme peut le faire l'Interpreteur GénÉPi.

Pour la suite, les utilisations de la solution IO Suite ou du prototype GénÉPi devraient permettre la constitution d'une base de connaissances regroupant les modèles de plusieurs situations de crise, à différent moment de la réponse. Si elle est utilisée, cette base de connaissance permettra à l'Interpreteur GénÉPi d'apprendre à déduire des relations de causes à effets via un raisonnement à partir de cas (« *Case based reasoning* ») (Bennani et al. 2017) ou des règles seuils au bénéfice du module Instantiate I.



Figure 2. Une des interfaces utilisateurs, présentant le modèle de la situation.

REMERCIEMENTS

Les auteurs tiennent à remercier les membres du projet GénÉPi pour leur aide et leurs conseils. Les travaux

présentés dans ce papier sont financés par l'Agence Nationale de la Recherche via le projet GÉNéPi (référence : ANR-14-CE28-0029).

REFERENCES

- Ackoff, R. L. (1989). "From data to wisdom." *Journal of applied systems analysis*, 16(1), 3–9.
- Agostinho, C., Pereira, J., Lorré, J.-P., Ghimire, S., and Benazzouz, Y. (2016). "A Distributed Middleware Solution for Continuous Data Collection in Manufacturing Environments."
- Bénaben, F., Lauras, M., Truptil, S., and Salatgé, N. (2016). "A Metamodel for Knowledge Management in Crisis Management." *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 126–135.
- Bennani, S., Maalel, A., Ghézala, H. B., and Abed, M. (2017). "Towards a Decision Support Model for the Resolution of Episodic Problems Based on Ontology and Case Bases Reasoning: Application to Terrorism Attacks." *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 1502–1509.
- Bézivin, J., and Gerbé, O. (2001). "Towards a Precise Definition of the OMG/MDA Framework." *16th IEEE International Conference Automated Software Engineering (ASE'2001)*, San Diego, USA, United States, 273–280.
- Caragea, C., Squicciarini, A. C., Stehle, S., Neppalli, K., and Tapia, A. H. (2014). "Mapping moods: Geo-mapped sentiment analysis during hurricane sandy." *Proceedings of the ISCRAM 2014 Conference*.
- Chalmers, S., Bothorel, C., and Picot Clemente, R. (2013). *Big Data - State of the Art*.
- Chandy, K., and Schulte, W. (2009). *Event processing: designing IT systems for agile companies*. McGraw-Hill, Inc.
- Dautun, C., and Lacroix, B. (2013). "Placer l'humain au cœur des crises." *Lettre d'information sur les risques et crises INHESJ*, (38), p10–19.
- Demchenko, Y., Membrey, P., Grosso, P., and de Laat, C. (2013). *Addressing big data issues in Scientific Data Infrastructure*. Piscataway, NJIEEE9781467364034.
- Endsley, M. R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors*, 37(1), 32–64.
- Endsley, M. R., and Jones, W. M. (2001). "A model of inter and intra team situation awareness: Implications for design, training and measurement. New trends in cooperative activities: Understanding system dynamics in complex environments." *CA*, 46–67.
- Etzion, O., and Niblett, P. (2010). *Event Processing in Action*. Manning Publications Co., Greenwich, CT, USA.
- Fan, W., and Bifet, A. (2013). "Mining Big Data: Current Status, and Forecast to the Future." *SIGKDD Explor. Newsl.*, 14(2), 1–5.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From data mining to knowledge discovery in databases." *AI magazine*, 17(3), 37.
- Fertier, A., Montarnal, A., Barthe-Delanoë, A.-M., Truptil, S., and Bénaben, F. (2016). "Adoption of big data in crisis management toward a better support in decision-making." *Proceedings of Conference on Information System for Crisis Response And Management (ISCRAM 16)*.
- Finn, R. L., Watson, H., and Wadhwa, K. (2015). "Exploring big 'crisis' data in action: potential positive and negative externalities." *Proceedings of the ISCRAM 2015 Conference*.
- Gantz, J., and Reinsel, D. (2011). "Extracting value from chaos." *IDC iview*, 1142(2011), 1–12.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015). "The rise of 'big data' on cloud computing: Review and open research issues." *Information Systems*, 47, 98–115.
- Kaisler, S., Armour, F., Espinosa, J. A., and Money, W. (2013). "Big Data: Issues and Challenges Moving Forward." *2013 46th Hawaii International Conference on System Sciences*, 995–1004.
- Koenig, K. L., and Schultz, C. H. (2010). *Koenig and Schultz's disaster medicine: comprehensive principles and practices*. Cambridge University Press.
- Kouadio, J., and Douvinet, J. (2015). "Les smartphones peuvent-ils aider à une meilleure remontée des données en cas de crue rapide pour améliorer les systèmes d'alerte?" *BSGLg*, 64, 57–68.
- Krishnan, K. (2013). *Data warehousing in the age of big data*. Newnes.
- Leskovec, J., Rajaraman, A., and Ullman, J. D. (2014). *Mining of massive datasets*. Palo Alto.
- Luckham, D., and Schulte, W. R. (2011). "Event Processing Glossary - Version 2.0." EPTS.
- Lukoianova, T., and Rubin, V. L. (2014). "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?" *Advances in Classification Research Online*, 24(1), 4.
- Meier, P. (2015). *Digital humanitarians: how big data is changing the face of humanitarian response*. Crc Press.
- Ministère de l'Intérieur. (2016a). "Lancement de l'application mobile SAIP." <http://www.interieur.gouv.fr/Archives/Archives-des-actualites/2016-Actualites/Lancement-de-l-application-mobile-SAIP>, <<http://www.interieur.gouv.fr/Archives/Archives-des-actualites/2016-Actualites/Lancement-de-l-application-mobile-SAIP>>.
- Ministère de l'Intérieur. (2016b). "Mise en oeuvre de l'application SAIP le soir du 14 juillet 2016." <http://www.interieur.gouv.fr/Actualites/Communiques/Mise-en-oeuvre-de-l-application-SAIP-le-soir-du-14-juillet-2016>, <<http://www.interieur.gouv.fr/Actualites/Communiques/Mise-en-oeuvre-de-l-application-SAIP-le-soir-du-14-juillet-2016>> (Jul. 21, 2016).
- Norheim-Hagtun, I., and Meier, P. (2010). "Crowdsourcing for crisis mapping in Haiti." *Innovations: Technology, Governance, Globalization*, 5(4), 81–89.
- Ohlhorst, F. J. (2012). *Big data analytics: turning big data into big money*. John Wiley & Sons.
- Power, D. J. (2014). "Using 'Big Data' for analytics and decision support." *Journal of Decision Systems*, 23(2), 222–228.
- Russell, S., and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Pearson, Upper Saddle River.
- Van der Aalst, W. M. (2014). "Data scientist: The engineer of the future." *Enterprise interoperability VI*, Springer, 13–26.
- Van der Aalst, W. M. P. (2016). *Process Mining - Data Science in Action*. Springer.
- Wang, H., and Wang, S. (2008). "A knowledge management approach to data mining process for business intelligence." *Industrial Management & Data Systems*, 108(5), 622–634.
- Wang, R. Y., and Strong, D. M. (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, 12(4), 5–33.

IDENTIFICATION DES RISQUES, OPPORTUNITÉS ET FACTEURS CRITIQUES DE SUCCÈS DE L'INDUSTRIE 4.0 POUR LA PERFORMANCE INDUSTRIELLE DES PME

Alexandre MOEUF

Exxelia Group – 1 rue des temps modernes, 77600,
Chanteloup-En-Brie, France,
alexandre.moeuf@exxelia.com

Adrien GERALDES

Exxelia Group – 1 rue des temps modernes, 77600
Chanteloup-En-Brie, France,
adrien.geraldes@exxelia.com

Robert PELLERIN

Polytechnique Montréal, Département de
mathématiques et de génie industriel – 2900 Boulevard
Edouard-Montpetit, Montréal, QC H3T 1J4, Canada,
robert.pellerin@polymtl.ca

Estefania TOBON-VALENCIA

Exxelia Group – 1 rue des temps modernes, 77600
Chanteloup-En-Brie, France,
estefania.tobon-valencia@exxelia.com

Samir LAMOURI

Arts et Métiers Paristech, LAMIH – 151 Boulevard de
l'Hôpital, 75013, Paris, France,
samir.lamouri@ensam.eu

RESUME : *Le concept d'industrie 4.0 présente de nombreux avantages pour les Petites et Moyennes Entreprises (PME) au regard d'autres méthodes de pilotage telles que le juste-à-temps ou le MRPII. Cependant, son adoption n'est pas triviale. En effet, l'adoption de stratégies alignées sur l'industrie 4.0 nécessite une expertise importante sur des technologies nouvelles. Notre objectif est ici d'identifier les risques, opportunités et facteurs critiques de succès pour l'exploitation de l'industrie 4.0 dans les PME. Lors de travaux précédents, il fut démontré que les cas empiriques d'industrie 4.0 étaient peu documentés dans la littérature scientifique et complexe à identifier dans l'industrie. Ainsi, nous proposons une étude prospective Delphi-Régnier par consultations d'experts. Nous montrons que la formation et l'expertise seront les clés pour l'exploitation de l'industrie 4.0, que le dirigeant a un rôle primordial et que l'industrie 4.0 est une opportunité de repenser les modèles d'affaires en vue de dégager de nouvelles valeurs ajoutées.*

MOTS-CLES : *PME, Industrie 4.0, Delphi-Régnier, Amélioration Industrielle.*

1 INTRODUCTION

En France, la définition des PME (Petites et Moyennes Entreprises) est régie par la Loi de Modernisation de l'Economie du 4 août 2008 (Décret d'application n° 2008-1354, article 51). Elles regroupent des entreprises dont le nombre de salariés est supérieur à 10 et inférieur à 250, dont le chiffre d'affaires annuel est inférieur à 50 millions d'euros ou le total de bilan n'excède pas 43 millions d'euros. Nous exploitons cette définition dans nos travaux.

Acteur majeur de l'industrie, les PME doivent répondre à des exigences clients toujours plus complexes. Plusieurs méthodes existent pour améliorer le pilotage industriel des PME et répondre au plus juste à la demande des clients, c'est le cas du juste-à-temps issu du Toyota Production System (Liker, 2006; Ohno, 1988) par exemple. Ce paradigme se base sur la standardisation de la production, la mise en ligne des flux et une production tirée par la demande client. Bien qu'ayant montré d'excellents résultats opérationnels dans les entreprises

(Shah and Ward, 2003; Womack and Daniel, 2003), cette méthode se révèle difficile à implanter dans les PME par manque d'expertise et de leadership (Achanga et al., 2006; Moeuf et al., 2016).

De même, les méthodes MRP et MRPII, supportés par le déploiement d'outils informatiques tels que les ERP (Deep et al., 2008) sont des méthodes complexes et coûteuses à implanter dans les PME. Elles se révèlent bien souvent trop rigides une fois mises en place (Cullinan et al., 2010).

Plus récemment est né le concept d'industrie 4.0 qui vise le pilotage de la chaîne logistique via la synchronisation en temps réel des flux pour éventuellement tendre à la fabrication de produits unitaires et personnalisés (Kohler and Weisz, 2016). Ce concept se base sur l'émergence de nouvelles technologies telles que l'infonuagique (*Cloud Computing*), l'internet des objets, les systèmes cyber-physiques et l'exploitation des données massives (*Big Data*) notamment. Ces technologies favorisent la transmission de l'information à travers l'ensemble du système piloté; cela permet une rationalisation des processus, mais également une synchronisation en temps réel des

processus face aux fluctuations de la demande. Il semble que ce concept soit moins coûteux et plus flexible que le Lean Manufacturing et les ERP, mais il repose sur des technologies que les PME peinent à exploiter.

Nous avons montré lors de nos travaux de recherche précédents que les cas empiriques de l'industrie 4.0 dans les PME étaient peu documentés au sein de la littérature scientifique (Moeuf et al., 2017b). Notre objectif est d'identifier les risques, opportunités et les facteurs critiques de succès (*CSF : Critical Success Factors*) de l'exploitation de l'industrie 4.0 dans les PME afin d'aider les décideurs et les cadres des entreprises dans leurs démarches.

Liée au faible nombre de cas documentés, une enquête de terrain serait complexe à mettre en œuvre. Les PME ayant mis en place l'industrie 4.0 sont peu nombreuses et complexes à identifier, ainsi nous visons l'atteinte de l'objectif par l'exploitation d'une méthode prospective Delphi, couplée aux abaques de Régnier.

Nous décrivons au sein de la section 2 le concept d'industrie 4.0 plus en détail. Nous décrivons les méthodes Delphi, les abaques de Régnier et l'organisation de l'étude au sein de la section 3. Nous présentons les résultats de l'étude dans la section 4, puis nous précisons les limites et les perspectives de l'étude dans la section 5. Nous concluons notre étude dans la section 6.

2 L'INDUSTRIE 4.0

Des groupes technologiques telles que l'internet des objets (Dijkman et al., 2015; Hao and Helo, 2015), les algorithmes d'analyse de données massives (Babiceanu and Seker, 2016; Chen et al., 2014; Mousannif et al., 2016) ainsi que l'infonuagique et l'amélioration des réseaux (Xu, 2012; Yue et al., 2015) sont récemment entrés dans le monde industriel favorisant l'émergence de nouveaux concepts de pilotage des processus. Le concept industrie 4.0 est né en Allemagne, à la foire à d'Hanovre en 2011 suite à une discussion entre les représentants de l'industrie, de la recherche, des syndicats et de l'état. Ce terme fait référence à la 4^e révolution industrielle, faisant suite à la mécanisation de l'industrie (1^{ère} révolution), à la création des lignes de production tirant profit des nouvelles sources d'énergie électrique (2^e révolution), puis à l'intégration des automates programmables (3^e révolution). L'industrie 4.0 vise l'intégration et la synchronisation de l'ensemble des ressources tout au long de la chaîne de valeur de valeur (Wang et al., 2016) :

- intégration horizontale : tout au long de la chaîne de valeur, favorisant la collaboration entre les entreprises et en connectant les unités géographiquement distantes ;
- intégration verticale : tout le long de la ligne hiérarchique, favorisant le pilotage des sous-systèmes à travers la flexibilité et la reconfiguration des réseaux de production ;

- intégration temporelle : tout au long du cycle de vie du produit, favorisant la personnalisation et la production unitaire.

À la base de ce concept, il y a une multitude de technologies (Ruessmann et al., 2015) :

- l'analyse de données massives (Big Data) ;
- la simulation numérique ;
- l'internet des objets ;
- les systèmes cyberphysiques ;
- l'infonuagique ;
- la réalité augmentée ;
- la cybersécurité ;
- les robots collaboratifs ;
- la communication entre machines.

Peu de cas d'application de l'industrie 4.0 dans les PME sont référencés dans la littérature scientifique (Moeuf et al., 2017a) et il paraît complexe d'identifier des cas d'applications au sein de l'industrie. Ainsi, pour atteindre notre objectif, nous avons réalisé une étude prospective par consultation d'experts par la méthode Delphi. Nous avons couplé cette étude avec les abaques de Régnier, nous détaillons cela dans la section suivante.

3 L'ETUDE DELPHI – REGNIER

Nous décrivons ici la méthode prospective par consultation d'expert dite méthode Delphi, puis nous détaillons les abaques de Régnier. Nous précisons par la suite l'organisation de l'étude.

3.1 La méthode Delphi

La méthode Delphi fait référence à l'oracle de Delphes, la Pythie étant remplacée par les experts. Dans les années 1950, à la *Rand Corporation*, O. Helmer et N. Dalkey développent une méthode prospective par consultation d'experts (Gordon and Helmer, 1964). L'objet de cette étude est de prévoir méthodiquement « l'orientation des tendances technologiques et scientifiques à long terme, et leurs effets probables sur les sociétés ».

La méthode Delphi se fonde sur deux principes :

- l'anonymat des résultats ;
- l'indépendance des jugements.

La méthode Delphi vise à supprimer l'effet néfaste de l'influence entre experts qui se crée lors des réunions en face à face. La finalité de l'étude Delphi est de mettre en évidence des convergences d'opinions d'experts et de dégager un accord sur des sujets précis (Maleki, 2009).

La consultation consiste à interroger un panel d'expert d'horizon divers, elle se distingue d'une simple interview en ce sens qu'elle implique le respect d'une série de démarche :

- l'élaboration d'un questionnaire soumis aux experts ;
- le dépouillement et le traitement des réponses ;
- à partir de ces réponses, l'élaboration d'un nouveau questionnaire soumis aux mêmes experts à

qui on révèle les réponses du premier questionnaire, et à qui il est demandé d'introduire des nuances ou de donner les raisons de leur choix ;

- le respect des principes précédents jusqu'à l'obtention de tendances communes et précises (consensus et dissensus).

Dans le cadre d'une étude Delphi classique, le traitement statistique des réponses impose que les questions abordées puissent être positionnées selon une notation comprise ou non dans un intervalle moyen. Ainsi, l'enquête est conditionnée pour chaque donnée à l'utilisation d'une échelle quantitative.

Dans le cadre de notre étude, les données sont essentiellement qualitatives, c'est pourquoi nous avons couplé l'étude Delphi aux abaques de Régnier dont nous détaillons le principe dans la sous-section suivante.

3.2 Les abaques de Régnier

L'abaque de Régnier est une technique originale qui exploite un panel de couleur pour recueillir les opinions des experts concernant des affirmations choisies dans des termes précis, concis, concrets et pertinents. Pour exprimer leur opinion face à une affirmation, les experts ont le choix parmi 7 couleurs :

- **vert** : l'expert est tout à fait d'accord avec l'affirmation ;
- **vert pâle** : l'expert est en accord avec l'affirmation ;
- **orange** : l'avis de l'expert est mitigé ;
- **rouge pâle** : l'expert est en désaccord avec l'affirmation ;
- **rouge** : l'expert n'est pas du tout d'accord avec l'affirmation ;
- **blanc** : l'expert ne peut pas répondre ;
- **noir** : l'expert ne veut pas répondre.

L'information ainsi collectée apparaît sous forme d'un diagramme coloré qui apporte une dimension complémentaire au langage écrit ou parlé. Nous détaillons l'organisation de l'étude dans la sous-section suivante.

3.3 Organisation de l'étude

Nous avons donc constitué une équipe de pilotage composée de 4 chercheurs du laboratoire des Arts et Métiers, d'Exxelia Groupe et de l'université Polytechnique de Montréal. L'équipe a ensuite piloté l'étude, nous détaillons les différentes étapes de l'étude dans les sous-sections ci-après.

3.3.1 La sélection des experts

L'un des points importants de la méthode Delphi est la constitution d'un panel d'experts. L'équipe de pilotage a donc en charge de définir le nombre des compétences, la composition et le fonctionnement du panel d'experts.

La liste d'experts comprend toutes personnes dont les savoirs et les expériences donnent les capacités à pouvoir développer le champ d'études. Le recrutement d'experts sur une base homogène risque d'obtenir des résultats biaisés par un courant de pensée. Il est donc conseillé de

sélectionner un panel d'expert hétérogène favorisant les avis différents (Maleki, 2009).

Le nombre d'experts est également un point important. Pourtant, selon les travaux de Van De Ven et Delbecq (1974), le nombre d'experts doit être déterminé en fonction des objectifs. Ils précisent dans leurs travaux qu'un groupe de 7 experts favorisent la génération d'idées. Les travaux de Ashton (1986) montre que la taille du groupe d'experts d'un sujet donné se situe autour de 11 experts. Maleki précise qu'il faut au moins 30 experts pour valider une étude Delphi, et lors des travaux de Héraud sur les technologies du futur, l'équipe de pilotage avait sélectionné 3388 experts (Héraud et al., 1997).

Autrement dit, le nombre d'experts ne s'impose pas a priori. Nous avons sélectionné 12 experts issus du monde de l'industrie 4.0, un choix favorisant l'émergence d'idée tout en considérant les travaux d'Ashton (1986). Afin de ne pas avoir un panel trop homogène, nous avons sélectionné :

- quatre experts académiques ;
- quatre experts de l'industrie 4.0 issus du monde industriel ;
- quatre dirigeants de PME.

Ce panel a répondu aux différents questionnaires soumis. Toutes les réponses étaient exploitables, aucun des experts n'ayant abandonné au cours de l'étude.

3.3.2 Création du questionnaire initial

L'initialisation du questionnaire pour soumettre aux experts est un élément clef, car il oriente de manière significative les domaines dans lesquels l'étude Delphi-Régnier apportera des réponses. L'objectif de l'étude est d'évaluer les risques, opportunités et facteurs critiques de succès pour l'exploitation de l'industrie 4.0 dans les PME. Pour initialiser le questionnaire, nous avons analysé ces mêmes éléments au sein d'autres concepts considérés connexes :

- les « *advance manufacturing technology* » (Dangayach and Deshmukh, 2005; Raymond, 2005) ;
- les **ERP** (Doom et al., 2010; M. Beheshti et al., 2014; Shaul and Tauber, 2012; Somers and Nelson, 2001) ;
- le « *Lean Manufacturing* » (Achanga et al., 2006) ;
- le **Six-Sigma** (Kumar et al., 2009) ;
- la **TQM** (Ismail Salaheldin, 2009; Saraph et al., 1989; Yusof and Aspinwall, 2000).

Nous avons également sélectionné les caractéristiques managériales des PME afin de croiser ces caractéristiques avec l'industrie 4.0 (Moeuf et al., 2016) :

- les ressources limitées ;
- l'absence de décomposition fonctionnelle ;
- le management de proximité ;
- la stratégie courte-vue ;
- le manque de procédures et de méthodes ;
- le manque d'expertise.

Suite à cette première étude, nous avons sélectionné les risques et les facteurs critiques de succès de l'implantation de ces concepts. Nous avons ensuite rempli le questionnaire avec les éléments propres à l'industrie 4.0.

3.3.3 Évolution du questionnaire

Le questionnaire initial composé de 24 affirmations a été envoyé aux experts. Suite à leurs réponses, les experts étaient invités à proposer une nouvelle affirmation à soumettre au panel d'expert. L'équipe de pilotage avait à charge de synthétiser et éventuellement reformuler les affirmations avant de les resoumettre au panel d'experts. Lors de nos échanges, les experts ont ainsi soumis un total de 36 affirmations complémentaires lors des deux premiers tours de l'étude.

Nous avons clôturé l'étude après 3 tours. Nous donnons les résultats de l'étude dans de la section suivante.

4 RESULTATS DE L'ETUDE DELPHI-REGNIER

Suite à l'étude Delphi-Régnier, l'équipe de pilotage a synthétisé les réponses aux questions des experts. Lors du premier tour de l'étude, les questions étaient préparées par l'équipe de pilotage, mais pour les tours suivants ce sont les experts qui ont proposé une liste d'affirmation à porter au jugement des autres experts de l'étude.

Cette approche a permis de faire évoluer le questionnaire vers quatre thématiques qui ont mobilisé les experts :

- l'organisation de la PME ;
- les technologies de l'industrie 4.0 ;
- l'implantation de l'industrie 4.0 ;
- les opportunités offertes par l'industrie 4.0.

Nous détaillons ces quatre thématiques dans la section suivante.

4.1 L'organisation

Plusieurs sous-thématiques ressortent de l'étude, nous les détaillons dans les sous-sections suivantes.

4.1.1 La stratégie courte-vue

Dans les PME, la stratégie courte-vue est l'une des caractéristiques managériales résultantes d'une surimplification du dirigeant sur les aspects opérationnels de l'entreprise (Mintzberg, 1982; Torres, 1999). Les experts constatent que les projets industrie 4.0 doivent s'inscrire dans une vision stratégique de l'entreprise, ils précisent que la stratégie courte-vue fait partie des risques pour les PME.

4.1.2 La prépondérance du dirigeant

La structure verticalisée de l'entreprise, centrée sur le dirigeant est ambivalente. Les experts déclarent que si le dirigeant est « *early-adopter* » et technophile, il favorisera l'implantation de l'industrie 4.0 en mobilisant les

ressources, sa position dominante sera alors un atout ; inversement s'il est opposé au projet, sa position dominante en fera un frein pour le projet.

Les experts conseillent aux dirigeants de se rapprocher des réseaux d'échanges entre dirigeants pour prendre conscience de l'intérêt de l'industrie 4.0.

4.1.3 Le manque de fonction experte

Les experts notent que le manque de fonction experte est un vrai risque pour les PME. Ils sont mitigés concernant le fait que les PME ont les connaissances pour exploiter des outils tels que le *Big Data* ou la simulation numérique.

Les experts préconisent de former les équipes pour assurer une montée en compétence des équipes des PME.

4.1.4 La courte ligne hiérarchique

Les experts notent que la proximité du dirigeant avec les équipes favorise la transparence et la clarté concernant les objectifs de projet industrie 4.0. D'autre part, l'horizontalité de la PME favorise la communication entre les employés et l'agilité de l'organisation, caractéristique nécessaire aux projets d'industrie 4.0. Enfin, les experts notent que l'alignement hiérarchique nécessaire à ces projets est favorisé par le peu de collaborateurs concernés au sein d'une ligne hiérarchique courte.

Les experts notent qu'il y a un risque que les technologies de l'industrie 4.0 soient perçues comme des moyens de surveillance pour les collaborateurs de l'entreprise.

4.2 Les technologies de l'industrie 4.0

4.2.1 La donnée au cœur de l'industrie 4.0

Les experts notent que la donnée est la clef de la réussite d'un projet industrie 4.0 pour les PME. Ils notent que les PME n'exploitent pas assez les données disponibles, et que les dirigeants ne considèrent pas les données comme des sources de valeur ajoutée. Les experts notent également que les technologies favorisant la génération, le stockage et l'exploitation des données sont les plus efficaces, c.-à-d. elles donnent de bons résultats d'amélioration pour un coût limité.

Les experts proposent d'orienter les PME vers l'exploitation de leurs données disponibles comme première étape vers l'industrie 4.0.

4.2.2 Les technologies support des flux de données

Les experts notent que les technologies supports du flux d'informations sont facilement accessibles, comme l'infonuagique par exemple. Cependant, les experts soulignent que les PME sous-considèrent la Cyber-Sécurité au sein de leur infrastructure. Enfin, les experts sont favorables à l'exploitation de la simulation par les PME comme technologie d'amélioration.

4.2.3 Les technologies support du flux physique

Les experts notent que les technologies visant l'amélioration du flux physique, tels que les systèmes cyberphysiques, la communication entre machines et les robots collaboratifs ne sont pas des clefs pour l'amélioration industrielle des PME. Ces technologies sont trop coûteuses et trop complexes à mettre en œuvre.

4.3 L'implantation de l'industrie 4.0

4.3.1 Le Leader de projet industrie 4.0

Les experts précisent que le Leader de projet industrie 4.0 nécessite des compétences multidisciplinaires, une ouverture d'esprit et une curiosité importante pour appréhender les multiples possibilités offertes par l'industrie 4.0.

4.3.2 La communication concernant le projet

Les experts notent que le risque pour la santé et la sécurité au travail des employés est maîtrisé. Mais les experts notent également que les outils de l'industrie 4.0 peuvent être perçus comme des moyens de surveillance du quotidien, ainsi ils préconisent de communiquer clairement sur les objectifs de tout projet d'industrie 4.0.

4.3.3 La formation versus le conseil

Les experts notent unanimement que les PME manquent d'expertise interne. Ils envisagent deux voies pour combler ce manque : la formation et/ou l'accompagnement par des consultants.

Les experts préconisent plusieurs étapes clefs nécessitant un accompagnement des PME :

- effectuer un audit de l'outil de production et une cartographie des flux ;
- effectuer un choix sur les technologies permettant de répondre aux problématiques soulevées.

Les experts préconisent de s'appuyer davantage sur des équipes académiques plutôt que sur des consultants extérieurs.

Les experts notent également un risque avec le conseil extérieur lié à la transmission du savoir, pas toujours assurée par les consultants.

Les experts notent toutefois que la formation interne favorise l'implantation de l'industrie 4.0. Les experts précisent également la possibilité de recruter les compétences manquantes au sein des différentes équipes des PME.

4.3.4 Simplifier les outils de l'industrie 4.0

L'une des pistes proposées par les experts est de simplifier les outils de l'industrie 4.0 pour les rendre plus accessibles aux PME.

4.4 Les opportunités offertes par l'industrie 4.0

Les experts s'accordent pour dire que l'industrie 4.0 est un facteur d'amélioration de la compétitivité des PME.

De même, les experts notent que l'industrie 4.0 peut aider à diminuer l'empreinte écologique des entreprises si celles-ci le souhaitent.

Au cours de notre étude, nous avons volontairement proposé 5 affirmations « fermées » aux experts. Toutes ces affirmations ont reçu des avis mitigés de la part des experts, nous pensons que cela traduit que l'industrie 4.0 présente de nombreuses possibilités, mais qu'aucune stratégie de passage vers l'industrie 4.0 ne semble prévaloir sur les autres. Il n'y a pas de « *One Best Way* » de l'industrie 4.0.

Enfin, les experts notent deux voies d'exploitation de l'industrie 4.0 : l'une visant à réduire les coûts de production, l'autre à repenser les modèles d'affaires dans le but de proposer de nouveaux produits et service aux clients.

5 LIMITES ET PERSPECTIVES

Nous précisons ici les limites de l'étude réalisée et les perspectives.

La première limite de l'étude concerne la méthode Delphi-Régner employée. Cette méthode est dépendante des experts sélectionnés, malgré la volonté de l'équipe de pilotage d'être neutre, le choix des experts comporte nécessairement un biais subjectif lié à l'identification des experts au sein des réseaux académiques et industriels exploités.

Le premier questionnaire était le résultat d'une analyse d'articles de recherche liés à des concepts proches par l'équipe de pilotage. Ce choix particulièrement structurant pour l'initialisation de l'étude comporte un biais subjectif dans les affirmations proposées aux experts. Cependant, à partir du second tour, ce sont les experts eux-mêmes qui ont enrichie et orienté l'étude par la proposition d'affirmation, ce biais est donc en partie compensé.

Enfin, la méthode des abaques de Régner est particulièrement adaptée à une étude de critères qualitatifs. Nous avons remarqué que lors de l'étude, nous étions restés sur des tendances et des estimations concernant l'industrie 4.0 dans les PME. Une étude plus quantitative serait un plus pour prolonger ces premiers travaux.

Ce travail permet aussi d'identifier une liste non exhaustive des risques, opportunités et facteurs critiques de succès liés à l'industrie 4.0 dans les PME. Notre étude se veut prospective en ce sens que la faiblesse des cas d'applications recensées ne permettait pas une analyse statistique. Pour effectuer ce travail, l'application d'une méthodologie de recherche par étude de terrain serait un moyen de confirmer, infirmer ou compléter la liste proposée.

De même, nous avons focalisé notre étude sur les PME. En France, la majorité des PME sont des entreprises dont le nombre d'employés est en moyenne proche de 25 (Chevalier, 2014), alors que la définition pousse à 250 employés. Il est probable que le comportement d'une

entreprise de 25 collaborateurs et de 250 collaborateurs soit différent face à des projets d'industrie 4.0.

Plusieurs éléments sont encore inconnus suite à nos travaux. Par exemple, est-ce que l'industrie 4.0 doit être considérée comme un concept de connectivité, où bien comme un ensemble de technologies accessibles servant de boîte à outils aux dirigeants ? Est-ce que l'industrie 4.0 doit être exploitée pour modifier en profondeur les modèles d'affaires et ainsi favoriser l'émergence de nouveaux espaces d'affaires (Kim and Mauborgne, 2005) ? Où bien, l'industrie 4.0 doit-elle servir à une augmentation de la compétitivité par la réduction des coûts de production et de fonctionnement des entreprises ?

6 CONCLUSION

Au cours de nos travaux, nous avons montré que l'industrie 4.0 dans les PME n'était pas triviale. L'expertise nécessaire et la formation résultante sont au cœur de la réussite d'un projet industrie 4.0 dans une PME. L'accompagnement des PME par des experts est vivement conseillé, du moment que cet accompagnement se solde par un transfert de connaissance rendant pérenne l'exploitation de l'industrie 4.0. D'autre part, le contexte spécifique des PME ne semble pas être un frein aux projets industrie 4.0. si les PME subissent le manque de moyen et le manque d'expertise, elle profite de l'agilité de l'organisation face aux changements à venir et de la capacité de l'entreprise à mobiliser l'ensemble des acteurs.

Nos travaux montrent que la donnée (*Data Analytics*), les outils de communication (*Cloud Computing*) et les outils de simulation sont les plus faciles à mettre en œuvre et sont les plus abordables.

Le rôle clef du dirigeant de la PME a été mis en évidence au cours de l'étude. Il est un maillon centralisant les pouvoirs, il peut donc mobiliser les acteurs autour d'un projet industrie 4.0. autant que bloqué toute initiative 4.0. Ainsi, il est nécessaire de convaincre le dirigeant avant tout lancement d'un projet industrie 4.0.

L'industrie 4.0. est un vecteur d'innovation et un espace d'entrepreneuriat pour les entreprises. C'est une occasion unique de proposer de nouveaux modèles d'affaires par l'exploitation de ces nouvelles technologies. Nous pensons que l'industrie 4.0 est un espace de créativité, d'innovation et de nouvelles opportunités.

Les PME présentent un avantage certain, elles sont créatives et possèdent des structures organisationnelles qui s'adaptent facilement à de nouveaux défis. Elles doivent favoriser la formation interne et la faculté d'innovation de leurs équipes.

Les recherches doivent explorer au-delà des groupes technologiques de l'industrie 4.0, les opportunités opérationnelles telles que les chantiers d'amélioration, les opportunités tactiques telles que les transformations d'usines permettant de produire plus efficacement, mais aussi les opportunités stratégiques visant de nouvelles perspectives de marché.

7 REFERENCES

- Achanga, P., Shehab, E., Roy, R., Nelder, G., 2006. Critical success factors for lean implementation within SMEs. *J. Manuf. Technol. Manag.* 17, 460–471. doi:10.1108/17410380610662889
- Ashton, R.H., 1986. Combining the judgments of experts: How many and which ones? *Organ. Behav. Hum. Decis. Process.* 38, 405–414. doi:10.1016/0749-5978(86)90009-9
- Babiceanu, R.F., Seker, R., 2016. Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. *Comput. Ind.* 81, 128–137. doi:10.1016/j.compind.2016.02.004
- Chen, M., Mao, S., Liu, Y., 2014. Big Data: A Survey. *Mob. Networks Appl.* 19, 171–209. doi:10.1007/s11036-013-0489-0
- Chevalier, F., 2014. Des PME aux ETI : les entreprises en croissance, 1re éditio. ed. de boeck, Paris.
- Cullinan, C., Sutton, S.G., Arnold, V., 2010. Technology monoculture: ERP systems, 'techno-process diversity' and the threat to the information technology ecosystem. *Adv. Account. Behav. Res.* 13, 13–30. doi:10.1108/S1475-1488(2010)0000013005
- Dangayach, G.S., Deshmukh, S.G., 2005. Advanced manufacturing technology implementation. *J. Manuf. Technol. Manag.* 16, 483–496. doi:10.1108/02656710210415703
- Deep, A., Guttridge, P., Dani, S., Burns, N., 2008. Investigating factors affecting ERP selection in made-to-order SME sector. *J. Manuf. Technol. Manag.* 19, 430–446. doi:10.1108/17410380810869905
- Dijkman, R.M., Sprekels, B., Peeters, T., Janssen, A., 2015. Business models for the Internet of Things. *Int. J. Inf. Manage.* 35, 672–678. doi:10.1016/j.ijinfomgt.2015.07.008
- Doom, C., Milis, K., Poelmans, S., Bloemen, E., 2010. Critical success factors for ERP implementations in Belgian SMEs. *J. Enterp. Inf. Manag.* 23, 378–406. doi:10.1108/17410391011036120
- Gordon, T., Helmer, O., 1964. Report on a Long-Range Forecasting Study, The Rand Corporation.
- Hao, Y., Helo, P., 2015. The role of wearable devices in meeting the needs of cloud manufacturing: A case study. *Robot. Comput. Integr. Manuf.* doi:10.1016/j.rcim.2015.10.001
- Héraud, J.A., Munier, F., Nanopoulos, K., 1997. Méthode Delphi : une étude de cas sur les technologies du futur. *Futuribles* 218, 33–53.
- Ismail Salaheldin, S., 2009. Critical success factors for TQM implementation and their impact on performance of SMEs. *Int. J. Product. Perform. Manag.* 58, 215–237. doi:10.1108/17410400910938832
- Kim, W.C., Mauborgne, R., 2005. *Stratégie Océan Bleu : comment créer de nouveaux espaces stratégiques.* Harvard Business School Publishing Corporation.

- Kohler, D., Weisz, J.-D., 2016. *Industrie : Les défis de la transformation numérique du modèle industriel allemand*. Paris.
- Kumar, M., Antony, J., Douglas, A., 2009. Does size matter for Six Sigma implementation? *TQM J.* 21, 623–635. doi:10.1108/17542730910995882
- Liker, J.K., 2006. *The Toyota Way Fieldbook, A practical guide for implementing Toyota's 4Ps*. The McGraw-Hill Companies, Inc, New York.
- M. Beheshti, H., K. Blaylock, B., A. Henderson, D., G. Lollar, J., 2014. Selection and critical success factors in successful ERP implementation. *Compet. Rev.* 24, 357–375. doi:10.1108/CR-10-2013-0082
- Maleki, K., 2009. *Méthodes quantitatives de consultations d'experts*, Editions P. ed. Publibook, Paris.
- Mintzberg, H., 1982. *Structure et Dynamique des organisations*, Vingt-trois. ed. Eyrolles.
- Moeuf, A., Lamouri, S., Pellerin, R., Eburdy, R., Tamayo, S., 2017a. Industry 4.0 and the SME: a technology-focused review of the empirical literature, in: 7th IESM Conference. Saaebrücken, Germany.
- Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., Barbaray, R., 2017b. The industrial management of SMEs in the era of Industry 4.0. *Int. J. Prod. Res.* 7543, 1–19. doi:10.1080/00207543.2017.1372647
- Moeuf, A., Tamayo, S., Lamouri, S., Pellerin, R., Lelievre, A., 2016. Strengths and weaknesses of small and medium sized enterprises regarding the implementation of lean manufacturing. *IFAC-PapersOnLine* 49, 71–76. doi:10.1016/j.ifacol.2016.07.552
- Mousannif, H., Sabah, H., Douiji, Y., Oulad Sayad, Y., 2016. Big data projects: just jump right in! *Int. J. Pervasive Comput. Commun.* 12, 260–288. doi:10.1108/IJPC-04-2016-0023
- Ohno, T., 1988. *The Toyota Production System — Beyond Large-Scale Production*, CRC Press.
- Raymond, L., 2005. Operations management and advanced manufacturing technologies in SMEs. *J. Manuf. Technol. Manag.* 16, 936–955. doi:10.1108/17410380510627898
- Ruessmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., Harnisch, M., 2015. Industry 4.0: the future of productivity and growth in manufacturing industries.
- Saraph, J. V., Benson, P.G., Schroeder, R.G., 1989. An Instrument for Measuring the Critical Factors of Quality Management. *Decis. Sci.* 20, 810–829. doi:10.1111/j.1540-5915.1989.tb01421.x
- Shah, R., Ward, P.T., 2003. Lean manufacturing: Context, practice bundles, and performance. *J. Oper. Manag.* 21, 129–149. doi:10.1016/S0272-6963(02)00108-0
- Shaul, L., Tauber, D., 2012. CSFs along ERP life-cycle in SMEs: a field study. *Ind. Manag. Data Syst.* 112, 360–384. doi:10.1108/02635571211210031
- Somers, T.M., Nelson, K., 2001. The impact of critical success factors across the stages of enterprise resource planning implementations, in: *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE Comput. Soc, p. 10. doi:10.1109/HICSS.2001.927129
- Torres, O., 1999. *Les PME*.
- Van De Ven, A.H., Delbecq, A.L., 1974. The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Acad. Manag. J.* 17, 605–621.
- Wang, S., Wan, J., Li, D., Zhang, C., 2016. Implementing Smart Factory of Industrie 4.0: An Outlook. *Int. J. Distrib. Sens. Networks* 2016, 1–10. doi:10.1155/2016/3159805
- Womack, J.P., Daniel, T.J., 2003. *Lean thinking*, First Free. ed. Simon & Schuster, Inc.
- Xu, X., 2012. From cloud computing to cloud manufacturing. *Robot. Comput. Integr. Manuf.* 28, 75–86. doi:10.1016/j.rcim.2011.07.002
- Yue, X., Cai, H., Yan, H., Zou, C., Zhou, K., 2015. Cloud-assisted industrial cyber-physical systems: An insight. *Microprocess. Microsyst.* 39, 1262–1270. doi:10.1016/j.micpro.2015.08.013
- Yusof, S.M., Aspinwall, E.M., 2000. Critical success factors in small and medium enterprises: Survey results. *Total Qual. Manag.* 11, 448–462. doi:10.1080/09544120050007760

IMPLEMENTING SCRUM TO DEVELOP A CONNECTED ROBOT

Diego Armando Díaz Vargas¹, Rui Xue², Claude Baron¹, Philippe Esteban¹,
Rob Vingerhoeds³, Citlalilh Y. A. Gutierrez Estrada⁴, Chao Liu²

¹LAAS-CNRS, Université de Toulouse, CNRS, INSA, UPS, 7 av. du Colonel Roche, 31400 Toulouse, France
{dadiavza, claud.baron, philippe.esteban}@laas.fr

²School of Economics and Management, Beijing University of Technology, 100 Ping Le Yuan, Chaoyang District, Beijing 100124, China,
{xuerui, liuchao}@bjut.edu.cn

³Institut Supérieure de l'Aéronautique et de l'Espace, 10 av. Edouard Belin BP 54032, 31055 Toulouse, France
rob.vingerhoeds@isae-superaero.fr

⁴Instituto Tecnológico de Toluca, Toluca, México,
citlaligh@yahoo.com

ABSTRACT: *Agile methods are receiving a growing interest from industry and these approaches are nowadays well accepted and deployed in software engineering. However, some issues remain to introduce agility in systems engineering. The objective of this paper is to show an agile management implementation in an educational project consisting in developing a connected mobile robot, and to evaluate the issues and benefits of adopting an agile approach. Among the most famous agile management methods, SCRUM has been chosen to lead this experiment. This paper first presents the project and how students traditionally manage it, then it describes how Scrum could be used instead. It evaluates the difficulties and interests to introduce agility in this project, and concludes on the ability of Scrum to design, test and progressively integrate the system, thus providing an operational prototype more quickly.*

KEYWORDS: *systems engineering, connected systems, agile management, integration, mobile robot, SCRUM.*

1 INTRODUCTION

Every year, the Institut National des Sciences Appliquées de Toulouse (INSA Toulouse, Université de Toulouse, France) proposes a project to master students with an objective to design and realize a remotely controlled LEGO Mindstorms robot. The six to eight students project team is supervised by a professor, in charge of explaining the robot mission and helping with the management of the three-months project. The students' assignments are to define the necessary resources and contributing systems, and to organize their work and their team in order to be able to deliver the robot on time.

The project includes the development of the robot as well as the writing of a complete report, in which technical and organizational aspects have to be detailed. Students traditionally follow the V-cycle model to develop the system. The project is divided into six stages: initiation, preliminary development, principal development, robot realization, test and close. Although the V-cycle is a widely used reference in the industry, with the emergence of agile approaches, many project managers have paid attention to agile management of projects, essentially in software engineering projects. Therefore, students must be trained to agile management before joining the industry.

This article aims at deploying, testing and evaluating the issues and benefits of an agile approach on this project, in order to determine if it could be a candidate to train stu-

dents to a basic practice of agile management. To this purpose, this paper considers and compares two options to implement agility.

The paper is organized as follows. Section 2 presents the way students are currently managing their project. Section 3 outlines the principles of the agile approach. Section 4 explains how the students could handle the project in an agile way. Section 5 concludes on the feasibility and interest of training students to an agile management on this project.

2 DEVELOPMENT OF A CONNECTED MOBILE ROBOT

The objective of the project is to deliver an autonomous remotely supervised wheeled robot. It includes the design and construction of the robot using LEGO building blocks. The robot can either be remotely operated or move autonomously on a delimited area. This area contains two distinct zones: a 'rest zone' where the robot is initially located, a 'camp' where the robot has several tasks to accomplish.

The mission of the robot is decomposed into three phases. It first has to reach the camp by autonomously following a line traced on the ground and to signal the operator that it arrived (send a 'bip'). There, an operator remotely takes the control of the robot to realize the tasks assigned to the robot. These tasks consist in dropping several wooden tokens that are embedded in a storage module in the robot, to different predefined zones. During the remotely controlled part of the mission the operator can visualize the

robot environment thanks to a camera in order to avoid the different traps on the area (crevasses, low light areas, low Wi-Fi/Bluetooth areas) to drive the robot following the best route. Once the job accomplished, the robot returns to the rest zone.

From a given set of robot requirements, students have to define a logical and physical architecture for the robot, to construct it and to test it against its specifications. They first have to organize, divide and plan the project before dealing with more technical aspects (system analysis, conception, integration, tests, etc.). The project team has three months to lead the project.

The project team defines six stages to structure the project, with 7 milestones (named M1 to 7 in Figure 1).

Initiation	Preliminary development	Principal Development	Robot realization	Test	Close
- Develop project charter	- System analysis - Architectural conception of system	- Detailed requirement definition - Architectural conception of system - Reference configuration - Integration - Validation - Verification	- Development of prototype - Improvement system performance	- Test - Maintenance	- Close project
M1	M2	M3	M4	M5	M6

Figure 1: The six stages of the project

Initiation: organization of the project group and definition of the responsibilities of each team member.

Preliminary development: definition of requirements, of alternatives for the robot architecture and of test cases.

Principal development: definition of the specifications and of the architectures of sub-systems; integration, validation and verification.

Robot realization: development of a robot prototype with LEGO bricks, improvement of the robot performance.

Test: students define test cases to verify the robot behavior and performance.

Close: students have to close the project before the deadline.

To lead the project, students use traditional methods, based on a predefined schedule, with cost constraints and quality objectives. They completely plan everything from the beginning, however wisely including some margins. The robot requirements are initially given to the students, as well as some physical constraints, such as the size and shape of tokens for instance. During the development stages, changes are not expected, or even allowed. This kind of management method is adapted to a V-model development cycle.

3 AGILE MANAGEMENT OF PROJECTS

Compared with the traditional management method used by the students, the agile management method attracts a lot of attention in industry because of several assets: flexibility, transparency, increased customer satisfaction,

close communication between developers, etc. This section first compares traditional and agile management, then outlines the main advantages of agile management, to end with a brief introduction to the three mostly used agile methods.

3.1 Traditional management vs Agile management

The traditional management was characterized by Chin (2004), Fowler (2000) and Highsmith (2009) as a methodology whose emphasis is on detailed planning and resistance to change. Schwaber (2007) corroborates that about 50 percent of the time is spent on requirements, architecture and specification, and that all this is done before even building any functionality when the project management team used traditional practices. Agile management is described as an approach that aims for flexibility, simplicity, iterating in short periods of time, and incrementally adding value (Boehm, 2002), (Chin, 2004), (Cockburn, 2002), (Cohn and Ford, 2003), (Highsmith, 200), (Misra et al., 2010). A comparison conducted by Shenhar and Dvir (2007) highlights the main differences between the traditional and agile management method. Hoda et al. (2010) also compare them with a software development focus, and highlight the main differences.

3.2 Advantages of agile management

Agile management now is widely spread and especially applied in software development. Its use allows companies to quickly react to project changes and to adjust their processes and organizations. The use of agile management can also help companies to improve their success rate (Augustine et al., 2005) (Serrador and Pinto, 2015). The case study that has been carried out in the research of Serrador and Pinto (2015) has demonstrated that the use of agile management method can help the companies to end projects on time, on budget, with a high customer satisfaction. Furthermore, agile project management can solve the problems induced by a change of requirements (Azanha et al., 2017). Azanha et al. also states that agile management practices can help companies in getting value for them and their customers by the benefits obtained with the agile approach usage, applied to systems development. Nowadays, project leaders and teams find themselves in an environment disrupted by exponential advances in technology and demands from customers for more immediate delivery of value, agile techniques and approaches effectively manage disruptive technologies (PMI, 2017). Comparisons between agile management and traditional management in literature mainly focus on the software development context. This paper analyzes the advantages of agile management in the context of a mobile connected robot, where software is embedded with physical components.

3.3 Introduction to the three main agile management methods

Different agile methodologies are currently practiced in the industry. This subsection briefly introduces the most widely used. Among them SCRUM has been chosen for this project; it will be presented with more details.

XP

XP is an agile software development methodology created by Kent Beck and first presented in 1999, in a book titled *Extreme Programming Explained: Embrace Change*. According to the author, “the goal of eXtreme Programming is outstanding software development at lower cost, with fewer defects, higher productivity, and much higher return on investment” (Beck, 2004). XP does not rely on planning, analyzing and designing for the far future. Instead, it allows these activities to be made throughout the whole development, in short cycles (Beck, 1999).

UPEDU

UPEDU is a software development process specialized for education, developed by Pierre-N. Robillard, Patrick d'Astous of the École Polytechnique de Montréal, and Philippe Kruchten from Rational Software. UPEDU is an academic customization of the Rational Unified Process (RUP). RUP is an iterative software development process framework created by the Rational Software Corporation, an IBM division since 2003 (Robillard et al., 2001). Many companies claimed that they have applied agile management on their projects. However, Schwaber et Sutherland suggests only 30% of scrum-adopting companies will become “excellent development organizations”. Change experts like Harvard Professor John Kotter (Kotter, 2009) regularly say 70% of major change efforts fail. Some of agile management method implementation in the companies only remain some aspects of agile management, such as its iteration or collaboration.

Scrum

Scrum was first proposed by Jeff Sutherland and Ken Schwaber in 1993 (Cervone, 2011), (Misra et al., 2010). It was originally designed for software development. Now the application of this method is not restricted to software development; it also can be applied to a large variety of domains, such as manufacturing, marketing, advertising agencies, architectural projects and banks (Cardozo et al., 2010). Among the three main agile management methods, Scrum is widely deployed in industry. In addition, Scrum is quite simple to train students without asking them to develop an advanced expert level beforehand. We thus chose Scrum to evaluate the interests of implementing agility into our educational project.

Scrum is founded on empirical process control theory, or empiricism. Empiricism asserts that knowledge comes from experience and making decisions based on what is known (Schwaber et al., 2017). Scrum is therefore defined as a framework which people can address complex adap-

tive problems, while productively and creatively delivering products of the highest possible value. This framework consists of scrum teams and their associated roles, events, artifacts and values, each component serves a specific purpose and is essential to Scrum’s success and usage. Below a brief description of the different Scrum aspects is given (adapted from Schwaber et al., 2017):

Scrum Values

The Scrum values are followed by everyone in the project. It enumerates five values:

- 1) V_1 - *Courage*, Scrum Team members have courage to do the right thing and work on tough problems,
- 2) V_2 - *Focus*, everyone focuses on the work of the Sprint and the goals of the scrum team,
- 3) V_3 - *Commitment*, people personally commit to achieving the goals of the Scrum Team,
- 4) V_4 - *Respect*, Scrum Team members respect each other to be capable, independent people, and
- 5) V_5 - *Openness*, the Scrum Team and its stakeholders agree to be open about all the work and the challenges with performing the work.

Scrum Events

Scrum Events are specifically designed to enable critical transparency and inspection. Each event is a formal opportunity to inspect and adapt something. All events are time-boxed events and they may ensure an appropriate amount of time without allowing waste in the process. Scrum enumerates five events:

- 1) **Sprint** – a time-box of one month or less, during which a usable and potentially releasable product increment is created; an increment is the sum of all the items completed during a sprint and the value of the increments of all previous Sprints. The Sprint is the heart of Scrum and a new Sprint starts immediately after the conclusion of the previous Sprint.
- 2) **Sprint Planning** – a time-box to a maximum of eight hours for a one-month Sprint. Sprint planning plans the work to be performed in each Sprint. This plan is created and maintained by the collaborative work of all members of the project. Sprint planning also defines the sprint goal, an objective set for the Sprint, that can be met through the implementation of the Product Backlog.
- 3) **Daily Scrum** – a 15-minute time-boxed event, held every day of the Sprint. Daily Scrum is used to inspect progress toward the Sprint Goal and to inspect how progress is trending toward completing the work in the Sprint Backlog.
- 4) **Sprint Review** – the inspection of the increment. During this inspection stakeholders review the

results of a Sprint. A Sprint Review is held at the end of the Sprint and adapts the Product Backlog if needed.

- 5) Sprint Retrospective – an opportunity for the team to review itself and create a plan for improvements to be done during the next sprint. Sprint Retrospective occurs after the Sprint Review and prior the next Sprint Planning.

The Scrum Team

Scrum Teams have as a characteristic “self-organizing and cross-functional”. Self-organizing teams choose how best to accomplish their work, rather than being directed by others outside the team. Cross-functional teams have all competencies needed to accomplish the work without depending on others. Furthermore, the Scrum Team delivers products iteratively and incrementally, maximizing opportunities for feedback. A Scrum Team comprises three entities:

- 1) a Product Owner: one person only and its decisions have to lead the entire team and the organization,
- 2) a Development Team: professionals who do the work of delivering a potentially releasable increment of product at the end of each sprint, and
- 3) a Scrum Master: it is responsible for promoting and supporting scrum framework, he does this by helping everyone understand practices, rules, values and scrum theory.

Scrum Artifacts

Scrum Artifacts represent work or value to provide transparency and opportunities for inspection and adaptation; they are designed to maximize transparency of key information. Product Backlog is considered as an ordered list of everything that is known to be needed in the product; the Product Owner is responsible for the Product Backlog. Sprint Backlog is the set of Product Backlog items selected for the Sprint, a plan for delivering the product increment and realizing the Sprint Goal; it makes the work that the Development Team identifies as necessary to meet the sprint goal visible. The Increment indicates the sum of all the Product Backlog items finalized during a Sprint, as well as the ones finished during all previous sprints. Product Backlog, Sprint Backlog and Increment are considered as artifacts.

Figure 2 shows the complete Scrum Framework and how each entity is distributed. The framework begins with the Product Backlog; the Product Owner defines all the features that the Development Team has to do and then to prioritize. The next step is the Scrum Planning; here the Product Owner and the Development Team plan which features of the Product Backlog have to be done and the time of each Sprint, which results in the Sprint Backlog. The Development Team has to do each Sprint defined in the Sprint Backlog. They have to hold daily meetings to

organize the work that has to be done. Sprint Review allows the Development Team to solve the problems that occurred during the sprint development.

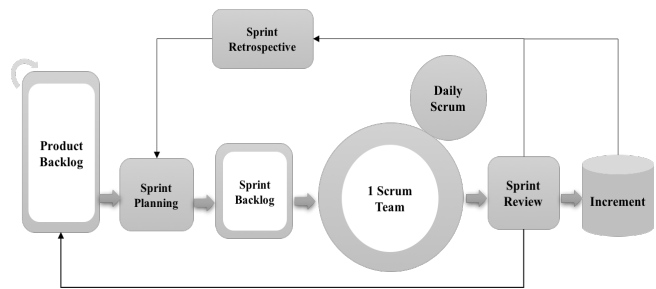


Figure 2: Global Scrum Framework (scrum.org, 2017)

4 USING AGILITY IN THE ROBOT PROJECT

Our proposal is to introduce agility in three project phases: principal development, robot realization and test phases (see Figure 1). This section considers and discusses several options to implement Scrum in these phases.

4.1 Implementing Scrum in the project

The robot mission is to move in a predefined environment, autonomously or remotely operated, as mentioned in Section 2. More specifically the robot shall achieve its mission following three successive phases:

- Phase 1: The robot shall move autonomously from the ‘rest zone’ to the ‘camp’ and send an audible signal once it arrives at the ‘camp’.
- Phase 2: The robot shall drop tokens in predefined zones under the operator directions, avoiding traps on the route (one crevasse, two low light areas and three low Wi-Fi/Bluetooth areas).
- Phase 3: The robot shall join the ‘rest zone’ under the operator commands from an operating terminal.

These three phases can be considered as a Product Backlog and can be used to initiate the Sprint Planning. Before starting the Sprint Planning, students should decide who is going to be the Product Owner, the Scrum Master and who will contribute to the Development Team; the goal of this role definition is to focus on the scrum framework recommendations. Students can define the global architecture of the robot using the given requirements, that allows them to identify the different interactions between stakeholders and the robot. The robot can be decomposed into three subsystems as shown in Figure 3: the robot structure and motion subsystem, the vision subsystem and the Human Machine Interface subsystem.

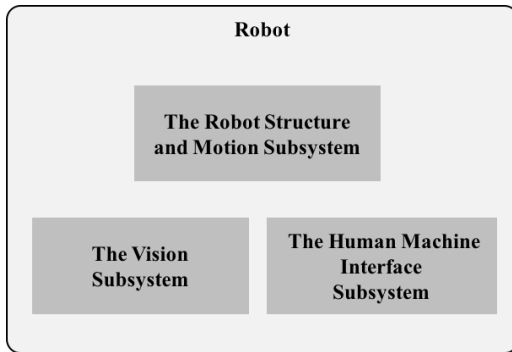


Figure 3: Global architecture of the robot

The robot structure and motion subsystem corresponds to the physical architecture of the robot with the motion subsystem; it has to answer all requirements of the specifications. The vision subsystem will allow an optimal view to the robot to accomplish its mission and the human machine interface subsystem will allow to remotely command the robot.

Then two different options can be considered for implementing the Scrum framework:

1. to define the sprints on the basis of the decomposition of the architecture into subsystems (Figure 3). Each sprint will address the development of one subsystem considered independently from the others, a supplementary sprint will address the integration of subsystems.
2. to define sprints on the basis of the phases of the robot mission. This means analyzing which parts of subsystems are invoked at each phase of the mission and define sprints aiming at incrementally developing these specific parts. Figure 4 shows the decomposition of each subsystem of the robot into parts. S1 and S2 are subsystems of the robot structure and motion subsystems, V1, V2 and V3 of the vision subsystem, HMI1 and HMI3 of the human machine interface subsystem. On the figure, we can see that the option taken here is to develop S1, V1 and HMI1 at one sprint, S2 and V2 at another sprint, and V3 and HMI3 in another sprint.

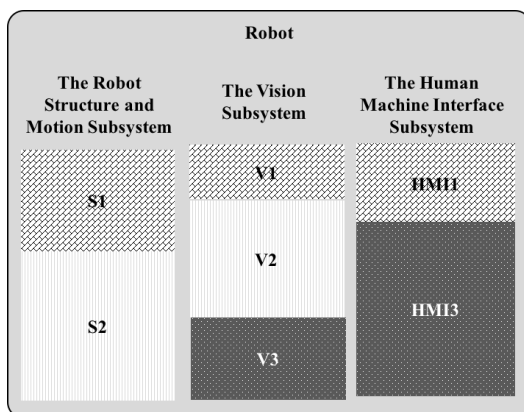


Figure 4: Distribution of the robot subsystems

These two options are presented here below.

4.1.1 Option 1

To define the sprints, the Product Owner has to prioritize the features inside the Product Backlog then plan each sprint with the Development Team. According to this project, we can distribute the Product Backlog using as a reference the global architecture defined in Figure 3, each sub-system is considered as a feature of the Product Backlog and shown in Figure 5.

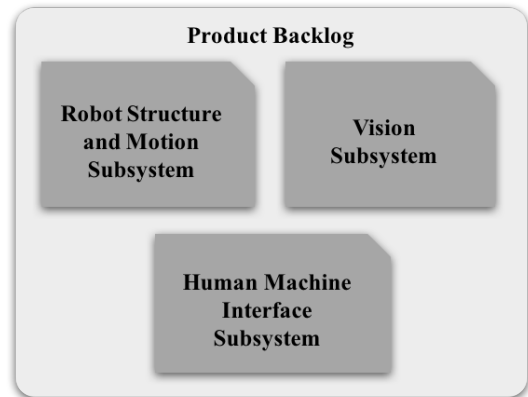


Figure 5: Product Backlog of the robot project

We thus propose four sprints, one sprint for each subsystem and a supplementary sprint for the assembly.

- **Sprint 1**: the first artifact is the architecture and motion system of the robot; it is the result of the first sprint.

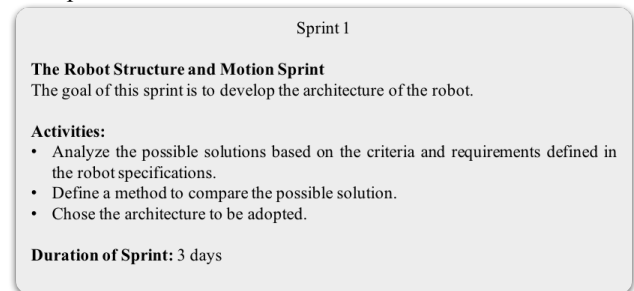


Figure 6: First sprint, the architecture and motion subsystem of the robot

- **Sprint 2**: the second artifact is the vision subsystem

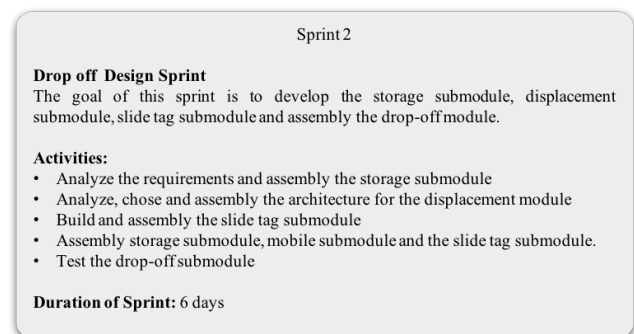


Figure 7: Second sprint, the vision subsystem

- **Sprint 3:** the human machine interface is the third increment of this sprint.

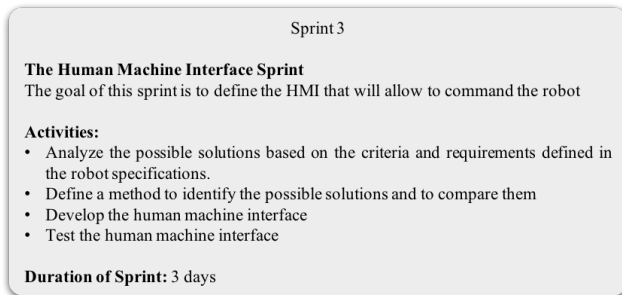


Figure 8: Third sprint, the human machine interface

- **Sprint 4:** this is the last sprint; the entire robot is the result of this sprint.

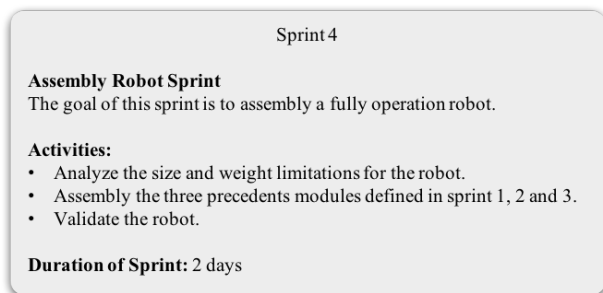


Figure 9: Fourth sprint, entire robot

The result of each sprint is considered as an increment, which means that once the first sprint is completed, the detail of the next phase (second sprint) is worked out, and so on. Each sprint has to follow a daily sprint meeting to know the status of the activities defined in each sprint, followed by the sprint review. If everything is OK, the development team can do the increment, otherwise the development team has to hold the sprint retrospective to correct or redefine the sprint. The proposal has to follow the framework described in Figure 2.

The focus point of this proposal is the architecture of the robot. The first three sprints follow a list of activities that allows to deliver and test a specific and independent part of the robot, each increment of these sprints are assembled in a fourth sprint to have a fully operational robot.

4.1.2 Option 2

This option proposes to identify which subsystem is invoked at each phase of the robot mission. To define the sprints, the Product Owner can consider the decomposition of the robot subsystems (Figure 4) as the Product Backlog then starts the sprint planning. Product Owner also has to prioritize the features inside the Product Backlog, then plans the sprints with the Development Team.

Once the sprints were defined, their development has to follow the framework described in Figure 2, in that case the increment is considered as the result of the integration

of each subsystem to accomplish each phase of the global mission, sprints have to follow a daily sprint meeting and the sprint review as stablish in the scrum framework. To illustrate how sprints can be defined in this option, we presented the first sprint as follows:

- **Sprint 1:** the first artifact of this sprint is the integration of three subsystems to accomplish Phase 1. It needs the subsystem S1 (from the robot Structure motion and subsystem), V1 (from the Vision subsystem), and HMI1 (from the human machine interface subsystem).

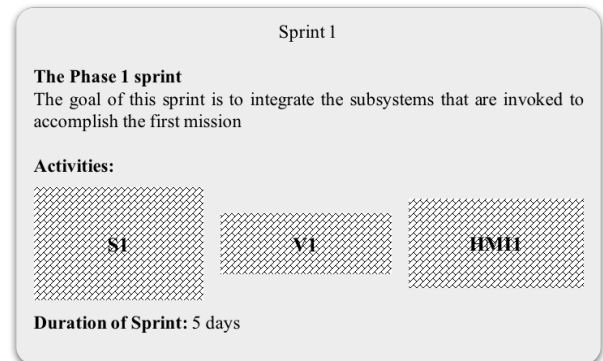


Figure 10: First sprint, the mobile module and HMI

Figure 10 shows the first sprint of this proposal. The activities were defined considering the subsystems needed and the integration of them is considered as the result of the first sprint (increment). The definition of the other two sprints should follow the same analysis.

This proposal establishes a different way from the first one. After the roles definition, Product Owner identifies which subsystems were involved to complete each phase of the global mission then defines each sprint. It allows to deliver artifacts that contains hardware and software at the same time.

4.2 Discussion

Section 4.1 shows two different conceptual ways to implement the Scrum Framework as an alternative to the traditional management methods, the goal of this implementation is to develop a mobile connected robot in an agile way. The implementation was proposed following three phases that the robot has to follow to accomplish a global mission. In fact, this work is focused in order to evaluate if the development of the project robot is a good case study to implement the scrum framework, considering that this project not only includes software development but also hardware integration.

From the implementation of agile management on this educational project, we can find that developing the robot with Scrum could help students to deliver functional modules in each increment and reduce time for testing as presented. For example, they delivered functional parts of the robot including hardware and software at the same time in

each sprint. The results of this proposal will be evaluated at the end of the school year.

This work also opens an additional discussion point relative to this investigation. The architecture and requirements were well defined before, but what if during the development that changes? As we know, requirements can be changed during the development; so, these changes may be discussed in the sprint review or in the sprint retrospective, but we do not really know if that will have a strong impact in the architecture. This point shall be evaluated next year, when agility will be implemented in new student projects.

Some difficulties could exist when trying to implement agile management methods on the mechanical development projects. For example, during the project, the project individual had more and more stress because of the self-management and self-organization rather than having direct orders from their superiors. From the management's view this creates uncertainty of project. The entire team is actually held accountable for the success or failure of a project even when the product owner is to be held responsible. It is not easy to divide the project into sprints according to the functions of product, this issue will be analyzed when students use the proposals established next year. Differences between the traditional management of the project and the implementation of an agile approach is another criterion to be considered as part of the analysis of this integration.

5 CONCLUSION

This article aimed at deploying, testing and evaluating the issues and benefits of an agile approach on a student project, in order to determine if this project could be a candidate to train students to a basic practice of agile management.

To this purpose, this paper considered and compared two options to implement agility. In the first option, four sprints were defined with a specific duration and related activities; at the end of each sprint, students can deliver a functional part of the robot and test it. In the second option, the sprint definition relied on the subsystems involved to achieve each phase of the global mission of the robot; it allows delivering software and hardware modules in parallel. During the sprint development, in both cases, close communication should be effective in order to cover all the requirements defined in the robot specifications. Specific roles should be considered to manage the evolution of the robot development; these roles allow the distribution of responsibilities during the development.

The paper concludes that implementing an agile method such as Scrum in this project during the system development could be interesting for students to improve cooperation among students, and quickly and efficiently develop the connected mobile robot, thus improving project performance. It also concludes that implementing Scrum in

this project could allow to train students to basics of agile management.

REFERENCES

- Azanha, A., Argoud, A.R.T.T., Camargo Junior, J.B.D., and Antonioli, P.D., 2017. Agile project management with Scrum: A case study of a Brazilian pharmaceutical company IT project. *International Journal of Managing Projects in Business*, 10(1), p.121-142.
- Augustine, S., Payne, B., Sencindiver, F. and Woodcock, S., 2005. Agile project management: steering from the edges. *Communications of the ACM*, 48(12), p.85-89.
- Beck, K. 1999. Embracing change with extreme programming. *Computer*, 32(10), p.70-77.
- Beck, K. 2004. *Extreme Programming Explained: Embrace Change*, 2nd ed. Addison-Wesley.
- Boehm, B., 2002. Get ready for agile methods, with care. *Computer*, 35(1), p.64-69.
- Bryde, D.J., 2003. Project management concepts, methods and application. *International Journal of Operations & Production Management*, 23(7), p.775-793.
- Cardozo, E.S., Neto, J.B.F.A., Barza, A., França, A.C.C. and da Silva, F.Q., 2010, April. SCRUM and Productivity in Software Projects: A Systematic Literature Review. In *EASE*.
- Cervone, H.F., 2011. Understanding agile project management methods using Scrum. *OCLC Systems & Services: International digital library perspectives*, 27(1), p.18-22.
- Misra, C. S., Kumar, V. and Kumar, U., 2010. Identifying some critical changes required in adopting agile practices in traditional software development projects. *International Journal of Quality & Reliability Management*, 27(4), pp.451-474.
- Chin, G., 2004. Agile project management: how to succeed in the face of changing project requirements. *AMACOM Div American Mgmt Assn*.
- Cockburn, A., 2002. Learning from agile software development—part one. *CrossTalk Oct*, p.10-14.
- Cohn, M. and Ford, D., 2003. Introducing an agile process to an organization [software development]. *Computer*, 36(6), p.74-78.
- Fowler, M. 2000. The new methodology, available at: www.martinfowler.com/articles/newMethodologyOriginal.html (accessed 19 December 2014).
- Highsmith, J., 2009. *Agile project management: creating innovative products*. Pearson Education.
- Hoda, R., Noble, J. and Marshall, S., 2010, May. Organizing self-organizing teams. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering—Volume 1* p. 285-294.
- Indelicato, G., 2012. Making sense of agile project management: balancing control and agility. *Project Management Journal*, 43(3), p.78-78.

- Kotter, J.P., 2009. Leading change: why transformation efforts fail. *Harvard Business Review*, 73(2).
- Kruchten, P., 2004. *The rational unified process: an introduction*. Addison-Wesley Professional.
- Project Management Institute (PMI), 2017. *Agile Practice Guide*
- Robillard, P. N., Kruchten, P. and d'Astous, P. 2001. Yoopeedoo (UPEDU): a process for teaching software process. *ResearchGate*, p. 18–26.
- Scheid, J. 2015. *Project Management Methodologies: How Do They Compare?* [online] <http://www.brighthubpm.com/methods-strategies/67087-project-management-methodologies-how-do-they-compare/>
- Schwaber, K., Sutherland, J., 2017. *The definitive guide to scrum: the rules of the game*. <http://www.scrumguides.org/docs/scrumguide/v1/scrum-guide-us.pdf>.
- Serrador, P. and Pinto, J.K., 2015. Does Agile work?—A quantitative analysis of agile project success. *International Journal of Project Management*, 33(5), p.1040-1051.
- Shahid, N., Khan, O.A., Anwar, S.K. and Pirzada, U.T., 2009. Rational unified process. *Online Notes on RUP*. http://ovais.khan.tripod.com/papers/Rational_Unified_Process.pdf. [Accessed: 19-Oct-2016].
- Shenhar, A.J. and Dvir, D., 2007. *Reinventing project management: the diamond approach to successful growth and innovation*. Harvard Business Review Press.
- “UPEDU”, [online]. Available at: <https://en.wikipedia.org/wiki/UPEDU> [Accessed 19 December 2017].

Inferring Extended deterministic Labelled Transition System models of Reactive System Behaviours

Luigi Novella, Manuela Tufo, Giovanni Fiengo

Department of Engineering
 Università del Sannio
 Piazza Roma 21
 82100 Benevento - Italy

luigi.novella@unisannio.it, manuela.tufo@unisannio.it, gifengo@unisannio.it

ABSTRACT: *Interest in reactive systems behaviour modelling is broadly growing both in academy and industry. The availability of such models enables the analysis, understanding, simulation and control of these systems and their interaction with the corresponding operating environment. The behaviour of reactive systems can often be specified using finite state machines as mathematical models. The manual construction of these models is a non-trivial task and requires a deep domain knowledge. System identification can be used in order to derive such models automatically from scenarios of interactions between the system and its environment. The paper proposes a model inference technique that automatically infers an Extended deterministic Labelled Transition System (ELTS) model of a reactive system during its activity. The approach treats the System Under Learning (SUL) as a black box. It consists of two different processes. The first one aims to explore and on-the-fly learn a raw ELTS model of the SUL. The second one consists of refining the model via inductive learning. The proposed state-merging based learning algorithm uses a novel heuristic approach according to the order in which states are chosen for merging. The approach's effectiveness is demonstrated through the modelling of a Graphical User Interface (GUI) of a mobile application.*

KEYWORDS: *Modelling, Machine Learning, Grammatical Inference.*

1 INTRODUCTION

A Discrete Event System (DES) is a dynamic system that evolves in agreement with the occurrence of a physical event. In the industry domain, the interest in the automated building of discrete-event models is broadly growing. These enable the analysis, understanding, simulation and control of the systems and their interaction with the corresponding operating environment. Referring to this context, the availability of DES model interests several applications such as reverse engineering for (partially) unknown systems, fault diagnosis, control or system verification. The process to inferring a model of the dynamics governing a system's interaction with its environment is called *system identification*. The identification of a DES consists of defining the mathematical model that describes its behaviour from the observation of the evolution of inputs and outputs and/or from other knowledge. Observations describe the data revealing the system activity and the models reproduce the observed behaviour. As shown in literature ((Cicchello and Kremer, 2003),(Walkinshaw et al., 2016),(Estrada-Vargas et al., 2017) and references therein), several approaches related to identification of DES are discussed. The first contributions

that represent the theoretical basis of DES identification procedures have been formulated in terms of grammar inference (Gold, 1967). In this context, the aim is to identify a language from a sample of accepted words: such methods are considered as *learning techniques* (Estrada-Vargas et al., 2010). The problem of grammar inference, as first identified by Gold in his fundamental paper (Gold, 1967), is to identify a language L given a set of sample sequences S . The set of samples must contain a set of positive samples S^+ that belong to the language and can optionally contain a set of negative samples S^- that do not. When L is a regular language, its grammar can be represented by a Deterministic Finite Automaton (DFA) and the problem can be re-interpreted as one that aims to identify the DFA that produces L . Many inference techniques based on the idea of (Gold, 1967) have been proposed in literature. Angluin's L^* technique (Angluin, 1987) constructs queries of various types that systematically explore the target system's behaviour, and in the process defines a complete DFA for the system. An alternative approach to Angluin's L^* technique is to shift the emphasis from the oracle that provides input during the learning process to the quality of the initial set of sample sequences S . State-merging tech-

niques are examples of such alternative approaches. They require less guidance from an oracle, but still can produce an accurate state machine on the condition that the initial set of samples S covers enough of the target machine. The Evidence Driven State Merging (EDSM) algorithm with blue-fringe ordering is considered to be the *state of the art* with respect to the inference of DFAs from positive and negative examples by merging of states. It won the Abbadingo competition (Lang et al., 1998) and was used as a baseline in (Walkinshaw et al., 2013). The algorithm starts from a tree-shaped automaton, the so-called *Prefix Tree Acceptor (PTA)*, that accepts the positive sample S^+ only, and successively merges states to generalize the induced language. There are a lot of variants of the same underlying algorithm. The order in which pairs of states are considered for merging is the key difference between them. In all cases, the generalization is controlled by the negative sample S^- to prevent merging incompatible states. The availability of negative information is theoretically motivated since positive and negative samples are required to identify in the limit any super-finite class of languages, including the regular language class (Gold, 1967). However, when few negative strings are available, one can rely on another kind of knowledge, typically provided by the application domain. For example, when some additional information is available about incompatibilities between states of the initial automaton, state-merging algorithms can easily be extended to avoid merging states that are known to be incompatible (Lambeau et al., 2008). The proposed method follows a black box identification approach to infer a behaviour model of reactive systems starting from observations obtained during the system activity. It allows to yield compact and expressive *Extended deterministic Labelled Transition System (ELTS)* models. The inference procedure is controlled by domain additional knowledge, unfeasible behaviours of the system are not provided. It consists of two different processes. The first one aims to explore and on-the-fly learn a raw ELTS model of the System Under Learning (SUL), the second one consists of refining the model via inductive learning. We use the basic steps of the EDSM algorithm as a basis for our ELTS inductive learning algorithm. It uses a novel heuristic approach according to the order in which pairs of states are chosen for merging that drastically reduces the number of required merge operations. The introduced heuristic approach of measuring the evidence of equivalence identifies incompatible pairs of states as soon as possible and without actually merging them. This is done using information provided by the state labelling function incorporated in the definition of ELTS. The paper is organized as follows. Section 2 describes the importance of *model learning* in the industry domain. In section 3 the synthesis of reactive systems' behaviour models is described as

an ELTS induction problem and the proposed overall model inference technique is illustrated. Finally, in section 4 the approach's effectiveness is demonstrated through the modelling of a mobile app GUI.

2 MODEL LEARNING

Modelling plays a key role during the system's development or during the analysis of an existing system. It aims to define a mathematical model representing the behaviour exhibited during the system operating. In DESs, the models are represented from abstract machines that reproduce the observed sequences of events. To build a model from observations, several approaches have been proposed in literature. Our solution has been addressed as a problem of grammatical inference for obtaining finite automata that represent a given language. This method is reported in literature as *model learning*. It represents an effective method for black box state machine models of reactive systems. Theoretical grammatical inference research continues to advance in many different directions such as the definition of new language classes being learned, the application of the learned models for the synthesis of control strategies (i.e. design of supervisors for complex systems), and the efficiency of the inference algorithms. In the recent years, beyond the academic communities, also the companies are interested in applying identification methods to industrial systems. These techniques must cope with large and complex systems that handle actual signals. The main applications regard the reverse engineering for unknown systems, fault diagnosis, control or system verification and validation. Model learning is emerging not only as a highly effective validation technique, with applications in several areas such as banking cards, network protocols, and legacy software but also as control methods for systems with complex behaviour and large state-spaces (Skoldstam et al., 2007).

3 REACTIVE SYSTEMS' BEHAVIOUR MODELS SYNTHESIS SEEN AS AN ELTS INDUCTION PROBLEM

The behaviour of reactive systems can often be specified using Finite State Machine (FSM) models (Sindhu, 2013). These models provide an overview of such systems, by describing the way in which they react to different inputs, and when they produce which output. Unfortunately, the construction of finite state machine models is often omitted during the development stage. An alternative to constructing these models manually, is to use system identification (or model synthesis/learning) in order to derive them automatically from scenarios of interactions between the system and its environment. Intuitively, reactive systems model synthesis tries to discover the lo-

gical structure underlying these sequences of events. This can be seen as a grammatical inference problem. Scenarios can be represented as strings over an alphabet of possible events and they can be generalized to form a language of acceptable behaviours. Whenever behaviours are modelled by FSMs, the problem becomes equivalent to FSM induction from positive examples only.

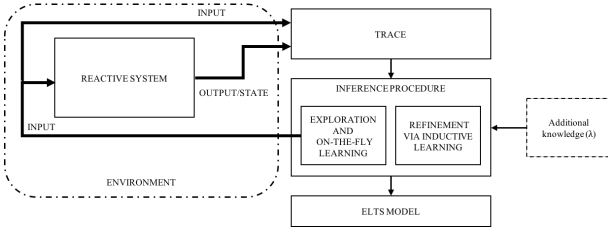


Figure 1 – System identification procedure

We restrict discussion to reactive systems that can be modelled as an ELTS. With reference to the scheme in figure 1, the overall proposed model inference technique treats the SUL as a black box and consists of two different processes. The first one aims to on-the-fly learn a hypothesis model of the reactive system behaviour during its activity. At each step, the inferred model will represent what has already been explored and new input sequences will be given to the system to learn unvisited states. This helps to build a deep and broad set of scenarios of interactions between the reactive system and its environment, that is a representative set of accepted behaviours. As soon as an inconsistency between the current ELTS hypothesis and the SUL is found, starting from the collected set of traces of system behaviour, the model is refined via inductive learning. The overall inference procedure is controlled by domain additional knowledge, unfeasible behaviours of the system (counterexamples) are not provided. The process terminates by means of an equivalence check between the complete inferred model and the SUL.

3.1 EXTENDED LABELLED TRANSITIONS SYSTEMS AND TRACES OF SYSTEM BEHAVIOUR

An ELTS can be defined as an extension of a Labelled Transition System (LTS) which introduces a *state labelling function* as in (Choi et al., 2013). Our mathematical formulation introduces also the concept of *language of an ELTS*. Formally an ELTS is a quintuple

$$ELTS = (Q, q_0, \Sigma, \delta, \lambda) \quad (1)$$

where:

- Q is a finite set of states,
- q_0 is the initial state,
- Σ is a finite input alphabet,

- $\delta : Q \times \Sigma \rightarrow Q$ is a partial (state) transition function (δ is not always defined). For a state $q \in Q$ and an input symbol $a \in \Sigma$, the transition function δ returns the state p that is reached when starting from q and processing the input symbol a ($\delta(q, a) = p$),
- $\lambda : Q \rightarrow \wp(\Sigma)$ is a state labelling function. $\lambda(q)$ denotes the set of inputs enabled at state q , that is a subset of Σ .

If δ is the partial transition function, then the extended transition function is denoted by $\hat{\delta}$. For a state q and a string w , the extended transition function $\hat{\delta}$ returns the state p that is reached when starting from state q and processing sequence w (Hopcroft et al., 2001). For the base case $\hat{\delta}(q, \epsilon) = q$ where ϵ denotes the empty string. For the inductive case, let w of the form xa , where a is a suffix of w , and x is the prefix then $\hat{\delta}(q, w) = \delta(\hat{\delta}(q, x), a)$.

Definition 1 Given some ELTS A , for a given state $q \in Q$, $L(A, q)$ is the language of A in state q , and can be defined as: $L(A, q) = \{w | \hat{\delta} \text{ is defined for } (q, w)\}$. The language of a ELTS A can be defined as: $L(A) = \{w | \hat{\delta} \text{ is defined for } (q_0, w)\}$.

The trace-encoding process and the concept of consistency are identical in principle to that described by (Choi et al., 2013). We presume that the interactions with a reactive system, modelled as an ELTS, can be characterized in terms of event names.

Definition 2 A trace of system behaviour t is a finite sequence of pairs of events and sets of enabled inputs at each traversed state q , starting from q_0 (initial state). Formally a trace t is an element of $(\Sigma \times \wp(\Sigma))^*$.

Definition 3 Given a set of traces T , L_T denotes the set of events labels in the traces in T (known as the ELTS alphabet Σ).

Definition 4 A trace $t = (a_1, \Sigma_1), \dots, (a_n, \Sigma_n)$ is consistent with a model ELTS $= (Q, q_0, \Sigma, \delta, \lambda)$ if and only if

$$\bigwedge_{i \in [1, n]} q_{i-1} \xrightarrow{a_i} q_i \wedge \lambda(q_i) = \Sigma_i, \quad q_1, \dots, q_n \in Q \quad (2)$$

Definition 5 A state in an ELTS is a frontier-state if there is no transition from the state on some input that is enabled on the state. Formally, a state q is a frontier-state if there exists $a \in \lambda(q)$ such that $q \xrightarrow{a} p$ is not true for any $p \in Q$.

3.2 EXPLORATION AND BEHAVIOUR MODEL ON-THE-FLY LEARNING

The exploration and on-the-fly learning procedure aim to automatically build a raw ELTS model of the

reactive system behaviour during its activity and to collect a large set of scenarios of interactions between the system and its environment. In this context, we refer to reactive systems that do not produce any output. The overall proposed model inference approach is similar to the one described in (Choi et al., 2013) where the authors address the problem of automatically generating sequences of test inputs for Android applications. In the referenced work the goal is to achieve the application code coverage quickly by learning and exploring an abstraction of the model of the GUI with minimal restart of the application. They propose an automatic technique, called *SwiftHand*. It combines *active learning* with testing to explore a running application and to on-the-fly learn a model of its GUI. Then it uses the execution traces on the generated inputs to refine the model via passive learning. Our inference procedure completely differs in the way the ELTS model is refined from the set of system behaviour traces T as shown with detail in section 3.3. Regarding the exploration and on-the-fly learning procedure, we start from the procedure described in (Choi et al., 2013). In order to generalize the overall proposed inference technique to identify the behavioural model of several reactive systems, we have modified it removing the heuristic used by *SwiftHand* to pick a state among the ones available for merging. In (Choi et al., 2013) the authors state that “*we discovered after some experimentation that if we prefer the nearest ancestor to other states, our merge operations are often correct*” so it is clear that this heuristic is strictly related to the specific SUL and to avoid passive re-learning in the future. Our on-the-fly learning procedure randomly picks a model-state for merging in the case of multiple model-states whose state labelling functions values match the set of enabled inputs of the reached SUL state. Even if a random choice could produce a not consistent ELTS, our proposed ELTS inductive algorithm refines it due to the proposed new measure of evidence the two states are equivalent. Each time the current trace t is not consistent with the current model hypothesis (Definition 4), the model is immediately refined from T via the ELTS inductive algorithm.

3.3 BEHAVIOUR MODEL REFINEMENT VIA INDUCTIVE LEARNING

As is the case with most behavioural model inference techniques, the one presented here takes a set of system execution traces as an input. We use the basic steps of the EDSM algorithm that aims to infer a DFA from positive and negative examples in the “blue-fringe” framework (Lang et al., 1998) as a basis for our ELTS learning algorithm. Unlike many similar machine learning techniques, we cannot presume the presence of negative examples. In our scen-

ario we only have system execution traces to work from, which are positive by definition (Walkinshaw et al., 2016). When few or no negative examples are available, one can sometimes rely on another kind of knowledge, typically provided by the application domain (Lambeau et al., 2008). The algorithm proposed uses a novel heuristic approach according to the order in which state pairs are chosen for merging. The basic idea is to consider the state labelling function, directly incorporated in the ELTS model, to measure the evidence that two states are equivalent.

3.3.1 The Algorithm

Algorithm 1 describes the state-merging based ELTS learning algorithm. It begins by arranging the set of traces into an initial ELTS in the form of a *Prefix Tree Acceptor (PTA)* (Walkinshaw et al., 2016). It is a tree-shaped ELTS whose states are the set of all prefixes of the traces in T (line 2). Starting off from the PTA and according to the blue-fringe framework, the algorithm iteratively chooses a pair of states deemed to be a suitable merge-candidate (lines 18-19), makes the merge (line 30) and promotes a state if no merge is possible (line 24) until no further pairs of equivalent states can be found, which indicates convergence at the final hypothesis ELTS. As in the blue-fringe framework, the merge always occurs between a *red* state and a *blue* state (line 36). They are always the roots of a tree. Formally, the red states correspond to states that have been analysed and which will not be revisited by the inductive learning process. They will be the states of the final ELTS. The blue states are the candidate states that have not been analysed yet and it should be from this set that a state is drawn in order to consider merging it with a red one (Lang et al., 1998; de la Higuera, 2010). The state merging frequently introduces non-determinism into the hypothesis, which can then be removed by the classical *determinization procedure* that recursively merges the children of the original nodes (Lang et al., 1998; Walkinshaw et al., 2016; Lambeau et al., 2008). The symmetrical merging operation, which requires determinization through a *cascade of merges* is replaced in our approach by the simpler asymmetric folding operation (*Fold*, lines 42-53) as in (de la Higuera, 2010). Our ELTS inductive learning algorithm avoids non-deterministic ELTSs. The quality of the learned model solely depends on the order in which pairs of states are chosen for merging. As an evidence-driven approach, the order by which pairs of states are considered for merging is determined here by a score computed in the *CalculateScore* procedure (lines 55-69). The pair of states with the associated highest score is iteratively chosen for merging. The higher is the score, the higher the evidence that the red-blue states are equivalent. In section 3.3.2, the novel heuristic approach according to the state-merging ordering is described.

Algorithm 1 ELTS Inductive Learning Algorithm

```

1: procedure REFINET( $T$ )
2:   ELTS  $\leftarrow$  BUILDPTA( $T$ )
3:   red  $\leftarrow$   $\{q_0\}$ 
4:   blue  $\leftarrow$   $\{\delta(q_0, a), a \in \Sigma\}$ 
5:   while blue  $\neq$   $\emptyset$  do
6:     promotion  $\leftarrow$  false
7:     bs  $\leftarrow$   $-\infty$ 
8:     for  $q_b \in$  blue do
9:       if not promotion then
10:        atleastonemerge  $\leftarrow$  false
11:        for  $q_r \in$  red do
12:          score  $\leftarrow$  CALCULATESCORE(ELTS,  $q_r, q_b$ )
13:          if score  $>$   $-\infty$  then
14:            atleastonemerge  $\leftarrow$  true
15:          end if
16:          if score  $>$  bs then
17:            bs  $\leftarrow$  score
18:             $\overline{q_r} \leftarrow q_r$ 
19:             $\overline{q_b} \leftarrow q_b$ 
20:          end if
21:        end for
22:        if not atleastonemerge then
23:          promotion  $\leftarrow$  true
24:          red, blue  $\leftarrow$  PROMOTE( $q_b, ELTS, red, blue$ )
25:        end if
26:      end if
27:    end for
28:    if not promotion then
29:      blue  $\leftarrow$  blue  $\setminus$   $\{\overline{q_b}\}$ 
30:      ELTS  $\leftarrow$  MERGE( $\overline{q_r}, \overline{q_b}, ELTS$ )
31:    end if
32:  end while
33:  return ELTS
34: end procedure
35:
36: procedure MERGE( $q, q', ELTS$ )  $\triangleright q \in red, q' \in blue$ 
37:   Let  $(q_f, a)$  be such that  $\delta(q_f, a) = q'$ 
38:    $\delta(q_f, a) \leftarrow q$ 
39:   return FOLD( $q, q', ELTS$ )
40: end procedure
41:
42: procedure FOLD( $q, q', ELTS$ )  $\triangleright q'$  being the root of a tree
43:   for  $a \in \Sigma$  do
44:     if  $\delta(q', a)$  is defined then
45:       if  $\delta(q, a)$  is defined then
46:         ELTS  $\leftarrow$  FOLD( $q, q', ELTS$ )
47:       else
48:          $\delta(q, a) \leftarrow \delta(q', a)$ 
49:       end if
50:     end if
51:   end for
52:   return ELTS
53: end procedure
54:
55: procedure CALCULATESCORE(ELTS,  $q_r, q_b$ )
56:   score  $\leftarrow$  0
57:   if  $\lambda(q_r) \neq \lambda(q_b)$  then
58:     score  $\leftarrow$   $-\infty$ 
59:     return score
60:   else
61:     while  $(t_1, t_2) \leftarrow$  EQUIVALENTTRANSITIONS( $q_r, q_b$ ) do
62:       score  $\leftarrow$  score + 1
63:        $d_1 \leftarrow \delta(q_r, t_1)$ 
64:        $d_2 \leftarrow \delta(q_b, t_2)$ 
65:       score  $\leftarrow$  score + CALCULATESCORE(ELTS,  $d_1, d_2$ )
66:     end while
67:   end if
68:   return score
69: end procedure

```

3.3.2 Heuristic State-Merging Ordering

The novel heuristic approach according to the state-merging ordering in the blue-fringe framework aims to identify as soon as possible the pair of states in the current ELTS hypothesis that are equivalent. As already mentioned before, the model induction learning procedure is governed by the ELTS state labelling function. States having the same state labelling function may be merged, states that disagree

on the state labelling function must not be merged. This clearly explains how the generalization process is possible even if there is no notion of negative examples. Our heuristic approach uses the information provided by the state labelling function to determine the state-merging ordering. This is performed by scoring each red-blue pair of states without actually merging it. Finally, the state pair with the highest score is deemed to be most likely to be equivalent. Let (q_r, q_b) be a candidate pair of states for merging, where q_r is a red state and q_b a blue one and let A be the current ELTS hypothesis. We define $L(A, q_r)$ and $L(A, q_b)$ as the languages of A in q_r and q_b respectively and $L_i = L(A, q_r) \cap L(A, q_b)$ as the set of common words w which can be recognized by A when starting in states q_r and q_b . Formally, the score associated to the pair of states (q_r, q_b) is computed as follows:

- a null score is associated if L_i is the empty set,
- provided that states q_r and q_b have the same state labelling functions, a positive score is associated if while processing all the w in L_i starting from q_r and q_b no target states with different state labelling functions are encountered in the outgoing paths. In this case the score is equal to the L_i cardinality ($|L_i|$),
- a negative score is associated when states q_r and q_b have different state labelling functions or in the case that, while processing all the w in L_i starting from q_r and q_b , for at least one of them, two target states with different state labelling functions are encountered in the outgoing paths. In this case the score is equal to $-\infty$.

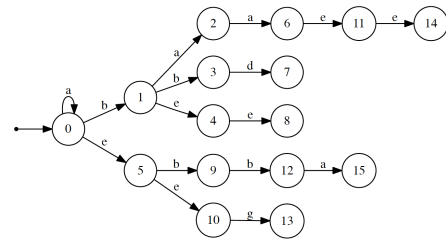


Figure 2 – Current ELTS hypothesis

For example, let $A = (Q, q_0, \Sigma, \delta, \lambda)$ be the current ELTS hypothesis shown in figure 2 and $\{0\}$, $\{1, 5\}$ the current sets of red and blue states respectively. Let $(q_r, q_b) = (0, 1)$ be a candidate pair of states for merging. The set of common words L_i recognized by A when starting from states 0 and 1 is $L_i = \{‘a’, ‘aa’, ‘aae’, ‘aaee’, ‘e’, ‘ee’, ‘b’\}$. In order to compute the score associated to the red-blue pair of states $(q_r, q_b) = (0, 1)$, each word w in L_i is processed starting from both q_r and q_b . These states have the same state labelling function and the initial score is set equal to zero. Let’s suppose to process the word

$w = 'aae'$, we obtain:

$$\hat{\delta}(0, 'aae') = 5, \quad \hat{\delta}(1, 'aae') = 11 \quad (3)$$

The encountered pair of states in the outgoing paths that lead to states 5 and 11, starting from states 0 and 1 have the same state labelling functions as it is shown in table 1, so the score is incremented by one.

A states	λ
{0, 1, 2, 3, 4, 5, 6, 11}	{'a', 'b', 'c', 'd', 'e', 'f'}
{8, 10, 14}	{'a', 'h', 'b', 'g'}

Table 1 – λ s of the encountered states from q_r and q_b

This is done for each word w in L_i . In this case the traversed states in the outgoing paths from states 0 and 1 have always the same λ , then a score equal to $|L_i| = 7$ is associated to the pair of states $(q_r, q_b) = (0, 1)$.

Let $(q_r, q_b) = (0, 5)$ be the other candidate red-blue pair of states for merging. The set of common words L_i recognized by A when starting from states 0 and 5 is $L_i = \{'bb', 'e'\}$. The score associated to this pair of states is equal to $-\infty$ because while processing the word $w = 'e'$ starting from states 0 and 5, the traversed states 5 and 10 disagree on λ . Finally, the selected red-blue pair of states is $(0, 1)$ and the inductive learning algorithm proceeds to merge it. The score associated to each red-blue pair of states of the current hypothesis is computed by the `CalculateScore` procedure. Informally speaking, it operates by counting the number of transitions in the outgoing paths from the two argument states that share the same labels (this is always finite because one of the states is always the root of a tree (Lang et al., 1998)) provided that the target states of such equivalent transitions agree on the state labelling function. As it is demonstrated, the proposed approach evaluates a red-blue merge score without actually performing it. The heuristic approach of measuring the evidence of equivalence identifies the incompatible states as soon as possible, before actually performing the merge. Differently from the heuristic approach according to the order in which state pairs are chosen for merging in (Choi et al., 2013), our approach results in a drastic decrease of the number of merges performed, especially in the case of large-sized target ELTS. In (Choi et al., 2013) the authors present a state-merging based learning algorithm with blue-fringe ordering and blocking constraints to infer an ELTS model of the subject system from a set of execution traces. A score is associated to each red-blue pair of states by merging it and without taking into account the information provided by the state labelling function applying the original blue-fringe ordering (Lang et al., 1998). As a matter of fact, the merging procedure first checks whether the selected pair of states agrees on the state labelling function, and then tries to merge them. If merging results in a

nondeterministic model, the algorithm tries to eliminate non-determinism by merging target states of non-deterministic transitions provided that the merged states have the same state labelling function. If at some point of the procedure, merging of two states fails because state labelling functions of the states are different, the algorithm unrolls the entire merging process for the original red-blue pair of states. In our approach, the score associated to each red-blue pair of states is computed without actually merging it and by taking into account the information provided by the state labelling function. The merge of a pair of states is performed only if the associated score is positive. An unroll procedure is therefore not needed.

4 IMPLEMENTATION AND EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the inference procedure illustrated in section 3 we consider as a reactive system the GUI of an Android application. A GUI for a mobile application or for a computer program can be thought of as a reactive system (Sindhu, 2013; Abadi et al., 2013). In particular, as our running example, we decide to infer the model of the *Sanity* app GUI. It is the same Android application used in (Choi et al., 2013) as running example for testing purposes. This is done to emphasize the improvements of the proposed ELTS inductive learning algorithm. Our implementation really treats the reactive system as a black box, no application code instrumentation is performed. The overall inference procedure is written in the Python language and makes use only of the Android Debug Bridge (ADB) command-line tool to control any device connected to the computer. Without loss of generality, we consider a reduced set of enabled user inputs at each app-state. For this reason a partial ELTS model, the one corresponding to the process for accepting the end-user licence agreement, of the *Sanity* app GUI is shown as the output of the ELTS model inference procedure (figure 5). Each screen shown in figure 3 represents a *Sanity* GUI app-state. For each one of these, an ELTS model-state is abstracted based on the set of user inputs enabled. We dynamically detect the set of enabled user inputs at each app-state during the application execution. Moreover, it is assumed that the *Back* and *Menu* buttons are present in all the application screens, then they are added to the set of enabled inputs for each app-state. The set of enabled user inputs at each app-state is modelled as the state labelling function λ for each model-state in the ELTS. The overall inference procedure installs and launches the *Sanity* app. The initial app-state is reached, where the set of enabled user inputs is $\{Ok, Cancel, ScrollUp, ScrollDown, Back, Menu\}$ (figure 3(a)). We presume that the interactions with the SUL can be characterized in terms of alphabet

symbols (Σ) that abstract the enabled user inputs at each app-state. In table 2, it is shown the correspondence between the ELTS alphabet Σ and the set of real user inputs. The initial app-state is abstracted as the initial model-state q_0 with $\lambda(q_0) = \{e, f, c, d, a, b\}$.

Σ	User inputs
'a'	Back
'b'	Menu
'c'	ScrollUp
'd'	ScrollDown
'e'	Ok
'f'	Cancel
'h'	ACTIVATE
'g'	CANCEL

Table 2 – Σ - real user inputs

At each step the algorithm selects a frontier state in the ELTS model to continue with the exploration. In this case, the algorithm picks the only frontier state q_0 and randomly selects one of its enabled inputs. Assume that the input e is selected and the corresponding real input 'Ok' is executed against the SUL, the resulting app-state (figure 3(b)) has the same set of enabled user inputs of the initial app-state therefore a merge is performed in the current ELTS model.

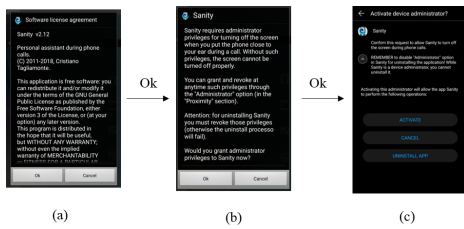


Figure 3 – Sanitty app-states

The app is now at an app-state (figure 3(b)) whose corresponding model-state is still q_0 (current model-state), so the algorithm can only pick this model-state as the next frontier-state to explore. A path from the current model-state to the selected frontier-state consists of the single transition e . After executing the corresponding user input (Ok) against the SUL, the algorithm discovers an inconsistency. The reached app-state (figure 3(c)) has a set of enabled user inputs ($\{ACTIVATE, CANCEL, Back, Menu\}$, input $UNINSTALL APP$ is not taken into account) different from the ones indicated by $\lambda(q_0) = \{e, f, c, d, a, b\}$ (table 2). The algorithm has discovered that the learned ELTS model is not consistent with the execution traces observed so far (Definition 4). In order to resolve the inconsistency, the proposed inductive learning algorithm refines the current ELTS model from scratch by using the collected set of execution traces (T). Figure 4 shows the output of the ELTS inductive learning algorithm. It represents the refined ELTS model consistent with the set of execution traces. Similar to figure 3, in figure 4

it is indicated the real user input Ok rather than the corresponding input alphabet 'e'.

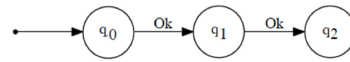


Figure 4 – ELTS refined model

Once the model is refined, the on-the-fly learning process continues exploring unvisited app-states of the SUL until the model is complete (no more frontier states are available). Each time an inconsistency is found during the SUL exploration, the ELTS inductive learning algorithm is executed to rebuild the model starting from the execution traces. Our goal is to infer a complete and consistent ELTS model representing the behaviour of the reactive system under learning. The ELTS model is complete as soon as there are not more frontier states (every transition from every model-state has been taken). However, to confirm that the complete ELTS model is equivalent to the SUL a model checking similar to the Angluin's L^* equivalence check has been implemented. It consists of selecting a sequence of user inputs that is not a subsequence of any collected execution trace. In the case the algorithm fails to find this sequence of inputs we can confirm that the inferred ELTS model is equivalent to the SUL. Otherwise, the algorithm executes the selected sequence of inputs and checks if the corresponding execution trace is consistent with the ELTS model. If an inconsistency is found then the model is re-learned from the updated set of execution traces. The final output of the overall inference procedure is shown in figure 5. It represents a partial ELTS model of the Sanitty app GUI.

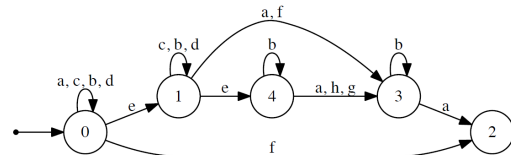


Figure 5 – Complete and consistent ELTS model of the Sanitty app GUI

The main advantage of the blue-fringe method is that it shrinks the pool of candidates for merging to the red-blue fringe (Lang et al., 1998). However, the main drawback of the EDSM strategy is due to the cost of evaluating each pair of mergeable states for each merge carried out. The proposed novel heuristic approach assigns a score to each red-blue pair of states without performing the merge operation. This is done by taking into account the domain knowledge encapsulated in the ELTS state labelling function. Having said that, in order to demonstrate the effectiveness of the proposed ELTS inductive learning algorithm we have measured the number of merge required to obtain the

target machine. This measure has been successively compared with the one that can be obtained by applying the approach described in (Choi et al., 2013) in the same experimental conditions. Our ELTS inductive learning algorithm executed only 12 merge operations in contrast with the 69 performed the ELTS passive learning algorithm presented in (Choi et al., 2013). The proposed heuristic approach according to the order in which states are chosen for merging drastically reduces the number of required merge operations because it does not perform incorrect merging due to the new measure of evidence that two states are equivalent.

5 CONCLUSION

Applications of system identification and machine learning techniques to automatically build reactive systems behavioural models are broadly growing both in academy and industry. In this paper, a black box identification technique with the aim of automatically inferring a behavioural ELTS model of the SUT during its activity is presented. Our ELTS inductive learning algorithm proposes a novel heuristic approach according to the state-merging ordering. The heuristic exploits domain knowledge to measure the evidence that two states are equivalent and identifies incompatible states as soon as possible to avoid incorrect merging. This results in a drastic decrease of the number of merges needed for obtaining a behavioural ELTS model equivalent to the SUT. In the future research, we want to combine our approach with supervisory control theory with the aim of solving the control problem of black box reactive systems. Moreover, even if the EDSM with blue-fringe ordering, which is behind our inductive learning algorithm, is considered as the state of the art with respect to the FSM inductive learning algorithms, more can be done from an empirical point of view. It is always possible to improve the model inference process exploiting the domain knowledge if available. Regarding this point, we intend making use of more domain knowledge.

REFERENCES

- Abadi, A., Feldman, Y. A. and Shagin, K. (2013), A screen-oriented representation for mobile applications, in ‘Proceedings of the 2013 Companion Publication for Conference on Systems, Programming, & Applications: Software for Humanity’, SPLASH ’13, ACM, New York, NY, USA, pp. 79–80.
- Angluin, D. (1987), ‘Learning regular sets from queries and counterexamples’, *Inf. Comput.* **75**(2), 87–106.
- Choi, W., Necula, G. and Sen, K. (2013), ‘Guided gui testing of android apps with minimal re-start and approximate learning’, *SIGPLAN Not.* **48**(10), 623–640.
- Cicchello, O. and Kremer, S. C. (2003), ‘Inducing grammars from sparse data sets: A survey of algorithms and results’, *J. Mach. Learn. Res.* **4**, 603–632.
- de la Higuera, C. (2010), *Grammatical Inference: Learning Automata and Grammars*, Cambridge University Press.
- Estrada-Vargas, A. P., López-Mellado, E. and Lesage, J. (2017), ‘A black-box identification method for automated discrete-event systems’, *IEEE Trans. Automation Science and Engineering* **14**(3), 1321–1336.
- Estrada-Vargas, A. P., López-Mellado, E. and Lesage, J.-J. (2010), ‘A Comparative Analysis of Recent Identification Approaches for Discrete-Event Systems’, *Mathematical Problems in Engineering* **2010**, 10.1155/2010/453254. 21 pages.
- Gold, E. M. (1967), ‘Language identification in the limit’, *Information and Control* **10**(5), 447–474.
- Hopcroft, J., Motwani, R. and Ullman, J. (2001), *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley series in computer science, Addison-Wesley.
- Lambeau, B., Damas, C. and Dupont, P. (2008), *State-Merging DFA Induction Algorithms with Mandatory Merge Constraints*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 139–153.
- Lang, K. J., Pearlmutter, B. A. and Price, R. A. (1998), *Results of the Abbadingo one DFA learning competition and a new evidence-driven state merging algorithm*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12.
- Sindhu, M. (2013), Algorithms and Tools for Learning-based Testing of Reactive Systems, PhD thesis, KTH, Theoretical Computer Science, TCS. QC 20130312.
- Skoldstam, M., Akesson, K. and Fabian, M. (2007), Modeling of discrete event systems using finite automata with variables, in ‘2007 46th IEEE Conference on Decision and Control’, pp. 3387–3392.
- Walkinshaw, N., Lambeau, B., Damas, C., Bogdanov, K. and Dupont, P. (2013), ‘Stamina: a competition to encourage the development and assessment of software model inference techniques’, *Empirical Software Engineering* **18**(4), 791–824.
- Walkinshaw, N., Taylor, R. and Derrick, J. (2016), ‘Inferring extended finite state machine models from software executions’, *Empirical Software Engineering* **21**(3), 811–853.

INGENIERIE SYSTEME AGILE PAR MODELES SIMULABLES

P. GICQUEL

CIL4Sys Engineering – Incubateur Telecom ParisTech -
9 rue Dareau, 75014 Paris
philippe.gicquel@cil4sys.com

RESUME : *Le logiciel devient partout prédominant. Les grands industriels de l'aéronautique, du ferroviaire de l'automobile peuvent-ils s'inspirer des progrès rapides de l'ingénierie logicielle et en retirer des principes efficaces pour leur propres démarches d'ingénierie, elles-mêmes de plus en plus concernées par la part logicielle ? Peut-on vraiment appliquer les démarches agiles en ingénierie système et si oui comment le faire ? Dans cet article l'auteur sélectionne quelques grands principes agiles et éclaire un chemin pour appliquer ces principes en ingénierie système. Un exemple de mise en œuvre est présenté. Enfin les gains qui peuvent être attendus de la démarche sont proposés.*

MOTS-CLES : *ingénierie système agile modélisation simulation exigences.*

1 INTRODUCTION

L'augmentation continue de l'importance de l'électronique dans les grands systèmes tels que les avions, les trains, les automobiles, a transformé les enjeux des projets de conception. L'interaction de ces systèmes avec leur environnement extérieur dans un monde de systèmes de systèmes va continuer à complexifier la conception des comportements. Comment ces grands systèmes sont-ils conçus ? Comment les grands industriels maîtrisent-ils cette augmentation de complexité ?

Alors que la conception mécanique met en œuvre des outils qui permettent de visualiser instantanément le résultat de conception (outils CAO pour visualiser la géométrie, maillages par éléments finis pour la tenue mécanique, ...), les interactions logiques entre les différents calculateurs électroniques restent décrites « à la main » avec des phrases. Le niveau de complexité atteint par les grands systèmes conduit les équipes d'ingénieurs à manipuler des dizaines de milliers voire des centaines de milliers d'exigences. Pendant les mois passés à rédiger ces exigences, les entreprises travaillent dans le noir, en aveugle, car si chaque ingénieur peut voir clair sur son périmètre, nul ne peut juger de la robustesse et de la cohérence de l'ensemble. Les managers ne savent pas juger de l'avancement réel du projet, les responsables marketing-produit ne peuvent pas vérifier si ce qui est écrit dans les exigences traduit correctement les attentes qu'ils avaient exprimées, les acheteurs ne savent pas si les cahiers des charges qui vont être envoyés aux fournisseurs ont la qualité requise. En se référant à la description en V du cycle de conception, on peut dire que s'agissant de la logique comportementale des systèmes, la descente est effectuée dans un tunnel non éclairé.

2 L'EXEMPLE DE L'AUTOMOBILE

L'enquête menée en 2016 en Europe auprès d'une dizaine de constructeurs automobiles et équipementiers de rang un a amené les conclusions suivantes :

- Tous utilisent des cahiers des charges textuels y compris pour décrire la logique de fonctionnement, les exigences textuelles sont ensuite stockées dans un outil de gestion des exigences (DOORS le plus souvent),
- Le cycle en V est le processus de référence en développement, l'utilisation de démarches agiles se limite à des expérimentations locales en développement logiciel,
- Les tests sont généralement définis par des équipes dédiées, à posteriori de l'écriture des spécifications par les concepteurs,
- L'ingénierie système appuyée sur les modèles est déployée localement, domaine par domaine, avec des méthodes et outils hétérogènes.

3 L'INGENIERIE LOGICIELLE COMME INSPIRATION

L'ingénierie logicielle a connu une explosion de la complexité depuis 30 ans. Le déploiement massif de la programmation objet a d'abord apporté une réponse technologique à cette complexité. Plus récemment, de nouvelles réponses ont émergé avec l'esprit « Agile », qui touchent cette fois à l'organisation et au processus de développement : SCRUM, Extreme Programming, ... autant de démarches qui déclinent les principes du « manifeste Agile » et ont été largement diffusées ces dernières années dans le monde de l'ingénierie logicielle. Quels sont les principes fondamentaux ? Sont-ils applicables à l'ingénierie système ?

Nous avons retenu quelques comportements clés de l'esprit agile pour lesquels nous allons éclairer plus loin comment les appliquer en ingénierie système :

- Se focaliser sur la valeur pour le client,
- Prioriser les activités pour créer la plus forte valeur au début,
- Itérer en boucles courtes pour pouvoir corriger en continu,
- Intégrer en continu,
- Tester en continu.

4 LE PROCESSUS OUTILLE

4.1 Le processus.

Pour mettre en œuvre ces principes pendant la phase d'écriture des exigences dans la descente du V, nous avons créé une démarche outillée qui utilise des modèles simulables et écrit automatiquement des phrases d'exigences à partir des diagrammes des modèles.

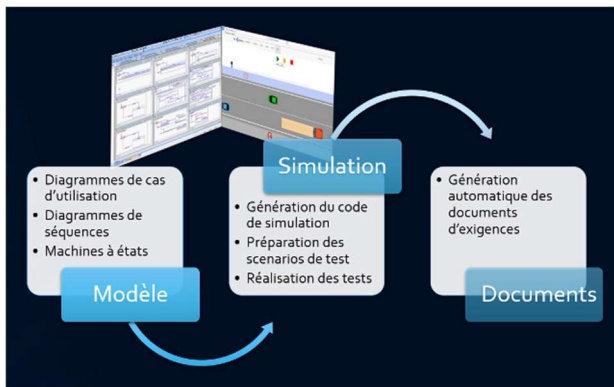


Figure 1 : le processus outillé

Ce processus s'appuie sur 2 principes fondamentaux :

- la visualisation dans un environnement virtuel de ce que l'ingénieur a écrit dans le modèle.
- la génération automatique des exigences à partir du même modèle dont on aura vérifié au préalable le comportement dans l'environnement virtuel.

Respecter ces deux piliers permet d'éclairer la conception en cours pour toute l'entreprise. Les collègues, le manager, le responsable produit, les acheteurs peuvent tous voir et comprendre ce que réalisera le produit en cours de spécification.

Afin de traiter ces deux piliers, le langage et les outils de modélisation sélectionnés doivent répondre à des enjeux multiples : d'une part décrire les comportements du système, les interactions dynamiques et les interfaces logiques entre les sous-systèmes et entre les composants qui vont réaliser les fonctions attendues ; en second lieu être en mesure de générer un code exécutable à partir de ce modèle afin de tester son comportement dans un environnement virtuel ; enfin, pouvoir générer automatiquement des exigences textuelles à partir des éléments du modèle afin d'être en mesure de s'insérer dans les processus existant d'échanges d'information entre départements et avec des tiers dans une relation contractuelle client-fournisseur.

4.2 Choix du langage et outil pour la modélisation

Nous ne souhaitons pas nous lancer dans cet objectif ambitieux d'outillage d'une ingénierie système agile en partant de la feuille blanche, mais plutôt choisir des outils existant puis les adapter à notre besoin.

Pour répondre à ces besoins nous avons choisi deux briques dans l'existant : la première brique est apportée par un outil de modélisation qui implémente le langage UML (Unified Modeling Language) et est capable de générer un code suffisamment complet pour simuler dans un environnement virtuel le comportement décrit ; la deuxième brique est un outil de modélisation complémentaire pour tous les algorithmes qui nécessitent des calculs mathématiques (pour lesquels UML n'est pas adapté), les modèles doivent là encore pouvoir être convertis en code exécutable.

Un langage graphique comme UML qui utilise des représentations symboliques sous formes de diagrammes (avec des boîtes, des flèches, des lignes, ...) est très adapté pour représenter dans une forme à la fois synthétique, rigoureuse et non ambiguë, des scénarios complexes d'échanges entre différents systèmes et objets. Ce type de représentation se montre à l'usage beaucoup plus efficace que d'écrire à la main un ensemble d'exigences textuelles. UML est un standard international largement déployé que ce soit pour le logiciel sous sa forme originelle UML ou dans forme dérivée pour l'ingénierie Système appelée SysML. De nombreux outils implémentent ces standards et permettent d'écrire des modèles conformes. Avec 14 diagrammes disponibles, UML permet de décrire toutes les interactions logiques entre différents systèmes ou objets, tant que l'on reste dans un domaine de logique discrète manipulant des structures de données, des états, des événements. Par contre UML n'est pas du tout fait pour décrire des modèles physiques et implémenter des algorithmes mathématiques. C'est pourquoi nous avons inséré dans notre solution un autre type d'outil pour couvrir cet autre besoin.

4.3 Choix d'une implémentation pour l'environnement de simulation.

De nombreux simulateurs sont disponibles sur le marché. Néanmoins nous n'avons pas identifié que les simulateurs existants pourraient répondre à notre besoin de test du code généré depuis nos modèles. C'est pourquoi notre choix a été de mettre en œuvre les technologies modernes du monde internet pour créer en javascript un environnement virtuel accessible via un navigateur web. Ce choix présente plusieurs intérêts :

- les nombreuses bibliothèques Open Source disponibles sur internet permettent de créer rapidement et à moindre coût un environnement virtuel au juste nécessaire pour notre besoin,
- la facilité offerte par ces technologies pour programmer la manipulation d'objets via des drag & drop, est une grande aide pour obtenir rapidement une interface rapide

et intuitive de construction des scénarios de test. Les corrections suite aux remarques des premiers utilisateurs (bêta-testeurs étudiants) ont pu être réalisées dans des temps très courts.

- L'accès à l'environnement de test dans un navigateur internet permet à un donneur d'ordre, client de la spécification, de créer ses propres scénarios de test et de les jouer sans avoir recours à des outils de conception sophistiqués.

4.4 Architecture de la chaîne outillée

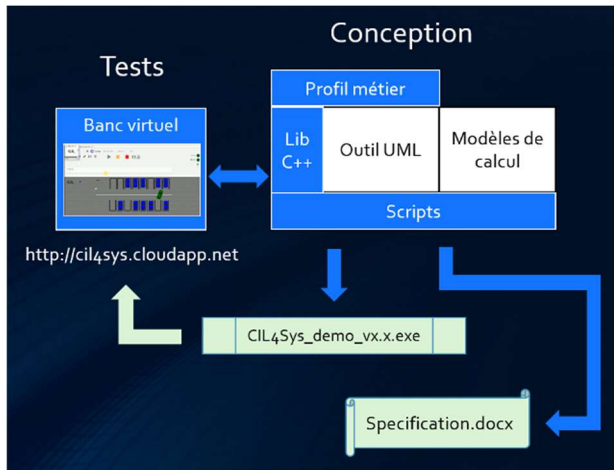


Figure 2 : Architecture de la chaîne outillée

L'outil utilisé pour les modèles de calcul est Matlab/Simulink. Deux outils UML ont un générateur de code compatible avec l'objectif poursuivi : PTC Integrity Modeler et IBM Rhapsody. L'un ou l'autre de ces deux outils peut être utilisé indifféremment au cœur de notre chaîne de conception.

4.5 UML n'est-il pas un langage trop compliqué pour les bureaux d'études mécatroniques ?

Une critique souvent faite à UML est sa propre complexité. Les diagrammes utilisés sont des représentations qui sont parfois difficiles à lire pour des personnes non entraînées. Par ailleurs, les outils laissent ouvertes toutes les possibilités de représentations offertes par le standard, ce qui peut conduire à un chaos si le déploiement n'est pas correctement préparé en amont avec une méthode très structurée.

Pour traiter ce risque nous avons mis en œuvre deux éléments majeurs :

- La paramétrisation des outils via des profils et avec une surcouche logicielle permet de guider les concepteurs dans le processus de modélisation, de les forcer à respecter les étapes attendues et à utiliser les bons objets de modélisation. Ecrire un modèle devient ainsi aussi naturel et plus rapide que d'écrire des exigences sous forme textuelle.

- La génération de code, et l'exécution dans un environnement virtuel de ce code généré, permet à l'ingénieur

de visualiser immédiatement le résultat de son travail et donc de se corriger en temps réel.

Pour vérifier l'efficacité des mesures mises en place, nous avons fait utiliser notre méthode en projet par des étudiants de 2^{ème} année de l'Ecole CentraleSupélec. Dans un premier projet trois étudiants ont modélisé et simulé une fonction automobile de créneau automatique. Dans un deuxième projet, deux étudiants ont modélisé et simulé une fonction de voiturier autonome. La présentation de ce deuxième exemple est l'objet du chapitre suivant.

5 UN EXEMPLE DE REALISATION : UN SERVICE DE VOITURIER AUTONOME

Les étudiants ont réalisé un modèle pour le parking et un modèle pour le véhicule. La simulation dans l'environnement virtuel montre la réalisation d'une collaboration entre le parking, qui décide de la place et choisit le chemin à suivre, et le véhicule qui suit la trajectoire proposée par le parking et réalise sa manœuvre de stationnement. Quelques diagrammes extraits des modèles sont présentés ci-après.

5.1 Le véhicule

Le processus présenté au §4.1 est appliqué à la modélisation du véhicule pour ce cas de parking autonome. Le premier diagramme UML utilisé est le diagramme de cas d'utilisation qui présente les interactions entre cas d'utilisations et acteurs.

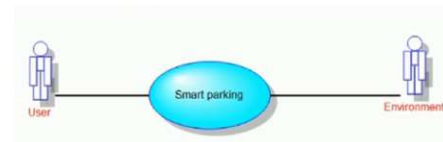


Figure 3 : Cas d'utilisation véhicule

Ensuite, pour chaque cas d'utilisation retenu, des diagrammes de séquence sont produits pour décrire les interactions entre le système étudié et les acteurs. De manière générique dans la démarche, les utilisateurs du système sont identifiés comme des acteurs en propre, et l'acteur « environnement » regroupe tous les autres objets en interaction. (Par exemple ici, le parking est dans l'environnement).

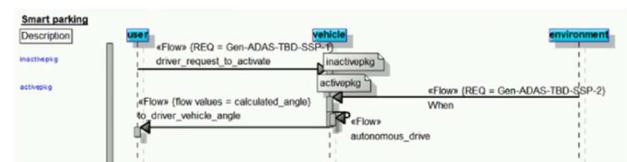


Figure 4 : Une séquence du véhicule

Une décomposition fonctionnelle du véhicule est ensuite réalisée et les séquences précédemment décrites au niveau du véhicule sont ensuite détaillées au niveau fonctionnel comme dans l'exemple donné ci-après.

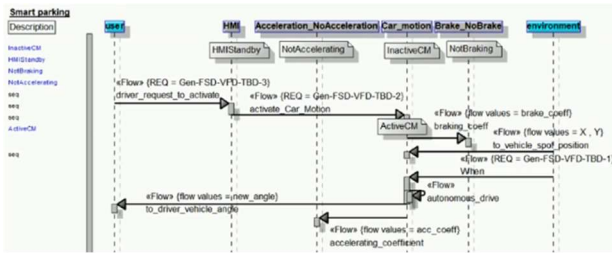


Figure 5 : Une séquence après décomposition fonctionnelle

Les flux qui interagissent avec l'utilisateur ou l'environnement ont été alloués vers une fonction ; le lien de traçabilité est conservé entre ces flux alloués et les « mêmes » flux dans le diagramme précédent au niveau du véhicule.

A partir de l'ensemble des diagrammes de séquences réalisés au niveau fonctionnel, des states machines sont générées automatiquement pour chaque fonction.

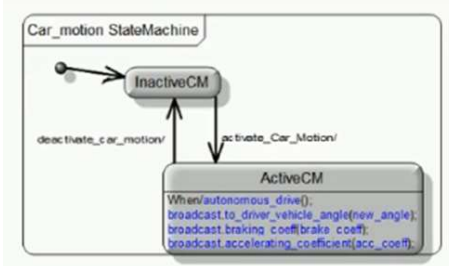


Figure 6 : Une machine à états véhicule

Ces states machines permettent de générer un code C++, compilable et exécutable pour simuler le comportement du modèle du véhicule en environnement virtuel.

5.2 Le parking

Les mêmes étapes sont suivies pour réaliser le modèle du parking.

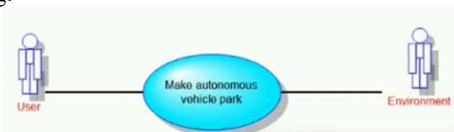


Figure 7 : Cas d'utilisation parking

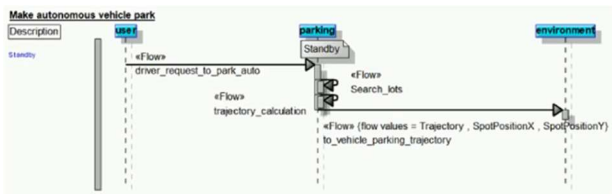


Figure 8 : Une séquence du parking

A noter : cette fois le véhicule est dans l'environnement du parking

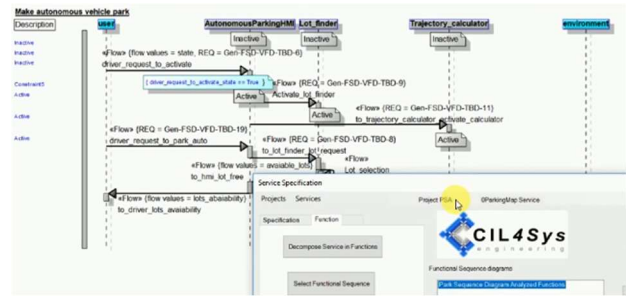


Figure 9 : Une séquence des fonctions parking

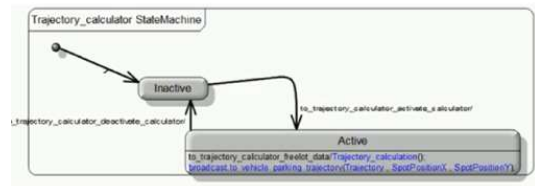


Figure 10 : Une machine à état du parking

A partir de l'ensemble des machines à état du parking une 2^{ème} code est généré puis compilé pour simuler le comportement du parking en environnement virtuel.

A noter : Les calculs de place disponible, de choix de la place la plus proche de la sortie, et de trajectoire, ont été réalisés dans des modèles Matlab/Simulink et intégrés au modèle UML via des scripts qui présentent au concepteur les entrées – sorties des modèles Matlab et lui permet de les interfacier dans sa description UML.

Le code Matlab correspondant est compilé avec le code issu du modèle UML.

5.3 La simulation

Les exécutables générés à partir du modèle sont lancés et connectés à l'environnement virtuel (ils ont été compilés avec une librairie websocket qui permet d'effectuer cette connexion)

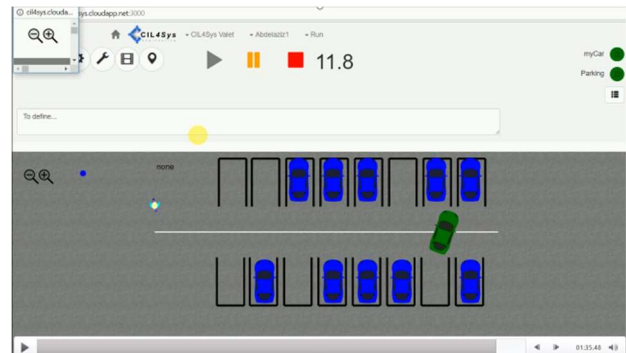


Figure 11 : Simulation voiturier autonome

L'appui sur la touche « Play » dans le simulateur lance le scénario.

Le film disponible à l'adresse donnée dans les références présente les diagrammes montrés dans ce chapitre et le scénario de simulation de la figure 11.

5.4 Constats

On peut constater que les premières simulations sont réalisées à un haut niveau d'abstraction, en fait dès le premier niveau de décomposition en grandes fonctions. A ce niveau-là peu importe le type des capteurs utilisés : le véhicule est supposé disposer de capteurs parfaits. La fonction de détection de l'environnement dans le véhicule traite des métadonnées envoyées par le simulateur ; ces métadonnées contiennent les positions et vitesses relatives des objets qui entourent le véhicule étudié.

Déjà, à ce premier niveau de décomposition, on peut vérifier grâce à la simulation qu'on a correctement décrit les scénarios opérationnels et les interfaces (flux fonctionnels) entre nos grandes fonctions. Si des erreurs sont commises dans la description, soit on n'arrivera pas à générer le code, soit on n'arrivera pas à compiler le code, soit le résultat vu dans l'environnement virtuel ne sera pas le résultat attendu. Forcer la simulation au plus tôt dans la conception est donc un outil de robustesse particulièrement puissant !

En continuant à détailler la conception sur un des axes fonctionnels chaque ingénieur en charge d'un domaine pourra insérer ses nouveaux éléments de conception dans le modèle global, générer à nouveau du code de simulation et tester. Dans une démarche traditionnelle de décomposition en fonctions puis sous-fonctions, les ingénieurs en charge d'une partie « enfouie » perdent rapidement de vue le contexte opérationnel. Dans la démarche agile proposée, chaque ingénieur est en mesure de visualiser l'impact de son travail sur le résultat opérationnel pour l'utilisateur final. Une telle possibilité est à la fois valorisante et le gage d'une grande robustesse de la conception car chacun aura à cœur de corriger en temps réel les dysfonctionnements constatés.

5.5 La génération d'exigences textuelles

Les cahiers des charges des différents objets de conception restent exprimés sous forme d'exigences atomiques, que cet objet soit le système dans son ensemble, une de ses fonctions, ou un de ses composants.

De nombreux arguments militent pour garder de tels cahiers des charges. Parmi ces arguments :

- Le poids de l'existant et de la réutilisation est important dans chaque nouveau système conçu.
- Les mécanismes contractuels s'appuient sur ces documents d'exigences, que ce soit entre départements au sein d'une entreprise ou entre une entreprise et ses partenaires ou ses fournisseurs.
- Certaines exigences sont difficilement modélisables et le langage naturel reste la meilleure manière de les exprimer.

Afin de respecter ce besoin de produire et partager des documents d'exigences, sans conduire les ingénieurs à devoir écrire deux fois la même chose sous deux formes

différentes, nous avons développé une génération automatique d'exigence à partir du modèle. Le principe est de s'accorder avec le donneur d'ordre sur le motif de phrase à utiliser pour écrire une exigence logique. Ensuite les transitions et activités des machines à état sont parcourues pour générer des exigences atomiques à partir du modèle. Le respect de bonnes pratiques de nommage pour les paramètres et attributs dans le modèle est important pour préserver la lisibilité des phrases générées.

Exemple d'exigence générée pour le parking :

The Parking shall check vehicle position AND calculate next move AND send the flow "send move" when the function is in the state "Busy" AND "vehicle_to_be_parked" is "True".

Exemple d'exigence générée pour le véhicule :

The vehicle shall move AND change its state to "vehicle_moving" when the function is in the state "vehicle_connected" AND the flow "request to move" is received.

Les exigences générées peuvent parfois paraître abruptes hors contexte, mais elles sont comparables à ce que des concepteurs écrivent à la main pour décrire de la logique, en présentant l'avantage d'une structure de phrase systématique, et d'une cohérence garantie entre exigences.

5.6 Compléments

Pour démontrer l'applicabilité de la démarche jusqu'en bas du cycle de conception, nous avons proposé à des étudiants de l'Université Pierre et Marie Curie de poursuivre le modèle de voiturier autonome pour aller jusqu'à la réalisation d'une maquette physique. Nous avons en parallèle un projet client dans lequel nous traitons avec un partenaire une architecture électronique jusqu'à sa réalisation.

6 ATOUTS DE LA DEMARCHE OUTILLEE

6.1 Intégration continue

Un objectif important dans la structure des modèles proposés est la capacité à intégrer progressivement des éléments de conception. Cette capacité démontrée dans le cadre de l'intégration de modèles Matlab/Simulink, peut aussi s'appliquer à l'intégration de modèles issus d'autres outils, à partir du moment où les modèles peuvent être encapsulés par leurs entrées / sorties logiques. Les différents éléments de conception peuvent être intégrés progressivement et séparément. A partir du moment où des premiers scénarios opérationnels ont été simulés, ils peuvent servir de référence et être rejoués avec un niveau de détail qui peut être beaucoup plus profond sur un domaine du modèle que sur un autre.

6.2 Déploiement incrémental

Un autre point fondamental de la démarche proposée est la capacité à s'insérer dans un processus de conception

traditionnel et à déployer de manière incrémentale. Les documents d'exigences générés automatiquement sont au format attendu habituellement dans l'entreprise utilisatrice. Ainsi, le déploiement et la formation à la nouvelle méthodologie peuvent être réalisés secteur par secteur sans remise en cause des projets en cours.

6.3 Unicité de la conception

L'utilisation de modèles de simulation créée à posteriori de l'écriture des exigences conduit souvent à créer un deuxième flux de conception qu'il peut être difficile de garder synchronisé avec le flux des exigences. Pendant que les ingénieurs en charge des exigences continuent à travailler et à les faire évoluer, les ingénieurs en charge de la simulation travaillent et font évoluer leur simulation de leur côté. Garder en permanence une cohérence entre les deux vues de ce qui devrait être une même conception nécessite de grands efforts de synchronisation. Au contraire dans la démarche proposée ici la simulation et les exigences sont deux vues d'un même modèle et restent cohérentes à chaque génération d'une nouvelle version du document d'exigence. Afin d'étendre cette cohérence aux exigences de performance, ces dernières qui restent exprimées sous forme textuelles, sont insérées dans des champs du modèle UML. Un deuxième intérêt de ce choix est que les exigences de performance sont attachées à l'intérieur de séquences dans des scénarios. Ainsi les contextes de test de ces performances sont prédéterminés et ne nécessiteront pas d'effort de réflexion et de rédaction supplémentaire au moment d'écrire le plan de tests.

6.4 Processus collaboratif

L'exemple de voiturier autonome présente aussi une caractéristique particulière. L'entreprise en charge du développement du parking et l'entreprise en charge du développement du véhicule seront très certainement des entreprises différentes. La capacité à travailler de manière collaborative, à se mettre d'accord sur les fonctionnalités attendues pour l'utilisateur, et à répartir les tâches entre partenaires, est très largement améliorée par la possibilité de visualiser les scénarios et d'avoir des premières simulations très tôt dans le projet. Des scénarios de référence peuvent être créés par les partenaires lors d'une phase initiale de travail en commun, puis chacun pourra travailler de son côté sur son architecture, sa conception et tester ses briques dans les scénarios. L'exemple peut être étendu à des scénarios mettant en œuvre des partenaires multiples, de tailles et notoriétés très différentes. La capacité pour les grands groupes à intégrer en cycle court les propositions de start-ups innovantes peut ainsi être démultipliée.

6.5 Intégration de la sûreté de fonctionnement

La modélisation et la simulation en environnement virtuel sont particulièrement bien adaptées pour supporter les activités d'analyse en sûreté de fonctionnement, pour implémenter et tester les safety concepts, les modes refuges, rechercher les modes communs, et toutes autres analyses de

sûreté de fonctionnement. Avoir un modèle exécutable avec lequel travailler est bien plus confortable et efficace que de devoir relire et comprendre des milliers de lignes d'exigence dans des documents. Pour faciliter la démarche les outils ont été adaptés afin de pouvoir coter les événements redoutés et les évaluations de risques. Des développements complémentaires sont engagés pour permettre de proposer les safety concepts directement dans le modèle de conception.

6.6 Application des principes agiles

Les principes clés identifiés en début d'article sont adressés par la démarche outillée proposée.

- La focalisation sur la valeur pour le client et la priorisation des développements à plus forte valeur ajoutée sont rendues possibles par la visualisation de la simulation par le responsable marketing – produit du projet, qui pourra ainsi réaliser en continu les arbitrages.
- Les itérations sont effectivement réalisées en boucle courtes : typiquement les simulations sont présentées à l'ensemble des parties prenantes du projet toutes les deux semaines.
- A chaque étape de conception détaillée un nouveau code de simulation peut être généré et testé dans les scénarios opérationnels de référence du projet, constituant ainsi les opérations d'intégration et test en continu pendant la descente du V.

En outre, une des grandes vertus du processus agile proposé est de rendre l'avancement de la conception compréhensible par toutes les parties prenantes, que ce soit au sein de l'entreprise (ingénieurs, managers, responsables produits, acheteurs) ou à l'extérieur (fournisseurs, partenaires, ...).

7 GAINS CONSTATES OU POTENTIELS

7.1 L'automatisation

Un client constructeur automobile, nous a proposé de nous mesurer en face à face avec un grand bureau d'ingénierie. L'exercice consistait en la rédaction sur une semaine du document de spécification d'une fonction automobile (la même fonction pour le bureau d'ingénierie concurrent et nous). Il a été réalisé sur le site de développement du client et le résultat a été jugé par un collègue d'experts chez le constructeur. Au final l'utilisation de notre démarche agile outillée nous a permis de réaliser avec moitié moins de ressources (un ingénieur contre deux chez le concurrent) un document jugé 30% meilleur par le client.

Cette expérience a permis de prouver que l'utilisation d'un outil UML ne constitue pas un handicap mais apporte au contraire une efficacité inégalée, à condition d'avoir correctement paramétré l'outil pour forcer une méthode stricte, rigoureuse, efficace. Les éléments d'automatisation que nous avons été en mesure de configurer dans les logiciels permettent d'éliminer les opérations sans valeur ajoutée, que ce soit des recopies d'informations, des

redondances inutiles ou de la mise en forme. Ainsi les ingénieurs peuvent concentrer leur effort sur l'essentiel et produire une conception d'emblée cohérente et robuste.

7.2 Les gains liés à la démarche agile

De nombreuses études montrent que le bouclage agile en cycle court permet des gains très significatifs. Une enquête menée par Le site savoiraagile.com présente les résultats d'une enquête menée en 2016.

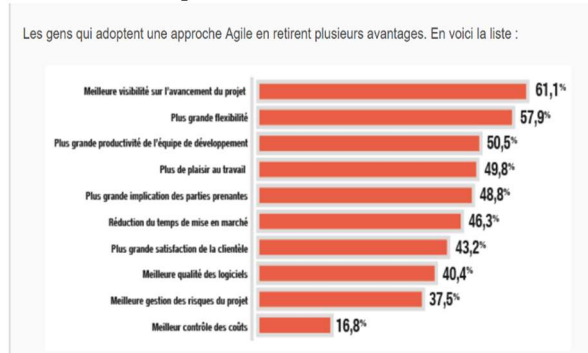


Figure 12 : Avantages d'une approche Agile

A travers les premiers projets réalisés, nous avons démontré que ces gains trouvés en ingénierie logicielle sont tout aussi accessibles en ingénierie système. Visibilité, flexibilité, productivité, plaisir, implication, réduction du Time to Market, satisfaction client, qualité, meilleure gestion des risques, meilleur contrôle des coûts : qui parmi les grands industriels ne rêve pas de progresser sur tous ces axes ?

7.3 La remontée du V

Les systèmes tels des avions, des trains, des automobiles la phase de remontée du V qui devrait être consacrée à la vérification et la validation de la réalisation physique, est pour une part trop importante occupée par la correction des incohérences de conception issues de la phase de spécifications. Les prototypes sont des objets chers ; maximiser leur temps d'utilisation pour leur vocation première et éviter de perdre du temps à juste leur permettre de fonctionner comme ils devraient, représente un enjeu économique majeur.

8 CONCLUSION

Depuis déjà de nombreuses années, les industriels utilisent des modèles pour rendre plus efficaces leur processus de conception. Le retour d'expérience sur le déploiement de modélisation est hétérogène. Sur les cas d'échecs, de nombreuses explications sont avancées : une préparation insuffisante des méthodes, une formation insuffisante des acteurs, des méthodes et outils trop complexes inadaptés au profil des acteurs, le surcoût pour réaliser les modèles en comparaison des démarches habituelles, ...

Trop souvent la démarche de modélisation vient effectivement s'ajouter à la méthode traditionnelle d'écriture des exigences. Dans ce cas, les modèles sont perçus par les ingénieurs comme un travail supplémentaire pour lequel les moyens ne sont pas toujours donnés.

Un élément très important dans la démarche présentée est que le modèle n'est pas un travail en plus, le modèle est la représentation première de la conception. De plus, une mesure objective chez un client industriel a démontré qu'elle permet de gagner du temps x homme par rapport à l'écriture à la main des exigences.

Le deuxième élément vraiment différenciant est la simulation du système dans son environnement opérationnel virtuel avant la génération automatique des exigences. Dans d'autres solutions du marché il faudrait modéliser le monde pour réaliser une telle simulation ; dans d'autres encore il faut créer à partir des exigences préalablement écrites un modèle dédié pour la simulation.

De part l'addition de ces deux éléments clés, la démarche présentée est singulière. Elle a été développée par une approche expérimentale de type essais-erreur en la faisant mettre en pratique par des étudiants dans le cadre de projets. Elle mériterait d'être complétée par un cadre scientifique formel si l'on veut notamment adresser les problématiques d'optimisation d'architecture ou les problématiques de preuve en sûreté de fonctionnement.

Le 21^{ème} siècle voit l'émergence de la domination du logiciel à tel point que certains acteurs du monde logiciel en viennent à penser que les architectes systèmes sont devenus inutiles. En réalité, c'est de la confrontation des deux mondes et des deux modes de pensée que peuvent émerger les solutions efficaces pour la conception industrielle de demain. Les industries traditionnelles peuvent s'inspirer et copier les progrès de l'ingénierie logicielle pour gagner en flexibilité et en efficacité, sans renier leurs acquis et leur expérience de plusieurs dizaines d'années dans la conception de grands systèmes mécaniques puis mécatroniques.

REMERCIEMENTS

Antoine Zuin, David Villemur, Abdelaziz Abassi les ingénieurs de CIL4Sys qui contribuent au quotidien à développer et expérimenter nos innovations en conception.

Santiago De Arquer Puig, Yunfei Qu, les étudiants de Centrale qui ont réalisé le projet de voiture autonome.

REFERENCES

Film voiture autonome

Abdelaziz Abassi – 18/10/2017

<https://www.youtube.com/watch?v=kVQG8tEzbow&list=PL65jVLkEHp8ZSZrTozGe2ENZnXFFhfXoo&loop=1>;

Enquête agilité

Marie-Eve Trempe - 04/11/2016

<http://savoiragile.com/2016/11/04/resultats-enquete-sur-letat-de-lagilite-en-2016/>

The Agile Manifesto - 2001

<http://agilemanifesto.org/principles.html>

A world in motion - SE Vision 2025 – INCOSE –

<https://www.incose.org/AboutSE/sevision>

Oct 13,2014

System Engineering Handbook –4th Edition,

INCOSE 2015

Compiled by David D.Walden, Gary J.Roedler, Kevin J.Forsberg, R Douglas Hamelin, Thomas M.Shortell

Philippe Gicquel – 16/08/2016

Livre blanc ingénierie système –

<http://cil4sys.com/whitepaper>

INTÉGRATION TACTIQUE-OPÉRATIONNELLE D'UNE PLANIFICATION HIÉRARCHIQUE AVEC FONCTION D'ANTICIPATION : ÉTUDE DE CAS DANS LE SECTEUR FORESTIER AU QUÉBEC

L. ESTIVALET, M. OUHIMMOU, A. CHAABANE

École de Technologie supérieure – 1100 rue Notre-Dame Ouest – H3C 1K3 Montréal – Canada

laureline.estivalet.1@ens.etsmtl.ca,

mustapha.ouhimmou@etsmtl.ca,

amin.chaabane@ens.etsmtl.ca

RESUME : *Le processus de planification dans le secteur forestier est complexe et des divergences apparaissent entre les solutions des modèles de différents niveaux de planification. Afin de diminuer celles-ci, plusieurs méthodes ont été développées : approches monolithiques, approches hiérarchiques, approches d'intégration de plusieurs niveaux de planification. Une méthode d'intégration tactique-opérationnelle dans le secteur forestier est présentée dans cet article. Deux modèles mathématiques existants ont été pris comme base pour réaliser cette intégration d'une approche hiérarchique : un modèle d'allocation des ressources et un modèle de localisation des chantiers de coupe. L'intégration a été réalisée avec une fonction d'anticipation basée sur la méthode de Schneeweiss. Cette méthode permet d'aider les gestionnaires du niveau supérieur à prendre des décisions afin qu'elles soient réalisables au niveau inférieur. Une étude de cas a été réalisée sur le processus de planification des ressources forestières dans la province du Québec, au Canada.*

MOTS-CLES : *Planification tactique-opérationnelle, Planification hiérarchique, Intégration, Fonction d'anticipation, Forestier.*

1 INTRODUCTION

Toute entreprise doit planifier son organisation et répondre à ses clients. Mais s'agit-il d'une planification « stratégique », « tactique » ou « opérationnelle » ? Et quelles sont les méthodes qui peuvent aider à optimiser leurs coûts ? Plusieurs niveaux de planification sont prévus avec pour chacune des objectifs différents. Ainsi la planification stratégique permet de prévoir des décisions à long-terme tel que l'allocation de ressources pour satisfaire les demandes ou tel que l'ouverture, la fermeture ou l'agrandissement d'usines ou d'entrepôts. Ces décisions permettent de prévoir des investissements sur le long terme. La planification tactique permet de mettre en pratique les décisions prises au niveau stratégique sur le moyen terme. Le dernier niveau de planification permet de prendre des décisions opérationnelles, c'est-à-dire de la planification court terme, souvent quotidienne. Pour les différents niveaux de planification, les informations nécessaires ne sont pas disponibles aux mêmes instants. Ainsi les décisions stratégiques sont prises avec des prévisions car les données nécessaires ne sont pas encore disponibles. Selon les types de décisions prises au niveau tactique, les informations seront disponibles ou il faudra prendre en compte des prévisions. Dès lors, la planification stratégique initiale sera modifiée pour correspondre davantage à la tendance au cours du temps.

L'exploitation des ressources naturelles est une part importante de l'économie canadienne. Dans un souci de développement durable, les décisions prises lors de la planification stratégique sont cruciales. Plus spécifiquement, dans le milieu forestier, la planification des ressources est une problématique de recherche d'optimisation de la valeur économique des ressources forestières.

Dans la province du Québec, le Ministère des Forêts, de la Faune et des Parcs (MFFP) détermine le niveau stratégique en donnant le volume de bois commercialisable par zones, tout en tenant compte de la possibilité forestière, et en accordant des contrats d'approvisionnement aux entreprises du secteur. Au niveau tactique, la sélection des chantiers et l'allocation des bois aux usines se font par zones. Cette planification se fait sur un horizon glissant de cinq ans. Pour atteindre l'objectif déterminé au niveau stratégique, il faut prendre en compte les contraintes liées à la dispersion spatiale, la demande, les capacités et les performances des usines, et les contraintes géographiques.

La planification stratégique contraint les solutions possibles au niveau tactique, et différentes planifications tactiques, ayant des objectifs différents, peuvent se contraindre entre elles. Une variation acceptable dans les objectifs stratégiques, soit un assouplissement des contraintes au niveau tactique pourrait apporter une valorisation des ressources forestières. Ainsi en acceptant de légères variations positives ou négatives sur les volumes

distribués aux différentes usines, la solution obtenue pourrait se révéler plus avantageuse d'un point de vue économique. De même, une fois les volumes alloués aux différentes entreprises, l'ouverture des chantiers de coupe dans les différentes zones entraînent des variations dans les volumes disponibles.

Cet article présente une méthode pour optimiser la valorisation des ressources forestières en minimisant les différences de volume de bois alloué à chaque usine par rapport à leur contrat d'approvisionnement, en prenant en compte les décisions d'ouverture des chantiers de coupe. La section 2 présente une mise en contexte. La méthodologie est développée dans la section 3. Enfin les résultats obtenus ainsi que les observations qui en découlent sont présentés dans la dernière section.

2 MISE EN CONTEXTE

À présent que la modélisation des objectifs de planification niveau par niveau est largement documentée, les chercheurs se penchent sur les méthodes d'intégration de plusieurs objectifs de planification. Les méthodes d'intégration doivent pallier aux divergences existantes dans les méthodes de planification utilisées actuellement. Ces divergences peuvent être d'ordre spatial, temporel ou liées aux contraintes des modèles.

Trois approches ont été identifiées par Andersson (2005) pour diminuer ces divergences : les approches monolithiques, les approches hiérarchiques et les approches *bottom-up*. Les approches monolithiques et *bottom-up* ont comme inconvénient de d'avoir besoin des données précises sur le niveau inférieur dès la première itération du modèle pour sa résolution. Les approches hiérarchiques permettent de résoudre les niveaux de planification l'un après l'autre et autorise les itérations dans leur résolution. Le temps de résolution est réduit par rapport à une approche de type monolithique.

La méthode avec fonction d'anticipation de Schneeweiss (2012) est une approche de type hiérarchique. Elle permet d'anticiper la réaction du niveau inférieur aux décisions du niveau supérieur. Cette méthode est intéressante à développer car le niveau supérieur peut analyser seul l'impact de ses décisions sur le niveau inférieur en anticipant le modèle de celui-ci. Et par conséquent, il peut vérifier si ses décisions débouchent sur une solution réalisable, ou non, pour la planification au niveau inférieur. Les données des deux modèles sont agrégées et désagrégées, ce qui peut entraîner des biais dans les solutions. Il faut donc porter une attention particulière lors de l'agrégation des données.

La planification dans le secteur forestier se fait de manière découplée, et des biais importants apparaissent entre les différents niveaux de planification. Plusieurs approches ont été étudiées : la collaboration entre les échelons d'une chaîne logistique (Abasian *et al.*, 2016), une méthode de génération de colonnes pour une planifi-

cation intégrée (Bouchard *et al.*, 2017), ou encore une méthode de planification intégrée avec fonction d'anticipation (Beaudoin *et al.*, 2008). La collaboration entre différentes parties est coûteuse en temps et en échange de données. Des problèmes humains peuvent également intervenir dans ce genre d'approche. La méthode de planification intégrée avec fonction d'anticipation permet, quant à elle, de limiter les échanges entre les différents niveaux à une validation de la modélisation du niveau inférieur anticipé et des lois de distributions des paramètres.

Parmi les problèmes les plus étudiés, les problèmes d'allocation et localisation sont largement représentés, ainsi que les problèmes de localisation-allocation. Cependant, dans le secteur forestier, l'allocation des ressources forestières se fait avant la localisation des chantiers de coupe. Or la modélisation d'un problème d'allocation-localisation ne semble pas avoir été étudiée, puisque la plupart des processus de planification sont modélisés au mieux par un problème de localisation-allocation. Ce problème d'allocation-localisation est étudié dans cet article avec une étude de cas dans le secteur forestier et en utilisant une méthode de planification intégrée.

Dans cette recherche, le modèle d'anticipation de Schneeweiss va être appliqué pour un cas dans le secteur forestier. Les modèles de Boukherroub *et al.* (2017) et de Ouhimmou *et al.* (2015) seront utilisés, respectivement, en tant que modèle pour le niveau supérieur et le niveau inférieur.

3 METHODOLOGIE

Un modèle de programmation linéaire mixte en nombres entiers pour la planification de l'allocation des ressources forestières aux entreprises et la planification de la localisation des chantiers à ouvrir pour chaque période est présenté dans cette partie. La méthode de Schneeweiss (2012) est utilisée pour l'intégration de deux niveaux de planification par une fonction d'anticipation. Le modèle présenté dans cette partie est basé sur les travaux de Boukherroub *et al.* (2017) et de Ouhimmou *et al.* (2015). Ces modèles sont utilisés comme base pour la méthodologie développée avec une fonction d'anticipation. La figure 1 représente le processus de prise de décision pour allouer les chantiers de coupe et les volumes de bois aux usines du secteur forestier dans la région du Québec. Les boucles de rétroaction sont représentées en pointillées pour signaler les endroits possibles d'amélioration. La boucle de rétroaction en rouge est celle étudiée par la méthodologie présentée dans ce chapitre.

Dans un premier temps les ensembles utilisés pour les deux parties du modèle intégré sont présentés. Ensuite, les paramètres, les variables de décisions et le modèle mathématique des parties pour l'allocation des

ressources et la localisation des chantiers de coupe sont présentés.

C_{em} Capacité de production pour une période de cinq ans de l'usine m pour les produits d'essence e

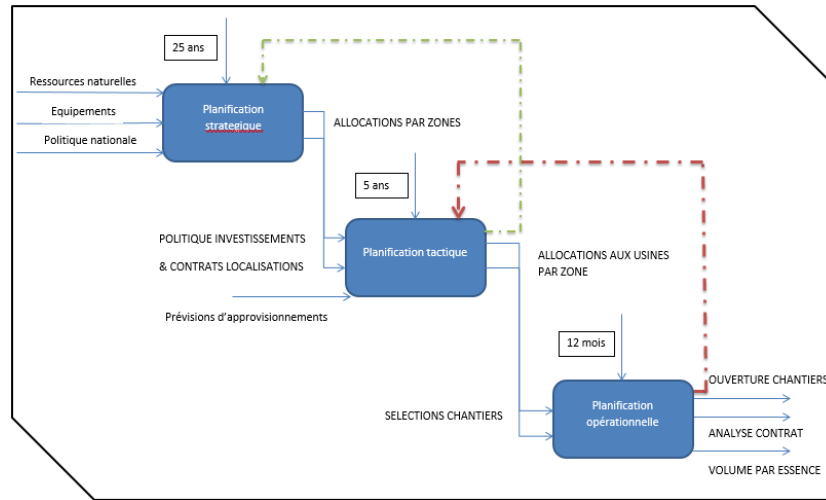


Figure 1 Processus de prise de décisions dans le secteur forestier dans la région du Québec

3.1 Ensembles

- $m \in M$ Ensemble des usines
- $u \in U$ Ensemble des unités d'aménagement
- $a \in A_{(u)}$ Ensemble des chantiers de coupe appartenant à l'unité d'aménagement u
- $e \in E$ Ensemble des essences de bois
- $q \in Q$ Ensemble des produits de qualité q
- $q \in Q_{(e)}$ Ensemble des qualités de bois appartenant à l'essence e
- $c \in C$ Ensemble des critères à considérer dans la fonction-objectif du modèle de localisation des chantiers de coupe
- $p \in P$ Ensemble des périodes de planification de localisation p sur l'horizon P (cinq périodes d'un an)

- v_{qm} Volume minimum de produits de qualité q qui doit être alloué à l'usine m pour une période de cinq ans
- S_{em} Score économique, valeur économique des produits d'essence e pour l'usine m

Le score de performance de Boukherroub *et al.* (2017) est défini ici uniquement par sa composante économique. Ce score est compris entre 0 et 1.

3.2 Modèle d'allocation des ressources

L'allocation des ressources se fait lors de la planification tactique au début de la période de 5 ans. Pour chaque année, le MFFP doit déterminer le volume à allouer aux entreprises ayant un contrat d'approvisionnement avec le ministère. Ce contrat détermine le volume annuel minimal et le volume annuel théorique à allouer à une entreprise pour chaque essence et qualité de bois. Ce premier modèle est basé sur celui de Boukherroub *et al.* (2017). Seules les considérations économiques du modèle de base seront prises en compte dans cet article.

3.2.1 Paramètres

- a_{qu} Possibilité de coupe pour une période de cinq ans pour les produits d'essence e et de qualité q de l'unité d'aménagement u

3.2.2 Variables

Les variables de décisions de ce modèle sont les quantités de produits de qualité et d'essence de bois allouées aux usines.

- X_{qmu} Quantité de produits de qualité q allouée à l'usine m venant de l'unité d'aménagement u
- Z_{emu} Quantité des essences de bois e allouée à l'usine m venant de l'unité d'aménagement u

3.2.3 Fonction-objectif et contraintes

Dans le modèle d'allocation, la somme des volumes alloués aux usines pondérés par leur score est à maximiser. Cette fonction-objectif permet d'allouer le bois aux usines proportionnellement à leur performance, l'objectif étant de satisfaire la demande des usines ayant des scores les plus forts (1). Par la suite, plusieurs contraintes sont à respecter dans ce modèle. Étant donné qu'il y a deux variables pour définir les volumes de bois, la contrainte (2) permet de les lier pour faire correspondre les quantités de produits alloués aux usines. Ces quantités doivent respecter la possibilité annuelle de coupe (3) et les contrats d'approvisionnement (5) et ne doivent pas excéder les capacités des usines (4). Les variables doivent être positives (6).

$$\max f_1 = \sum_{m \in M} \sum_{e \in E} (s_{em} \times \sum_{u \in U} Z_{emu}) \quad (1)$$

$$Z_{emu} = \sum_{q \in Q(e)} X_{qmu} \quad \forall e \in E, \forall m \in M, \forall u \in U \quad (2)$$

$$\sum_{m \in M} X_{qmu} \leq a_{qu} \quad \forall q \in Q(e), \forall u \in U \quad (3)$$

$$\sum_{u \in U} Z_{emu} \leq c_{em} \quad \forall e \in E, \forall m \in M \quad (4)$$

$$\sum_{u \in U} X_{qmu} \geq v_{qm} \quad \forall q \in Q, \forall m \in M \quad (5)$$

$$Z_{emu}, X_{qmu} \geq 0 \\ \forall q \in Q, \forall e \in E, \forall m \in M, \forall u \in U \quad (6)$$

3.3 Modèle de localisation des chantiers de coupe

Une fois l'allocation des ressources réalisées, le MFFP détermine quel chantier ouvrir pour la récolte. Pour ce modèle, il est nécessaire de connaître les volumes alloués à chaque entreprise du modèle précédent. Ce deuxième modèle est basé sur le modèle de Ouhimmou *et al.* (2015). Cependant, la contrainte de répartition spatiale n'est pas prise en compte dans cet article afin de simplifier l'intégration des deux niveaux de planification.

3.3.1 Paramètres

Les paramètres estimés sont le volume disponible de bois pour chaque chantier de coupe ainsi que la valeur moyenne des critères pour chaque chantier de coupe. Ces paramètres sont dits anticipés, car leur valeur cherche à anticiper le résultat réel. Les valeurs des variables de décisions pour les quantités à allouer aux usines du modèle précédent deviennent un paramètre, puisqu'il faut respecter ces quantités à allouer dans le choix des chantiers de coupe à ouvrir. La capacité annuelle de production et le volume minimum à allouer de chaque usine sont les mêmes paramètres que pour le modèle d'allocation des ressources aux usines. Une valeur cible pour chaque critère a également été étudiée. Ces valeurs cibles sont définies par les usines clientes.

\hat{W}_{aeqp}	Volume disponible anticipé de bois d'essence e et de qualité q disponible à la coupe dans le chantier de coupe a (m^3) pour la période p
\hat{b}_{ca}	Valeur moyenne anticipée du critère c pour le chantier de coupe a (ha)
c_{em}	Capacité de production pour une période de cinq ans de l'usine m pour les produits d'essence e

v_{qm}	Volume minimum de produits de qualité q qui doit être alloué à l'usine m pour une période de cinq ans
t_{cm}	Cible à atteindre pour chaque critère c pour l'usine m
X_{qmu}^*	Solution du modèle précédent pour la quantité de produits de qualité q allouée à l'usine m venant de l'unité d'aménagement u
Z_{emu}^*	Solution du modèle précédent pour la quantité des essences de bois e allouée à l'usine m venant de l'unité d'aménagement u

3.3.2 Variables

F_{ameqp}	Flux total de bois de l'essence e et de qualité q (m^3) du chantier de coupe a vers l'usine m pour la période p
R_{aeqp}	Volume résiduel de bois de l'essence e et de qualité q (m^3) du chantier de coupe a pour la période p
Y_{ap}	Variable binaire, égale à 1 si le chantier de coupe a est choisi pour la production pour la période p , 0 sinon
S_{cpm}^+	Variable d'écart positive pour le critère c pour l'usine m pour la période p
S_{cpm}^-	Variable d'écart négative pour le critère c pour l'usine m pour la période p

3.3.3 Fonction-objectif et contraintes

Dans le modèle de localisation, la somme des écarts, positifs et négatifs, de chaque critère est à minimiser (7). Les critères possibles sont les suivants : coût d'ouverture des chantiers de coupe, coût d'achat des ressources, volume par tige, volume par hectare et coût de transport. Pour optimiser la localisation des chantiers de coupe à ouvrir, il faut respecter les contraintes de capacité de ces chantiers (8) tout en respectant les contraintes de demandes minimales et maximales pour chaque usine cliente (9 et 10). Le volume résiduel est égal à la différence entre la somme des volumes de bois envoyés aux usines pour chaque essence et qualité de bois et la capacité des chantiers de coupe pour ces mêmes essences et qualités de bois. Chaque critère a des valeurs cibles définies par les différentes parties intéressées, la contrainte (11) permet de les respecter tout en définissant les variables d'écarts pour chaque critère. Un chantier de coupe ne peut être ouvert au maximum qu'une seule fois sur la période P (12). Les variables d'écart, de flux et de volume résiduel doivent être positives (13, 14 et 15) tandis que les variables d'ouverture des chantiers doivent être binaires (16).

$$\min f_2 = \sum_{c \in C} \sum_{p \in P} \sum_{m \in M} (s_{cpm}^+ + s_{cpm}^-) \quad (7)$$

$$\sum_{m \in M} F_{ameqp} = \hat{w}_{aeqp} \times Y_{ap} - R_{aeqp}$$

$$\forall a \in A_{(u)}, \forall e \in E, \forall q \in Q_{(e)}, \forall p \in P \quad (8)$$

$$v_{qm} \leq \sum_{a \in A_{(u)}} \sum_{p \in P} \sum_{e \in E} F_{ameqp} \quad \forall m \in M, \forall q \in Q_{(e)} \quad (9)$$

$$\sum_{a \in A_{(u)}} \sum_{p \in P} \sum_{q \in Q_{(e)}} F_{ameqp} \leq c_{em} \quad \forall e \in E, \forall m \in M \quad (10)$$

$$\left(\sum_{a \in A_{(u)}} \sum_{e \in E} \sum_{q \in Q_{(e)}} (\hat{b}_{ca} \times Y_{ap} \times \hat{w}_{aeqp}) - (t_{cm} \times F_{ameqp}) \right) + S_{cpm}^+ - S_{cpm}^- = 0$$

$$\forall c \in C, \forall m \in M, \forall p \in P \quad (11)$$

$$\sum_{p \in P} Y_{ap} \leq 1 \quad \forall a \in A \quad (12)$$

$$F_{ameqp} \geq 0$$

$$\forall q \in Q_{(e)}, \forall e \in E, \forall a \in A_{(u)}, \forall p \in P, \forall m \in M \quad (13)$$

$$R_{aeqp} \geq 0$$

$$\forall q \in Q_{(e)}, \forall e \in E, \forall a \in A_{(u)}, \forall p \in P \quad (14)$$

$$S_{cpm}^+, S_{cpm}^- \geq 0 \quad \forall c \in C, \forall m \in M, \forall p \in P \quad (15)$$

$$Y_{ap} \in \{0;1\} \quad \forall a \in A_{(u)}, \forall p \in P \quad (16)$$

3.3.4 Contraintes de couplage

Afin de lier les deux modèles précédents, il faut considérer les variables du premier modèle comme des paramètres pour le second modèle. Les quantités de bois d'essence e de qualité q coupées dans le chantier a allouées aux usines m doivent être égale à celles des produits de même qualité q allouées aux usines lors de la même période t (17). Idem avec les quantités d'essence e et de qualité q allouées entre les chantiers de coupe a et les usines m et celles des produits de même essence e allouées aux usines lors de la même période t (18).

$$\sum_{a \in A_{(u)}} \sum_{q \in Q_{(e)}} \sum_{p \in P} F_{ameqp} = Z_{emu}^*$$

$$\forall m \in M, \forall e \in E, \forall u \in U \quad (17)$$

$$\sum_{a \in A_{(u)}} \sum_{e \in E} \sum_{p \in P} F_{ameqp} = X_{qmu}^*$$

$$\forall m \in M, \forall q \in Q_{(e)}, \forall u \in U \quad (18)$$

La solution obtenue donne l'ensemble des chantiers de coupe à ouvrir ainsi que les volumes de bois de chaque chantier de coupe à allouer aux usines de façon à minimiser les écarts des valeurs de critère pour chaque usine.

Avec ces résultats, il est possible d'aider les gestionnaires lors de la prise de décisions. En effet, avec une méthode de génération de scénarios, telle que présentée par Beaudoin *et al.* (2008), il est possible de déterminer les solutions les plus performantes. Pour cela, il faut effectuer une analyse statistique afin de déterminer les solutions performantes sur des critères définis par les gestionnaires. Dans la méthode développée par Beaudoin *et al.* (2008), les solutions sont évaluées sur leur faisabilité opérationnelle et sur leur robustesse face aux autres scénarios.

4 ETUDE DE CAS

Une étude de cas a été réalisée afin d'obtenir des résultats préliminaires sur l'utilisation d'une méthode de planification intégrée dans le secteur forestier au Québec. Des hypothèses de travail ont été posées afin de limiter le temps de collecte de données.

4.1 Hypothèses de travail

La planification des ressources forestières se fait sur des zones géographiques agrégées. Plusieurs peuplements ont été agrégés pour former des superficies de plus grande taille ayant des caractéristiques similaires. Les unités d'aménagement correspondent aux zones étudiées lors de la planification tactique, et donc lors de l'allocation des ressources aux usines. Chaque unité d'aménagement est composée de plusieurs secteurs d'intervention, eux-mêmes composés de plusieurs chantiers de coupe qui sont étudiés lors de la planification opérationnelle, et donc lors de la localisation des chantiers de coupe. Chaque région du Canada possède plusieurs dizaines d'unités d'aménagement. Cependant, dans cette étude de cas, simulée et inspirée du contexte du secteur forestier au Québec, seules deux unités d'aménagement ont été étudiées. Chaque unité d'aménagement est divisée en une dizaine de chantiers de coupe.

Pour chaque zone géographique, un volume de bois est déterminé selon l'essence et la qualité du bois. Quatre essences ont été étudiées : SEPM (Sapins, Épinettes, Pins gris et Mélèzes), Thuyas, Peupliers et Feuillus. Ces essences peuvent avoir trois qualités différentes : bois de sciage, bois de placage et bois de pâtes à papier (BS, BP, BPP). Cependant, toutes les qualités ne sont pas disponibles pour toutes les essences. Les essences et les qualités étudiées dans cette étude de cas sont celles présentes au Québec.

La possibilité forestière est déterminée pour chaque essence et qualité de bois, et pour chaque unité d'aménagement. Les possibilités forestières de chaque

essence et qualité de bois ont été générées de manière à obtenir suffisamment de bois pour fournir les usines au-delà des volumes minimums qu'elles exigent.

Le volume disponible pour chaque essence et qualité de bois, et pour chaque chantier de coupe a été estimé. Ces volumes ont été générés selon une loi normale, ayant pour moyenne une valeur proche de la possibilité forestière divisée par le nombre de chantiers de coupe dans l'unité d'aménagement, et pour l'écart-type une valeur aléatoire comprise entre 1 et 10.

Les performances des usines ont été considérées identiques pour chaque essence de bois. La capacité annuelle de chaque usine et les volumes annuels minimums à allouer à chaque usine ainsi que les valeurs cibles et les valeurs moyennes anticipées du critère du modèle de localisation des chantiers de coupe ont été générés aléatoirement, dans cette étude de cas. Le critère étudié est le coût d'approvisionnement. Il correspond aux coûts liés au transport du bois entre les chantiers de coupe et les usines, ainsi qu'aux coûts engendrés par les opérations de coupe. Par conséquent, les coûts d'approvisionnement varient selon les usines et les chantiers de coupe, puisqu'ils dépendent de leurs emplacements et des routes les reliant.

4.2 Méthode de génération de scénarios

Lors de l'utilisation d'une fonction d'anticipation, il est utile de générer des scénarios afin d'aider les gestionnaires à la prise de décisions. En effet, ne pouvant prédire le futur, la génération de scénarios permet de se rapprocher de celui-ci grâce à des analyses statistiques. La méthode de Beaudoin *et al.* (2008) pour la génération des scénarios lors de l'intégration des niveaux de planification a été adaptée pour correspondre au problème étudié.

Dans un premier temps, des scénarios ont été créés par la génération des valeurs pour les paramètres du modèle mathématique du niveau supérieur selon leurs lois de distribution. Ces scénarios ont ensuite été étudiés avec le modèle du niveau supérieur. Dans cette étude, il s'agit du modèle d'allocation des ressources. Pour chaque scénario, une solution a été obtenue. Ces solutions ont ensuite été soumises à deux études en parallèle. La première étude a soumis chaque solution aux autres scénarios, en dehors du scénario duquel elle est obtenue, pour vérifier sa faisabilité dans ceux-ci. La seconde étude a transmis chaque solution du modèle du niveau supérieur en tant que décisions au modèle anticipé du niveau inférieur pour vérifier sa faisabilité au niveau inférieur. Dans cette étude, le niveau inférieur est le modèle de localisation des chantiers de coupe. À partir de ces deux études, une analyse statistique de la situation peut être réalisée afin de déterminer la robustesse et la faisabilité de chaque solution.

4.3 Résultats et analyse

Chaque scénario a été codé avec le logiciel de programmation AMPL, qui utilise le solveur d'optimisation CPLEX v.12.7. Les scénarios du modèle d'allocation des ressources sont composés de 120 variables et 111 contraintes. Les scénarios du modèle de localisation des chantiers de coupe avec la méthode d'intégration sont composés de 4950 variables et 1025 contraintes, tandis que les scénarios du modèle de localisation des chantiers de coupe sans la méthode d'intégration ont 4950 variables et 1485 contraintes. Le temps de résolution de chaque scénario a été de quelques secondes.

4.3.1 Résultats

Six scénarios ont été générés pour le modèle d'allocation des ressources. Seules les valeurs de la possibilité forestière pour chaque qualité de bois varient selon les scénarios. Le modèle d'allocation des ressources cherche à maximiser le produit des volumes alloués aux usines par le score de performance de celles-ci. Les résultats obtenus sont présentés dans le tableau 2. Les trois premiers scénarios ont leur fonction-objectifs égales car la capacité des usines est arrivée à saturation, c'est-à-dire que les volumes alloués aux usines utilisent au maximum la capacité des usines. Cependant, il n'est pas possible de comparer les résultats numériques des fonctions-objectifs entre les différents scénarios. En effet, les paramètres de chaque scénario diffèrent pour les possibilités forestières. Par conséquent, les scénarios 4, 5 et 6 ne sont pas moins performants que les trois premiers scénarios, mais ont des possibilités forestières différentes.

Scénario	1	2	3	4	5	6
Fonction-objectif (m ³)	7005	7005	7005	6295	6815	6925

Tableau 2 : Résultats de la fonction-objectif en m³ pour les six scénarios du modèle d'allocation des ressources

Afin d'obtenir l'analyse de sensibilité pour les solutions obtenues du modèle d'allocation, les solutions de chaque scénario ont été testées avec les paramètres des cinq autres scénarios. Aucune solution réalisable n'a été trouvée. Deux interprétations peuvent être réalisées. Premièrement, la variation entre les scénarios est très grande, et les solutions trouvées ne peuvent satisfaire que le scénario dont elles sont tirées. Ceci ne permet pas une grande flexibilité de la part des gestionnaires. Deuxièmement, il se peut que les variations générées aléatoirement dans cette étude de cas ne soient pas représentatives de la réalité. Dans une étude de cas réelle, il faudra vérifier les lois de distributions des possibilités forestières par essence et qualité de bois.

Pour chaque scénario du modèle d'allocation des ressources, dix sous-scénarios du modèle de localisation des chantiers de coupe ont été réalisés. Dans chaque sous-scénario, les valeurs cibles du critère sont identiques, mais les valeurs moyennes anticipées du critère sont

différentes pour les dix sous-scénarios. Pour chaque scénario et sous-scénario, les volumes disponibles anticipés par chantier de coupe ont des valeurs différentes, puisqu'elles ont été générées aléatoirement selon une loi de distribution normale ayant des paramètres variant selon les scénarios.

Le modèle de localisation des chantiers de coupe cherche à minimiser les écarts entre les valeurs cibles et les valeurs moyennes anticipées du critère pour chaque usine. Les scénarios pour lesquels des sous-scénarios sont non réalisables sont le premier et le cinquième scénario. Ces sous-scénarios ne peuvent satisfaire la demande des usines, c'est-à-dire que les volumes générés par essence et par qualité de bois ne peuvent satisfaire l'allocation des volumes de bois aux usines.

Pour des fins de comparaison, un onzième sous-scénario a été créé. Ce sous-scénario représente la situation où les gestionnaires n'utiliseraient pas le modèle avec la fonction d'anticipation (I), mais utiliseraient la planification hiérarchique classique (H), c'est-à-dire que les solutions du modèle d'allocation des ressources sont directement transmises au modèle de localisation des chantiers de coupe en tant que décisions. Le modèle mathématique utilisé est similaire à celui de localisation des chantiers de coupe avec les contraintes de couplage. Des paramètres ont été ajoutés ainsi que des contraintes afin de respecter les décisions prises lors de la résolution du modèle d'allocation des ressources. Les paramètres binaires suivants ont été ajoutés afin de garantir la stabilité des solutions : 1) un paramètre qui détermine si les unités d'aménagement sont ouvertes et 2) un paramètre de liaison entre les unités d'aménagement et les chantiers de coupe. La contrainte de la demande minimale du modèle de localisation des chantiers de coupe avec la planification intégrée a été modifiée pour la planification hiérarchique : le volume minimum à allouer aux usines est maintenant égal à la somme des volumes alloués aux usines pour chaque essence et qualité de bois pour l'ensemble des unités d'aménagement. Les contraintes de couplage ne sont pas présentes dans la planification hiérarchique, mais une contrainte a été ajoutée afin de respecter les volumes déterminés lors du modèle d'allocation des ressources. Deux contraintes ont été ajoutées pour la planification hiérarchique : 1) une contrainte qui ouvre un chantier de coupe uniquement si l'unité d'aménagement auquel il appartient est ouverte, et 2) une contrainte qui borne la somme des volumes alloués aux usines pour tous les chantiers de coupe par la possibilité forestière pour chaque essence et qualité de bois si les unités d'aménagement sont ouvertes.

Pour les sous-scénarios résolus avec une méthode de planification hiérarchique (H), les volumes disponibles pour chaque chantier de coupe sont égaux à la moyenne des volumes disponibles anticipés des dix premiers scénarios. Pour des fins de comparaison, il en est de même pour tous les autres paramètres du modèle de localisation des chantiers de coupe résolu avec une planification

hiérarchique, c'est-à-dire que pour chaque scénario d'allocation des ressources, les paramètres pour le onzième sous-scénario sont égaux à la moyenne des paramètres des dix premiers sous-scénarios. Les résultats de la fonction-objectif pour les sous-scénarios de planification hiérarchique pure et la moyenne des fonctions-objectifs des sous-scénarios avec fonction d'anticipation. Dans le cas où des sous-scénarios sont non réalisables, les moyennes sont réalisées sur les sous-scénarios réalisables. Les sous-scénarios ne pouvant satisfaire l'allocation des bois décidée lors du niveau supérieur sont considérés comme non réalisables. Comme la somme des écarts du critère du coût d'approvisionnement entre sa valeur cible et sa valeur réelle est à minimiser, les scénarios avec une planification intégrée (I) où tous les sous-scénarios sont réalisables obtiennent en moyenne un meilleur résultat que les scénarios avec une planification hiérarchique (H) (voir Figure 2).

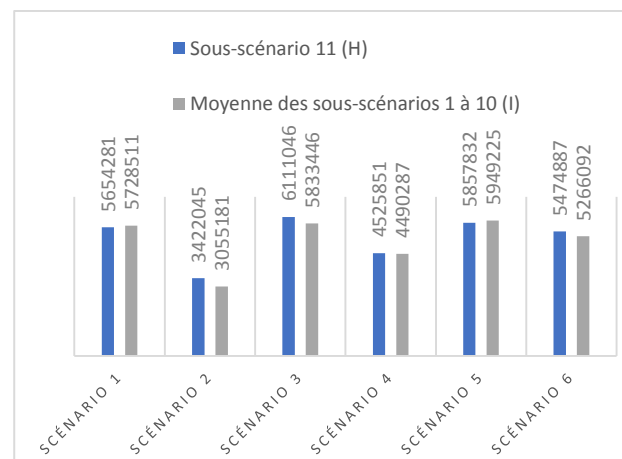


Figure 2 : Graphique comparatif des résultats de la fonction-objectif en \$ pour les sous-scénarios de planification hiérarchique pure (H) et la moyenne des sous-scénarios avec fonction d'anticipation (I)

Pour chaque sous-scénario, les informations dans le tableau 4 ont été compilées à partir des variables de décisions des solutions des sous-scénarios avec fonction d'anticipation. Les mêmes informations pour les sous-scénarios sans fonction d'anticipation sont présentées dans le tableau 5. Pour chaque scénario, le sous-scénario sans fonction d'anticipation ouvre en moyenne autant, ou plus, de chantiers de coupe que les sous-scénarios avec fonction d'anticipation. Pour chaque scénario, le volume moyen alloué aux usines par chantier est quasiment identique pour les sous-scénarios avec la planification intégrée (I) et pour le sous-scénario avec la planification hiérarchique (H). Les volumes moyens coupés par chantier de coupe sont dans le même ordre de grandeur pour les deux types de sous-scénarios. Cependant, les scénarios 1 à 4 ont un volume plus important lors des sous-scénarios sans fonction d'anticipation, contrairement aux scénarios 5 et 6 qui ont un volume plus important lors des sous-scénarios avec fonction d'anticipation. Les coûts d'approvisionnement sont plus faibles par m³ pour

les sous-scénarios avec fonction d'anticipation et ayant tous les sous-scénarios réalisables que pour les sous-scénarios sans fonction d'anticipation. Par conséquent, les gestionnaires peuvent choisir des solutions qui en moyenne aboutissent à de meilleurs résultats en utilisant une planification intégrée plutôt qu'une planification hiérarchique.

4.3.2 Analyse

Le scénario 5 a un seul sous-scénario de réalisable avec la fonction d'anticipation et le sous-scénario sans fonction d'anticipation est réalisable. Dans ce cas, il y a de fortes probabilités que les décisions prises lors du niveau d'allocation des ressources ne puissent être réalisables lors de la planification de localisation des chantiers de coupe. Les gestionnaires devraient chercher à modifier les paramètres ou les contraintes de ce scénario de façon à avoir davantage de sous-scénarios réalisables. Le sous-scénario avec la planification hiérarchique est réalisable ici, car ces paramètres sont égaux à la moyenne des paramètres des dix sous-scénarios résolus avec la planification intégrée. Les paramètres ainsi générés sont donc moins contraignants et permettent de trouver une solution réalisable avec une planification hiérarchique.

Le scénario 1 a trois sous-scénarios sur dix de non réalisables. Les informations supplémentaires dans les tableaux 4 et 5, ainsi que le résultat des fonctions-objectifs (figure 2), montrent que ce scénario est en moyenne plus coûteux dans les sous-scénarios avec fonction d'anticipation que dans le sous-scénario sans fonction d'anticipation. Si ces informations ne suffisent pas à prendre une décision, les gestionnaires devraient réaliser des études complémentaires : création de nouveaux sous-scénario, création d'un scénario *worst-case*, variation des paramètres du scénario initial, etc. Ces études complémentaires permettraient d'avoir une meilleure analyse des scénarios et de leur impact sur leur faisabilité opérationnelle. Par exemple, un scénario *worst-case* permettrait d'évaluer la performance des décisions sur l'allocation des ressources pour chaque scénario pour la localisation des chantiers de coupe dans le pire des cas (manque important de certaines essences et qualités de bois, ouverture de tous les chantiers de coupe, etc.) et de prévoir les décisions qui seraient à prendre pour minimiser l'impact de ces changements sur la planification.

Pour les quatre autres scénarios, il semble raisonnable de penser que les solutions du modèle d'allocation des ressources permettront d'avoir des solutions réalisables lors du modèle de localisation des ressources. Les décisions prises lors de l'allocation des ressources semblent robustes face aux variations du volume de bois disponible.

Dans les tableaux 4 et 5, le nombre de chantiers ouverts en moyenne pour chaque scénario est proche du nombre de chantiers disponibles dans cette étude de cas, soit vingt chantiers de coupe. Ceci est dû aux volumes de bois disponibles pour chaque chantier de coupe. En effet, certains bois avec une essence et une qualité précise sont

très demandés par les usines, comme le bois de SEPM avec la qualité de bois de sciage. Or, pour ces essences et qualités de bois, la somme des volumes disponibles sur les vingt chantiers de coupe est très proche de la somme des volumes demandés par les usines.

Scénario	1	2	3	4	5	6
Moyenne du nombre de chantiers ouverts	19	13	19	20	20	18
Volume moyen alloué aux usines par chantier ouvert (m ³)	350,30	350	350	314,80	341	346,30
Volume moyen coupé par chantier ouvert (m ³)	1 138	779	1 196	868	1 127	1 168
Coût moyen d'approvisionnement (\$/m ³)	55,27	46,20	50,89	53,38	57,22	54,56

Tableau 4 : Informations supplémentaires sur les solutions obtenues pour chaque sous-scénario avec une planification intégrée (I)

L'écart entre le volume moyen coupé par chantier de coupe et le volume moyen alloué aux usines signifie qu'il y a un grand volume résiduel. Ceci est dû à un volume disponible non balancé par essence et qualité de bois, et par une importante variation de la demande par essence et qualité de bois. En effet, les volumes disponibles anticipés du bois par essence et par qualité ont été générés en fonction d'une loi de distribution ne tenant pas en compte la réelle répartition de la population des essences de bois selon les chantiers de coupe. De plus, la demande par essence de bois varie fortement, ce qui a un impact sur le nombre de chantiers de coupe à ouvrir pour satisfaire cette demande. Or, pour chaque chantier de coupe ouvert, les essences et qualités de bois qui ne sont pas demandées par les usines doivent être tout de même coupées, ce sont les volumes résiduels.

Le coût moyen d'approvisionnement pour chaque scénario reste proche de la valeur de 50\$, ce qui est pertinent puisque les valeurs cibles pour chaque usine sont comprises entre 50\$ et 55\$. Cependant les variations observées entre les coûts moyens d'approvisionnement pour chaque scénario sont dues aux choix d'ouverture des chantiers de coupe et aux variations dans la génération aléatoire des valeurs moyennes anticipées des coûts d'approvisionnement pour chaque scénario.

Scénario	1	2	3	4	5	6
Nombre de	20	13	19	20	20	20

chantiers ouverts						
Volume moyen alloué aux usines par chantier ouvert (m ³)	350,20	538,80	368,70	314,70	340,70	346,20
Volume moyen coupé par chantier ouvert (m ³)	1 141	780	1 197	870	1 125	1 048
Coût moyen d'approvisionnement (\$/m ³)	53,94	50,51	53,05	53,90	55,31	56,01

Tableau 5 : Informations supplémentaires sur les solutions obtenues pour chaque sous-scénario avec une planification hiérarchique (H)

5 CONCLUSION

Les méthodes actuelles de planification, aussi bien les approches monolithiques que les planifications hiérarchiques, créent des divergences dans les solutions des différents niveaux de planification. Les méthodes d'intégration de plusieurs niveaux de planification permettraient de réduire ces divergences. Dans cet article, une méthode d'intégration tactique-opérationnelle d'une planification hiérarchique a été développée et appliquée à une étude de cas dans le secteur forestier. Cette méthode a utilisé des versions de deux modèles mathématiques existants dans le processus de planification des ressources forestières : un modèle d'allocation des ressources et un modèle de localisation des chantiers de coupe. Ces deux modèles font partie d'une planification hiérarchique qui entraîne des divergences entre leurs solutions. Ces divergences peuvent être d'ordre spatial, temporel ou liées aux contraintes elles-mêmes. Afin de réduire ces divergences, la méthodologie de Schneeweiss (2012) a été utilisée pour introduire une fonction d'anticipation lors de la planification du niveau supérieur pour anticiper la réaction du niveau inférieur à ses décisions.

Cette méthode d'intégration de deux niveaux de planification avec fonction d'anticipation est un outil d'aide à la prise de décisions et ne peut être comparée à une optimisation de modèle mathématique, comme l'a précisé Beaudoin *et al.* (2008). La méthodologie présentée dans cet article permet de générer des scénarios afin de déterminer lors de la planification de l'allocation des ressources si les décisions prises seront réalisables lors de la localisation des chantiers de coupe. Dans le cas contraire, les gestionnaires pourront s'appuyer sur une étude statistique afin de déterminer les risques d'obtenir une solution non réalisable au niveau inférieur.

Les résultats obtenus permettent de converger plus rapidement vers une solution réaliste pour le décideur. Dans cette perspective, cette méthode peut être expérimentée plus précisément avec les acteurs du monde forestier afin d'approcher plus justement les gains potentiels en termes d'exécution de prises de décisions, ainsi qu'au niveau des résultats d'exploitation. Par ailleurs, cette expérimentation permettra d'identifier plus rigoureusement le nombre de variables et de contraintes générées afin de perfectionner la modélisation et le nombre de scénarii à générer pour les décideurs. Enfin, cette méthode pourra être déployée sur des secteurs d'activités demandant de prendre en compte des contraintes opérationnelles sur une planification stratégique.

La méthodologie proposée dans cet article pourrait à terme s'inscrire dans le cadre des recherches sur l'industrie 4.0. En effet, l'industrie 4.0 cherche à lier toutes les parties d'une même chaîne logistique par un système d'information commun. Dans cette optique, les méthodes d'intégration de plusieurs niveaux de planification sont une étape importante pour la modélisation des objectifs communs à toute la chaîne logistique.

REFERENCES

- Abasian F., M. Rönnqvist, P. Marier, JG. Gill-Couture, and D. Fjeld, 2016. The transportation game. *Communication présentée à la 11th International Conference on Modeling, Optimization and Simulation (MOSIM 2016)*, Montreal.
- Andersson, Daniel, 2005. *Approaches to integrated strategic/tactical forest planning*. Thèse de license, Swedish University of Agricultural Sciences, Umeå.
- Beaudoin, D., J. M. Frayret, and L. Lebel, 2008. Hierarchical forest management with anticipation: An application to tactical-operational planning integration. *Canadian Journal of Forest Research*, 38(8), 2198-2211
- Bouchard, M., S. D'Amours, M. Rönnqvist, R. Azouzi, and E. Gunn, 2017. Integrated optimization of strategic and tactical planning decisions in forestry. *European Journal of Operational Research*, 259(3), 1132-1143.
- Boukherroub, Tasseda, Luc LeBel, and Angel Ruiz, 2017. A framework for sustainable forest resource allocation: a Canadian case study. *Omega*, 66, 224-235.
- Ouhimmou, Mustapha, Samir Haddad, and Tasseda Boukherroub, 2015. Decision support system for selection of harvest areas and wood allocation to mills. *15th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 2015)*, vol. 48(3), p. 628-633.
- Schneeweiss, Christoph (2012). Distributed decision making. *Springer Science & Business Media*

INVENTORY LOCATION ROUTING PROBLEM: ISSUES, MODELS AND METHODS

Sanaa Lerhlaly^{1,2}, Maria Lebbar^{3,5}, Sohaib Afifi¹, Hamid Allaoui¹, Driss Ouazar^{2,4}

¹ LGI2A, University of Artois, France

² LASH, Mohammadia School of engineering, Mohammed V University, Rabat, Morocco

³ SYSCOM, Rabat School of Mines, Morocco

⁴ University Mohammed VI Polytechnique, Benguerir, Morocco

⁵ MASI, Mohammadia School of engineering, Mohammed V University, Rabat, Morocco

ABSTRACT: Facility location, vehicle routing and inventory management are the three primary sub-problems involving the design of a distribution network design. Researchers have recently realized the importance of integration of all the three components of the supply chain management into one model. The inventory-location-routing problem (ILRP) consists in opening a subset of depots, assigning customers to them, determining vehicle routes, quantities delivered and inventory policies, in order to minimize the total cost (location, routing and inventory). Since, according to our knowledge, there's no survey for the ILRP, this paper analyzes most of studies of the last two decades on the field of ILRP. We summarize the ILRP aspects such as: data type, time horizon, single or multiple echelons, inventory policy and place, one or multiple products and the fleet composition. The proposed models in the literature for the ILRP and the solution approaches are also listed.

KEYWORDS: Inventory location routing problem, distribution network design, solution methods.

1 INTRODUCTION

Supply chain management is a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses and stores; so that we produce and distribute the right product at the right quantities, to the right location, and at the right time, in order to minimize system costs while satisfying the clients' demand (Simchi-Levi et al., 2003). This involves a set of management activities like purchasing, inventory control, production, sales and distribution. Generally, decisions for logistics and distribution can be categorized into three levels depending on the investment requirement, the time horizon and frequency of the decisions involved. It's about the strategic, tactical, and operational decisions. The classical approach for the optimization of logistics systems is the decomposition of the system into sub problems according to these decision levels.

The facility location problem (strategic level) comprises fixing number, location and capacity of warehouses, which aims to meet customer demand at lower cost. This problem was reviewed by (Drezner, 1995; Mel et al., 2009; Owen and Daskin, 1998).

The vehicle routing problem (tactical or operational level) consists of the routes construction for several vehicles from a depot to a set of geographically distributed points in order to satisfy, with a minimal cost, the required demands in the network and return to the depot when the capacity of vehicle and/or some other criteria is met. In 1983, Bodin et al. presented a first classification of routing problems. Then several states of the art have been made in this direction including Drexl (2013), Laporte (2009), Laporte and Osman (1995) and more recently Lin et al. (2014) for the green VRP. The solving

techniques used for this problem were summarized by Laporte (1992), Laporte et al. (2000), and (Toth and Vigo, 2002) for the routes of limited capacity vehicles.

The approach of decomposition has many advantages such as simplification of modeling and problem solving techniques. However, this can often lead to a sub optimality of solutions because it doesn't take into account the whole issues of the problem (Prodhon, 2006). Several researchers studied the interdependence among the three sets of decisions, and the benefits of their integration (Perl and Sirisoponsilp, 1998). Fig.1 illustrates problems considering the integration of logistics levels decisions.

In the literature, many papers integrate two of levels decisions. On one hand, Location-Routing Problem (LRP) combines two decision levels, a strategic level which is the facility location and operational or tactical level related to vehicle routing problem (VRP). Salhi and Rand (1989) highlighted for the first time the benefit of integration of these types of decisions. Laporte (1992) presented a first classification of this problem. Another survey has been made in this direction by Nagy and Salhi (2007). Seven years after, Prodhon and Prins (2014) update this survey by reviewing the literature published from 2007 and papers not discussed by Nagy and Salhi (2007). Drexl and Schneider (2014) presented in another survey variants and extensions of the standard LRP.

On the other hand, Inventory-Routing Problem (IRP), integrates inventory management (Tactical level) and vehicle routing (operational level). It was surveyed in Baita et al. (1998), Kleywegt et al. (2002), and Vidovic et al. (2014).

Finally, the integration of facility location problem and inventory management is named Inventory-Location Problem (ILP); Daskin et al. (2002) and Shen (2005) studied different variants of ILP. A review on ILP models is provided in Farahani et al. (2015).

Recently, the progress of operations research and computer science encourages researchers to try to solve more complex and large scale problems, and thus have a more global vision of supply chain problems.



Fig.1: Structure of literature linked to the intersection of logistics problems

This will help to anticipate interactions between decisions levels. In this sense, the Inventory Location-Routing Problem (ILRP) can be described as an approach to integrate the three components (facility location, vehicle routing and inventory management) into one model, in order to optimize the supply chain design and minimize the total costs (Guerrero et al., 2013).

In this paper we will emphasize the issues, models and methods related to the combination of location, routing and inventory management. Thus, we present (in chronological order) some works identified in the literature relating to ILRP.

The keywords used for data collection include “Location”, “Routing”, “Inventory”, “Supply chain”, “Design”. Two combinations of these keywords were used including: (1) Inventory AND Location AND Routing, (2) Supply chain AND Design. We collected journal articles (conference papers excluded), from Scopus database, for the defined search terms.

The remainder of this paper is organized as follows. Section 2 provides an overview of the problems dealing with the integration of the ILRP. In Section 3, formulations and models for ILRP are presented, along with resolution approaches in Section 4. Finally, Section 5 presents the conclusions.

Table 1 summarizes respectively, for each paper, the ILRPs aspects, type of model, the solution method and the size of the largest instances solved (if available).

2 CLASSIFICATION OF ILRPs

In this section we present the different problems studied in the literature in the context of the integration of the ILRP according to different characteristics, namely: the data type (deterministic, stochastic or fuzzy), time horizon, single or multiple echelon, inventory policy and place, one or multiple products and the fleet composition.

Lui and Lee (2003) are among first authors to deal with the ILRP. They studied a location-routing problem (LRP) including nonlinear inventory costs for a single product and multi-depot problem with unlimited capacity and stochastic demand. The same problem was presented by Liu and Lin (2005) and decompose it into a depot location- allocation problem and a routing and inventory problem.

Ambrosino and Scutella (2005) presented a four levels distribution network consisting of supply points, central depots, regional facilities and demand points. Authors considered two types of customers; regular customers, and big customers (big demand). Each client has a deterministic demand and served with a heterogeneous fleet over a multi-period planning horizon.

Shen and Qi (2007) studied a single-period, three-layer supply chain with one or more suppliers, a set distribution centers (DCs) and a set of customers. Customers have uncertain demand and each open DC is served directly by the supplier and distributes products to customers. Each customer orders at the beginning of the period and each DC orders inventory from the supplier using a (r, Q) policy, with service level constraints. The same problem was studied by Javid and Azad (2010).

A single-product, multi-period ILRP with stochastic demand was studied by Zhang et al. (2008). Clients must place an order at their assigned facility once every period. The facilities have a facility-specific lead time and plan their stock levels taking into account fixed ordering and variable holding and shortage costs.

Reza Sajjadi and Hossein Cheraghi (2011) handled a multi-products and third-party logistics under the fixed order interval inventory policy, considering a stochastic demand and a homogeneous fleet.

A multi-period ILRP with deterministic demand and homogeneous unlimited fleet of vehicles was studied by Guerrero et al. (2013) and Guerrero et al. (2015). The authors considered three-layer supply chain with one plant, a set of potential depots, and a set of retailers.

Ronald G. Askin and al. (2013) presented a framework for designing a multi-product integrated supply chain network. Each product originating from its unique production plant, and delivered directly to retailers or with using intermediate warehouses for shipment consolidation and/or inventory pooling prior to final delivery to retail demand points.

Dehui Chen and al. (2014) studied a ILRP in B2C E-Commerce distribution system, which consists of a single supplier, multiple distribution centers and customers.

Customers and distribution centers demand is considered fuzzy random variables. Each distribution center follows (T, r_j) periodic review inventory policy. (In this policy, the inventory level is observed only at discrete points in time T (review period length). At each review point the decision maker may place a replenishment order if the observed inventory level is less than or equal to the reorder point.)

Nekooghadirli et al. (2014) studied a bi-objective location-routing-inventory (LRI) that considers multi-period, stochastic demands, multi-product and probabilistic travelling time among clients.

Ying Zhang et al. (2014) considered supply chain network with multiple depots and geographically dispersed clients. Each customer has deterministic but period-variable demand, delivered by a homogeneous fleet of capacitated vehicles. The depot serves as coordinator of the replenishment process and they act as “break-bulks” or transshipment points.

Liu and al (2015) studied a three-phase supply chain distribution system for e-commerce, consisting of one supplier, a set of retailers, and a single type of product with continuous review (Q, r) inventory policy. (In this policy when the inventory level gets to or below a reorderpoint, a fixed quantity Q is ordered to the supplier).

Lerhlaly et al. (2016) studied a framework for designing a multi-period ILRP for hazardous materials considering costs, risks and routing’s CO_2 emissions. Clients have a non-constant deterministic demand. The hazardous product is delivered from depots to clients by a capacitated heterogeneous fleet. Each vehicle visits at most a fixed number of clients and return to the starting point.

Zhalechian et al. (2016) studied a sustainable design of a closed-loop supply chain network design by considering economic, environmental and social measures.

Hiassat et al.(2017) studied a multi-period inventory location routing problem with non-constant deterministic demand for perishable product. They studied the distribution (with homogeneous fleet of vehicles) of a single product from a single manufacturer to a set of retailers, through a set of candidate warehouses that can be located. The shelf-life of perishable product was measured by the number of time periods.

Rayat et al. (2017) handled a bi-objective ILRP that considers multi-product, multi-period and disruption risks.

Vahdani et al. (2017) studied a multi-period and multi-commodity three level relief chain for humanitarian logistics network design under uncertainty.

A three echelon multi-period ILRP for perishable products was studied by Zahra Rafie-Majd et al,(2018). The customers’ demand is stochastic and follows normal distribution with certain mean and standard deviation.

3 MODELS

In this section we discuss the models proposed in the literature to formulate the ILRPs.

Liu and Lee (2003) and Liu and Lin (2005) build a mathematical model for ILRP with nonlinear inventory costs. The aim is to minimize the total cost of selected depots, transportation, and also three inventory costs at the clients: ordering, holding, and shortage costs for an order-up-to-level policy. An extension of this model was proposed by Reza Sajjadi and Hossein Cheraghi (2011) to add a third-party logistics that offers a flexible excess space for selected warehouses.

Ambrosino and Scutella (2005) proposed two mathematical formulations for static and dynamic cases: a three-index arc-variable formulation and a flow formulation.

A non-linear integer program model was studied by Shen and Qi (2007) in order to integrate the location, the inventory, and the routing approximate costs. Authors assume that the transportation costs and the inventory costs present economies of scale under which the average unit cost decreases as the total volume of activity increases, this result in nonlinear terms in the model. The route costs are estimated using a continuous approximation model. The objective is to open a subset of DCs and to assign customers to them in order to minimize the costs of location, transportation and inventory at the DCs.

Zhang et al.(2008) presented an arc-variable based MIP model in order to minimize the sum of facility opening, vehicle routing, and inventory facilities costs.

Javid and Azad (2010) applied Shen and Qi’s model to find simultaneously the decisions of location, allocation, inventory at depots and then they calculate the exact vehicle routing cost from each opened distribution center to its assigned customers routing. The model is formulated as a mixed integer convex program.

Guerrero et al.,(2013) and Guerrero et al (2015) presented a MILP models and two sets of valid inequalities to strengthen it, in order to make decisions on facility configuration, vehicle routing, the shipment quantities from the supplier to facilities and from facilities to retailers, and the sequence in which retailers are replenished. The cost includes the facility opening costs, transportation costs, ordering costs and inventory costs at facilities and customers, including an obsolescence penalty cost.

Dehui Chen and al.,(2014) developed a fuzzy random mixed 0-1 integer programming model of ILRP in B2C E-Commerce distribution system. The aim is to determine the number and location of distribution centers, vehicle routing for its allocated clients, optimal order quantity and the reorder point of each open Distribution center.

Nekooghadirli et al.,(2014) presented a mixed integer nonlinear programming (MINLP) model to minimize the total cost and maximize mean time for delivering commodities to clients.

Liu and al.,(2015) proposed a nonlinear integrated programming model based on forward and reverse logistics networks about stochastic LIRP under e-supply chain environment. The aim is to determine the quantity,

locations, order times, and order size of merchandise centers and arrange the routes

A MIP model was used by Ghorbani and Jokar (2016) in order to determine: the location of depots from a given predetermined set of candidate locations, the set of customers allocated to each opened depot, inventory policy and route vehicles to transmit customers' demand such that the total system cost is minimized.

Lerhlaly et al. (2016) presented a MIP bi-objective model in order to minimize the total cost and the routing CO₂ emissions.

A mixed-integer linear programming model (MILP) was introduced by Hiassat et al. (2017) in order to determine the location and number of selected depots, the travelling routes, the quantities to deliver to retailers, and the level of inventory at each retailer.

A two multi-objective two-stage mathematical models were developed by Vahdani et al (2017), taking into account the hard time windows limit and the reliability of the route. The first phase concerns decisions for warehouses and distribution centers location and determining the amount of inventory. While in the second phase, the vehicle routing and distribution of goods from the distribution centers to the affected areas was decided considering the limitation of the timing windows.

Rayat et al. (2017) presented a bi-objective mixed-integer nonlinear programming ILRP model in order to minimize the total network cost and the failure costs related to disrupted distribution centers.

Zahra Rafie-Majd et al. (2018) proposed a MINLP model in order to minimize the total costs, including the costs of opening DCs, expected waste cost, fuel cost, driver cost, setup cost, holding safety stock by distribution centers, orders by distribution centers to the supplier, and inventory holding cost at customers and DCs.

4 SOLUTION APPROACHS

The ILRP belongs to the class of NP-hard problems as it encompasses three NP-hard problems. The complexity of the problem makes it very hard to solve medium and large instances to proven optimality. Therefore, a few authors designed exact methods. Approximation algorithms are often used. In addition, some articles suggested a combination of exact and heuristic procedures leading to a heuristic defined as a matheuristic (Guerrero et al. (2013), Guerrero et al. 2015).

Lui and Lee (2003) used two-phase heuristic method to solve the problem. The first phase is to find a better initial solution based on route-first, location-allocation second approach, while the second is to improve the initial one. For the same model, Liu and Lin (2005) proposed a hybrid heuristic based on the Tabu search (TS) and simulated annealing (SA). The computational results of the proposed algorithm indicate better solution than what it was proposed by Liu and Lee (2003).

Ambrosino and Scutella (2005) used a linear programming relaxation for the ILRP and find optimality using commercial solver (Cplex).

In Shen and Qi (2007) the problem is formulated, and solved with using Lagrangian relaxation embedded in branch-and-bound. The algorithm is tested on instances with 40–320 customers.

Zhang et al. (2008) proposed a mixed genetic algorithm to solve the ILRP. To compute the total inventory costs, the authors developed a formula in each period. The saving algorithm determines vehicle routes for each opened facility. Individuals are selected for reproduction with a roulette-wheel procedure. A two-point crossover operator and a mutation operator that randomly exchanges two values in the encoding bit-string are used.

Javid and Azad (2010) proposed, to solve the problem, an exact method and then a heuristic method decomposed into constructive stage and improvement stage. In the constructive stage a randomly generated initial solution is built. In the improvement stage the location/allocation and routing decisions are iteratively tackled in two separate phases by means of a hybrid of SA and TS to improve the initial solution in each phase. The algorithm is tested on instances with 40–400 customers and 50 DCs. Compared to Shen and Qi's model, the expected saving varies between 16–28% for 40 customers and 10–18% for 400 customers.

Reza Sajjadi and Hossein Cheraghi (2011) presented a solution approach based on simulated annealing that solves randomly generated instances of 40 products, 350 clients, and one period.

An algorithm for a multi-period ILRP with deterministic demand was presented by Guerrero et al. (2013). They proposed a matheuristic which aims at alternatively solving a supply chain design problem with estimated distribution costs, using exact methods, and determining the routing decisions, using heuristic procedures. The algorithm optimizes globally the components of the problem by sharing information between decisions levels. Guerrero et al. (2015) investigated a "relax and price" heuristic for the same problem. They coordinate the generation of new columns in the pricing problem and update of Lagrangian multipliers. The results of computational experiments on randomly generated ILRP instances, show important cost saving and efficient computation if compared to the previous heuristic.

Askin et al. (2013) implemented an evolutionary genetic algorithm to solve an inventory location model considering multi-commodity and distribution planning decisions.

Ying Zhang et al. (2014) proposed a hybrid metaheuristic solution to solve an ILRP.

To solve the studied problem, Liu and al (2015) developed a pseudo-parallel genetic algorithm integrating simulated annealing (PPGASA).

Zhalechian et al. (2016) proposed a hybrid approach including a stochastic-probabilistic programming method to cope with the uncertainty and to address the multiple conflicting objectives of the proposed model.

Ghorbani and Jokar (2016) used a hybrid heuristic based on the simulated annealing and imperialist competitive algorithm (IC-SA) to solve their model. The proposed algorithm is compared with simulated anneal-

ing (SA) algorithm in small and large size instances. The results show that the (IC-SA) algorithm outperforms the (SA) algorithm in terms of solution quality and CPU time.

Hiassat et al.2017 developed a Genetic Algorithm approach to find an optimal or near optimal solution. An archived multi-objective simulated annealing meta-heuristic solution approach was proposed by Rayat et al. (2017) to solve their model.

Zahra Rafie-Majd et al,(2018) proposed a Lagrangian Relaxation Method to solve the ILRP and determine the lower bound; and a heuristic algorithm was used to feasibility the result of the Lagrangian Relaxation Algorithm and determine the upper bound.

Authors	Product (multi/single)	Period (multi/single)	Inventory		Demand (det/st/fuzzy)	FC	Model	Objective	Solution Method	Size of prob
			Rp policy	Place						
Liu and Li (2003)	single	single		C	St	Hom	NLP	Mono	route first-locate second	
Liu and Lin (2005)	single	multi		C	St	Hom	NLP	Mono	Sequential/improv. Stage	
Ambrosino and Scutellà(2005)	single	multi		D/C	det	Het	LP	Mono	Exact	
Shen and Qi(2007)	single	single	(r,Q)	D	St	Hom	NLP	Mono	Branch-and-bound	
Zhang et al.,(2008)	single	multi		D	St	Hom	MILP	Mono	mixed GA	
Javid and Azad(2010)	single	single	(Q,R)	D	St	Hom	NLP	Mono	TS and SA	
Sajjadi and Cheraghi(2011)	multi	single		C	St	Hom	NLP	Mono	SA	40 Prod/350C
Ahmadi-Javid and Seddighi (2012)	single	single		D	det	Hom	MINLP	Mono	three phase heuristic	
Guerrero and al.,(2013)	single	multi	order-up to level	D/C	det	Hom	MILP	Mono	Iterative Metaheuristic	15C/5 D/7P
Ronald G. Askina and al.,(2013)	multi	single		D/C	det	Het	NLP	Mono	GA	
Dehui Chen and al.,(2014)	single		(T, rj)	D	Fuzzy	Hom	fuzzy random MIP	Mono	two phase TS	
Ying Zhang and al.,(2014)	single	multi	Flexible	C	det	Hom	LP	Mono	SA-Hybrid-ILRP	300C/25D/7P
N. Nekooghadrili(2014)	multi	multi	(Q, R)	D	St	Het	MINLP	bi-objective	(MOICA), MOPSA, NSGA-II,PAES	
Guerrero and al.,(2015)	single	multi		D/C	det	Hom	MILP	Mono	Metaheuristic (Relax and Price)	5D/15 C/7P
liu and al.,(2015)	single	single	(Q,R)	C	St	Hom	NLP	Mono	PPGASA	
Lerhlaly et al.,(2016)	single	multi		C	det	Hom	MILP	bi-objective	Exact	20C/5 D/5P

M. Zhalechian and al.,(2016)	multi	multi	(Q,R)	D	St	Het	MINLP	Multi	hybrid approach	
Ghorbani and Jorkar (2016)	multi	multi		D	det	Hom	MILP	Mono	IC-SA	
Hiassat and al., (2017)	single	Multi		C	det	Hom	LP	Mono	GA	
Farnaz Rayat,et al.,(2017)	multi	multi		D	St	Het	MINLP	Multi	AMOSIA	
Vahdani et al. (2017)	multi	multi		D	St	Het	MIP	Multi	NSGAI – MOPSO	
Zhao and Ko (2017)	single	single		C		Het	MILP	bi-objective	TOPSIS	
Zahra Rafie-Majd et al.,(2018)	multi	multi	(Q,R)	D/C	St	Het	MINLP	Mono	Lagrangian Relaxation Algorithm	

Table1. A summary of ILRP's studies

dt: Deterministic, st: Stochastic, FC: Fleet composition, D: Depot; C: Clients; Het: Heterogeneous; Hom: Homogeneous; LP: Linear programming; NLP: Non linear programming; Prob: Problem; P: Period; Prod: Product.

CONCLUSION

In this paper we analyze most of studies during the last two decades concerning the ILRP. We summarize the ILRP aspects such as: data type, time horizon, single or multiple echelons, inventory policy and place, one or multiple products and the fleet composition. We list also the different types of mathematical formulations proposed in the literature for the ILRP and the solution approaches to resolve it.

This enumeration unveils same number of papers on stochastic and deterministic; and only one paper on fuzzy ILRPs. The multi-objective ILRPs has received a little attention in comparison with the mono-objective ILRPs. In fact, at the beginning, many studies on the ILRP considered only one objective function to be minimized, namely the total cost of opened depots, routing and inventory. However, for application fields like the hazardous materials or also humanitarian logistics, there are conflicting optimization criteria leading to multi-objective mathematical programming approaches.

Only six papers consider the heterogeneous fleet. The most of works develops heuristics approaches; on the other hand, too few exact methods have been studied. In addition, new methods are used combining exact and heuristic methods.

In the present paper we can notice that there is a clear trend towards considering more complex problems like: multi-objective ILRPs, multi-echelon ILRPs, ILRPs with uncertainty...

This paper can provide useful information for researchers and motivate further studies in the inventory location routing problem.

For future research opportunities, we propose the following areas:

- Modelling more real situations: this confirms the applicability of ILRP, and helps practitioner to realise the

importance of using ILRP models and methods to solve their conception problems.

-Benchmarks for solution efficiency: Although the ILRP is new, it has received a considerable amount of attention. According to our knowledge, there are no available standard benchmark instances for the ILRP, making difficult to properly judge solution quality. Furthermore, only some papers ((Liu and Lee (2003), Liu and Lin (2005)); (Guerrero et al.,(2013), Guerrero et al (2015)); (Shen and Qi (2007), Javid and Azad (2010))) give computational comparisons with their peers.

-Developing algorithms for large-scale problems.

REFERENCES

- Ahmadi Javid, A. and N. Azad, 2010. Incorporating location, routing and inventory decisions in supply chain network design. *Transportation Research Part E: Logistics and Transportation Review*, 46(5), p. 582 - 597.
- Ahmadi Javid,A., and A.Seddighi.2012. Alocation-routing inventory model for designing multisource distribution networks. *Engineering Optimization*,44, p.637–656.
- Ambrosino, D. and M.G. Scutella, 2005. Distribution network design: New problems and related models. *European Journal of Operational Research* , 165(3), p. 610–624.
- Archetti, C., L. Bertazzi, G. Laporte and MG. Speranza, 2007. A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation Science*, 41, p. 382–91
- Askin, Ronald G., I. Baffo, and M.Xia, 2014. Multi-commodity warehouse location and distribution planning with inventory consideration. *International*

- Journal of Production Research*, 52(7), p. 1897-1910.
- Baita, F., W. Ukovich, R. Pesenti, and D. Favaretto, 1998. Dynamic routing-and inventory problem: a review. *Transportation Research: Part A , Policy and Practice*, 32(8), p. 585-598.
- Bodin, L.D., B.L. Golden, A.A., Assad, and M.O. Ball, 1983. Routing and scheduling of vehicles and crews. The state of the art. *Computers and Operations Research* , 10(2), p. 63-211.
- Chen, D., D. Chen, D. G. Sun, and G. Liu, G, 2014. Combined Location Routing and Inventory Problem of E-Commerce Distribution System with Fuzzy Random Demand. *International Journal of Hybrid Information Technology*, 7. p. 429-442.
- Daskin, M. S., C. R. Coullard, and Z. J. M. Shen, 2002. An inventory-location model: Formulation, solution algorithm and computational results. *Ann. Oper. Res* , 110(1), p. 83–106.
- Drexl, M. and M. Schneider, 2014. A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, 241(2), p. 283–308.
- Drexl, M., 2013. Applications of the vehicle routing problem with trailers and transshipments. *European Journal of Operational Research* , 227(2), p. 275 - 284.
- Drezner, Z. , 1995. *Facility location: A survey of applications and methods*, Springer, New York.
- Farahani, R. Z., H. Rashidi Bajgan, B. Fahimnia, and M. Kaviani, 2015. Location-inventory Problem in Supply Chains: A Modelling Review. *International Journal of Production Research*, 53 (12), p. 3769–3788.
- Ghorbani, A. and M.R.A. Jokar, 2016. A hybrid imperialist competitive-simulated annealing algorithm for a multisource multi-product location-routing-inventory problem. *Comput. Ind. Eng*, 101, p. 116–127.
- Guerrero, W.J., C. Prodhon, N. Velasco, and C.A. Amaya, 2013. Hybrid heuristic for the inventory location-routing problem with deterministic demand. *International Journal of Production Economics*, 146(1), p. 359–370.
- Guerrero, W.J., N. Velasco, C. Prodhon, and C.A. Amaya, 2015. A Relax-and-Price Heuristic for the Inventory-Location-Routing Problem. *International Transactions in Operations Research*, 22(1), p. 129–148.
- Hiassat, A., A .Diabat, and I.Rahwan, 2017. A genetic algorithm approach for location-inventory-routing problem with perishable products. *Journal of Manufacturing Systems*, 42, p. 93–103
- Ho, T.S., S.H. Chung, and H.Y. Lam, 2014. Survey of Green Vehicle Routing Problem: Past and future trends. *Expert Systems with Applications*, 41, p. 1118–1138.
- Kleywegt, A., V.S. Nori, and M.W.P. Savelsbergh, 2002. The stochastic inventory routing problem with direct deliveries. *Transportation Science*, 36(1), p. 94-118.
- Laporte, G. and H. Osman, 1995. Routing problems: A bibliography. *Annals of Operations Research* , 61(1), p. 227-262.
- Laporte, G., 2009. Fifty Years of Vehicle Routing. *Transportation Science* , 43(4), p. 408-416.
- Laporte, G., 1992. The Vehicle Routing Problem: An overview of exact and approximate algorithms. *European Journal of Operational Research* , 59(3), p. 345-358.
- Laporte, G. , M. Gendreau, , J. Potvin, and F. Semet, 2000. Classical and modern heuristics for the vehicle routing problem. *International transactions in operational research* , 7(4-5), p. 285-300.
- Lerhlaly, S., M. Lebbar, H. Allaoui, D. Ouazar, and S. Afifi, 2016. An Integrated Inventory Location Routing Problem Considering CO2 Emissions. *Contemporary Engineering Sciences*, 9(7), p. 303-314.
- Liu, B., H. Chen, Y. Li, and X. Liu, 2015. A pseudo-Parallel genetic algorithm integrating simulated annealing for stochastic location-inventory-routing problem with consideration of returns in E-commerce. *Discrete Dynamics in Nature and Society*, 2015 (2015),
- Liu, S.C. and S.B. Lee, 2003. A two-phase heuristic method for the multi-depot location routing problem taking inventory control decisions into consideration. *The International Journal of Advanced Manufacturing Technology*, 22, p. 941–950.
- Liu, S. and C. Lin, 2005. A heuristic method for the combined location routing and inventory problem. *The International Journal of Advanced Manufacturing Technology*, 26, p. 372-381.
- Majd, Z.R., S.H.R. Pasandideh, and B. Naderi, 2018. Modelling and solving the integrated inventory-location-routing problem in a multi-period and multi-perishable product supply chain with uncertainty: Lagrangian relaxation algorithm. *Computers & Chemical Engineering*, 109(4), p. 9-22.

- Melo, M.T., S. Nickel, and F. Saldanha-da-Gama, 2009. Facility location and supply chain management - A review. *European Journal of Operational Research*, 196(2), p. 401-412.
- Nagy, G. and S. Salhi, 2007. Location-routing, issues, models, and methods: A review. *European Journal of Operational Research*, 117(2), p. 649-672.
- Nekooghadirli, N., R. Tavakkoli-Moghaddam, R., V. Ghezavati, and S. Javanmard, 2014. Solving a new bi-objective location-routing-inventory problem in a distribution network by meta-heuristics. *Comput. Indust. Eng.* 76, p. 204-221.
- Owen, S.H. and M. S. Daskin, 1998. Strategic facility location: A review. *European Journal of Operational Research*, 111(3), p. 423-447.
- Perl, J. and S. Sirisoponsilps, 1988. Distribution networks: Facility location, transportation and inventory, *International Journal of Physical Distribution and Materials Management*, 18 (6), p. 18-26.
- Prins, C., C. Prodhon, C.R. Wolfler, 2004. Nouveaux algorithmes pour le problème de localisation et routage sous contraintes de capacité. In: Dolgui, A. and S. Dauzère-Pérès (eds) MOSIM'04, vol 2, p. 1115-1122, Ecole des Mines de Nantes, Lavoisier.
- Prodhon, C., 2006. *Le Problème de Localisation-Routage*, PhD Thesis, University of Technology of Troyes, Troyes, France.
- Prodhon, C. and C. Prins, 2014. A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1), p. 1-17.
- Rayat, F., MM. Musavi, and A.B. Amiri, 2017. Bi-objective reliable location-inventory-routing problem with partial backordering under disruption risks: A modified AMOSA approach. *Applied Soft Computing*, 59, p. 622-643.
- Salhi, S. and G.K. Rand, 1989. The effect of ignoring routes when locating depots. *European Journal of Operational Research*, 9(2), p. 150-156
- Shen, Z.J., 2005. A multi-commodity supply chain design problem. *IIE Transactions*, 37(8), p. 753-762.
- Shen, Z.J.M. and L. Qi, 2007. Incorporating inventory and routing costs in strategic location models. *European Journal of Operational Research*, 179(2), p. 372-389.
- Simchi-Levi D, P. Kaminsky, and E. Simchi-Levi, 2003. *Designing and managing the supply chain: concepts, strategies and case studies*. New York: Irwin/McGraw-Hill.
- Toth, P. and D. Vigo, 2002. Relaxations and exact approaches for the capacitated vehicle routing problem. *Discrete Applied Mathematics*, 123(1-3), p. 487-512.
- Vahdani, B., D. Veysmoradi, F. Noori, and F. Mansour, 2017. Two-stage multi-objective location-routing-inventory model for humanitarian logistics network design under uncertainty. *Int J Disast Risk Reduct*,
- Vidovic, M., D. Popovic, and B. Ratkovic, 2014. Mixed integer and heuristics model for the inventory routing problem in fuel delivery. *Int. J. Production Economics*, 147(C), p. 593-604.
- Zhao, J. and Ke. Ginger Y, 2017. Incorporating inventory risks in location-routing models for explosive waste management. *International Journal of Production Economics*, 193(C), p.123-136.
- Zhang, Y., M.Qi.L. Miao, and E. Liu, 2014. Hybrid metaheuristic solutions to inventory location routing problem. *Transport. Res. Part E: Logist. Transport. Rev*, 70(0), p. 305-323.
- Zhalechian, M., R. Tavakkoli-Moghaddam, B. Zahiri, and M. Mohammadi, 2016. Sustainable design of a closed-loop location-routing-inventory supply chain network under mixed uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 89(C), p. 182-214.

IWODEMO: UN NOUVEL ALGORITHME D'ÉVOLUTION DIFFÉRENTIELLE HYBRIDE POUR L'OPTIMISATION MULTI-OBJECTIFS

Caroline GAGNÉ, Mathieu FOURNIER, Aymen SIOUD, Marc GRAVEL

Département d'Informatique et Mathématique
Université du Québec à Chicoutimi
Québec, Canada

caroline.gagne@uqac.ca, mathieu.fournier1@uqac.ca, aymen.sioud@uqac.ca, marc.gravel@uqac.ca

RESUME : *L'hybridation de métaheuristiques vise à obtenir une meilleure performance en combinant les forces de plusieurs stratégies afin de tirer profit de la synergie produite. L'optimisation multi-objectifs ne fait pas exception à cette tendance. Un nouvel algorithme hybride, IWODEMO, impliquant l'évolution différentielle pour résoudre des problèmes multi-objectifs avec variables continues est proposé dans ce papier. Cette hybridation est réalisée avec l'algorithme de la mauvaise herbe afin d'équilibrer les capacités d'exploration et d'exploitation des solutions. Les résultats obtenus montrent l'efficacité de IWODEMO qui obtient de meilleurs résultats que les autres algorithmes de la littérature (NSDE, DEMO et ADE-MOIA) utilisés en comparaison en termes de convergence, de dispersion des solutions sur le front Pareto-optimal et de temps de calculs. Afin de proposer une étude comparative fiable et équitable, les algorithmes de la littérature ont été reproduits et testés dans les mêmes conditions expérimentales.*

MOTS-CLES : *Optimisation multi-objectifs, problèmes continus ZDT, évolution différentielle, algorithme de la mauvaise herbe, hybridation, métaheuristiques.*

1 INTRODUCTION

De nombreux problèmes d'optimisation rencontrés dans l'industrie, aussi bien en optimisation combinatoire qu'en optimisation continue, nécessitent le traitement simultané de plusieurs objectifs souvent contradictoires. Dans de tels contextes, l'optimal recherché n'est plus un point unique mais un ensemble de solutions dites de compromis aussi appelé front *Pareto-optimal*. Par conséquent, on peut définir deux buts principaux en optimisation multi-objectifs (MO): (i) proposer un ensemble de solutions de compromis dont la qualité est aussi proche que possible du front Pareto-optimal; et (ii) proposer un ensemble de solutions de compromis aussi diversifié que possible. Atteindre ces deux buts en un temps raisonnable est un défi important pour tout algorithme MO (Deb *et al.*, 2002).

Les méthodes de résolution sont habituellement conçues pour l'optimisation d'un seul objectif. Elles doivent alors être suffisamment flexibles et performantes pour permettre d'en modifier la structure afin de traiter plusieurs objectifs. Ces dernières années, l'adaptation de méthodes métaheuristiques s'est avérée une alternative gagnante pour l'optimisation simultanée de plusieurs objectifs. Plusieurs extensions des métaheuristiques furent proposées pour résoudre des problèmes MO. Entre autres, les extensions de l'algorithme génétique (Holland, 1975), du système immunitaire artificiel (Farmer *et al.*, 1986) et de l'évolution différentielle (ED) (Storn et Price, 1997) ont inspiré respectivement les algorithmes Non-Dominated Sorting Genetic Algorithm-II (NSGAI) (Deb *et al.*,

2002), Multiobjective Immune System Algorithm (MISA) (Coello Coello et Cortés, 2005) et Differential Evolution for Multiobjective Optimization (DEMO) (Robic et Filipic, 2005).

Les problèmes MO avec variables continues sont présents en industrie. Il arrive que les ingénieurs doivent optimiser en même temps plusieurs objectifs contradictoires. En génie électrique, un des problèmes consiste à déterminer le meilleur flux électrique dans un système tout en voulant diminuer les risques. Ce problème peut être traité par une métaheuristique (Nikoofard *et al.*, 2012). D'autres problèmes dans le domaine de la bio-informatique, de la robotique, de l'ingénierie chimique, de la reconnaissance d'imagerie numérique sont aussi traités par des métaheuristiques (Das et Suganthan, 2011). Les problèmes avec variables continues demandent une exploration particulière des solutions, puisque l'algorithme doit pouvoir approcher le plus précisément possible les valeurs réelles optimales. L'algorithme ED, qui est reconnu pour avoir une bonne capacité de recherche globale (Lin *et al.*, 2015), propose de considérer les solutions comme des vecteurs et d'utiliser le déplacement vectoriel pour explorer l'espace de solutions d'un problème. L'algorithme de la mauvaise herbe (IWO) (Mehrabian et Lucas, 2006) s'inspire de la prolifération des graines dans un environnement. L'analogie avec la nature permet de traduire la *r-sélection* et la *k-sélection* afin de faire de l'exploration en début de résolution pour ensuite faire davantage d'exploitation. Pour de plus amples informations sur la variété de métaheuristiques pour ce type de problèmes, le lecteur peut se référer aux états de l'art (Boussaïd *et al.*, 2013, Fister *et al.* 2013).

Considérant le théorème du « No Free Lunch » (Wolpert et Macready, 1997), aucune stratégie d'optimisation générale ne peut être assez performante pour être globalement la meilleure pour tous les problèmes. Une idée intéressante afin de se rapprocher de cet idéal consiste à utiliser l'hybridation (Raidl, 2015). La motivation principale de l'hybridation est d'obtenir une meilleure performance en combinant les forces de plusieurs métaheuristiques afin de tirer profit de la synergie produite. L'optimisation MO ne fait pas exception à cette tendance. Par exemple, plusieurs algorithmes MO hybrides impliquant l'ED ont été proposés dans la littérature. NSDE (Iorio et Li, 2005) est hybride entre l'algorithme génétique et l'ED. Les résultats empiriques démontrent que cette hybridation permet à NSDE d'obtenir de meilleurs résultats que NSGAI sur des problèmes de rotation (Iorio et Li, 2005). ADE-MOIA (Lin *et al.*, 2015) est un hybride entre un système immunitaire artificiel et l'ED offrant également une bonne performance. Pour sa part, GISMOO utilise l'ED afin de pouvoir traiter des variables continues plutôt que des variables discrètes et obtient de bons résultats en comparaison avec d'autres algorithmes de la littérature (Zinflou *et al.*, 2012). En ce qui concerne l'hybridation entre IWO et ED, il semble exister un seul algorithme pour l'optimisation de problèmes uni-objectif (Basak *et al.*, 2013).

La contribution principale de ce papier consiste à proposer un nouvel algorithme performant hybride impliquant l'ED pour résoudre des problèmes MO continus. Cette hybridation est réalisée avec l'algorithme IWO afin d'équilibrer les capacités d'exploration et d'exploitation des solutions. Ce nouvel algorithme, IWODEMO, doit ainsi viser les deux buts des métaheuristiques MO, soit de converger vers le front Pareto-optimal et de contenir des solutions diversifiées. Au meilleur de notre connaissance, les algorithmes IWO multi-objectifs ne proposent actuellement pas de comparaison avec des méthodes ED (Kundu *et al.*, 2011, Nikoofard *et al.*, 2012). Un autre volet de la contribution consiste donc à réaliser une étude comparative dans des conditions expérimentales équitables. Pour ce faire, l'algorithme proposé est comparé à trois algorithmes de la littérature (NSDE, DEMO et ADE-MOIA) pour la résolution du même groupe d'instances, pour une puissance de calculs équivalente et avec le maximum de composantes communes. Dans la suite de ce papier, la Section 2 présente les principaux algorithmes MO hybrides avec l'ED. La Section 3 décrit IWODEMO, un nouvel algorithme hybride entre l'ED et l'IWO pour résoudre des problèmes MO. Le détail des expérimentations numériques réalisées est présenté à la Section 4 et l'analyse des résultats à la Section 5. La conclusion et directions futures de recherche font l'objet de la Section 6.

2 ALGORITHMES ED POUR DES PROBLEMES MO CONTINUS

Suite à la présentation sommaire de l'algorithme ED, cette section présente trois algorithmes de la littérature faisant intervenir l'ED pour l'optimisation MO : NSDE, DEMO et ADE-MOIA.

2.1 Notions de base concernant l'algorithme ED

À l'origine, l'ED (Storn et Price, 1997) a été proposé pour l'optimisation d'un seul objectif et est reconnu pour avoir une bonne capacité de recherche globale (Lin *et al.*, 2015). L'ED est résumé par l'algorithme de la Figure 1.

```

Générer aléatoirement  $POP_0$ 
Tant que le critère d'arrêt n'est pas atteint
  Pour chaque solution  $\vec{X}_i$  ( $i=1, \dots, N$ ) de  $POP_g$ 
    MUTATION - Création d'un mutant appelé vecteur donneur  $\vec{V}_i$ 
      Sélectionner trois solutions aléatoires  $\vec{X}_{r_1}$ ,  $\vec{X}_{r_2}$  et  $\vec{X}_{r_3}$ 
      Calculer la nouvelle solution à l'aide de perturbations :
        
$$\vec{V}_i = \vec{X}_{r_1} + F \cdot (\vec{X}_{r_2} - \vec{X}_{r_3}) \quad (1)$$

    CROISEMENT - Création d'un descendant appelé vecteur d'essai  $\vec{U}_i$  :
        
$$u_{ij} = \begin{cases} v_{ij} & \text{si } (rand[0,1] \leq CR) \text{ ou } (j = j_{rand}) \text{ avec } j=1, \dots, D \\ x_{ij} & \text{sinon} \end{cases} \quad (2)$$

    Évaluer le vecteur d'essai  $\vec{U}_i$ 
    SÉLECTION : Conserver la meilleure solution entre  $\vec{U}_i$  et  $\vec{X}_i$  pour la
    prochaine génération  $POP_{g+1}$ 
  Retourner la meilleure solution trouvée
  
```

Figure 1: Pseudo-code de l'algorithme ED avec la mutation ED/rand/1/bin

Cet algorithme évolutionnaire débute par la génération aléatoire d'une population initiale POP_0 de taille N selon les bornes de chaque variable. Une solution $\vec{X}_i = [x_{i,1}, \dots, x_{i,j}, \dots, x_{i,D}]$, souvent appelée vecteur, est représentée par D variables de décision. Ces variables sont des nombres réels bornés entre $x_{i,j_{min}}$ et $x_{i,j_{max}}$. Après l'initialisation, l'algorithme effectue une séquence de mutations, de croisements et de sélections sur les solutions. La phase de mutation se concrétise par la création d'un mutant \vec{V}_i , aussi appelé vecteur donneur, à partir d'un vecteur cible \vec{X}_i . L'Équation (1) à la Figure 1 présente la mutation ED/rand/1/bin qui consiste à ajouter une différence pondérée de trois autres vecteurs, \vec{X}_{r_1} , \vec{X}_{r_2} et \vec{X}_{r_3} , pris aléatoirement dans la population aux composantes du vecteur cible \vec{X}_i . Un facteur de pondération F possédant une valeur réelle conseillée entre $[0.4, 1]$ est utilisé. Plusieurs autres formes de mutations peuvent être appliquées. Toutefois, ED/rand/1/bin est reconnu pour converger vers le front Pareto-optimal tout en gardant une bonne diversité de solutions (Das et Suganthan, 2011). À ce moment, il y a deux vecteurs considérés : le vecteur cible \vec{X}_i et le vecteur donneur \vec{V}_i . Une phase de croisement binaire fait le lien entre ces deux vecteurs. Elle produit un nouveau vecteur, appelé vecteur d'essai \vec{U}_i . Pour chaque variable $j=1, \dots, D$, le vecteur d'essai prend soit la variable du vecteur cible, soit celle du vecteur donneur. Pour ce faire, la probabilité de croisement CR est utilisée. Ce paramètre est déterminé au départ de l'algorithme et il se situe entre $[0, 1]$. Le croisement du vecteur d'essai est réalisé selon l'Équation (2) à la Figure 1, où j_{rand} est un entier tiré aléatoirement dans l'intervalle $[1, D]$ afin d'assurer le maintien d'au moins un gène du vecteur donneur. Enfin, la phase de sélection permet de constituer la prochaine population POP_{g+1} . La sélection se fait par compétition entre la solution cible \vec{X}_i et son descendant \vec{U}_i , le meilleur étant conservé. Ce processus mène à la création d'une nouvelle population de taille identique pour la prochaine itération.

2.2 NSDE (Iorio et Li, 2005)

NSDE correspond à l’algorithme NSGAI (Deb *et al.*, 2002) avec la mutation ED/rand/1/bin comme méthode de croisement. La Figure 2 présente les principales composantes de l’algorithme NSDE. L’algorithme débute par la création d’une population aléatoire de taille N qui est triée en front. À chaque itération, les concepts classiques de l’algorithme génétique sont appliqués: sélection, croisement, mutation et remplacement. L’hybridation avec l’ED intervient lors de la sélection des parents et de leur croisement. Lors de la sélection, un seul parent est sélectionné par tournoi binaire pour être considéré comme vecteur cible \vec{X}_i . Il produit un vecteur d’essai \vec{U}_i selon une mutation ED/rand/1/bin ajoutant deux perturbations à \vec{X}_i . Ensuite, le vecteur d’essai est muté à l’aide d’une mutation polynomiale (Deb et Agrawal, 1994). De cette manière, N nouvelles solutions sont créées et placées dans la population enfant. Les deux populations, courante et enfant, sont combinées et triées en front pour le remplacement. Dans un même front, les solutions sont triées selon le facteur d’isolement. Les N meilleures solutions sont conservées pour la prochaine itération.

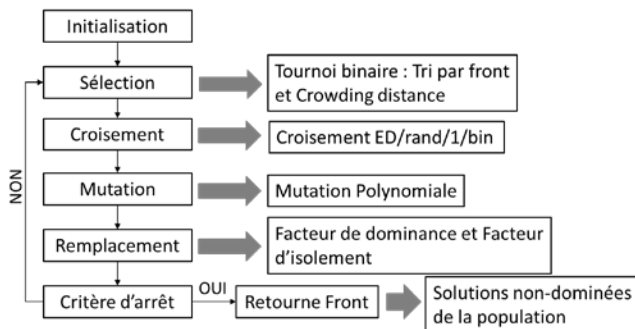


Figure 2: Principes généraux de l’algorithme NSDE

2.3 DEMO (Robic et Filipic, 2005)

DEMO est un algorithme purement ED qui introduit des stratégies afin de résoudre des problèmes MO. Cet algorithme évolutionnaire modifie une population de vecteurs sur plusieurs itérations. La Figure 3 présente les principaux points définissant DEMO. La sélection consiste à utiliser les vecteurs cibles \vec{X}_i de la population de taille N afin de produire un vecteur d’essai \vec{U}_i pour chacun à l’aide d’un croisement de type ED/rand/1/bin. Au niveau du remplacement, DEMO utilise les notions de dominance au sens Pareto pour déterminer quelle solution est conservée dans la population entre le vecteur cible, le vecteur d’essai ou les deux. Par la suite, une réduction de la population est alors possible afin de conserver une population de taille N . Pour ce faire, la population est triée en front et la *crowding distance* est calculée pour chaque solution de la population. Les N meilleures solutions sont alors conservées.

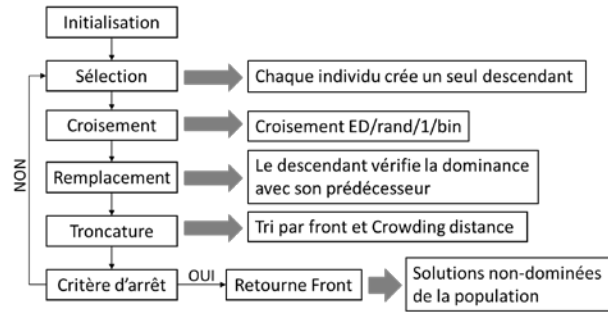


Figure 3: Principes généraux de l’algorithme DEMO

2.4 ADE-MOIA (Lin *et al.*, 2015)

ADE-MOIA propose une hybridation entre une version adaptative de l’ED et un système immunitaire artificiel. Un résumé de l’algorithme est proposé à la Figure 4.

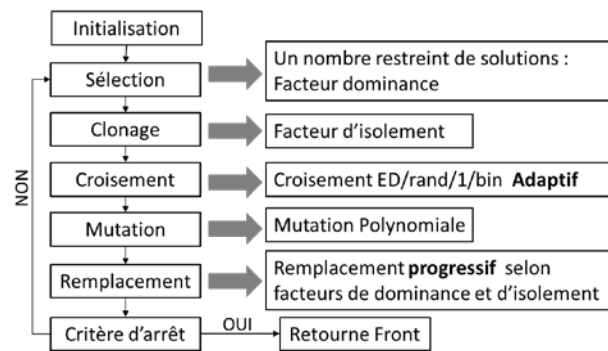


Figure 4: Principes généraux de l’algorithme ADE-MOIA

Au départ, la population initiale est générée de manière aléatoire et contient N solutions. Ces solutions sont ensuite divisées en deux groupes : les solutions non-dominées sont placées dans l’ensemble *EXA* et les solutions dominées dans l’ensemble *DA*. À chaque itération, un nombre restreint NC (inférieur à N) de solutions non-dominées de l’ensemble *EXA* sont sélectionnées pour produire des descendants. Ces NC solutions produisent un total N clones. Le nombre de clones produits à partir d’une solution est proportionnel à son facteur d’isolement dans l’ensemble *EXA*. Pour générer de nouvelles solutions, les clones sont d’abord modifiés grâce au croisement ED/rand/1/bin pour lequel les paramètres CR et F sont rendus adaptatifs au fil des itérations. De plus, l’utilisation des deux ensembles *EXA* et *DA* apporte une modification à la génération du vecteur d’essai \vec{U}_i par rapport à la version originale de ED/rand/1/bin. La différence se situe au niveau de la sélection des trois vecteurs aléatoires \vec{X}_{r1} , \vec{X}_{r2} et \vec{X}_{r3} . Les deux premiers vecteurs sont sélectionnés dans l’ensemble *EXA* et le troisième vecteur est sélectionné dans l’ensemble *DA*. Cette sélection particulière a pour but d’ajouter de la diversité dans les solutions proposées (Lin *et al.*, 2015). Suite à la modification d’un clone par le croisement, il est ensuite muté à l’aide d’une mutation polynomiale. Le clone reçoit alors une perturbation sur certaines de ses variables. Lorsque la population enfant

contient N nouvelles solutions, celle-ci est combinée avec l'ensemble EXA . Les solutions identiques sont supprimées. Ensuite, les solutions sont distribuées dans les ensembles EXA et DA . Si la taille de EXA excède N , les solutions sont triées d'après la *crowding distance*. La solution la moins isolée est éliminée. Le processus d'élimination est répété jusqu'à ce qu'il reste seulement N solutions non-dominées dans EXA . Cette stratégie de *remplacement progressif* permet de conserver une bonne diversité de solutions dans la population. Elle représente toutefois un calcul coûteux.

3 PROPOSITION D'UNE HYBRIDATION ENTRE L'ED ET L'ALGORITHME DE LA MAUVAISE HERBE : IWO DEMO

Cette section présente un nouvel algorithme pour résoudre des problèmes MO continus : IWO DEMO. Ce nouvel algorithme est une hybridation entre l'ED et l'algorithme de la mauvaise herbe qui est brièvement présenté.

3.1 Analogie de l'algorithme IWO

L'algorithme de la mauvaise herbe (IWO – *Invasive Weed Optimization*) fut introduit en 2006 (Mehrabian et Lucas, 2006). Les auteurs proposent un algorithme évolutionnaire qui s'inspire du concept de prolifération de la mauvaise herbe dans l'environnement. Il existe une compétition entre les colonies de plantes dans la nature. Certains types de plantes prolifèrent beaucoup de graines sur de longues distances pour occuper des terrains qui sont imprévisibles, ces plantes sont classées dans la *r-sélection*. D'autres plantes prolifèrent leurs graines à courte distance pour exploiter leur environnement et renforcer leur colonie. Ils font partie de la *k-sélection*. Dans IWO, les solutions sont appelées des plantes et les descendants des graines. La population est modifiée lorsque les plantes \vec{X}_i prolifèrent des graines \vec{W}_i dans l'espace des solutions en utilisant les concepts de *r-sélection* et *k-sélection*. Au début de l'algorithme, les plantes prolifèrent les graines sur de longues distances pour établir des colonies diversifiées. Plus les itérations avancent, plus les plantes déposent leurs graines près d'elles. Ainsi, une graine est un clone de la plante mère où chaque variable de décision reçoit une modification plus ou moins importante en rapport avec une déviation standard σ_g . La valeur de σ_g diminue au fil des itérations afin de passer de l'exploration à l'exploitation. IWO utilise un autre concept de la nature pour assurer une bonne convergence. Plus une plante est forte, plus elle prolifère de graines. À chaque itération, une plante produit un certain nombre de graines proportionnel à sa valeur de la fonction objectif. De plus, il existe une compétition entre les colonies de plantes dans la nature. Certaines colonies s'éteignent, tandis que d'autres colonies continuent à prendre de l'expansion et à s'adapter à l'environnement. Dans l'algorithme, ce concept est représenté par un remplacement élitiste classique à la fin de chaque itération pour conserver seulement les meilleures solutions.

Dans la littérature, il existe déjà quelques méthodes IWO multi-objectifs, comme NSIWO (Nikoofard *et al.*, 2012). En fait, l'algorithme présente le même processus de résolution que le NSGAI, mais utilise la prolifération des graines comme méthode de croisement. Il existe aussi l'algorithme IWO_MO (Kundu *et al.*, 2011) qui conserve le plus fidèlement possible le mécanisme de résolution de la mauvaise herbe. Pour déterminer la qualité des solutions, il utilise la Fuzzy dominance au lieu de la dominance au sens Pareto. Ce calcul permet de trier la population et ainsi déterminer quelles solutions accèdent à la prochaine itération. Le rang d'une solution influence aussi la quantité de graines à proliférer.

3.2 IWO DEMO

L'algorithme IWO DEMO est une hybridation entre les algorithmes ED et IWO. La Figure 4 présente le pseudo-code de l'algorithme. La population initiale contient N solutions aléatoires. Sur plusieurs itérations, les solutions de la population courante produisent des descendants. Dans IWO DEMO, un nombre exact de N descendants sont produits lors d'une itération. Comme dans IWO, une solution produit un certain nombre de descendants en rapport avec son rang dans la population triée. Deux facteurs sont utilisés à cette fin. Un tri est d'abord effectué selon un facteur de dominance correspondant à un tri par front classique comme dans NSGAI (Deb *et al.*, 2002). Ensuite, les solutions d'un même front sont triées d'après un facteur d'isolement correspondant à la *crowding distance*. Une fois la population triée, la solution qui occupe le premier rang (la plus isolée) produit les premiers descendants. Puis, les prochaines solutions de la population produisent à leur tour des descendants jusqu'à ce que N descendants soient produits. À chaque itération g , une solution \vec{X}_i produit un certain nombre de descendants $ND_{i,g}$ proportionnel à sa valeur de la fonction objectif ($f(\vec{X}_i)$) en lien avec la meilleure ($MaxFit_g$) et la pire solution ($MinFit_g$) dans la population actuelle comme le présente l'Équation (3) à la Figure 4. Des paramètres de l'algorithme servent également à limiter le nombre maximum ($MaxG$) et le nombre minimum ($MinG$) de descendants pouvant être produits par une solution.

Deux types de descendants peuvent être produits à partir d'une solution \vec{X}_i : les vecteurs d'essai \vec{U}_i et les graines \vec{W}_i . Dans le premier cas, un vecteur cible \vec{X}_i produit un vecteur d'essai \vec{U}_i selon la mutation adaptative ED/rand/1/bin empruntée à l'algorithme ED. Il est à noter que le paramètre CR s'adapte au fil des itérations. Le principe vient de la *r-sélection* et *k-sélection* de IWO. Au départ, CR est assez élevé ce qui implique que le vecteur d'essai reçoit davantage de variables du vecteur donneur. Plus les itérations passent, plus CR diminue. Ce principe, utilisé dans IWO et dans ADE-MOIA, permet de faire de l'exploration en début de résolution pour ensuite faire davantage d'exploitation. Le paramètre CR passe de 0.8 à 0.1 durant la résolution. Les deux vecteurs considérés, \vec{X}_i et \vec{U}_i , passent ensuite à la sélection. Celle-ci consiste à comparer le vecteur

d'essai au vecteur cible. S'il domine le vecteur cible, il le remplace directement dans la population. Si aucune dominance n'est identifiée entre les deux solutions, le vecteur d'essai est ajouté à la population comme dans DEMO (Robic et Filipic, 2005). Si le vecteur d'essai est dominé, il est supprimé.

```

Générer aléatoirement  $POP_g$  de taille  $N$  et évaluer les solutions  $\vec{X}_i$ 
TRI de  $POP_g$  : selon les facteurs de dominance et d'isolement
Tant que le critère d'arrêt n'est pas atteint
  Tant que  $NbDescendants < N$ 
    Prendre dans l'ordre le prochain vecteur  $\vec{X}_i$  de  $POP_g$ 
    Calculer le  $ND_{i,g}$  pour  $\vec{X}_i$  :
    
$$ND_{i,g} = \left[ \left( \frac{f(\vec{X}_i) - MinFit_g}{MaxFit_g - MinFit_g} \cdot (MaxG - MinG) \right) \right] + MinG \quad (3)$$

    Pour chacun des  $ND_{i,g}$  descendants de  $\vec{X}_i$ 
      CROISEMENT - Création d'un vecteur d'essai  $\vec{U}_i$  selon la
      mutation adaptative ED/rand/1/bin
      SÉLECTION - 3 cas possibles
      a) Si  $\vec{U}_i$  domine  $\vec{X}_i$ , il le remplace
      b) Si  $\vec{U}_i$  est dominé par  $\vec{X}_i$ , supprimer  $\vec{U}_i$  et passer à la
      PROLIFÉRATION
      c) Si aucune relation de dominance, ajouter  $\vec{U}_i$  à  $POP_g$ 
      PROLIFÉRATION - (Seulement si  $\vec{U}_i$  est dominé)
      Planter une graine  $\vec{W}_i$  selon la prolifération IWO :
      
$$\vec{w}_{i,j} = \vec{x}_{i,j} + N(0, \sigma_g^2) \text{ pour } j = 1, \dots, D \quad (4)$$

      Faire la SÉLECTION avec  $\vec{W}_i$  (au lieu de  $\vec{U}_i$ )
      TRI de  $POP_g$  : selon les facteurs de dominance et d'isolement
      Suppression des solutions identiques dans  $POP_g$ 
      RÉDUCTION progressive de la  $POP_g$  à  $N$  solutions
  Retourner les solutions non-dominées
  
```

Figure 5: Pseudo-code de l'algorithme IWODEMO

Dans le deuxième cas, IWODEMO produit une graine \vec{W}_i à partir de la plante \vec{X}_i selon l'Équation (4) à la Figure 4 où j correspond aux variables de décision et $N(0, \sigma_g^2)$ correspond à une valeur aléatoire tirée selon une distribution normale de moyenne zéro et de variance égale à la déviation standard au carré. Cette création d'un descendant correspond au principe de prolifération d'une plante empruntée à l'algorithme IWO à la différence qu'une probabilité de mutation s'ajoute à l'équation : seulement les variables de décision qui passent le critère de mutation reçoivent une perturbation. Les deux vecteurs considérés, \vec{X}_i et \vec{W}_i , passent ensuite à la sélection. Par contre, si la graine est dominée par le vecteur cible, elle est supprimée et la résolution se poursuit sans repasser par la prolifération.

La solution \vec{X}_i continue la génération de ses $ND_{i,g}$ descendants. Il se peut toutefois que la solution ait été remplacée par un descendant. Dans ce cas, le descendant produit lui-même d'autres descendants. Au total, un maximum de N descendants est produit lors d'une itération, incluant les vecteurs d'essai et les graines. L'algorithme ADE-MOIA (Lin *et al.*, 2015) produit aussi plusieurs descendants à partir d'une solution, par contre ceux-ci sont placés dans une population enfant. IWODEMO permet le remplacement immédiat d'une solution ce qui encourage une convergence plus rapide. Suite à la création des descendants, les solutions identiques sont supprimées dans POP_g . La population est à nouveau triée et une réduction progressive assure le maintien des N meilleures.

4 EXPERIMENTATIONS NUMERIQUES

Cette section présente une étude comparative de l'algorithme proposé, IWODEMO, avec les principaux algorithmes hybrides faisant intervenir l'ED pour la résolution de problèmes MO. Plus spécifiquement, ces algorithmes de la littérature sont NSDE, DEMO et ADE-MOIA. Cette section présente les instances, les métriques et les conditions expérimentales utilisées afin de réaliser les expérimentations numériques.

4.1 Instances ZDT

Pour réaliser l'expérimentation numérique, les instances-test ZDT introduites par Zitzler *et al.* (2000) sont privilégiées. Ce groupe d'instances est peut-être l'une des références en MO les plus largement utilisées dans la littérature. Le Tableau 1 présente le détail de ces instances avec respectivement le nom de l'instance, les bornes et le nombre de variables de décision, les fonctions objectif et la nature du front de Pareto. Il est possible de remarquer que les instances ZDT contiennent deux fonctions objectif. Les solutions réalisables sont en encodage réelles et possèdent 30 (ZDT1, ZDT2, ZDT3) ou 10 (ZDT4, ZDT6) variables de décision. Les fronts Pareto-optimaux contiennent un ensemble infini borné de solutions. Dans l'espace de décision, les solutions Pareto-optimales de toutes les instances possèdent les caractéristiques suivantes : $0 \leq x_1 \leq 1$ et $x_j = 0$ pour $j=2, \dots, D$ où D est le nombre de variables de décision. En observant le Tableau 1, il est possible de remarquer que les instances ZDT offrent une diversité au niveau de l'espace de solutions. En effet, les solutions Pareto-optimales présentent des formes particulières entraînant ainsi la complexité de résolution de ce groupe d'instances.

4.2 Reproduction des algorithmes de la littérature

Parmi les principaux algorithmes hybrides faisant intervenir l'ED pour résoudre efficacement des problèmes MO, on retrouve NSDE (Iorio et Li, 2005), DEMO (Robic et Filipic, 2005) ou ADE-MOIA (Lin *et al.*, 2015). La comparaison de la performance avec ces différents algorithmes n'est toutefois pas simple à réaliser. En effet, les articles originaux présentant ces algorithmes emploient généralement une implémentation informatique diversifiée et une évaluation de la performance qui peut être difficilement comparable.

Afin de proposer une étude comparative fiable et équitable, les algorithmes de la littérature NSDE, DEMO et ADE-MOIA ont été reproduits et testés dans les mêmes conditions expérimentales. Pour ce faire, une reproduction respectant les indications de l'article original a d'abord été réalisée pour assurer l'exactitude de chaque algorithme. Le calcul de la métrique d'origine a permis de déterminer que la reproduction est fidèle à l'algorithme original. Le lecteur peut se référer à (Fournier, 2016) pour les détails à ce sujet. Cela permet ensuite la réalisation de l'étude comparative proposée dans ce papier.

ZDT	Var.	Fonctions objectif	Nature front Pareto
1	$x \in [0,1]^{30}$	$f_1(x) = x_1$ $f_2(x) = g(x) \left(1 - \sqrt{x_1/g(x)}\right)$ $g(x) = 1 + \frac{9}{J-1} \sum_{j=2}^J x_j$	Front Pareto convexe
2	$x \in [0,1]^{30}$	$f_1(x) = x_1$ $f_2(x) = g(x) \left(1 - (x_1/g(x))^2\right)$ $g(x) = 1 + \frac{9}{J-1} \sum_{j=2}^J x_j$	Front Pareto non-convexe
3	$x \in [0,1]^{30}$	$f_1(x) = x_1$ $f_2(x) = g(x) \left(1 - \sqrt{x_1/g(x)} - \frac{x_1}{g(x)} \sin(10\pi x_1)\right)$ $g(x) = 1 + \frac{9}{J-1} \sum_{j=2}^J x_j$	Front Pareto discontinu
4	$x \in [0,1] \times [-5,5]^{10}$	$f_1(x) = x_1$ $f_2(x) = g(x) \left(1 - \sqrt{x_1/g(x)}\right)$ $g(x) = 1 + 10(J-1) + \sum_{j=2}^J (x_j^2 - 10 \cos(4\pi x_j))$	Plusieurs fronts Pareto locaux
6	$x \in [0,1]^{10}$	$f_1(x) = 1 - \exp^{-4x_1} \sin(6\pi x_1)^6$ $f_2(x) = g(x) \left(1 - (f_1(x)/g(x))^2\right)$ $g(x) = 1 + \frac{9}{J-1} \sum_{j=2}^J x_j$	Densité faible de solutions près du front Pareto.

Tableau 1 : Instances-test avec variables réelles sans contrainte (ZDT) utilisées dans cette étude

4.3 Métriques

Un survol des travaux dans la littérature démontre que l'évaluation de la performance d'un algorithme MO est souvent une tâche ardue. D'un côté les métriques utilisées par les auteurs varient considérablement. D'un autre côté, la recherche de solutions pour un algorithme MO est double en visant à la fois la qualité des solutions ainsi que leur diversité. En ce sens, l'utilisation d'une seule métrique permet difficilement d'évaluer simultanément ces deux aspects. Dans ce papier, l'évaluation de la performance des algorithmes est donc évaluée à l'aide de deux métriques : la convergence γ et la diversité Δ . Le front Pareto-optimal de chaque instance doit être connu pour le calcul de ces métriques. Les fronts Pareto-optimaux composés de 500 solutions uniformément espacées mise en ligne par S. Huband sont utilisés (<http://www.scis.ecu.edu.au/research/wfg/datafiles.html>).

La métrique de convergence γ offre une évaluation de la distance entre l'ensemble obtenu des solutions non-dominées P et les solutions du front Pareto-optimal P^* . Le calcul de cette métrique γ est réalisé selon l'Équation (5) où d_i est la distance euclidienne dans l'espace objectif entre la solution $i \in P$ et le membre le plus proche dans P^* . Plus la valeur de cette métrique est petite, meilleure est la convergence vers le front de Pareto-optimal. Lorsque toutes les solutions obtenues reposent exactement sur les solutions de P^* , cette métrique prend une valeur de zéro.

$$\gamma = \frac{\sum_{i=1}^{|P|} d_i}{|P|} \quad (5)$$

La deuxième mesure de performance utilisée est la métrique de diversité Δ . Celle-ci évalue l'étendue de la propagation réalisée parmi les solutions non dominées. Le calcul de cette métrique est réalisé selon l'Équation (6) où d_i est la distance euclidienne dans l'espace objectif entre la solution $i \in P$ et le membre le plus proche dans P^* , et \bar{d} indique la moyenne de ces distances. Les paramètres d_f et d_l représentent respectivement la distance euclidienne entre les solutions extrêmes de P^* et la solution limite de l'ensemble obtenu P . Plus la valeur de cette métrique est petite, meilleure est la diversité de l'approximation.

$$\gamma = \frac{d_f + d_l + \sum_{i=1}^{|P|-1} |d_i - \bar{d}|}{d_f + d_l + (P-1) * \bar{d}} \quad (6)$$

4.4 Conditions expérimentales et paramètres

Pour réaliser l'étude comparative de IWODEMO avec trois algorithmes de la littérature, les conditions expérimentales sont celles des articles NSGAI, GISMOO et DEMO (Deb *et al.*, 2002, Robic et Filipic, 2005, Zinflou *et al.*, 2012). Les métriques de convergence et de diversité présentées dans les tableaux de résultats représentent les moyennes de dix résolutions. Un maximum de 25 000 évaluations de solutions est accordé aux algorithmes pour offrir une approximation d'un front Pareto-optimal. Pour IWODEMO et ADE-MOIA qui ont comme critère d'arrêt selon un nombre d'itérations, 25 000 évaluations correspondent à 250 itérations. Les algorithmes sont implémentés en C++ sous Visual Studio 2010. Les expérimentations numériques sont effectuées sur un ordinateur portable ASUS avec un processeur Intel 2.4 GHz et 8G de mémoire RAM. L'ordinateur utilise Windows comme système d'exploitation. Les paramètres des algorithmes sont généralement ceux proposés par les articles originaux. Ces paramètres sont détaillés au Tableau 2.

NSDE:

- N : 100, Nombre d'évaluations max: 25000
- Probabilité de mutation : $1/D$
- CR : 0.3 (Sauf pour ZDT4 et ZDT6 où $CR = 0.1$)
- F : 0.5

DEMO:

- N : 100, Nombre d'évaluations max: 25000
- CR : 0.3 (Sauf pour ZDT4 et ZDT6 où $CR = 0.1$)
- F : 0.5

ADE-MOIA:

- N : 100, Nombre d'itérations: 250
- Probabilité de mutation : $1/D$
- CR : Adaptif [0.1, 0.9], F : Adaptif [0.1, 0.9]

IWODEMO:

- N : 100, Nombre d'itérations: 250
- Probabilité de mutation IWO : 0.3
- CR : Adaptif [0.1, 0.8], F : 0.5
- $MaxG$: 2, $MinG$: 1
- Déviation initiale (σ_{ini}) : $(MaxFit_0 - MinFit_0) \cdot 0.2$
- Déviation finale : $\sigma_{ini} \cdot 0.01$

Tableau 2 : Paramètres des algorithmes NSDE, DEMO, ADE-MOIA et IWODEMO

5 RESULTATS ET ANALYSE

Le Tableau 3 présente les moyennes de la métrique de convergence γ calculées à partir des fronts Pareto-optimaux contenant 500 solutions tel que proposé par les auteurs de DEMO (Robic et Filipic, 2005). Dans les tableaux suivants, les valeurs ombragées des métriques correspondent aux meilleurs résultats pour une instance.

IWODEMO obtient la meilleure valeur moyenne de convergence pour toutes les instances à l'exception de ZDT4 pour laquelle l'algorithme DEMO présente une qualité de solution supérieure. Une égalité entre DEMO et IWODEMO est également observée pour l'instance ZDT6. Il est possible de remarquer que tous les algorithmes offrent des approximations qui convergent bien vers les fronts Pareto-optimaux.

	NSDE	DEMO	ADE-MOIA	IWODEMO
ZDT1	1,25E-03	1,06E-03	1,10E-03	1,02E-03
ZDT2	8,84E-04	7,86E-04	8,15E-04	7,56E-04
ZDT3	1,25E-03	1,23E-03	1,25E-03	1,20E-03
ZDT4	1,15E-03	9,76E-04	1,64E-03	1,04E-03
ZDT6	6,22E-04	6,02E-04	5,54E-03	6,02E-04

Tableau 3 : Métrique convergence γ pour les instances ZDT (fronts Pareto-optimaux de 500 solutions)

Afin d'étendre cette analyse de la qualité des solutions, la précision de la métrique de convergence peut être augmentée en considérant des fronts Pareto-optimaux à plus de 500 solutions. La définition mathématique des instances à l'étude permet ce calcul. Le Tableau 4 présente de nouveau les moyennes de la métrique de convergence γ pour les instances ZDT avec des fronts Pareto-optimaux de 25 000 solutions. IWODEMO obtient cette fois la meilleure métrique de convergence pour toutes les instances ZDT. L'instance ZDT4 offre maintenant l'avantage à IWODEMO plutôt que DEMO. Il est également intéressant de souligner que tous les algorithmes présentent une métrique de convergence plus près de zéro.

	NSDE	DEMO	ADE-MOIA	IWODEMO
ZDT1	2,28E-04	1,23E-05	1,09E-04	4,80E-06
ZDT2	1,50E-04	1,45E-05	6,33E-05	3,91E-06
ZDT3	9,05E-05	2,71E-05	1,40E-04	2,66E-05
ZDT4	6,72E-04	1,39E-05	1,01E-03	4,89E-06
ZDT6	2,40E-05	2,39E-05	4,79E-03	2,31E-05

Tableau 4 : Métrique convergence γ pour les instances ZDT (fronts Pareto-optimaux de 25000 solutions)

Le Tableau 5 présente, pour chaque algorithme, la moyenne des dix métriques de diversité Δ calculées pour chaque instance ZDT. La taille des fronts Pareto-optimaux n'a pas d'influence sur le calcul de la diversité. Seulement les deux solutions Pareto-optimales extrêmes sont nécessaires. ADE-MOIA obtient la meilleure diversité en ZDT1 et ZDT2, tandis que IWODEMO obtient la meilleure diversité en ZDT3, ZDT4 et ZDT6. IWODEMO et ADE-MOIA sont les deux seuls algorithmes qui utilisent un remplacement élitiste progressif. C'est-à-dire que pour

obtenir la prochaine population, ils éliminent les pires solutions une par une, en calculant à chaque fois le facteur d'isolement. Le processus est plus coûteux, mais les résultats confirment que ce principe permet de diversifier davantage les solutions.

	NSDE	DEMO	ADE-MOIA	IWODEMO
ZDT1	0,32	0,36	0,11	0,12
ZDT2	0,33	0,36	0,10	0,12
ZDT3	0,34	0,37	0,19	0,17
ZDT4	0,30	0,33	0,20	0,13
ZDT6	1,31	0,34	0,32	0,12

Tableau 5 : Métrique diversité Δ pour les instances ZDT

Afin de permettre une comparaison plus complète, les temps d'exécution pour résoudre les instances ZDT sont comptabilisés. Le Tableau 6 présente la moyenne des temps d'exécution en secondes pour résoudre les instances ZDT. Ces temps montrent que les algorithmes prennent tous en moyenne moins d'une seconde. DEMO et IWODEMO sont légèrement plus rapides que NSDE et ADE-MOIA pour la résolution des instances ZDT.

	NSDE	DEMO	ADE-MOIA	IWODEMO
ZDT1	0,4	0,2	0,4	0,2
ZDT2	0,4	0,2	0,4	0,2
ZDT3	0,5	0,2	0,4	0,3
ZDT4	0,6	0,2	0,7	0,2
ZDT6	0,5	0,2	0,7	0,3

Tableau 6 : Temps moyen d'exécution (en secondes)

6 CONCLUSION

Dans cet article, nous avons proposé IWODEMO, un nouvel algorithme MO Pareto, hybridant des concepts issus de l'ED avec des propriétés inspirées de IWO. Les performances de IWODEMO ont été analysées à l'aide d'instances MO bien connues dans la littérature (ZDT). Les résultats obtenus montrent que l'approche combine efficacement les forces des deux méthodes ayant fait l'objet de l'hybridation. En effet, IWODEMO obtient en général de meilleurs résultats que les autres algorithmes de la littérature utilisés en comparaison en termes de convergence, de dispersion des solutions sur le front Pareto-optimal et de temps de calculs. Des expérimentations supplémentaires ont été réalisées mais n'ont pu être présentées par manque d'espace. Ces résultats supplémentaires sur un autre groupe d'instances-test (DTLZ) confirment la performance obtenue.

Une conclusion naturelle de ces résultats expérimentaux est que IWODEMO représente une alternative efficace pour résoudre des problèmes à objectifs multiples. En particulier, ces résultats mettent en évidence l'importance des mécanismes de diversification et d'intensification de la recherche de solutions sur la performance globale d'un algorithme MO. La phase de prolifération empruntée à IWO et les paramètres adaptatifs de IWODEMO semblent être un schéma d'hybridation complémentaire à l'ED qui est

reconnu pour avoir une bonne capacité de recherche globale. Il est également important de noter que IWO DEMO est un algorithme générique Pareto, donc il peut être étendu et adapté pour résoudre d'autres problèmes MO. Dans la suite de la recherche, il est donc prévu de tester IWO DEMO en augmentant le nombre de fonctions objectif. De plus, il serait intéressant de vérifier l'efficacité de l'algorithme sur des problèmes de l'industrie.

REMERCIEMENTS

Le soutien financier du CRSNG (CANADA) a rendu possible la réalisation de cette recherche.

REFERENCES

- Basak, A., D. Maity and S. Das (2013). "A differential invasive weed optimization algorithm for improved global numerical optimization." *Applied Mathematics and Computation* **219**(12): p. 6645-6668.
- Boussaïd, I., J. Lepagnot and P. Siarry (2013). "A survey on optimization metaheuristics." *Information Sciences* **237**: p. 82-117.
- Coello Coello, A. C. and N. C. Cortés (2005). "Solving Multiobjective Optimization Problems Using an Artificial Immune System." *Genetic Programming and Evolvable Machines* **6**(2): p. 163-190.
- Das, S. and P. N. Suganthan (2011). "Differential Evolution: A Survey of the State-of-the-Art." *IEEE Transactions on Evolutionary Computation* **15**(1): p. 4-31.
- Deb, K. and R. B. Agrawal (1994). "Simulated binary crossover for continuous search space." *Complex Systems* **9**(3): p. 1-15.
- Deb, K., A. Pratap, S. Agarwal and T. Meyarivan (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE Transactions on Evolutionary Computation* **6**(2): p. 182-197.
- Farmer, J. D., N. H. Packard and A. S. Perelson (1986). "The immune system, adaptation, and machine learning." *Phys. D, tome 2, no 1-3, ISSN 0167-2789*: p. 187-204.
- Fister, J. I., X. S. Yang, I. Fister, J. Brest and D. Fister (2013). "A Brief Review of Nature-Inspired Algorithms for Optimization." *arXiv preprint arXiv:1307.4186*.
- Fournier, M. (2016). *IWO DEMO : Hybridation entre l'évolution différentielle et l'algorithme de la mauvaise herbe*, Master thesis, Université du Québec à Chicoutimi.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Harbor.
- Iorio, A. and X. Li (2005). Solving Rotated Multi-objective Optimization Problems Using Differential Evolution. *AI 2004: Advances in Artificial Intelligence*. S. B. H. G. Webb and X. Yu. **3339**: p. 861-872.
- Kundu, D., K. Suresh, S. Ghosh, S. Das, B. K. Panigrahi and S. Das (2011). "Multi-objective optimization with artificial weed colonies." *Information Sciences* **181**(12): p. 2441-2454.
- Lin, Q., Q. Zhu, P. Huang, J. Chen, Z. Ming and J. Yu (2015). "A novel hybrid multi-objective immune algorithm with adaptive differential evolution." *Computers & Operations Research* **62**: p. 95-111.
- Mehrabian, A. R. and C. Lucas (2006). "A novel numerical optimization algorithm inspired from weed colonization." *Ecological Informatics* **1**(4): p. 355-366.
- Nikoofard, A. H., H. Hajimirsadeghi, A. Rahimi-Kian and C. Lucas (2012). "Multiobjective invasive weed optimization: Application to analysis of Pareto improvement models in electricity markets." *Applied Soft Computing* **12**(1): 100-122.
- Raidl, G. (2015). "Decomposition based hybrid metaheuristics." *European Journal of Operational Research* **244**: p. 66-76.
- Robic, T. and B. Filipic (2005). DEMO: Differential Evolution for Multiobjective Optimization. *EMO, Guanajuato, Mexico, Springer Lecture Notes in Computer Science*
- Storn, R. and K. Price (1997). "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces." *Journal of Global Optimization* **11**(4): p. 341-359.
- Wolpert, D. and W. Macready (1997). "No free lunch theorems for optimization." *IEEE Transactions on Evolutionary Computation* **1**(1): pp. 67-82.
- Zinflou, A., C. Gagné and M. Gravel (2012). "GISMOO: A New Hybrid Genetic/Immune Strategy for Multiple Objective Optimization." *Computers & Operations Research* **39**(9): p. 1951-1968.
- Zitzler, E., K. Deb and L. Thiele (2000). "Comparison of multiobjective evolutionary algorithms: Empirical results." *Evolutionary Computation* **8**: p. 173-195.

LE PILOTAGE DES APPROCHES INTÉGRÉES DANS LA PLANIFICATION ET LE CONTRÔLE DE LA PRODUCTION

Angela Patricia VELASCO ACOSTA, Christian MASCLE

Département de Génie Mécanique, École Polytechnique de Montréal
C.P. 6079, Succursale Centre-ville, Montréal (Québec) H3C 3A7, Canada
angela-patricia.velasco-acosta@polymtl.ca, christian.mascle@polymtl.ca

RESUME : *Les approches de planification et de contrôle de la production ont évolué dans l'industrie manufacturière afin de satisfaire les besoins changeants de l'environnement de la fabrication. Une partie des entreprises ont fait le passage des méthodes à flux poussé à des méthodes à flux tiré, d'autres ont intégré une nouvelle méthode à celle déjà adoptée dans le but d'améliorer leurs performances. C'est ainsi que les approches Lean Six Sigma et demand-driven MRP ont été conçues, mais l'adoption avec succès de celles-ci nécessite l'expertise et le jugement du planificateur. Nous avons développé un modèle pour chacune des deux approches pour évaluer leur performance en matière de niveau de stocks et de niveau de service. La simulation de la production caractérisée par des particularités de l'industrie aéronautique a été effectuée sous le logiciel ARENA. Le modèle LSS proposé est validé à partir de la comparaison des deux approches. Les résultats obtenus justifient la nécessité d'utiliser un cadre de fonctionnement reproductible dans les industries. Les cadres des approches intégrées ou hybrides proposées permettent au moins une réduction des stocks de 24% et une augmentation du niveau de service de 38%.*

MOTS-CLES : *planification et contrôle de la production, Lean Six Sigma, demand-driven MRP, simulation, modélisation.*

1 INTRODUCTION

Un système de production est la combinaison des procédures et des systèmes d'information nécessaires pour assurer le flux de matière à travers les différentes étapes de production dans une usine manufacturière (Perry, 1980).

Le progiciel de gestion intégré (PGI, en anglais *Enterprise Resource Planning ERP*) est le système le plus utilisé dans ce contexte parce qu'il s'agit d'un système d'information qui regroupe des procédures. Cependant, les caractéristiques sous lesquelles le système fonctionne ne sont pas applicables à toutes les industries, principalement parce qu'il utilise la logique de la planification des ressources d'une entreprise (PRP, en anglais *Manufacturing Requirements Planning MRPII*) pour la planification et le contrôle de la production.

Aussi, les approches créées par les japonais, telles que le juste-à-temps (JAT, en anglais *just-in-time JIT*) qui fonctionne sous la pensée *Lean* (Womack et Jones, 1996), ont pris aussi place dans l'industrie manufacturière. Bien que le *JIT* a été conçu comme une stratégie générale de fonctionnement, il permet à l'aide du Kanban ou de ses extensions, une approche de pilotage. Ces approches d'adoption facile dans l'atelier ont prouvé qu'elles permettaient des gains considérables et immédiats pour une organisation. Cependant, certaines faiblesses de ces sys-

tèmes sont apparues à l'usage. Notamment, il était demandé de réduire leurs inventaires, en même temps qu'ils devaient augmenter la quantité des commandes livrées selon les dates prévues et réduire les délais de production (Kumar et Meade, 2002).

Même si des approches qui rassemblent les meilleurs attributs de plusieurs systèmes ont été proposées depuis plus de 20 ans, elles n'ont seulement commencé à être implantées que récemment. Ceci a été possible grâce à l'apparition de la plus récente approche dans cette catégorie, nommée *demand-driven MRP*, qui semble fournir au planificateur les étapes à suivre pour leur implémentation. Cette approche montre des gains par la réduction des stocks et l'augmentation du niveau de service.

Ces approches représentent une très bonne solution, car elles adaptent les règles d'opération et des systèmes aux nouveaux scénarios sans modifier les règles parfois obsolètes et logiques déjà existantes. Mais, comment est-il possible de piloter ces systèmes hybrides?

Le but de cet article est de proposer un modèle de l'approche *Lean Six Sigma*, reconnue dans la littérature comme l'approche la plus performante dans l'industrie aéronautique (Akbulut-Bailey *et al.*, 2012) et d'évaluer sa performance en matière de stocks et de niveau de service en vue de vérifier le cadre de fonctionnement proposé. Une comparaison des paramètres de performance

est effectuée avec l'approche *DDMRP* qui commence à occuper une certaine place dans les industries.

La littérature existante nous montre les avantages obtenus lors d'une bonne implémentation des méthodes, mais par contre la manière dont le planificateur doit le faire, d'après son interprétation et l'étendue de ses connaissances.

2 L'ÉVOLUTION DE LA PLANIFICATION ET LE CONTRÔLE DE LA PRODUCTION

Au fil du temps, il a été possible de distinguer trois catégories de méthodes ou d'approches de la planification et de contrôle de la production (figure 1): les méthodes à flux poussé, les méthodes à flux tiré et les approches hybrides.

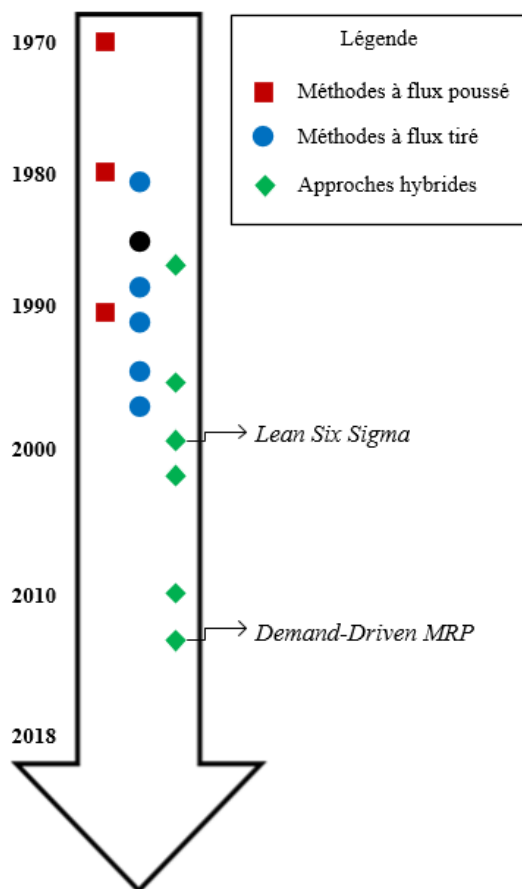


Figure 1 : Ligne de temps des approches

L'apparition des méthodes suit l'évolution de la complexité de l'environnement manufacturier dans le temps, l'évolution des capacités et des techniques informatiques et l'évolution des méthodes managériales.

2.1 Méthodes à flux poussé

En 1970, Orlicky a révolutionné la planification de la production en introduisant la planification des besoins en matière (PBM, en anglais *Material Requirements Planning MRP*). Au cours du temps, l'environnement des industries a changé, donc la méthode *MRP* a dû évoluer

aussi pour satisfaire ces exigences. C'est ainsi qu'en 1980, le *MRPII* a été développé. En 1990, le progiciel de gestion intégré *ERP* a offert l'aide d'un système d'information performant (Elragal et Haddara, 2012).

La principale caractéristique de ces méthodes dites à flux poussé énonce que les composants, les produits intermédiaires et les produits finis sont poussés en production, en réponse à une commande de fabrication, décidée sur des prévisions chiffrées, et non parce qu'il existe un besoin effectif et identifié (Benton et Shin, 1998).

À mesure que les approches de planification et de contrôle de la production ont évolué, leurs concepteurs ont adopté et intégré les derniers développements en système d'information (Rondeau et Litteral, 2001). Le besoin d'intégrer ces systèmes a pris de l'importance au-delà du simple traitement de l'information (O'Brien et Marakas, 2010).

Les systèmes d'information de fabrication développent des liens pour rendre l'information nécessaire disponible (Umble *et al.*, 2001) et favoriser l'intégration des activités internes pouvant s'étendre jusqu'à l'intégration des clients et des fournisseurs (Rondeau et Litteral, 2001). C'est ainsi que les méthodes de cette catégorie sont considérées comme des systèmes d'information et utilisées pour planifier et contrôler toutes les activités manufacturières dans l'organisation (Flapper *et al.*, 1991).

Pour intégrer les lacunes identifiées entre le progiciel de gestion et le contrôle de la fabrication, les systèmes d'exécution de la fabrication ont été développés (en anglais *Manufacturing Execution Systems MES*). Ils permettent d'assurer le pilotage et le suivi, en temps réel, des diverses étapes de la production (Liu *et al.*, 2006). Cependant, l'énorme quantité d'informations à manipuler, à centraliser et à stocker provenant de plusieurs sources incluant des informations qui ne seront pas utilisées (Powell *et al.*, 2013), pose un problème dont la solution n'est pas si évidente.

2.2 Méthodes à flux tiré

Le Kanban, issu du juste-à-temps, fait partie des principales méthodes dans cette catégorie dont les auteurs ont fortement critiqué la performance des approches à flux poussé en affirmant que le flux est déclenché et tiré par le client. C'est-à-dire que les commandes de fabrication ne sont pas passées, jusqu'au moment où il y a un besoin réel.

Ils visent la simplification des processus afin d'éliminer le gaspillage, qui représente la pensée *Lean* (García *et al.*, 2016). Cependant, ils ne sont pas performants quand il y a beaucoup de variabilité.

Plusieurs chercheurs considèrent la théorie des contraintes (en anglais *theory of constraints TOC*) comme une méthode à flux tiré. Elle peut aussi être implémentée en flux poussé.

2.3 Approches hybrides

La divergence des méthodes à flux poussé et à flux tiré a permis aux chercheurs, d'abord d'identifier et de comparer leurs avantages et leurs faiblesses, puis de penser à une possible intégration des deux points de vue, d'où la création de la catégorie hybride.

Les premières approches hybrides proposées dans la littérature ont rassemblé seulement deux méthodes (e.g. *MRP – JIT*, *MRP – TOC*, *Lean Six Sigma*). Actuellement, les approches regroupent plus de deux méthodes (e.g. *demand-driven MRP*).

2.3.1 Lean Six Sigma - LSS

Conçue en 2000, elle est définie comme une méthode d'amélioration. Néanmoins, elle rentre dans la catégorie d'approche hybride parce qu'elle couvre la combinaison du *Lean* et du *Six Sigma* (Akbulut-Bailey *et al.*, 2012).

Le *Lean* est une philosophie, mais elle regroupe des stratégies telles que le *Kanban* qui permet le pilotage de la planification et le contrôle de la production.

En implémentant les principes et les outils des deux approches (Carreira et Trudell, 2006), *LSS* permet d'accroître la production, de réduire les coûts, d'améliorer les délais de livraison, de réduire les stocks et d'améliorer la satisfaction de la clientèle en même temps (George *et al.*, 2004). De plus, le processus est amélioré en éliminant les défauts (i.e. les éléments qui ne répondent pas aux besoins des clients), la variation et en se concentrant sur le flux.

Akbulut-Bailey *et al.* (2012) ont été les premiers à démontrer le succès de l'implantation du concept de la méthode dans une entreprise de l'industrie aéronautique. Ce succès s'est affirmé par des réductions de stocks, une augmentation de la production, une amélioration de la qualité et un niveau de service élevé, d'où la pertinence de cette méthode.

Il reste beaucoup de questionnement concernant l'adoption de cette méthode. Est-ce qu'il s'agit d'implémenter les deux approches en parallèle ?

2.3.2 Demand-Driven MRP - DDMRP

La *DDMRP* est considérée comme une approche hybride puisqu'elle regroupe les aspects encore pertinents du *MRP* et de la distribution des ressources (en anglais *distribution resource planning DRP*), la visibilité et le flux tiré retrouvé dans le *Lean* et le *TOC*, et la réduction de la variation perçue dans le *Six sigma* (Ptak et Smith, 2016). Le principe fondamental de la *DDMRP* est de protéger les flux de matière et d'information qui doivent être directement en rapport avec la demande.

La seule façon de bloquer le passage ou l'amplification de la variation entre des composants dépendants du système est le découplage (Ptak et Smith, 2016). La forme de découplage utilisée, caractéristique de l'approche, est le stock tampon (en anglais *buffer*).

L'approche comprend cinq étapes : le positionnement stratégique des stocks, la détermination intelligente de

profils des buffers et de niveaux de stock, l'ajustement de buffers variables, la planification tirée par la demande et l'exécution visible et en collaboration (Ptak et Smith, 2016).

Étant une approche assez récente, la littérature existante se limite à évaluer la pertinence de l'approche dans une industrie graphique (Ihme et Stratton, 2015) et à comparer la *DDMRP* au *MRPII* dans différents scénarios de variabilité (Miclo *et al.*, 2016) et au *MRP* dans une industrie automobile (Shofa et Widarto, 2017).

Toutes ces études de cas ne précisent pas comment l'approche a été mise en place.

En somme, les deux approches hybrides *Lean Six Sigma* et *demand-driven MRP* ne fournissent pas un cadre technique détaillé de fonctionnement permettant de les implémenter dans n'importe quelle industrie. La littérature montre que la recherche n'en est encore qu'à l'étape préliminaire puisque les travaux réalisés se limitent à proposer l'intégration des principes pour le cas du *LSS*, et à présenter les étapes globales pour le cas du *DDMRP*. Notre propos est ici de comparer ces deux méthodes par une étude de cas, qui n'a pas encore été proposée.

3 ÉTUDE DE CAS – USINE AÉRO

L'usine AÉRO est une entreprise fictive puisque les informations réelles ne sont pas publiables et que les données ont été générées pour pouvoir simuler la production la plus proche de l'industrie aéronautique avec laquelle nous travaillons. La principale particularité de cette industrie est la complexité de leurs unités de production quant à leurs niveaux de nomenclature. C'est ainsi que notre étude de cas est basée sur la production des produits P-jj. Cette production intègre un certain degré de complexité lors la planification et le contrôle de la production.

3.1 Description de l'usine

AÉRO fabrique 40 types de produits différents pendant une année de production de 48 semaines. La nomenclature des produits P-jj comprend quatre niveaux, ce qui est une nomenclature faible pour le contexte aéronautique, mais qui a l'avantage d'être explicite dans un article (figure 2).

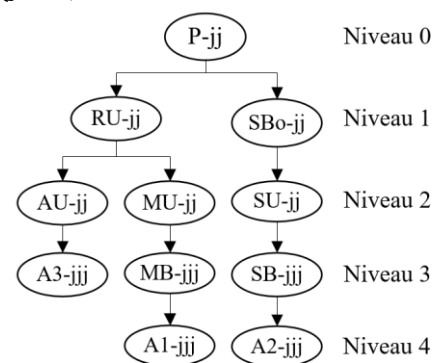


Figure 2 : Nomenclature de P-jj

conteneur et le nombre de cartes de *Kanban* respectif, ces derniers jouant un rôle fondamental. En effet, ces ajustements (*Analyze* du *DMAIC*) permettent d'améliorer le processus (*Improve* du *DMAIC*).

3.3 Modèle DDMRP

Notre modèle a été développé à partir des cinq étapes présentées par Ptak et Smith (2016).

La première étape est le positionnement stratégique des stocks qui a été effectué en utilisant la matrice *BOM – DLT*. 28 *buffers* ont été placés parmi les 123 possibilités de la nomenclature du produit. Les *buffers* ont été positionnés selon deux critères : les matières premières, qui sont favorisées, car le coût de stockage est moins élevé que les autres, et le *DLT* (i.e. la plus longue séquence non protégée dans la nomenclature d'un produit) qui ne doit pas dépasser les 10 jours ouvrables.

Ensuite, les *buffers* ont été classifiés en 9 profils, caractérisés par de longs délais d'approvisionnements. Enfin, le calcul de ses zones (i.e. verte, jaune et rouge) a été réalisé pour chacun des *buffers* en utilisant les équations de Ptak et Smith (2016). La description détaillée de ces deux étapes est présentée dans Velasco *et al.* (2017).

Lors de la troisième étape, la planification pilotée par la demande a été simulée en utilisant le logiciel *ARENA* pour les deux types d'articles identifiés – *buffer* et sur commande (figure 5). Les deux étapes précédentes sont fondamentales puisqu'elles servent d'information d'entrée pour cette démarche.

Après 45 répliques, les premiers résultats obtenus n'étaient pas satisfaisants, c'est ainsi que 7 *buffers* parmi les 28 ont été ajustés à travers le facteur *ADU* (en anglais *Average Daily Usage*) pendant la quatrième étape. Finalement, les alertes assignées pour l'exécution ont permis au planificateur de réduire les ruptures de stocks.

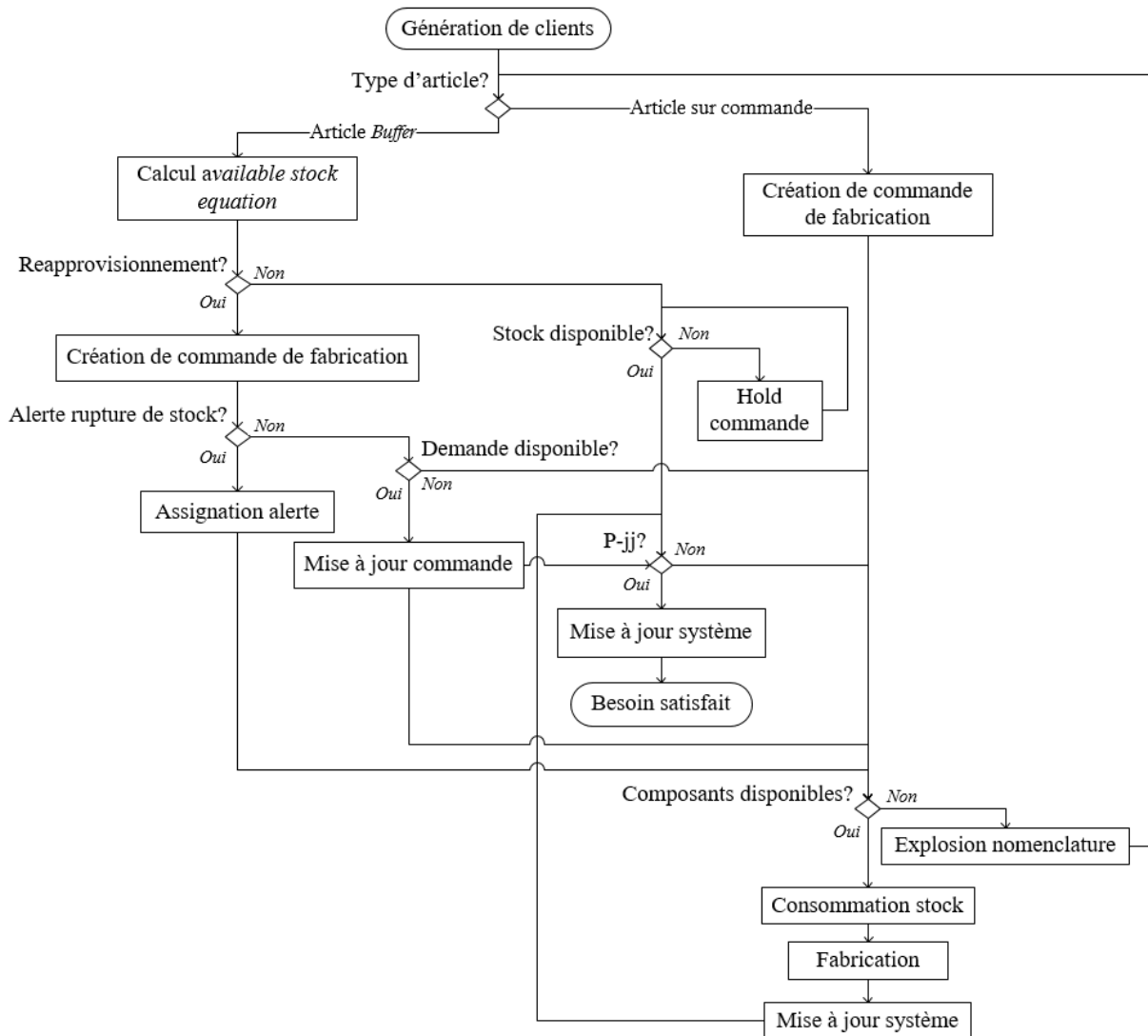


Figure 5 : Planification pilotée par la demande – DDMRP

4 RÉSULTATS

La performance de ces deux modèles simulés sous *ARENA* a été évaluée à partir du niveau de stocks et du niveau de service.

4.1 Le niveau de stocks

Deux paramètres ont été analysés : le stock physique et la rupture de stock. Le stock initial de l'usine AÉRO est de 26 641 unités entre les produits P-jj et les composants bufférisés pour le modèle *DDMRP*

Au cours de la simulation, la moyenne des stocks s'est maintenue autour de 19 031 unités, ce qui représente une réduction de 28,5% de stock physique. Dans le cas du modèle *LSS*, AÉRO compte initialement 28 715 unités de stock, qui équivalent à 662 conteneurs des 123 articles. La moyenne de stock se maintient autour de 21 791 unités, équivalente à une réduction de 24,1%.

Concernant la visibilité des ruptures de stock, nous avons pu identifier et prendre des actions correctives pour 7 *buffers* qui étaient en rupture de stock grâce au graphique d'aires empilées du modèle *DDMRP* (voir Velasco *et al.*, 2017). Ce qui n'est pas le cas pour le modèle *LSS*. La simplicité du tableau de cartes *Kanban* permet d'observer lorsque les cartes se rapprochent du point de rupture de stock. Cependant, ce modèle *LSS* ne permet pas de réaliser des ajustements en temps réel. Le e-*Kanban* pourrait être avantageux dans ce cas.

Dans la simulation, ces articles ont été considérés comme prioritaires lors de la fabrication. Afin d'éviter ceci, nous devons ajouter des conteneurs de *Kanban* ou ajuster la taille des conteneurs, ce qui implique plus de stocks.

En résumé, même si les deux modèles obtiennent au moins 24% de réduction du stock total, les résultats sont plus favorables pour le modèle *DDMRP* en matière de visibilité et d'emplacement physique.

En comparant les produits stockés des deux modèles (tableau 1), nous remarquons que le modèle *DDMRP* prévoit plus d'unités en stock, qui pourraient ne pas être utilisées. Donc, le modèle *LSS* offre aussi des gains en matière de stock.

Produit	LSS	DDMRP	Différence
P-06	125	208	40%
P-07	86	124	30,6%
P-08	70	107	34,6%
P-09	42	49	14,3%
P-10	27	27	0%

Tableau 1 : Produits P-jj stockés (en unités).

4.2 Le niveau de service

Ce paramètre a été mesuré en fonction des délais d'approvisionnement et de la date prévue de livraison de la commande.

La moyenne du délai d'approvisionnement du modèle *DDMRP* a été obtenue autour de 18 jours pour les produits P-jj. Ce qui fait un total de 14 jours de retard. Tandis que sous *LSS*, les jours de retard sont en moyenne de 2 jours (tableau 2).

Si nous analysons seulement les produits bufférisés du modèle *DDMRP* et du modèle *LSS* (tableau 2), les résultats de ce dernier sont toujours plus intéressants (i.e. 6 jours contre 2 jours de retard respectivement). Néanmoins, des améliorations devront être envisagées afin de réduire le nombre de jours de retard des commandes.

En jours	État initial	DDMRP		LSS
	P-jj	Σ P-jj	P-jj <i>buffer</i>	P-jj
Délai d'approvisionnement	29,0	17,6	12,1	2,4
Date prévue	16,4	3,7	5,8	0,2
Retard	-12,6	-14,1	-6,3	-2,2

Tableau 2 : Délais des produits P-jj.

Produit	État initial		DDMRP		LSS	
	Date prévue	Délai	Date prévue	Délai	Date prévue	Délai
P-36	17,8	28,2	3,3	1,2	0,2	1,6
P-37	17,8	26,1	3,0	0,7	0,2	2,9
P-38	17,8	36,5	3,6	2,2	0,2	3,5
P-39	17,7	36,1	2,5	1,8	0,1	1,9

Tableau 3 : Particularités des modèles *DDMRP* et *LSS*.

D'autre part, les produits P-36, P-37, P-38 et P-39 (tableau 3) qui n'étaient pas bufférisés sous *DDMRP*, mais qui ont été découplés à un seul niveau de nomenclature comme dans le modèle *LSS*, ont été fabriqués plus rapidement que leur date prévue et que le délai de *LSS*.

En somme, le délai est réduit d'au moins 38% de l'état initial, ce qui signifie que le niveau de service pourrait augmenter considérablement dans les deux cas.

Or, le niveau de service du modèle *LSS* est plus élevé que celui de *DDMRP* puisque les délais d'approvisionnement sont plus courts même s'il faut ajuster les paramètres pour éliminer les retards de commandes.

5 CONCLUSION

Notre modèle *LSS* proposé représente le premier cadre documenté du fonctionnement dans la planification et le contrôle de la production. Il est aussi le premier travail qui compare deux approches intégrées (i.e. *LSS* et *DDMRP*).

Les résultats du modèle *LSS* reflètent une performance satisfaisante puisque les objectifs en matière de niveau de stock et de niveau de service ont bien été atteints comparativement à l'état initial. Cependant, le fonctionnement du modèle doit encore être validé en matière du nombre d'unités par conteneur de *Kanban* et de critères tels que le point de commande et le point de rupture de stock.

Les deux approches visent une réduction des stocks. L'avantage du modèle *DDMRP* est qu'il permet au planificateur d'avoir une visibilité sur les ruptures de stocks et de pouvoir agir en temps réel en utilisant l'ajustement des *buffers*. Alors que le modèle *LSS* utilisant un *Kanban* permet de fabriquer des produits dans des courts délais, mais sans aucune visibilité. Néanmoins nous envisageons d'utiliser un *e-Kanban*, *CONWIP*.

Une remarque intéressante est le fait que la complexité de la nomenclature exige un grand emplacement physique pour stocker les conteneurs de *Kanban*. C'est ainsi que la planification et le contrôle de la production de 123 articles, pour une demande de 4000 produits par mois, (i.e. environ 140 unités par jour) ne semblent pas si évidents.

Les développements dans le progiciel *ERP* permettent d'ajouter des logiciels facilitant l'adoption des modèles, mais la synchronisation et la standardisation de l'information n'est pas précisée. Ceci pourra définir en grande partie la performance du modèle.

Selon Li et Li (2000), le problème, que les concepteurs de systèmes d'information doivent analyser, est comment intégrer les principes de la fabrication dans les systèmes d'information, sans nécessairement ajouter des logiciels. Les concepts de l'industrie 4.0 pourraient faciliter cette intégration.

Les méthodes évaluées dans cet article font partie d'un projet de recherche visant la conception d'une méthode *ERP* – hybride intégrant les principes *Lean* dans l'architecture de l'*ERP* comme étape initiale.

REFERENCES

Akbulut-Bailey, A. Y. , J. Motwani, and E. M. Smedley, 2012. When Lean and Six Sigma converge: a case study of a successful implementation of Lean Six Sigma at an aerospace company. *International Journal of Technology Management*, 57(1/2/3), p. 18.

Benton, W. C. and H. Shin, 1998. Manufacturing planning and control: The evolution of MRP and JIT integration. *European Journal of Operational Research*, 110(3), p. 411- 440.

Carreira, B., and B. Trudell, 2006. *Lean Six Sigma that works a powerful action plan for dramatically improving quality, increasing speed, and reducing waste*. New York: American Management Association.

Elragal, A. and M. Haddara, 2012. The Future of ERP Systems: look backward before moving forward. *Procedia Technology*, vol. 5, p. 21-30

Flapper, S. D. P., J. Wijngaard, and G. J. Miltenburg, 1991. Embedding JIT into MRP. *International Journal of Production Research*, 29(2), p. 329- 341.

García A., J.L., A.A. Maldonado M., D. J. Prieto L., J. Blanco F., A. Gil L., and E. Jiménez M., 2016. Main benefits obtained from a successful JIT implementation. *The International Journal of Advanced Manufacturing Technology*, 86(9-12), p. 2711-2722.

George M. L., B. Kastle, and D. Rowlands, 2004. *Qu'est-ce que le Lean Six Sigma?* Maxima-L. Du Mesnil éd.

Gross, J. M., and K.R. McInnis, 2003. *Kanban made simple: demystifying and applying Toyota's legendary manufacturing process*. AMACOM American Management Association.

Ihme, M. and R. Stratton, 2015. Evaluating Demand Driven MRP: a case based simulated study. *International Conference of the European Operations Management Association*, p. 1-10.

Kumar, S. and D. Meade, 2002. Has MRP run its course? A review of contemporary developments in planning systems. *Industrial Management & Data Systems*, 102(8), p. 453- 462.

Li, H. and L. X. Li, 2000. Integrating Systems Concepts into Manufacturing Information Systems. *Systems Research and Behavioral Science*, 17(2), p. 135- 147.

Liu, Z, G. Li, and L. Zhu, 2006. The Study of Key Technologies in Manufacturing Management-Control Integration. *2006 International Conference on Computational Intelligence for Modelling Control and Automation, and International Conference on Intelligent Agents Web Technologies and Internet Commerce (CIMCA-LAWTIC'06)*, p.190-196. IEEE.

Miclo, R., F. Fontanili, M. Lauras, J. Lamothe, and B. Milian, 2016. An empirical comparison of MRPII

- and Demand-Driven MRP. *International Federation of Automatic Control IFAC-PapersOnLine*, vol. 49, p. 1725-1730.
- O'Brien, J., and G. Marakas, 2010. *Management Information Systems*, McGraw-Hill/Irwin.
- Perry P. F., 1980. Optimising production, inventory and procurement control in manufacturing industry: two case studies of companies with computerised management information systems. *IEE Proceedings A - Physical Science, Measurement and Instrumentation, Management and Education - Reviews*, 127(8), p. 575- 583.
- Powell, D., E. Alfnes, J. O. Strandhagen and H. Dreyer, 2013. The concurrent application of lean production and ERP: Towards an ERP-based lean implementation process. *Computers in Industry*, 64(3), p. 324- 335.
- Ptak, C., and C. Smith, 2016. *Demand Driven Material Requirements Planning (DDMRP)*. Industrial Press, Inc.
- Rondeau, P. J. and L. A. Litteral, 2001. Evolution of manufacturing planning and control systems : From reorder point to enterprise resource planning. *Production and Inventory Management Journal*, 42(2), p. 1- 7.
- Shofa, M. J. and W. O. Widyarto, 2017. Effective production control in an automotive industry: MRP vs. demand-driven MRP. *AIP Conference Proceedings*, vol. 1855, p. 20004-1-9.
- Umble, M., E.Umble, and L. Von Deylen, 2001. Integrating enterprise resources planning and theory of constraints: a case study. *Production and Inventory Management Journal*, 2nd Quarter(2), p. 43-49.
- Velasco, A., C. Mascle, and P. Baptiste, 2017. L'applicabilité de l'approche demand-driven MRP dans un environnement complexe. *12ème édition du Congrès International de Génie Industriel*. Compiègne, France
- Womack, J. P., and D. T. Jones, 1996. *Lean thinking: banish waste and create wealth in your corporation*. Simon & Schuster.

L'INDISPENSABLE EVOLUTION DES « TRANSPORT MANAGEMENT SYSTEM » POUR LE PILOTAGE DES FLUX DE PRODUITS SENSIBLES

**A. Quentin SCHOEN, B Matthieu LAURAS,
C. Franck FONTANILI, D. Sébastien
TRUPTIL**

IMT Mines Albi – Campus Jarlard – 81013 ALBI

- A. quentin.schoen@mines-albi.fr,
- B. matthieu.lauras@mines-albi.fr,
- C. franck.fontanili@mines-albi.fr,
- D. sebastien.truptil@mines-albi.fr

E. André MACHADO-ALVES

Etablissement Français du Sang – Avenue de Grande
Bretagne – 31000 Toulouse

E. andre.machado-alves@efs.sante.fr

RESUME : *Les Transport Management Systems (TMS) se sont développés depuis une trentaine d'années avec pour objectif de supporter les activités de transport des colis. Ces systèmes, devenus de plus en plus sophistiqués afin de s'adapter à des pratiques de plus en plus complexes se sont aussi fragmentés en « briques fonctionnelles » activées ou non selon les besoins des clients. Malgré tout, les flux de produits considérés sensibles restent difficiles à maîtriser et les TMS ne répondent pas ou mal au besoin exprimé. Aussi, considérant les mutations profondes en cours dans le domaine de la logistique et les technologies émergentes, il apparaît indispensable de mettre en place une nouvelle approche pour développer des systèmes de pilotage et surveillance des flux de transport de produits sensibles. L'Etablissement Français du Sang est pris en exemple de potentiel terrain d'application.*

MOTS-CLES : *Transport Management System, agilité, pilotage, surveillance, produit sensible*

INTRODUCTION

Depuis une trentaine d'années, afin de supporter les activités de transport, des logiciels ont fait leur apparition. Aujourd'hui appelés Transport Management System (TMS) ils étaient initialement utilisés en majorité pour (Marchet *et al.*, 2009):

- Soutenir la planification des transports et optimiser les chargements (Mason *et al.*, 2003)
- Faciliter les échanges de données entre le terrain et le système informatique, automatiser certains flux de colis (Lambert *et al.*, 1998)
- Surveiller les véhicules et les sites de distribution (état, consommations, temps d'arrêts)

En même temps que se développaient ces TMS, dans les années 2000, l'environnement logistique s'est peu à peu complexifié avec de plus en plus d'acteurs intervenant dans les opérations logistiques (Marchet *et al.*, 2009). Avec la démocratisation d'Internet et de technologies basées sur le réseau, de nouvelles fonctionnalités ont pu apparaître dans les TMS et les capacités de suivi des véhicules se sont renforcées. Les moteurs d'optimisation se sont développés intégrant toujours plus de paramètres et permettant une visualisation toujours plus claire des solutions proposées. Ce type de logiciel permet de soutenir le transport de milliers de colis/conteneurs à travers le monde. Toutefois, et malgré toutes les fonctionnalités proposées, le transport reste dans certains cas, difficile à maîtriser et peut alors être qualifié de « sensible ».

Le transport peut être jugé sensible pour 2 raisons :

- L'étape de transport présente un risque pour le produit en lui-même (perte, vol, endommagement du produit, de l'environnement, etc.).
- L'étape de transport est sensible non pas parce que le produit l'est mais parce qu'une dégradation de la qualité de service de cette étape (délai, dommages sur le produit) a de lourdes conséquences (financières, vitales) sur le reste de la chaîne.

Dans le premier cas nous pouvons inclure des produits :

- Chers, comme des appareils électroniques à haute valeur ajoutée, des métaux précieux.
- Vitaux, comme des organes, des produits sanguins, certains médicaments.
- Dangereux, comme des produits chimiques et/ou potentiellement contaminants (bactériologiques, radioactifs, etc.).
- Rares comme des œuvres d'art.
- Fragiles, comme des produits sensibles aux chocs, vibrations, températures, ou des êtres vivants.

Dans le second cas, n'importe quel produit est susceptible d'être concerné dès lors que celui qui en demande l'acheminement considère que les impacts en cas de retard ou d'endommagement du produit seront lourds. Ceci est particulièrement vrai lors d'une crise humanitaire par exemple ou certains produits a priori banals deviennent indispensables à des milliers de personnes (eau, nourriture, médicaments, etc.). Ces deux catégories ne sont évidemment pas disjointes.

Autant les fonctionnalités apportées par les TMS qui existent actuellement sur le marché peuvent a priori suffire à la gestion et au pilotage du transport de produits

communs, autant elles ne suffisent pas dès lors qu'un besoin d'agilité et de surveillance accrue entre en jeu. De plus, des lacunes au niveau de la réelle efficacité et efficience des transports organisés par les TMS apparaissent alors même que l'environnement logistique se complexifie. Ces lacunes risquent de s'accroître vu les tendances technologiques et d'usages qui se développent.

Aussi, il nous apparaît nécessaire de nous interroger sur l'approche à adopter pour développer les TMS qui serviront dans les années à venir. L'objectif de cette étude est de définir une nouvelle orientation pour le développement de telles applications et de soulever les axes de recherches sous jacents. La première partie est consacrée aux fonctionnalités existantes actuellement, leurs objectifs et leurs limites, en particulier pour le transport de produits sensibles. La seconde partie développe les tendances technologiques et d'usages attendues dans les années à venir, leurs impacts et opportunités sur les activités logistiques et dresse un bilan du transport des produits sensibles dans ces conditions. La troisième partie s'attache quant à elle à proposer un cadre conceptuel pour soutenir les principales évolutions à engager. Enfin, la quatrième prend pour exemple l'Etablissement Français du Sang où ce type de question se pose concrètement.

1 LES TRANSPORT MANAGEMENT SYSTEMS (TMS) – UNE EVOLUTION NECESSAIRE

1.1 Les premiers TMS

Comme (Giannopoulos, 2004), l'énonce les logiciels de pilotage des transports servent à fournir « *“an enabling tool” for the safe and efficient operation of Freight Transport systems* ». Ces logiciels sont nés dans les années 90 et se sont développés dans les années 2000 avec de plus en plus de fonctionnalités. En effet, des projets nationaux et européens pour développer ce type de système capable de supporter les activités logistiques ont alors vu le jour (Perego *et al.*, 2011). Le transport étant à l'interface de nombreux acteurs différents (expéditeur, transporteur, destinataire, transitaire, etc.) qui n'appartiennent pas forcément à la même entité la coordination (Perego *et al.*, 2011) et l'interfaçage sont devenus rapidement des problématiques à traiter.

Comme l'explique (Marchet *et al.*, 2009), ce fut d'autant plus le cas que l'organisation du transport de fret s'est complexifiée dans les années 2000. Le nombre important d'acteurs impliqués dans le transport d'un conteneur entre le moment où celui-ci est rempli et le moment où il est vidé rend les opérations logistiques de plus en plus complexes. Ceci est d'autant plus vrai entre plusieurs pays et/ou avec des changements de mode de transport (transporteur ferroviaire, maritime, routier, gestionnaire de plateforme multimodale, de stockage provisoire, affréteur, etc.), sans considérer la gestion du dernier kilomètre pour certains produits. Comme (Giannopoulos, 2004) l'avait anticipé, les TMS se sont retrouvés peu à peu les uns à côté des autres, en mesure de gérer la

supply chain de leur entreprise mais avec des difficultés pour échanger des informations entre eux et collaborer. Ces discontinuités dans l'échange d'information tout le long de la chaîne logistique posent le problème de la prise de décision efficace et efficiente avec suffisamment d'informations fiables.

1.2 Les TMS aujourd'hui

Similairement aux autres logiciels du marché utilisés pour gérer les opérations logistiques, les TMS ont évolué peu à peu vers des modes comportant plusieurs briques fonctionnelles activées ou non selon les besoins de chaque client, qu'ils soient « transporteurs » ou « chargeurs » (i.e. en demande de transport).

Les fonctions que l'on trouve aujourd'hui classiquement chez les éditeurs de solutions TMS sont les suivantes :

- Configuration des sites et zones (les nœuds du réseau)
- Configuration des tournées de transport
- Configuration des prestataires sous-traitants
- Mise à l'expédition de colis, édition des documents de transport et documents douaniers
- Optimisation et simulation des tournées de transport
- Suivi des étapes de transport des colis, surveillance des véhicules
- Bourse au fret avec mise en concurrence des différents prestataires potentiels pour une même commande, création de devis
- Gestion de la facturation et suivi client
- Gestion de la logistique inverse (retour produits et bacs vides)
- Suivi de la qualité, des anomalies, KPI
- Connexion et configuration des matériels mobiles (douchettes, PDA, capteurs)
- Maintenance et suivi du parc de véhicules, des consommations, gestion RH
- Solution Warehouse Management System simplifiée

Ces systèmes sont utilisés tous les jours pour piloter et surveiller les transports de colis à travers le monde. Au travers d'Echanges de Données Informatisées (EDI) et grâce à des Application Program Interfaces (API), les données circulent d'un système à l'autre de façon plus ou moins efficace et automatisée entre acteurs de la chaîne logistique. Ces interfaces entre « briques fonctionnelles » d'acteurs divers et l'accès en mode SaaS sous forme « Plug and Play » ont permis la démocratisation de ces logiciels, non plus réservés aux entreprises qui peuvent se les offrir (licence et maintenance) mais potentiellement à n'importe quel acteur logistique (Harris *et al.*, 2015).

Nous avons réalisé un benchmark sur 10 entreprises qui proposent des briques fonctionnelles de TMS et/ou des licences. La figure 1 est une synthèse de ce benchmark avec la mise en avant des fonctionnalités décrites précédemment et leur occurrence. Il en ressort que les parties optimisation et simulation de tournées sont moins représentées que les autres. L'optimisation est souvent sous-traitée à des moteurs d'optimisation extérieurs, les algo-

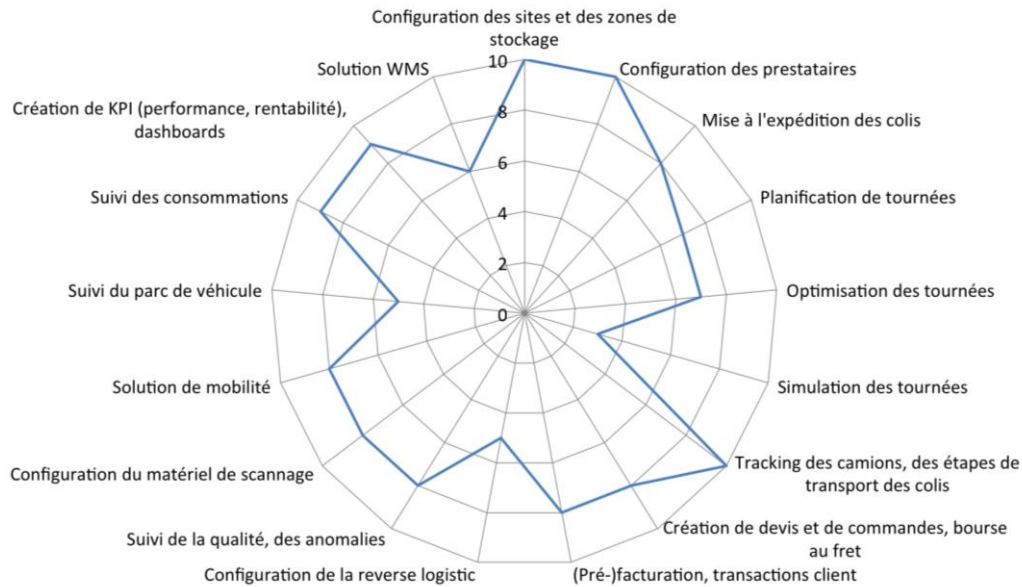


Figure 1 : Radar d’occurrences des fonctionnalités courantes des TMS

rithmes étant très spécifiques. La simulation n’est pas encore très développée, probablement également car cela demande des développements bien particuliers. La Reverse Logistics (gestion des contenants vides et retours produits) est souvent considérée comme étant à la limite des Warehouse Management Systems (WMS) d’où sa faible présence. Enfin la planification de même que la gestion des véhicules ne sont parfois pas proposées, en particulier dans le cas de TMS uniquement destinés à des chargeurs qui ne se servent que de la bourse au fret.

1.3 Avantages et limites

L’un des objectifs premiers des TMS est de maximiser le chargement des véhicules en optimisant les tournées (régulières ou non) de transports vis-à-vis des colis/conteneurs à transporter. Ceci a pour but de réduire les coûts tout en respectant qualité du transport et délais d’acheminement souhaités. L’autre objectif est d’assurer la traçabilité du transport des colis. Ceci permet de limiter les risques de perte/vol, de suivre le colis et en cas de transport dans des conditions particulières de prouver qu’il a toujours été stocké dans un environnement adapté (chaîne du froid par exemple).

Cette approche basée sur la notion de contenu/contenant (produits dans colis, colis sur palette, palette dans conteneur, conteneur sur camion, etc.) a l’avantage d’automatiser et d’optimiser l’allocation des colis aux véhicules. De plus, elle permet de ne suivre que ce qui est nécessaire. En effet, dès lors que l’on regroupe des colis dans un véhicule, cela signifie qu’ils partagent un certain nombre de paramètres (site de départ, d’arrivée, de transit, conditions de transport, type de produits contenus, etc.). Si on valide la présence d’un colis dans l’enceinte d’un véhicule, en suivant l’état et la localisation de l’enceinte on suit l’état des unités qu’elle contient. Ceci est valable également pour de potentielles décisions en cas de problème. L’ensemble des colis est donc assimilable une unité que l’on suit, le véhicule.

Malheureusement, ces systèmes montrent peu à peu leurs limites, en particulier pour le transport de colis sensibles. Nous listons ci-dessous quelques-unes des principales limites :

- Malgré tous les efforts des modules d’optimisation, les véhicules sont rarement pleins et transportent beaucoup de vide (Montreuil, 2011).
- L’optimisation des transports prend en compte les transporteurs avec lesquels chaque société est en marché et non pas tous les véhicules disponibles et intéressés autour d’elle (autres sociétés, particuliers, transports en commun, etc.).
- L’agilité (Barthe-Delanoë *et al.*, 2013) est limitée car dès lors que l’on met en place ce système de contenu/contenant seul le véhicule est suivi pendant le transport, le contenu étant vu comme une boîte noire. En cas d’événement imprévu il est parfois difficile de remonter aux produits transportés, les TMS ne fournissant des informations que sur la liste d’unités suivies et le véhicule, et pas chaque produit.
- La multimodalité reste difficile (Montreuil, 2011) à atteindre et est en général évitée quand c’est possible, la normalisation informatique et physique des colis étant encore à parfaire. Il apparaît souvent plus sûr de mettre un colis dans un véhicule fermé entre l’expéditeur et le destinataire que de lui faire « subir » plusieurs ruptures de charges.
- Le suivi des conditions de transport (température, humidité, chocs, etc.) au niveau de l’enceinte du véhicule n’est pas toujours suffisant pour des produits sensibles. Les chocs, la diffusion thermique, sont autant de paramètres non traçables au niveau du véhicule mais utiles voir nécessaires au niveau des colis.

Pour le transport de produits sensibles, ces lacunes ont un impact direct. En effet, la surveillance de chaque produit est indispensable afin d’en assurer le bon acheminement et pour le protéger de l’environnement extérieur (température, humidité, risque de vol, risque sur

l'environnement, etc.). En cas d'urgence (produits médicaux, gestion de crise, retard important) ou d'événement imprévu pendant le transport (panne véhicule, trafic, demande urgente), l'agilité est nécessaire afin de protéger les produits et de les acheminer au plus vite là où cela est nécessaire.

La surveillance fine et l'agilité ne sont actuellement pas des points forts des TMS alors qu'elles sont absolument nécessaires au transport de produits sensibles. Dans la suite nous considérons l'efficacité comme le rapport des résultats sur les moyens et l'efficacités le rapport des résultats sur les objectifs (Gibert, 1980). Nous venons de voir que l'efficacité est limitée malgré toutes les optimisations possibles. L'efficacité a un coût, avec en général des transports rapides, dédiés, point à point. Ce coût est autant environnemental, financier que social.

Aussi, cette vision centrée sur le véhicule n'est probablement plus aussi pertinente qu'avant, l'environnement logistique ayant poursuivi sa complexification alors que les fondements des TMS n'ont pas ou peu évolués. Cette complexification et les lacunes soulevées risquent d'ailleurs d'être de plus en plus criantes dans les années à venir.

2 LES MUTATIONS PROFONDES DE L'ENVIRONNEMENT LOGISTIQUE

Plusieurs axes de recherche, tant académiques qu'industriels, tendent à montrer que dans les années qui viennent la logistique devrait être affectée par des transformations majeures. En effet, plusieurs innovations et changement d'habitudes devraient impacter directement la façon avec laquelle nous produisons et transportons ce dont nous avons besoin. Ci-dessous, une dizaine d'hypothèses sont mises en avant. Les premières, concernant la captation et l'exploitation des données, vont impacter toute la *supply chain*. La seconde partie, concernant le réseau de distribution s'intéresse aux nœuds du réseau et à leurs liens alors que la troisième concerne l'évolution des colis qui circulent sur ce réseau. Les impacts potentiels et les besoins que ces tendances font naître dans les logiciels de pilotage de transports, en particulier pour les produits sensibles, sont aussi abordés.

2.1 Capter, accéder et exploiter les données

Ces 10 dernières années, le développement de technologies et d'applications web a révolutionné peu à peu la façon dont nous créons, transmettons et accédons aux données d'une part, les exploitons d'autre part.

Pour ce qui est de les capter, l'émergence de l'internet des objets (IoT) et d'appareils toujours plus petits et puissants permet de connecter n'importe quel objet au réseau, n'importe quand et n'importe où (Harris *et al.*, 2015). L'apparition des smartphones, des montres connectées, et de tous les autres objets connectés à destination du grand public permettent de suivre l'état et connaître l'environnement d'une infinité d'objets. Capter ces

données sur le terrain, les transmettre via Internet et les traiter permet de surveiller et piloter ces objets ou leur environnement. La RFID a déjà conquis de nombreux secteurs de la logistique et les possibilités offertes entre autres par les puces de géolocalisation avec des technologies type RTLS ou LoRa devrait rapidement accroître leur utilisation.

Concernant le stockage et le traitement des données transmises, le « cloud computing » offre depuis plusieurs années déjà la possibilité de stocker ces données dans des data centers et d'y accéder depuis n'importe quel appareil connecté à Internet. Ceci a contribué à développer une inter-connectivité qui offre de nouvelles perspectives. Par exemple, l'échange de données entre entreprises, bien que toujours imparfait, s'en trouve grandement simplifié ; chaque acteur n'a qu'à se connecter au serveur pour récupérer les informations pour lesquelles il possède un droit d'accès. Cette simplification pousse à l'inter-connectivité, inhérente aux activités logistiques entre plusieurs acteurs, qu'ils soient privés ou publics. Cette capacité à échanger des données rapidement et à collaborer peut être déterminante en cas de prises de décisions rapides basées sur des éléments terrain.

Grâce entre autres au cloud computing, de nombreuses applications sont maintenant proposées sous forme de service avec le développement des modes « Software as a Service (SaaS) » (O'Sullivan, 2007), (Weber, 2010). Ceci permet à n'importe qui d'accéder à un service via Internet modulo un loyer. Les prestataires d'applications, entre autres logistiques (TMS, WMS, OMS, etc.), proposent maintenant leurs logiciels sous forme de briques fonctionnelles avec des entrées et des sorties, briques que l'on décide ou non d'activer. Ceci permet tout d'abord une grande flexibilité dans l'utilisation des fonctionnalités d'une application, les services étant activés à la demande. Cela donne accès à ces applications à de petites et moyennes entreprises qui ne pouvaient pas se permettre d'acheter et de maintenir un logiciel (Harris *et al.*, 2015). Chaque acteur se concentre sur son métier en profitant de ce dont il a besoin au travers d'un accès « plug and play » qui doit lui offrir le service dont il a besoin.

Ces services utilisent la plupart du temps des données qui doivent être traitées, exploitées pour en extraire une information exploitable. Avec le développement de l'IoT sont apparues les problématiques de Big Data. En effet, de plus en plus d'appareils sont maintenant capables de capter et transmettre des données du terrain qu'il faut être en mesure d'exploiter avec des volumes et des débits importants. (Wamba *et al.*, 2015) définit alors 5 paramètres liés au Big Data (les « 5 V ») :

- Volumétrie (quantité de données)
- Vitesse (débit de données)
- Variété (diversité des types de données)
- Véracité (niveaux de confiance dans les données récoltées)
- Valeur (interprétation des données vis-à-vis de leur valeur)

La capacité des acteurs de la chaîne logistique à tirer profit de ces données en extrayant des informations ex-

exploitables capables de soutenir et alimenter un système d'aide à la décision sera à l'avenir un facteur clé de succès. En effet, capter ces données est une chose mais des techniques telles que le Process Mining, le Machine Learning, et des algorithmes d'optimisation suffisamment efficaces et pertinents seront nécessaires pour réellement supporter la chaîne logistique.

Ces technologies ont commencé déjà à impacter le grand public avec smartphones, objets connectés, publicité ciblée, et proposition de films, musiques, adaptés à nos goûts. Le secteur de la logistique n'est pas en reste et devrait rapidement adopter ce type de services tant d'un point de vue opérationnel que décisionnel.

2.2 Prendre en charge, transporter et distribuer

Nous abordons ici 4 tendances qui devraient offrir de nouvelles opportunités en transformant les réseaux de distribution et devraient imposer de nouveaux besoins aux logiciels de pilotage et suivi du transport.

Tout d'abord la multi modalité, c'est-à-dire l'utilisation d'a minima 2 modes de transport différents (route, ferroviaire, aérien, etc.) pour transporter des biens (UN, 1980), pourraient se développer rapidement dans les années à venir. En effet, (Harris *et al.*, 2015) expliquent qu'en dépit des contraintes sur les emballages, le transport et les conditions de stockage, la multi modalité permet de tirer profit de chaque moyen de transport utilisé. Ceci permet une grande flexibilité, le colis étant transporté sur chaque segment par le moyen a priori le plus adapté. Toutefois cela implique une normalisation, a minima physique et a priori informatique, pour que le colis supporte les ruptures de charges et puisse être pris en charge sur chaque segment. En cas d'urgence, cela revient à tirer parti des meilleurs moyens de transports disponibles et ne pas être freiné par des contraintes physiques, informatiques ou administratives qui impactent la performance de l'ensemble du réseau.

Ce premier point s'accompagne assez logiquement d'une séparation, d'un découpage des activités logistiques entre plusieurs acteurs. L'apparition d'entreprises 3PL et 4PL voir 5PL (Hosie *et al.*, 2007) qui prennent en charge peu à peu les activités logistiques d'entreprises dont ça n'est pas le métier initie cette décomposition. Comme le mode SaaS dans l'informatique, des services proposés par des entreprises ou des particuliers viennent soutenir et modifier la chaîne logistique (Uber, Deliveroo, Blablacar, etc.) comme (DHL, 2016) le prévoit. Sur chaque segment de transport potentiel nous devrions alors assister à une mise en concurrence de tous ces acteurs, chacun proposant un certain service à un certain coût. Si un ou plusieurs systèmes informatiques supportent cette activité, ceci permettrait lors d'un besoin de transport de profiter de tous le réseau des véhicules qui circulent avec une capacité de chargement disponible. En plus de gains d'efficience, un gain d'efficacité devrait apparaître, sans compter une flexibilité accrue. Le colis devient l'objet

suivi sur le réseau et dont la route s'adapte suivant son besoin et les possibilités qui l'entourent, quels que soient les divers acteurs en jeu. Pour que ceci soit envisageable, il est donc nécessaire de rendre le colis suffisamment autonomes pour bien délimiter les responsabilités et le router parmi les différents acteurs.

Parallèlement, nous assistons à la personnalisation de la distribution des produits qui sont livrés au plus proche du client et suivant ses envies comme énoncé par (DHL, 2016) : « anytime, anywhere delivery model ». La mise en place d'infrastructures de proximité partagées et donc l'apparition de nouveaux usages devraient soutenir ce type de réseau comme les « smart lockers » (Faugere et Montreuil, 2016), des espaces de stockages partagés (Franklin et Spinler, 2011), et le covoiturage de colis. Ces nouvelles habitudes et matériels devraient grandement aider et supporter le transport, en particulier sur le dernier kilomètre. Via ce maillage fin à haute fréquence de prise en charge potentielle, la livraison « n'importe où, n'importe quand, de n'importe quel produit » devient une possibilité. Ces technologies et usages sont encore une fois basés sur le partage de ressources (stockage temporaire, covoiturage) et la collaboration. Si nous voulons conserver voire améliorer la traçabilité des colis ils se doivent donc d'être là encore suffisamment autonome pour enregistrer ce qui leur arrive et communiquer avec leur environnement.

Enfin, l'apparition des véhicules autonomes non plus dans des environnements « protégés » à l'intérieur des entreprises mais sur la voie publique (Van Meldert et De Boeck, 2016) pourrait rapidement révolutionner notre façon de transporter des biens. Dans un premier temps ces véhicules sont attendus dans des environnements relativement contrôlés. La variabilité et les interférences pendant le trajet y sont faibles, comme sur des autoroutes. Puis peu à peu nous pouvons imaginer que ces véhicules pourront rouler dans des environnements plus complexes. L'utilisation de ces véhicules disponibles a priori 24h/24h et 7j/7j peut permettre de diminuer le temps de prise en charge entre le moment où un besoin apparaît et le moment où le produit part réellement du site expéditeur. De plus, l'utilisation de drones devrait accroître grandement l'efficacité de tout le système, les livraisons urgentes et/ou de courtes distances étant a priori facilitées. Il est alors possible de livrer quasiment n'importe où, n'importe quel produit, efficacement et à toute heure.

2.3 Aspect colis

Pour finir, nous décrivons 4 hypothèses orientées sur les produits à transporter qui devraient également modifier les besoins en pilotage de *supply chain*.

Premièrement, comme (DHL, 2016) le mentionne, la tendance actuelle est à la personnalisation des produits selon les désirs du client. D'une production de masse uniforme l'industrie parvient au travers de configurations

à des produits finis de plus en plus personnalisés. Alors même que la satisfaction client peut a priori augmenter, le produit répondant précisément au besoin, cela complexifie la gamme de produits et par ricochet la *supply chain*. En effet, en cas de dommage ou de perte d'un colis lors du transport, si les produits contenus étaient uniques il n'est a priori pas possible de les remplacer par d'autres contenus dans le même véhicule. Ceci impacte directement le client qui devra attendre une nouvelle livraison voire fabrication ou choisir un autre produit. La question de la traçabilité fine au niveau de chaque colis sur toute la chaîne de distribution apparaît de plus en plus centrale afin d'assurer un taux de service suffisant.

Deuxièmement, et cela est lié à la personnalisation, le développement de l'additive manufacturing ou impression 3D devrait à terme modifier les flux de produits. Comme (Boon et Van Wee, 2017) l'expliquent, l'apparition de centre d'impression / personnalisation proches des consommateurs devraient diminuer le transport de produits finis sur de longues distances et au contraire augmenter celui de matières premières et produits « semi-finis » relativement standards. Cela induit alors une nouvelle décomposition de la chaîne de production et pourrait supporter la croissance des produits personnalisables. En devenant de plus en plus unique chaque produit devient de plus en plus sensible et donc nécessite d'être tracé finement.

Troisièmement, les flux de produits froids augmentent avec l'apparition de nouvelles technologies pour les emballages, les véhicules et le contrôle de la chaîne du froid. Celui est valable pour les produits agroalimentaires mais aussi pour certains médicaments et vaccins qui doivent être conservés à température maîtrisée quel que soit l'environnement extérieur. Conjointement, pour des raisons sanitaires, de nouvelles règles de traçabilité et de surveillance de ces produits sont apparues (Bogataj *et al.*, 2005) afin d'assurer la qualité du produit livré au destinataire. Une surveillance accrue des produits et de leurs conditions de transport est là encore nécessaire. Comme mentionné précédemment, être en mesure de faire le lien entre contenants et contenus est ici essentiel afin de tracer les produits et réagir avec agilité aux aléas (demande urgente, perte, dommage sur le produit).

Enfin, une demande accrue de traçabilité et de transparence de la part des consommateurs et des pouvoirs publics engage les industriels à fournir de plus en plus de données. Ces données, concernant l'origine du produit, ses conditions de fabrication et de distribution, deviennent même des arguments de vente (respect de l'environnement, des salariés qui conçoivent le produit, lieu de fabrication, etc.) qu'il faut donc fournir avant même l'achat du produit. La demande de justification dans l'utilisation des ressources, entre autres lorsqu'il s'agit de dons (logistique humanitaire) ou d'argent public (logistique hospitalière par exemple) rejoint aussi

cette demande accrue de traçabilité fine. Pour pouvoir finement tracer ces nombreux paramètres au niveau de chaque colis il apparaît alors indispensable de les capter et de les analyser lors de chaque étape afin de les fournir dès la livraison du colis. Ceci rejoint les possibilités offertes par les services précédemment décrits.

Quelle que soit la technologie ou la nouvelle organisation adoptée il est essentiel d'imaginer et d'anticiper dès maintenant les fonctionnalités dont nous aurons besoin pour surveiller et piloter ces activités dans un futur proche. Des fonctionnalités sont a priori manquantes dans la figure 1, et à envisager pour les futurs TMS.

2.4 Transporter des produits sensibles demain

Considérant ce que proposent les TMS du marché actuellement, leurs lacunes vis-à-vis du transport de produits sensibles, et les tendances logistiques nous pouvons dresser la conclusion suivante :

Actuellement :

- Les transports de produits sensibles ne sont pas agiles alors qu'ils le nécessitent.
- La surveillance des colis est soit disponible en temps réel mais imprécise, soit détaillé au colis mais disponible a posteriori du transport.
- L'efficacité est limitée, l'utilisation de transports dédiés étant souvent privilégiée pour assurer sécurité et efficacité.

Dans les années qui viennent :

- Le nombre de produits considérés sensibles et/ou à tracer finement devrait augmenter.
- L'environnement logistique tend à devenir de plus en plus collaboratif, permettant une plus grande efficacité. Avec de très nombreux acteurs en jeu sur un même acheminement, devrait alors être offert plus d'agilité et si les interfaces sont bien définies, d'efficacité via l'inter-modalité.
- Les réseaux logistiques deviennent capillaires, offrant la possibilité de connecter rapidement et facilement l'expéditeur au destinataire où qu'ils soient.
- L'automatisation des tâches logistiques devrait se poursuivre ce qui requiert a priori une normalisation physique et informatique des colis.
- De nouvelles possibilités pour capter, transmettre et exploiter les données offrent et offriront des opportunités majeures pour détecter les imprévus et supporter la prise de décisions.

Aussi, pour faire face à ces lacunes actuelles qui ne feront a priori que se renforcer dans l'environnement futur, alors même que de nouvelles opportunités sont à saisir, il apparaît nécessaire d'adopter une nouvelle approche pour définir les TMS. Nous en développons une dans la partie suivante, plus centrée sur les colis.

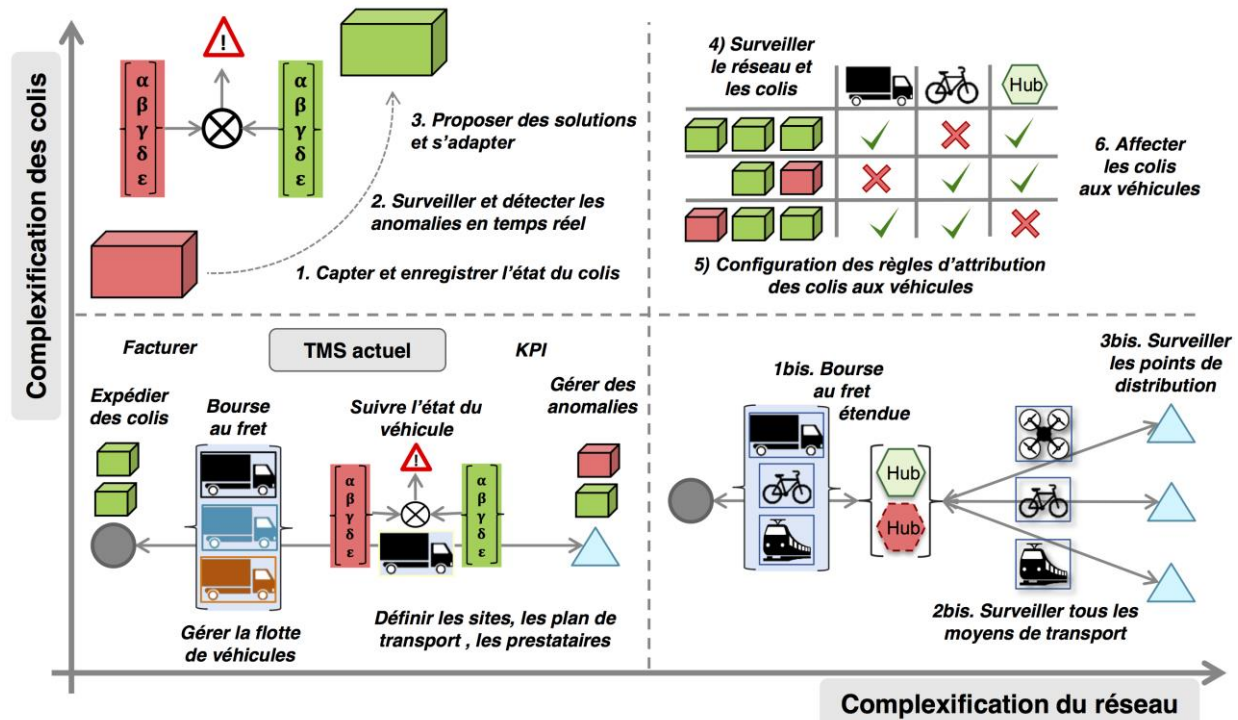


Figure 2 : Cadre de développement des futurs logiciels de pilotage et surveillance des flux de transport

3 EVOLUTIONS A APPORTER AUX TMS ACTUELS

Ainsi, une nouvelle approche est à envisager, les systèmes de pilotage et surveillance des transports actuels devant évoluer d'une vision *véhicule-centrée* vers une vision *colis-centrée*. En effet, dans un environnement « hyper-connecté » et qui essaie le plus possible de collaborer, il semble intéressant de voir les véhicules comme un support sur lequel circule des colis. De la même manière que l'information encapsulée circule sur le réseau Internet avec ses connexions et ses routeurs, les colis normalisés physiquement et informatiquement devraient être en mesure d'être déplacés par le réseau d'équipements et de véhicules. Cette idée est développée par Benoit Montreuil (Montreuil, 2011).

Pour cela, il apparaît nécessaire de développer d'une part l'autonomie des colis et d'autre part la surveillance et le pilotage du réseau. Ceci devrait permettre à terme d'attribuer et router automatiquement les colis tout le long de leur parcours. La figure 2 schématise ce cadre et met en avant des axes de recherches principaux :

- **Prise en compte et surveillance des colis :** La capacité à collecter des données au niveau du colis, soit directement via des capteurs intégrés soit via des capteurs externes et la possibilité d'en déduire une information exploitable est un premier point. Quel degré d'autonomie pour les colis ? Quelle centralisation de la prise de décision ? Quel système d'information pour supporter le système d'aide à la décision ? Quelle normalisation physique et informatique ? sont autant de questions encore à aborder.

- **Analyse du réseau de transport :** La récolte et la mise en commun de données concernant les acteurs potentiellement intéressés pour prendre en charge un colis à un moment « t » et un endroit « p » sera un autre chantier à étudier. Tant sur le plan technique qu'économique des modèles doivent être imaginés pour ouvrir le plus possible ce système et conserver efficacité et agilité dans la prise en charge des colis. Une/Des plateforme(s) de bourse au fret étendue(s) sont alors à concevoir conceptuellement, la variabilité des acteurs (taille, structure, activité) étant a priori importante.
- **Surveillance du réseau de transport :** La surveillance du réseau de véhicules (ou capacité de déplacement de charge) ainsi que du réseau de distribution et de stockage sont des prérequis indispensables à la mise en place d'une telle bourse au fret. Lancer un colis sur le réseau est une chose, s'assurer qu'il puisse être pris en charge et livré dans de bonnes conditions à l'étape suivante en est une autre.
- **Piloter les colis sur le réseau :** Des règles d'attribution des colis aux moyens de transports disponibles sont à créer, le routage du colis pas à pas suivant les moyens disponibles en conservant efficacité et efficacité est à imaginer. Suivant les contraintes de chaque colis et des colis environnants, et suivant les moyens disponibles dans cet environnement, un système de répartition des colis aux moyens apparaît modélisable. Cette capacité à avancer segment par segment en tirant profit à chaque instant des moyens disponibles autour du colis devrait accroître l'agilité et l'efficacité tout en conservant un haut niveau d'efficacité.

4 PERSPECTIVES, TERRAIN D'APPLICATION POTENTIEL

La situation décrite ci-dessus, les problèmes soulevés et le besoin d'agilité se retrouvent très bien dans le transport de produits sanguins par l'Établissement Français du Sang (EFS). Ces derniers sont vitaux pour le receveur, le respect de leurs conditions de transports et de stockage est donc essentiel. Le tableau ci-dessous présente les différents produits et leur durée de vie :

Produits	Durée de vie	Température
Sang total	Moins de 48h	18-24°C
Tubes échantillon	Moins de 48h	2-10°C
Plaquettes	5 jours	20-24°C agité
Globules rouges	42 jours	2-10°C
Plasma	1 an	< -25°C

Tableau 1 : Durée de vie et température de conservation des principaux produits transportés par l'EFS

Ces contraintes fortes impliquent une très grande réactivité pour ne pas perdre de produits. La notion « anytime, anywhere delivery model » se retrouve ici avec une obligation de résultat dans la mise à disposition n'importe où n'importe quand du/des produit(s) sanguin(s) le(s) plus adapté(s) au besoin (type, quantité, groupe sanguin). La nécessaire répartition géographique des sites de l'EFS pour être présent partout implique alors chaque année de transporter plus de 2 million de colis dont plus de la moitié à température dirigée (étude interne). Le système décrit précédemment offrirait des gains d'agilité et de surveillance pertinente pour l'EFS afin de piloter les flux et supporter la prise de décisions.

Pour développer le cadre conceptuel décrit, une phase approfondie de tests et de simulations est nécessaire pour définir comment exploiter en temps réel ces données afin de construire des informations exploitables. Des capteurs actifs directement insérés dans les colis et au niveau des véhicules lors de ces expériences devrait nous permettre de confirmer/infirmier notre modèle et de mettre en avant plusieurs axes de développement.

5 REFERENCES :

Barthe-Delanoë, A. M., Lauras, M., Truptil, S., Bénaben, F., & Pingaud, H. (2013, September). A platform for event-driven agility of processes: A delivery context use-case. In *Working Conference on Virtual Enterprises* (pp. 681-690). Springer, Berlin, Heidelberg.

Bogataj, M., Bogataj, L., & Vodopivec, R. (2005). Stability of perishable goods in cold logistic chains. *International Journal of Production Economics*, 93, 345-356.

Boon, W., & Van Wee, B. (2017). Influence of 3D printing on transport: a theory and experts judgment based conceptual model. *Transport Reviews*, 1-20.

DHL,

http://www.dhl.com/content/dam/downloads/g0/about_u/logistics_insights/dhl_logistics_trend_radar_2016.pdf

Faugere, L., & Montreuil, B. (2017). Hyperconnected Pickup & Delivery Locker Networks.

Franklin, R., & Spinler, S. (2011). Shared warehouses—Sharing risks and increasing eco-efficiency. *International Commerce Review*, 10(1), 22-31.

Giannopoulos, G. A. (2004). The application of information and communication technologies in transport. *European Journal of Operational Research*, 152(2), 302-320.

Gibert, P. (1980). *Le contrôle de gestion dans les organisations publiques*. Paris, Editions d'Organisation.

Harris, I., Wang, Y., & Wang, H. (2015). ICT in multi-modal transport and technological trends: Unleashing potential for the future. *International Journal of Production Economics*, 159, 88-103.

Hosie, P., Egan, V., & Li, Y. (2007). *Drivers of fifth party logistics (5pl) service providers for supply chain management*.

Lambert, D. M., Cooper, M. C., & Pagh, J. D. (1998). Supply chain management: implementation issues and research opportunities. *The international journal of logistics management*, 9(2), 1-20.

Marchet, G., Perego, A., & Perotti, S. (2009). An exploratory study of ICT adoption in the Italian freight transportation industry. *International Journal of Physical Distribution & Logistics Management*, 39(9), 785-812.

Mason, S. J., Ribera, P. M., Farris, J. A., & Kirk, R. G. (2003). Integrating the warehousing and transportation functions of the supply chain. *Transportation Research Part E: Logistics and Transportation Review*, 39(2), 141-159

Montreuil, B. (2011). Toward a Physical Internet: meeting the global logistics sustainability grand challenge. *Logistics Research*, 3(2-3), 71-87.

O'sullivan, D. (2007). Software as a service: developments in supply chain IT. *Logistics and Transport Focus*, 9(3), 30-3.

Perego, A., Perotti, S., & Mangiaracina, R. (2011). ICT for logistics and freight transportation: a literature review and research agenda. *International Journal of Physical Distribution & Logistics Management*, 41(5), 457-483

Van Meldert, B., & De Boeck, L. (2016). *Introducing autonomous vehicles in logistics: a review from a broad perspective*.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246.

Weber, T.: Cloud computing for business goes mainstream: <http://www.bbc.co.uk/news/10097450> (accessed 11.01.2018)

L'INNOVATION ORGANISATIONNELLE AU SERVICE DE L'AUTONOMISATION EN SANTE.

PY. TRAYNARD, M. DORSA

Pôle de Ressources - Ile de France - en Education
Thérapeutique du Patient - 16bis Rue Lauzin -75019
Paris
pierre-yves.traynard@poletp.fr, madorsa@poletp.fr

M. LONDRES

[Im]Patients Chroniques&Associés
14 rue Scandicci – 93508 - Pantin
muriel.londres@chronicite.org

RESUME : *Le Pôle de Ressources-Ile de France-en Education Thérapeutique du Patient (Pôle ETP) vient, dans un contexte d'augmentation des maladies chroniques, en appui de projets d'éducation thérapeutique et d'accompagnement au plus près des personnes malades. Améliorer leurs capacités d'agir, les impliquer dans une vision globale, pluri partenariale, coopérative passe par une dynamique de co-construction et d'innovation permanentes, une co-production de méthodes et d'outils y compris numériques avec et au plus près des patients. Nous présentons en premier le modèle d'activité qui aide le Pôle ETP à développer une compréhension collective du « quoi, pourquoi, comment ? » utile pour élaborer une forme innovante de coordination de projet. Puis nous développons les étapes itératives aboutissant à sa mise en œuvre (émergence des actions les plus opérationnelles, prototypage de solutions, évaluation de leurs usages par les patients). En appliquant une démarche coopérative visant des transformation/innovations potentielles, le Pôle ETP utilise des techniques participatives facilitant le partage et la mise en commun des savoirs issus d'expériences personnelles, communautaires et organisationnelles. Nous illustrons ces méthodes en décrivant concrètement comment un projet de recherche action crée les conditions environnementales pour que les propositions de solutions, y compris de la e-santé, n'échappent pas aux bénéficiaires finaux.*

MOTS-CLES : *Maladies chroniques, organisation innovante, pilotage agile, coproduction, autonomisation, évaluation*

1 INTRODUCTION ET CONTEXTE

Le Pôle de Ressources-Ile de France-en Education Thérapeutique du Patient (Pôle ETP) a été créé en 2013 par des « acteurs de terrain » et avec le soutien de l'Agence Régionale de Santé. En appui au développement de pratiques d'Education Thérapeutique (ETP), il a un rôle d'expert, catalyseur et médiateur contribuant à développer un environnement favorable à leur appropriation par les professionnels et les patients, avec un accent fort mis sur leur équité d'accès. Pour la réalisation de ses missions principales il cherche à développer et mettre à la disposition de ses « usagers » (professionnels, patients, institutionnels) des ressources méthodologiques et documentaires visant à soutenir leur travail de mise en œuvre d'activités d'ETP au plus près des patients suivant leurs besoins.

La création du Pôle ETP (Traynard PY *et al.*, 2011) tient d'une part à l'accroissement exponentiel du nombre de personnes vivant avec une ou plusieurs maladies chroniques et (ou) avec un handicap qui ont besoin d'ETP. En effet la prévalence des maladies chroniques augmente de façon continue et touche aujourd'hui près de 35% de la population française (Assurance Maladie, 2017) avec un gradient d'inégalités sociales qui s'aggrave (Delpierre C. *et al.*, 2016). Elles impactent la vie quotidienne des personnes concernées qui se trouvent investies dans leurs

propres soins, dans les décisions et adaptations utiles pour à la fois se mettre en sécurité et continuer à vivre leur vie personnelle, professionnelle et sociale. D'autre part, en conséquence, le nombre de professionnels de santé qui doivent mettre en œuvre des pratiques d'ETP de qualité est à soutenir (Traynard PY et Gagnayre R, 2013).

Ce constat interroge les modalités de leur prise en charge au long cours, notamment en obligeant à reconsidérer la place qu'y occupe l'ETP mais aussi « l'accompagnement à l'autonomie en santé ». Posée comme un droit des patients depuis la Loi HPST (2009), l'ETP vise à aider les personnes malades et les professionnels de santé à structurer un espace de partenariat pour affronter deux ordres d'enjeux concomitants : ceux du patient qui vit avec une ou des maladie(s) chronique(s) et ceux du soignant qui ne peut pas à le guérir. Tout en ne constituant pas une obligation légale mais dans une même perspective d'empowerment individuel et collectif, l'accompagnement au long cours en est le corollaire nécessaire (Saout C, 2015), reconnu en tant que tel par la loi de modernisation de notre système de santé en 2016 qui a autorisé un dispositif national d'expérimentation.

De nombreux travaux montrent de façon consistante que tels qu'ils sont conçus aujourd'hui et bien que visant l'encapacitation des personnes (Bravo P. *et al.*, 2015 ; Lager,

G. *et al.*, 2013 ; Pélicand J *et al.*, 2009) la majorité des dispositifs d'ETP parviennent insuffisamment à prendre en compte l'ensemble des impacts des maladies chroniques au fil de la vie des personnes concernées, notamment dans les champs du médico-social et de leur vie personnelle. A cela plusieurs raisons : 1- les programmes d'ETP ont été construits et mis en œuvre avec un tropisme majoritairement hospitalier ; 2- les difficultés d'accès restent fortes, notamment pour « ceux qui en ont le plus besoin », pour des raisons financières, socio culturelles et de littératie en santé (Margat A, 2017), d'inégalités territoriales (Lombrail P *et al.*, 2013), et les offres d'ETP sont finalement mal identifiées par les professionnels et les patients (Fournier C. *et al.*, 2017) ; 3- les représentants de patients sont trop rarement associés à la conception et à la mise en œuvre des activités d'ETP.

Les tentatives d'organiser de façon structurée et moins « descendante » les soins chroniques se heurtent à des résistances internes, à des défauts d'analyses de pratiques, et de collaboration pluriprofessionnelle de nombre de praticiens (Hung D *et al.*, 2007). Or, la littérature scientifique souligne l'importance de soutenir l'ouverture aux changements et à l'innovation, d'aller vers une approche pluridisciplinaire dans la conception et la délivrance des soins, d'impliquer les personnes malades dans les processus décisionnels, de prendre en compte l'environnement social comme un facteur majeur (Bossy D, 2016 ; Rogers A *et al.*, 2015).

Dans le même temps, l'analyse des dispositifs existants suggère les directions à prendre pour répondre à ces défis :

- 1-Mettre en relation la réorganisation des interventions médicales (postures adaptées, inclusion des pratiques éducatives) avec une plus forte implication des personnes malades dans leurs rôles de pairs aidants et accompagnants ;
- 2- Améliorer les capacités d'agir des personnes malades par un accès à des ressources intersectorielles leur permettant d'éviter les accidents aigus dus à la maladie et/ou à leurs traitements, de bénéficier de dispositifs de soins et d'éducation diversifiés tenant compte de leurs différences interpersonnelles et répondant à leurs difficultés personnelles (Bossy D *et al.*, 2016) ;
- 3- Agir dans le territoire de proximité (de vie) de la personne concernée ce qui, face au besoin croissant de soins, d'aides et de coordination entre les acteurs professionnels, associatifs et sociaux, apparaît être un moyen essentiel pour replacer l'individu, usager des soins, comme acteur principal des choix qui doivent lui être proposés (Coldefy M et Lucas-Gabrielle V, 2012). Un maillage territorial adéquat est nécessaire pour permettre de répondre au mieux aux risques de ruptures des parcours médico-socio-éducatifs.
- 4- Co-concevoir des méthodologies permettant de tester un usage intelligent des offres de téléservices (mise en réseau, télésurveillance, bouquets de services) et d'objets connectés, en fonction d'un vrai couplage entre les besoins réels des patients et les besoins du système de soins et non pas que dans une vision essentiellement médico-économique encore trop souvent pensée pour et à la place des patients. Le défaut d'organisation professionnelle dans l'usage des nouvelles technologies de type connectées, le manque de coopération entre industriels, patients

et professionnels de santé pour assurer une fiabilité et une sécurité des usages, une méthodologie d'étude non adaptée aux usages, constituent les principaux freins à leur développement (Simon P, 2017).

Accompagner les personnes concernées par une maladie chronique ne peut donc se concevoir comme une simple addition d'actions (information, éducation, conseil, soutien) mais selon des modalités de coordination originale fondée sur une vision globale, pluri partenariale et coopérative et visant l'« encapacitation » des personnes à faire leurs propres choix de santé. Dans ce cadre l'accompagnement se comprend comme toute intervention soutenant ce processus chez des personnes concernées par une vulnérabilité en santé (Saout C, 2015).

2 CATALYSER LA COPRODUCTION DE SERVICES VISANT A L'ENCAPACITATION DES PERSONNES.

Dans ce contexte complexe à la fois en termes d'objectifs et d'actions, de territoires d'intervention avec un nombre très important d'interlocuteurs, d'émergence de nouveaux outils, le Pôle ETP a identifié les stratégies, méthodologies et outils susceptibles de servir au mieux le but final de l'appui au développement des pratiques éducatives et d'accompagnement, i.e que les personnes malades puissent y accéder et en bénéficier.

Sa conviction est que la réussite de ses missions passe par sa capacité à favoriser une dynamique de co-construction et d'innovation permanentes visant une co-production de méthodes et d'outils, avec et au plus près des personnes malades.

En adéquation avec ces principes, les méthodes de travail du Pôle ETP s'inscrivent dans des analyses de pratiques éducatives interdisciplinaires et interprofessionnelles, des formation-action, des techniques de coconstruction, Design Thinking et conception et rédaction participative (techniques de « sprint écriture », scénarios, story board...), y compris avec l'utilisation de techniques numériques.

Le Pôle ETP souhaite inscrire sa contribution dans l'ingénierie des organisations innovantes dans les systèmes de santé et le pilotage agile des processus sanitaires au service d'une conception coopérative et personnalisée des prises en charge de patients.

2.1 Un exemple illustrant la démarche du Pôle

Pour illustrer de façon concrète ses méthodes de travail et les outils qu'il développe, nous avons choisi de décrire la manière dont le Pôle ETP est intervenu avec ses partenaires associatifs et professionnels pour co-construire le projet de recherche-action Aceso (Accompagnement Coopératif Evolutif et Solidaire), en co-concevoir la coordination et participer aux premières étapes de sa mise en oeuvre.

Aceso est un projet collectif rassemblant 25 partenaires (associations de patients, organisations du secteur médico-social, structures professionnelles, le Forum Living

Labs Santé Autonomie, le Laboratoire Education et Pratiques de Santé EA 3412 – U.Paris 13). Il est porté par [Im]Patients Chroniques & Associés, coalition d'associations de personnes vivant avec une maladie chronique, et co-coordonné par le Pôle ETP. Il vise à rendre opérationnels des programmes d'accompagnement innovants, co-construits, efficaces et inclusifs, en partant de la mise en évidence des pratiques existantes chez les partenaires. Il constitue un des projets inclus dans l'expérimentation nationale de dispositifs d'accompagnement à l'autonomie en santé (Ministère de la santé, 2016), choisis après appel à projets financés pour 5 ans (2017-2021).

Dans cet article nous présenterons tout d'abord le modèle qui a servi à élaborer cette forme innovante de coordination. Puis nous décrirons les différentes étapes de l'intervention du Pôle ETP aboutissant à la phase d'écriture de la réponse collective à l'appel à projet et au début de sa mise en œuvre, en en présentant les principaux résultats. Nous discuterons ces derniers ainsi que l'usage fait du modèle conceptuel.

3 LE MODELE CONCEPTUEL

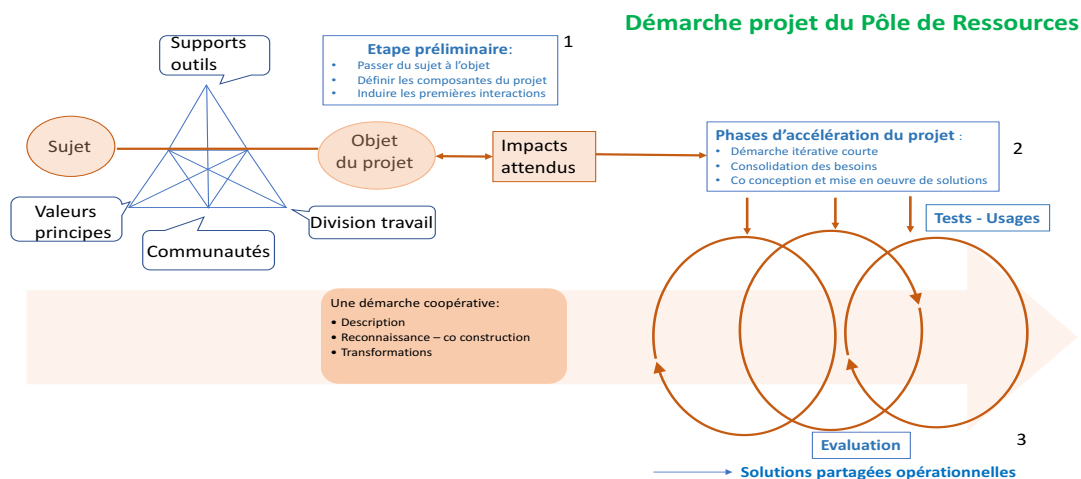
Nous en faisons état dans la mesure où il constitue un des outputs de l'intervention du Pôle ETP. En effet, l'usage de ce modèle (que le Pôle a utilisé pour reconfigurer en partie ses objectifs et modalités d'action (Boucomont A, 2017) est apparu utile lorsqu'il s'est agi de faire travailler les partenaires pressentis durant la phase initiale d'écriture du projet dans la mesure où les objectifs construits suggéraient fortement des transformations de leurs pratiques.

Il s'agit d'un modèle de l'activité développé par Engström (Engström Y, 2001) à partir des travaux initiaux de Vygotsky (Vygotsky I.S.1987) et Leont'ev (Leont'ev A.N,1978) portant sur l'activité humaine en général. Pour ces chercheurs qui travaillaient dans le domaine de la psychologie, toutes les activités humaines sont médiées par des artefacts (outils, supports médiateurs) culturellement élaborés et par des relations sociales. Ils insistent en particulier sur le rôle du langage mais aussi de schémas, diagrammes, dessins porteurs du contexte social et qui le

transmettent à l'individu. Leont'ev (Leont'ev A.N,1978) a élargi cette conception individualiste de l'activité en y intégrant une dimension collective.

Engeström l'a adapté de manière spécifique aux collectifs de travail pour décrire les activités au sein d'une organisation et notamment la manière dont les innovations peuvent ou non s'y développer. Son approche des systèmes d'activités est étroitement reliée à la notion d'apprentissage (Gilbert *et al.*, 2013). L'usage d'un outil suppose son appropriation qui entraîne un changement chez l'utilisateur, sujet de l'activité, mais il peut aussi changer l'objet même de l'activité avec de nouveaux buts à satisfaire dans la mesure où l'outil porte en lui des buts implicites, l'héritage culturel mis par ses concepteurs et qui réfère à un contexte d'usage. Engström a élargi cette approche organisationnelle ce qui l'a conduit à préciser les éléments spécifiques à l'analyse du travail au sein d'une organisation : l'existence de communautés régies par des valeurs (Alter N, 2010), des systèmes de règles formelles et informelles (Crozier M et Friedberg E, 2014), une division du travail (modèle Fordien).

On aboutit alors au triangle du système d'activités (Figure 1). Il inclut le sujet (le groupe et sa thématique), l'objet (ou objectif principal), les principes et valeurs, la communauté des acteurs, la division du travail, et les artefacts (supports et outils médiateurs). Les outils sont des médiateurs entre le sujet et l'objet, les règles entre sujet et communauté et la division du travail entre l'objet et la communauté. Ces composantes doivent être examinées ensemble pour comprendre les systèmes d'activités en place et leur dynamique d'évolution. Le passage du sujet à l'objet se construit au regard des impacts attendus par les partenaires et non plus seulement en fonction d'une déclinaison d'objectifs prédéfinis. Les actions envisagées peuvent alors être hiérarchisées et soutenues par l'identification des outils (existants ou à créer) ce qui en facilite la mise en œuvre et l'opérationnalité. Cette étape aide à pointer les éléments sociaux (valeurs, communautés, organisation) et matériels (supports et méthodes préexistants) utiles pour développer une compréhension du « quoi, pourquoi, comment ? » d'un groupe en action (Postma C *et al.*, 2012).



Ce qui est essentiel dans cette approche est de savoir mener de front et de constamment relier le travail conceptuel avec l'usage des outils qui le portent.

Le mode d'intervention du Pôle ETP, congruent avec cette conception des transformation/innovations potentielles, est fait d'étapes itératives (phases dites d'accélération de projet) : facilitation de l'émergence des actions les plus opérationnelles, prototypage des solutions en veillant à la prise en compte des différents contextes d'action et à ce que leurs usages par les bénéficiaires soient évalués (figure 1). Le partage et la mise en commun des savoirs issus d'expériences personnelles, communautaires et organisationnelles reposent alors sur des temps d'écriture collective (Bremner S. 2010). Leur mise en œuvre de façon synchrone (regroupements) et/ou asynchrone (supports numériques partagés), les techniques participatives (Brain Storming, Design Thinking), les alternances de rythme (séquentiel, parallèle, sommatif), mobilisent les partenaires, aident à la régulation des controverses et facilitent la construction de savoirs collectifs (réflexivité collective et métacognition).

Pour l'étape de mise en œuvre, nous nous appuyons en particulier sur l'ingénierie de démarche des Living Labs Santé et Autonomie (Picard R, 2017). La construction de la dynamique de projet s'appuie sur la production de données issues des besoins identifiés, un processus de décision de passage d'une étape à l'autre (montage – idée – mise en œuvre – conception évaluative), des itérations autant que de besoin jusqu'à l'obtention des objectifs et l'ajustement des moyens en conséquence. On retrouve les principes d'une démarche coopérative de type « Agile » (Hoda R *et al.*, 2010) reposant sur une organisation horizontale, une planification fondée sur la dynamique de production, des modes décisionnels démocratiques, des méthodes et dispositifs de communication facilitant le partage des connaissances, un développement basé sur les résultats des tests réalisés.

4 LES RESULTATS

4.1 Première étape : La phase d'écriture du projet

Au cours de l'étape préliminaire d'Aceso (Juin-Aout 2016) deux séminaires et trois réunions de travail ont été réalisés pour la construction d'une réponse collective à l'appel à projets.

Le premier séminaire a cherché à créer les conditions/ingrédients nécessaires pour solidifier l'engagement futur des candidats dans le projet envisagé. Il été consacré à la présentation réciproque des partenaires, à un premier recensement des besoins en termes d'accompagnement de chacun d'entre eux et à la construction d'un positionnement commun sur trois questions : « En quoi l'accompagnement nous concerne » ; « Ce que nous voulons montrer » ; « Ce sur quoi nous pourrions être évalués ».

En a résulté la mise en évidence de principes directeurs et des actions à privilégier dans le cadre du projet :

- S'accorder sur l'enjeu central de la transformation des pratiques (des soignants et des patients) ;
- Cibler la vulnérabilité comme un contexte de départ ;

- Être attentifs aux interactions entre les différents processus caractérisant l'accompagnement : professionnels de santé autour de la maladie, professionnels et patients ensemble autour de l'assistance à la personne, patients vers les professionnels de santé autour des ressorts du « vivre avec la maladie » ;

- Elaborer une grille commune de description des structures partenaires et des pratiques d'accompagnement ;

- S'efforcer d'augmenter les ressources mobilisées (veiller à l'élargissement de l'ETP), repérer les compétences transversales mobilisables dans les actions d'accompagnement ;

- Accepter que l'évaluation soit co-construite « en marchant » avec le but de mesurer in fine le gain en pouvoir d'action des bénéficiaires mais aussi la qualité de vie et l'efficacité.

Ce séminaire a permis de s'accorder sur : la division du travail entre partenaires au sein du projet, la création d'un comité de pilotage, la nature des supports et outils communs envisageables (calendrier, support numérique collaboratif), autant d'éléments nécessaires pour entamer l'écriture collective du projet.

Le deuxième séminaire visait la co-rédaction du projet, suivant une méthode d'animation et un déroulé structurés : temps de réflexion sur le sujet et son champ lexical, écriture individuelle des mots-clés, groupements thématiques visant à construire l'architecture du texte. Cette technique a permis au groupe de clarifier ses intentions, ses valeurs communes, et les attentes, rôles et engagements de chacun pour les cinq prochaines années. Les membres du comité d'écriture ont finalisé la rédaction au cours de trois réunions de travail sur un « pad » partagé assurant la transparence et la validation collective.

Ces travaux ont abouti à la conception des 4 objectifs : 1- élaborer une méthodologie commune permettant la co-construction de modalités d'accompagnement pérennes ; 2- élaborer de nouveaux outils et supports contributifs aux nouveaux programmes d'accompagnement ; 3- faire émerger un cadre commun d'évaluation de leurs résultats et d'impacts ; 4- partager et diffuser des savoirs et des savoir-faire issus du projet.

Entre-temps, un descriptif des programmes et actions d'accompagnements déjà existants, a été produit par chaque partenaire sur une grille définie en commun.

4.2 Deuxième étape : le début de la mise en œuvre.

Le troisième séminaire Aceso (Janvier 2017) a été consacré à l'organisation générale et au démarrage de l'expérimentation. Cette étape de la démarche s'est efforcée de :

- Consolider les analyses d'une part des besoins des personnes malades, d'autre part des actions existantes en cherchant à recueillir les valeurs soutenant les pratiques des acteurs de terrain. Cette explicitation constitue le socle à partir duquel la coopération peut s'envisager entre les parties prenantes du projet et les patients, un préalable à toute accélération de projet. Elle va à l'encontre des schémas classiques de logique descendante du management des organisations (Alter N. 2010),

- Lever les ambiguïtés sur les impacts attendus, qui doivent être envisagés sous l'angle d'hypothèses de départ à

tester ce qui focalise l'évaluation sur l'analyse des contextes, conditions et processus d'implantation des activités (Moore G *et al.*, 2015) ;

- Favoriser l'émergence de solutions raisonnables et opérationnelles coproduites par les organisations partenaires et qu'elles s'engagent à mettre en œuvre dans leur contexte spécifique.

Un temps de réflexion a été dédié à relever « les idées et les envies » des partenaires, par exemple : Fabriquer et insuffler de la force aux acteurs ; Être toujours situés par rapport au terrain ; Ne pas être dans les règles de l'art, mais dans le dérèglement de l'art ; Tenir compte des besoins essentiels de l'existence ; Viser l'autonomie en valorisant le savoir-vivre des usagers...

Lors de cette rencontre, les partenaires ont co-écrit une « charte des principes et des valeurs » du projet : sprint écriture à partir de mots-clefs, rédaction d'un texte martyr imprimé sur place, publié sur le site collaboratif, puis complété et validé par l'ensemble des partenaires.

L'élaboration des axes d'action a conduit à l'organisation de quatre groupes de travail thématiques :

1-« Guide du Routard » : produire, par des visites mutuelles, de l'interconnaissance mise en commun, créer un « portrait » de chaque structure partenaire et une cartographie des actions d'accompagnement existantes ;

2-« Glossaire » : construire une définition de l'accompagnement à partir des différentes conceptions des partenaires et de la littérature ;

3-« Communication » : faciliter le flux d'informations sur le projet dans chaque structure partenaire et travailler la communication vers l'extérieur (professionnels de santé et personnes concernées) ;

4-« Bourse de situations et solutions » : analyser et échanger sur les situations complexes vécues par les partenaires, les solutions mises en place et la coconstruction d'actions innovantes.

Dans ce cadre, la première action proposée a consisté en la mise en place de « visites » entre partenaires, chaque structure étant visitée une fois et en visitant deux autres. La rédaction de notes d'étonnement, inspirée de la « pédagogie d'étonnement » (Thievenaz J., 2013), a été appliquée par les visiteurs et les visités : dégager les éléments saillants et ce que les visiteurs ont compris des actions d'accompagnement de la structure visitée, à partir de ce qu'ils ont entendu/vu et de ce qui faisait sens pour eux. Cette approche a facilité l'interconnaissance, suscité des apprentissages communs et, surtout, déclenché un premier mouvement de réflexivité des partenaires sur leurs pratiques. Tout le matériel produit au cours d'un premier cycle de 11 visites a été rassemblé dans des « kits de découverte », premier outil opérationnel structurant les visites futures (chaque kit contenait les notes d'étonnement, les résumés des situations présentées, des photos prises pendant la visite, ainsi que les descriptifs des structures avec leurs emplacements sur une carte).

Après 10 mois du lancement du projet, **un quatrième séminaire Aceso** a eu lieu en octobre 2017 pour réfléchir sur le chemin parcouru et les visites d'étonnement, reconstruire les missions des groupes de travail en action et pla-

nifier la suite des travaux. La journée a débuté par une déambulation entre les « kits de découverte », inspirée des « voyages apprenants » (Billaudeau V et Poutier E, 2015), ponctuée de « souvenirs du futur », récits de parcours de visiteurs et de visités envisageant les perspectives d'application de ces apprentissages dans les pratiques futures. Ce parcours a préparé les participants à un travail de partage des conceptions des uns et des autres sur l'accompagnement.

A partir des résultats d'analyse d'un questionnaire auprès des partenaires et par des techniques d'élaboration progressive les participants ont sélectionné une gamme de notions fortes, partagées par tous, sur ce qui est et ce qui n'est pas de l'accompagnement. Conformément à son action spécifique le groupe « Glossaire » contribuera à la stabilisation de la définition d'accompagnement dans le projet Aceso en réfléchissant plus finement au fil du projet, aux notions controversées et à celles dépendantes de la « posture d'accompagnement », afin de présenter un champ lexical précis aussi bien que des postures adéquates ou à éviter.

4.3 Le travail sur l'évaluation

La dernière partie de ce quatrième séminaire a été consacrée au démarrage de la réflexion sur l'auto-évaluation d'Aceso, en partant d'un accord sur la proposition d'une évaluation caractérisée par les dimensions suivantes :

- Participative, transversale et inclusive : s'efforçant d'associer dans la mesure du possible un maximum de partenaires à ses différentes phases, afin de déterminer collectivement la valeur des solutions proposées par les parties prenantes (valeurs de chaque partie et valeur commune) ;

- Pluridimensionnelle et donc multicritères : gestion du projet, organisation, méthodologies et usages, techniques (et technologies) mises en œuvre, économie ;

- Processuelle : décrivant les apprentissages des différents partenaires au cours du projet, les conditions qui les rendent possibles et les changements de pratiques en résultant ;

- Formative et donc accompagnatrice dans la durée.

Simultanément, le groupe a dû se positionner par rapport à l'évaluation nationale auxquels sont soumis tous les projets participant à l'Appel à Projets. Pour construire ce cadre d'évaluation nationale, une procédure DELPHI (méthode de consensus visant à organiser la consultation d'experts sur un sujet précis) a été organisée et un questionnaire portant sur les critères susceptibles d'être retenus a été envoyé à chaque projet sélectionné.

Pour y répondre, le groupe de travail en charge de l'évaluation d'Aceso a imaginé une méthode permettant d'aboutir à une réponse collective représentant au mieux les attendus de l'ensemble des partenaires. L'analyse des réponses au questionnaire apportées par les différents participants, a permis de faire émerger les consensus mais aussi les zones de divergence reflétant l'hétérogénéité de jugement sur les actions existantes. Elle a aidé les partenaires à s'interroger sur les critères pertinents d'évaluation de leurs pratiques spécifiques d'accompagnement au sein d'Aceso. Mais aussi à se questionner ensemble sur les types de pratiques d'accompagnement futures et à

commencer à se préparer à les confronter avec celles des autres projets...

5 ELEMENTS DE REFLEXION

La complexité de la maladie chronique crée la nécessité d'outiller l'espace d'un partenariat patient-soignant dont l'effectivité est nécessaire dès lors qu'on vise une transformation du système de soins favorable à un engagement plus actif des malades et au plus proche de leurs besoins.

La démarche de coopération utilisée ici et visant à la co-production apparaît en phase avec la démarche d'ETP et d'accompagnement qui, par ses dimensions collaboratives entre professionnels de santé et patients, entraîne que chacun puisse gagner en pouvoir d'agir. En conformité avec le modèle d'Engström, cette démarche vise à aboutir à des transformations mutuelles, ce qui ne peut se réaliser que si elle s'appuie sur 3 dynamiques complémentaires (Needham C, 2009) :

- La description des actions menées par les parties prenantes dans leur contexte afin d'établir leur potentiel collaboratif ;
- La reconnaissance que les actions décrites ont une valeur et créent des valeurs nouvelles par leur contribution aux objectifs communs et partagés ;
- La transformation des pratiques et services résultant d'une redistribution des pouvoirs et contrôles au travers du développement de dispositifs plus centrés sur l'empowerment des personnes bénéficiaires.

La méthodologie utilisée pour répondre collectivement à l'appel à projets a eu des effets positifs sur la constitution du groupe et de son identité. Dans notre exemple, la première formulation de l'objet a été celle de l'accord entre des partenaires d'horizons divers mais tous motivés par le triple challenge de l'empowerment individuel (Bravo P *et al.*, 2015), communautaire (Wallerstein N. 2006) et organisationnel (Israel B *et al.*, 1994). Les méthodes et outils-soutiens coopératifs ont facilité les interactions à l'origine des premières propositions de solutions, contribuant à définir de plus en plus précisément ce que les actions d'accompagnement peuvent être. Des outils spécifiques ont été mis à disposition. Le site collaboratif contient toutes les informations sur le projet et les outils permettant de travailler ensemble et à distance : cartographie des partenaires, description des programmes d'accompagnement existant, agenda partagé, tableau de bord, un mur d'expression libre etc. Ce site est évolutif, incrémenté par les partenaires en fonction de l'évolution du projet et des besoins identifiés dans chaque étape.

Les visites d'étonnement, kits de découverte, voyages apprenants sont autant d'éléments de reconnaissance des actions que les partenaires mettent déjà en oeuvre. Leur analyse collective contribue à les valoriser au travers du processus de définition de l'accompagnement dans le projet Aceso. Cette approche par l'analyse collective des situations collant aux réalités de la vie quotidienne avec une maladie chronique, abordées ici lors des visites, semble s'imposer en étant portée par un courant naissant des démarches qualitatives. Elle demeure trop souvent qualifiée de « science molle » en décalage avec les analyses d'in-

tervention « hors contextes » propres à la médecine fondée sur les preuves. Pourtant nous avons montré que les démarches de coproduction en santé peuvent s'appuyer sur des modèles reconnus de l'activité humaine, qu'elles peuvent être explicitées, leurs offres bien identifiées et décrites, les processus de constructions de solutions rendus explicites pour pouvoir être partagés (Durose C, 2014). Elles participent à aider les personnes malades et celles qui les entourent à construire un environnement capacitant, leur permettant de développer de nouvelles compétences et connaissances, d'élargir leurs possibilités d'action, leur degré de contrôle sur leur tâche et sur la manière dont ils la réalisent, c'est-à-dire leur autonomie (Falzon P, 2005). Ces évolutions représentent des changements de paradigme conséquents qui nécessitent des appuis, des espaces de médiation, des méthodologies facilitant les expérimentations locales de terrain à la recherche des solutions les plus opérationnelles pour les bénéficiaires, notamment les plus vulnérables.

C'est ce positionnement qui a permis au Pôle ETP d'investir une démarche de projet « hybride » empruntant aux courants de pensée coopérative. Il ne s'agit pas d'offrir une nouvelle déclinaison d'un modèle de prestation standard ni de faire une offre de solution magique universelle. Ce qui est visé d'abord est la façon de mettre les partenaires d'un projet dans des dispositions facilitant la co-construction et la co-production. Dans cette démarche deux principes sont essentiels. Le premier est la reconnaissance du fait que les contributions novatrices doivent être issues des besoins et réalités du terrain, avec la pleine participation des bénéficiaires. Le second est de faire de la diversité des parties prenantes non un problème mais un atout pour la création du bien commun ce qui constitue aussi un de ses défis (Needham C, 2009).

Ces résultats interrogent leurs conditions de généralisation. Il y a indéniablement une condition motivationnelle et un volontariat pour être et agir en tant que partenaire d'un projet. On ne peut pas exiger que tous soient en phase avec sa temporalité. Accepter que certains s'éloignent, plus observateurs, tandis que d'autres y reviennent, plus acteurs, est un des défis auxquels la coordination doit faire face. Alternier les séminaires pléniers avec des travaux de groupe thématiques, assurer la permanence et la transparence des productions de propositions (supports numériques coopératifs), veiller à ce que chaque partie prenante « trouve son compte » (balance entre les apports au bien commun et les retours pour ses propres activités), sont autant de défis pour maintenir le « concernement » des acteurs. On retrouve ici de manière située, la notion d'environnement facilitant et potentiellement capacitant (Falzon P, 2005) qui crée la possibilité d'accueillir des propositions de solutions plus techniques à type de services et d'objets connectés, dont on pourrait alors garantir que leur processus de création inclut les bénéficiaires dès leur conception.

6 CONCLUSION

Dans un contexte de bouleversement des politiques de santé, les changements de paradigme provoqués par le besoin d'accompagner « l'empowerment » des personnes malades obligent les acteurs de santé à redéfinir l'objet de

leurs interventions, à les resituer dans des contextes de ressources et de collaborations interdisciplinaires qui favorisent la prévention et la réadaptation durables. S'appuyer sur l'existant, sur les relations et les réseaux informels ou organisés, inclure les personnes concernées sont des conditions pour que les transformations puissent s'opérer et se maintenir dans le temps.

Entre recherche action, recherche participative et recherche interventionnelle, la démarche projet du Pôle ETP s'appuie sur des principes et des concepts qui fondent et rendent effective la coopération entre professionnels, patients et institutionnels. En visant que les bénéficiaires puissent gagner en autonomie suivant leurs besoins propres, elle doit alors être en mesure de créer les conditions environnementales pour que les propositions de solutions, y compris de la e-santé, ne leur échappent pas.

REMERCIEMENTS

Pour ses contributions et relecture : M Naiditch

Aux référents des partenaires d'Aceso : E Henry, N Coupeaux (ACSBE); B Wade, B Ginameau (Basiliade); C Terra (CMS Nanterre-FNCS); N Chalhoub, S Brunet (Cité de la santé); I Ben Amor (Comité des Familles); JL Cousineau (CORDIA); F Said Dauvergne (FeMasIf); R Picard (Forum LLSA); C Hervé (Laboratoire d'Ethique – Paris 5); K Vescovacci, Pascal Revault (Comede); P Lombrail, R Gagnayre (LEPS-Paris 13); M Safwate (Migrations Santé); A Gawtarnik (Paris Diabète); P Voizard (Revesdiab); F Abdou (URACA); A Vié (AFA); C Chau-meil (AFD); G Duport, T Sannié (AFH); J C Malaize (AFSEP); E Hermant, V Pihet (DINGDINGDONG); C Robert (Fibromyalgie France); R Aubert (France Psoriasis); G Labat, O Basse (FNAIR)

REFERENCES

Alter N. 2010. Donner et prendre. La coopération en entreprise. *La Découverte*

Assurance Maladie. 2017. *Présentation de la Cartographie médicalisée des dépenses de santé*. Disponible sur : https://www.ameli.fr/fileadmin/user_upload/documents/CP_Cartographie_des_depenses_et_des_pathologies_Vdef.pdf

Billaudeau V. et Poutier E. 2015. Une démarche innovante de recherche-action : le voyage apprenant. *Journée LITEM - Enjeux et perspectives de la publication en sciences de gestion*, 26 Mars, Evry.

Bossy D. *et al.* 2016. Institutional logic in self-management support: coexistence and diversity. *Health and Social Care in the Community*, 24(6), p.191-200.

Boucomont A. 2017. Change laboratory: a tool for policy development? A pilot study on health policy in France. *Consulting and coaching for change*, HEC Paris – Saïd Business School, Oxford University

Bravo P. *et al.* 2015. Conceptualising patient empowerment: a mixed methods study. *BMC Health Services Research*, 15:252.

Bremner S. 2010. Collaborative writing: Bridging the gap between the textbook and the workplace. *English for Specific Purposes*, 29, p.121-132.

Coldefy M et Lucas-Gabrielli V. 2012. Le territoire, un outil d'organisation des soins et des politiques de santé ? Évolution de 2003 à 2011. *QES IRDES*, 175.

Crozier M et Friedberg E. 2014, l'acteur et le système. *Points Essais*

Delpierre C. *et al.* 2016. Environnement social précoce, usure physiologique et état de santé à l'âge adulte : un bref état de l'art. *BEH*, n°16-17.

Durose C. *et al.* 2014. Evaluating co-production: pragmatic approaches to building the evidence base. *For Co-production Panel, Political Studies Association Conference*

Engström Y. 2001. Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14(1), p.133–156.

Falzon P. 2005. Ergonomics, knowledge development and the design of enabling environments. *Proceedings of the Humanizing Work and Work Environment HWWE'2005 Conference*, 10-12 December, Guwahati, India, p.1-8.

Fournier C *et al.* 2017. Comment les médecins généralistes décrivent la dimension éducative de leurs pratiques. *Santé éducation, Spécial Congrès*, 27(1), p.7-9.

HCAAM, *Innovation et système de santé*, Rapport 2016.

Hoda R. *et al.* 2010. How much is just enough: some documentation patterns on Agile projects. *Proceedings of the 15th European Conference on Pattern Languages of Programs*

Hung D. *et al.* 2007. Rethinking Prevention in Primary Care: Applying the Chronic Care Model to Address Health Risk Behaviors, *The Milbank Quarterly*, Vol. 85, No. 1, pp. 69–91

Israel B. *et al.* 1994. Health education and community empowerment: Conceptualizing and measuring perceptions of individuals organizational, and community control. *Health Education Quarterly*, 21 (2), 149-170

Lagger G. *et al.* 2013. Questions d'éthique soulevées par la pratique de l'ETP, *Educ Ther Patient/Ther Patient Educ*, 5(2), p.409-414.

Levesque J-F. *et al.* 2009. Barrières et éléments facilitant l'implantation de modèles intégrés de prévention et de gestion des maladies chroniques. *Prat Organ Soins*, 40(4), p.251-265.

Leont'ev A.N. 1978, Activity, Consciousness, and Personality, *Englewood Cliffs: Prentice-Hall*.

- Lombrail P. *et al.* 2013. Les conditions de l'accès de l'ETP en ville : les principes, comment faire ? Traynard P-Y et Gagnayre R. *Education Thérapeutique des Patients en ville et sur les territoires*, Paris : Ed Maloine.
- Margat A. 2017. *ETP : Étude de l'évaluation quadriennale des programmes d'éducation thérapeutique autorisés en Île de France (entre 2010 et début 2011)*. Disponible sur : <https://www.iledefrance.ars.sante.fr/etp-etude-de-levaluation-quadriennale-des-programmes-autorises-en-ile-de-france-entre-2010-et-debut>
- Ministère de la santé. 2016. *Expérimentation de dispositifs d'accompagnement des patients*. Disponible sur : http://solidarites-sante.gouv.fr/IMG/pdf/fiche_26.pdf
- Moore G. *et al.* 2015. Process evaluation of complex interventions. UK Medical Research Council (MRC) guidance, *BMJ*, 134p.
- Needham C. 2009. Co-production: an emerging evidence base for adult social care transformation. *Research Briefing*, 31, p.1-24.
- Ostrom E. 1996. Crossing the great divide: co-production, synergy and development. *World Development*, 24(6), p.1073-1087.
- Pélicand J. *et al.* 2009. Observance, auto-soin(s), empowerment, autonomie : quatre termes pour questionner les enjeux de l'éducation du patient dans la relation de soins. *ADSP*, 66, p.20-23.
- Picard R. 2017. La co-conception en Living Lab santé et autonomie 1. *Coll Santé, technologies et société*, Paris : ISTE.
- Postma C. *et al.* 2012. Social Theory as a Thinking Tool for Empathic Design, *Design Issues*. 28(1), p.30-50.
- Rogers A *et al.* 2015. Meso level influences on long term condition self-management: stakeholder accounts of commonalities and differences across six European countries. *BMC Public Health* 15:622
- Saout C. 2015. Cap santé : Rapport en vue du cahier des charges des expérimentations des projets d'accompagnement à l'autonomie prévues par le projet de loi de modernisation de notre système de santé. France. *La Documentation Française*. Disponible sur : http://solidarites-sante.gouv.fr/IMG/pdf/20_07_15_-_RAPPORT_-_M-_Saout.pdf
- Simon P. 2017. Les leçons apprises des principales études sur les objets connectés en télémédecine et santé mobile. *European Research in Telemedicine/La Recherche Européenne en Télémédecine*, 6(2), p.67-77.
- Thievenaz J. 2013. Le rôle de l'étonnement dans la construction de l'expérience. *Education Permanente*, 197, p.113-124.
- Traynard PY et Gagnayre R. 2013. Education Thérapeutique du Patient en ville et sur les territoires, *Maloine*
- Traynard PY *et al.* 2011. De la nécessité d'un pôle régional de ressources en éducation thérapeutique des patients en ambulatoire, en Ile de France. *Congrès de la SFSP, Lille*
- Vygotsky L. S. 1987, *Thinking and Speech*. *New York: Plenum*.
- Wallerstein N. 2006. What is the evidence on effectiveness of empowerment to improve health? *Copenhagen, WHO Regional Office for Europe. Health Evidence Network report*.

MISE EN ŒUVRE D'UN SIMULATEUR DEVS UTILISANT DES RÉSEAUX DE PETRI TEMPORISÉS SUR FPGA

Clément Foucher, Vincent Albert

LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France
clement.foucher@laas.fr, vincent.albert@laas.fr

RÉSUMÉ : Une étape du *Rapid Control Prototyping* peut consister à exécuter une version simulée d'un système de commande en lien avec un environnement réel. Dans ce cadre, le simulateur doit assurer une exécution en temps réel de la simulation de la commande afin de permettre un fonctionnement opérationnel. Cet article propose une mise en œuvre matérielle d'un simulateur à événements discrets qui, si elle ne garantit pas encore le temps réel, permet d'accélérer fortement la simulation pour tendre vers cet objectif. Il s'agit d'un travail en cours s'inscrivant dans notre volonté d'assurer un flot complet de génération d'un système depuis des modèles dans une approche d'ingénierie dirigée par les modèles.

MOTS-CLÉS : Simulation à événements discrets, Réseaux de Petri, Accélération matérielle, Rapid Control Prototyping, Temps réel, FPGA

1 INTRODUCTION

Les noyaux de simulation peuvent être découpés en deux grandes catégories : les simulateurs à temps discret, et les simulateurs à événements discrets. Dans une simulation à temps discret, le temps évolue par pas fixes prédéterminés. Dans une simulation à événements discrets, l'évolution du temps simulé est variable, déterminée par le temps auquel les événements de simulation se produisent.

La simulation à événement discrets peut comporter des avantages en termes de précision et/ou de durée d'exécution de la simulation, selon les cas considérés. Ainsi, dans un modèle dans lequel les variables évoluent lentement, il est possible d'obtenir des pas de simulation longs, et donc moins de pas et ainsi une durée d'exécution réelle plus courte pour arriver au même temps simulé. À l'inverse, si les valeurs des variables se mettent à changer plus rapidement, les événements arriveront plus souvent, conduisant à une précision de simulation plus fine.

Néanmoins, certaines précautions sont à prendre, comme le fait que des événements peuvent se produire en temps nul, conduisant dans le pire des cas à une boucle en temps zéro, et donc à l'impossibilité d'évoluer au delà d'un certain temps simulé. Par ailleurs, la variabilité des pas de temps conduit à une difficulté à prédire la durée réelle de la simulation, celle-ci pouvant être très différente en fonction des vitesses d'évolution des variables au sein du modèle.

Nous étudions la simulation à temps discret au travers

d'un environnement de simulation que nous avons conçu : Project DEVS (Albert et Foucher, 2018). Project DEVS est basé sur le formalisme DEVS (*Discrete Event System Specification*) (Zeigler et al., 2000), qui permet de définir de manière mathématiquement formelle des modèles et leur sémantique d'exécution. Notre simulateur est constitué d'un cœur de simulation, historiquement développé en Java, qui génère des classes représentant les modèles, puis les compile et les exécute pour simuler ces modèles. Project DEVS dispose d'une interface utilisateur (*Graphical User Interface* – GUI), également en Java, que nous nommons ProDEVS.

Nous adjoignons actuellement à cet outil la possibilité de distribuer la simulation sur différentes plateformes d'exécution, au travers d'une génération de code en langage C, qui permet une exécution logicielle plus efficace qu'en Java, mais également en langage VHDL, qui permet la génération de circuits logiques dédiés, déployables notamment sur FPGA. La génération matérielle poursuit deux buts. Le premier objectif est un but d'optimisation de la durée d'exécution. En effet, pour les modèles disposant d'un parallélisme inhérent, une exécution matérielle permet de tirer pleinement parti de ce parallélisme. Le second objectif est plus prospectif : il s'agit de faciliter le développement du contrôle d'un système dans un objectif de *Rapid Control Prototyping* (RCP) consistant à générer le code destiné au produit fini directement depuis les modèles de haut niveau.

Pour cela, une étape intéressante consiste à coupler un contrôle simulé à un environnement réel, afin de

vérifier dès la création des modèles de haut niveau que leur comportement est correct. Dans ce cas, il est nécessaire d'avoir une exécution de la simulation du contrôle qui répond à des temps de simulation bornés afin de permettre une exécution en temps réel. Dans ce contexte, améliorer la durée d'exécution d'une simulation permet de faciliter l'intégration du modèle simulé dans un environnement réel.

Dans cet article, nous détaillons la mise en œuvre du simulateur matériel dans Project DEVS, et les premiers résultats que nous avons obtenus. Il s'agit d'un travail en cours, certaines parties de la génération de code n'étant pas finalisées, mais les premiers résultats se révèlent encourageant.

Dans la section 2, nous détaillons les motivations nous ayant poussés à développer une mise en œuvre matérielle. La section 3 passe en revue différents travaux déjà réalisés sur la mise en œuvre matérielle de Réseaux de Petri, que nous utilisons comme modèles intermédiaires. La section 4 présente la mise en œuvre que nous proposons, tandis que la section 5 détaille les résultats obtenus. Enfin, dans la section 6, nous concluons sur ce développement tout en indiquant les pistes actuellement suivies par nos travaux sur ce sujet.

2 MOTIVATIONS

L'approche de l'ingénierie dirigée par les modèles (*Model-Based System Engineering* – MBSE) consiste à développer d'abord des modèles de haut niveau, puis à raffiner ceux-ci jusqu'à des modèles plus précis, pour terminer par le système final. La traçabilité entre tous les niveaux est une contrainte majeure qui permet de s'assurer que les modifications effectuées au plus haut niveau d'abstraction se répercutent sur les modèles de plus bas niveau. Dans le cadre de l'ingénierie système, les exigences et les spécifications formant le cahier des charges peuvent même parfois être considérées comme le modèle de plus haut niveau.

Une approche de prototypage virtuel (*Virtual Prototyping* – VP) consiste à modéliser la totalité du produit fini, incluant le contrôle et l'éventuelle partie physique du dispositif, ainsi que son environnement. Pour le système de contrôle, la génération automatisée constitue l'approche dite de *Rapid Control Prototyping* et permet de passer directement des modèles au code exécutable. Dans ce contexte, la co-simulation entre les modèles du contrôle et les modèles de son environnement peut évoluer au cours du cycle de production pour que les modèles du contrôle puissent communiquer avec un environnement physique réel.

Dans ce cadre, il est nécessaire que le simulateur qui porte les modèles du contrôle soit à même d'assurer

que l'évolution du temps simulé suive le temps réel. Pour cela, il est nécessaire de borner l'exécution de la simulation, et de s'assurer que la puissance de calcul soit suffisante pour assurer ces bornes. Or, Project DEVS a été historiquement développé en Java, alors que la gestion du temps réel n'était pas envisagée dans l'outil. Le niveau d'abstraction de Java rend particulièrement difficile, voir impossible en raison des processus non maîtrisés tels le *garbage collector*, la garantie d'un temps réel.

L'objectif de notre mise en œuvre matérielle est d'exploiter au maximum les capacités de parallélisme et d'optimisation propres aux systèmes numériques matériels pour accélérer la simulation. Par système numérique matériel, on entend un circuit logique dédié à une application, permettant de prendre en compte ses spécificités, par opposition à un traitement logiciel effectué sur un processeur séquentiel générique. Les circuits numériques matériels permettent deux améliorations principales par rapport à un traitement logiciel.

Tout d'abord, le parallélisme. En effet, dans un circuit logique, des traitements indépendants peuvent être effectués de manière simultanée à chaque cycle d'horloge, là où un processeur ne permet l'exécution que d'un seul calcul à la fois. Dans des systèmes multi-cœurs et/ou multiprocesseurs, certaines opérations peuvent évidemment être parallélisées, mais la granularité reste au niveau du processus ou du thread, là où le matériel peut descendre au niveau du calcul élémentaire. De plus, même dans les systèmes disposant de plusieurs unités de calcul, celles-ci restent en nombre limité, à l'exception notable des architectures many-cores et des GPU.

Par ailleurs, le second apport principal de l'utilisation d'une architecture logique dédiée consiste en la personnalisation à l'envie de l'architecture en elle-même. En effet, un processeur est limité par sa genericité, les cycles d'exécution consistant à aller chercher l'instruction à utiliser et les données (*fetch*), effectuer le calcul (*execute*), puis stocker le résultat (*store*), souvent en passant par un *pipeline* comportant de nombreux niveaux. En revanche, un circuit dédié ne dispose pas de cette limitation, le circuit étant directement conçu pour réaliser l'opération en question.

Les circuits dédiés peuvent être réalisés sous une forme définitive lorsque ceux-ci ont vocation à être utilisés à grande échelle. On parle alors d'ASIC (*Application-Specific Integrated Circuit*). Ils peuvent également être réalisés sous forme reprogrammable sur des composants de type FPGA (*Field-Programmable Gate Array*). Les FPGA offrent l'avantage d'un cycle très court entre la conception du circuit et sa disponibilité, pouvant aller de quelques minutes à quelques heures selon la complexité du

circuit à synthétiser. Un FPGA restant un circuit générique, il offre des performances moindres qu'un ASIC, mais qui restent tout de même extrêmement compétitives par rapport à un processeur, notamment pour des applications utilisant massivement le parallélisme.

Notre objectif dans cet article est de proposer une mise en œuvre de couples modèles/simulateur pour des modèles au formalisme DEVS. Ainsi, pour un modèle saisi dans l'interface ProDEVS, il s'agit de générer automatiquement un composant matériel comprenant le modèle associé à un simulateur afin d'être utilisable de manière autonome pour réaliser la simulation du modèle au sein de l'environnement Project DEVS.

3 TRAVAUX ANTÉRIEURS

Notre approche de génération des modèles de simulation se base sur des modèles intermédiaires en Réseaux de Petri temporisés. En effet, nous utilisons les Réseaux de Petri (RdP) pour permettre une vérification formelle des modèles DEVS, notamment afin de vérifier que ceux-ci ne disposent pas de boucle infinie en temps zéro. Pour cela, nous générons automatiquement des modèles en RdP représentant les modèles DEVS saisis dans l'interface ProDEVS. La vérification de la correspondance entre les modèles DEVS et leur équivalent en RdP est réalisée par bisimulation. Nous avons précédemment présenté ces travaux dans (Albert et Foucher, 2017). Ce sont donc ces modèles intermédiaires que nous utilisons comme base pour la génération matérielle. L'étude des travaux antérieurs que nous avons réalisés balaie donc le domaine de la génération matérielle depuis les Réseaux de Petri.

Dans (Węgrzyn et Węgrzyn, 2011), deux approches pour la mise en œuvre de RdP en VHDL sont proposées. La première méthode utilise un *process* général, synchrone, comprenant la totalité de la description du RdP. Celui-ci est découpé en trois blocs : le premier permet la définition des transitions à tirer, le second met à jour le marquage des places et le troisième met à jour les sorties. Ces blocs communiquent entre eux par des variables. Bien que cela produise un code VHDL correct et synthétisable, cette méthode produit un seul fichier dont la taille croît avec la taille du RdP. Par ailleurs, le détail de la représentation en places et transitions n'apparaît pas explicitement, celle-ci étant noyée dans un flot de variables.

La seconde méthode proposée représente les places et les transitions sous la forme de composants, les liant par des signaux représentant des arcs. Cette méthode préserve la visibilité de la représentation places/transitions dans le code final, produisant une vue du modèle proche de la représentation du RdP.

De plus, cette méthode est plus proche de la représentation basée composants que l'on retrouve fréquemment dans les circuits numériques. Il semble s'agir de la méthode généralement utilisée dans la littérature, comme suggéré dans (Gomes et al., 2007). C'est donc cette seconde méthode que nous utiliserons pour la représentation.

Dans (Soto et Pereira, 2005), les auteurs présentent une méthode pour mettre en œuvre les RdP dans un FPGA. Leur description met l'accent sur la réduction de la consommation de ressources matérielles. Ainsi, il s'agit de minimiser le nombre de CLB (*Configurable Logic Bloc*) utilisés en mutualisant ceux-ci pour les places du RdP. Néanmoins, cela apporte une complexité intrinsèque au code généré, et notre approche n'est pas pour l'instant axée vers l'optimisation des ressources. Ainsi, si cet article apporte des points de vue intéressants, nous n'irons pas jusqu'à mettre en œuvre cette optimisation.

Dans l'article (Bukowiec, 2013), les auteurs proposent une manière de mettre en œuvre un RdP en utilisant une approche GALS (*Globally Asynchronous-Locally Synchronous* – Globalement asynchrone, localement synchrone). Ceci permet d'avoir différentes parties d'un RdP utilisant chacune une horloge indépendante des autres. La synchronisation est effectuée entre les domaines d'horloge en utilisant des buffers. Cet article porte un regard très intéressant sur le découpage en sous-réseaux et l'indépendance des sous-réseaux entre eux.

4 LA PLATEFORME MATÉRIELLE

Pour notre mise en œuvre, nous avons besoin d'un circuit programmable pour la réalisation du couple modèle/simulateur, mais également d'une passerelle permettant la communication du système avec l'environnement afin de pouvoir commander celle-ci à distance directement depuis l'interface ProDEVS. Pour cette partie, dont les performances ne sont pas aussi cruciales que l'exécution du modèle en lui-même, nous avons choisi d'utiliser une mise en œuvre logicielle afin de disposer des outils existant pour la gestion du réseau, notamment la pile TCP/IP qui permet de disposer immédiatement de communications standardisées par le biais de *sockets*.

Pour cela, nous avons besoin d'une plateforme permettant aisément de mélanger le matériel au logiciel, et facilitant l'accès à un réseau. Nous avons donc choisi la plateforme de développement ZedBoard (Avnet, 2018). En effet, celle-ci comporte un système sur puce (*System on Chip* – SoC) Zynq (Xilinx, 2017), qui comprend un système matériel reprogrammable (FPGA) et un processeur situés sur un même circuit. Elle dispose également de toute la connectivité nécessaire à un branchement

réseau, notamment une interface Ethernet.

Ainsi, comme présenté sur la Figure 1, notre mise en œuvre est constituée de deux parties, une logicielle s'exécutant sur un processeur et une matérielle fonctionnant sur le FPGA, les deux communiquant via un bus AXI.

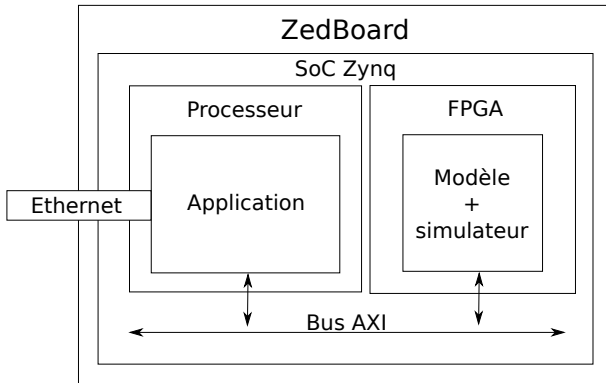


Figure 1 – Architecture de la plateforme matérielle.

4.1 Le bloc logiciel

Le bloc logiciel fonctionne sur un système d'exploitation FreeRTOS (Real Time Engineers Ltd., 2018), système ouvert permettant le déploiement de logiciels multi-threadés, accompagné de la bibliothèque LwIP (Dunkels, 2018), qui fournit une pile réseau TCP/IP. Le système logiciel que nous avons développé est générique, c'est-à-dire que le cœur est commun à tous les déploiements que nous effectuons depuis Project DEVS. Ce cœur commun comporte la partie communication avec l'interface, reposant sur un protocole développé sur le socle des sockets TCP/IP, et permettant de recevoir des ordres de l'interface, et d'envoyer des résultats de simulation lorsque ceux-ci sont disponibles.

À côté de ce cœur générique, des fichiers de configurations sont générés automatiquement par Project DEVS, qui décrivent les caractéristiques du modèle, sous la forme de fichiers d'en-tête qui sont automatiquement inclus par le cœur qui les utilise pour se configurer. Ces fichiers décrivent l'interface du modèle simulé, c'est-à-dire ses sorties, et les registres associés à chacune d'entre elle afin que le cœur soit à même de monitorer leur état, et ainsi de transmettre les valeurs générées sur les sorties à l'interface ProDEVS.

Le cœur dispose de différents threads. Parmi ceux-ci, on note principalement un serveur TCP/IP qui attend les commandes de l'interface, un thread de monitoring du circuit matériel vérifiant en permanence si des sorties ont été émises par le simulateur et un thread client TCP/IP qui émet les résultats en direction de l'interface. Tous ces threads communiquent entre eux

par des boîtes aux lettres, ce qui permet un fonctionnement asynchrone, les threads ne se déclenchant que lorsqu'ils ont un travail à effectuer. Ainsi, le thread de monitoring du circuit matériel ne démarre que lorsque la simulation est lancée, et le thread client TCP/IP ne se déclenche que lorsqu'il y a des données à émettre.

4.2 Le bloc matériel

Le bloc matériel constitue l'élément principal de notre mise en œuvre : le simulateur. Comme indiqué précédemment, le bloc contient à la fois le modèle simulé et le simulateur associé, rendant le bloc totalement autonome en termes d'exécution. Comme présenté dans (Albert et Foucher, 2017), nous utilisons les Réseaux de Petri temporisés comme pivot dans nos descriptions de modèles DEVS. Ceci nous permet de garantir que le modèle utilisé pour la validation formelle est le même que celui utilisé dans la simulation.

Le simulateur mis en œuvre est donc un Réseau de Petri temporisé, composé de plusieurs sous-réseaux correspondant aux modèles atomiques DEVS. La hiérarchie des modèles couplés est mise à plat, celle-ci n'étant pas nécessaire à la simulation proprement dite. Ainsi, tous les composants atomiques se retrouvent au même niveau dans le simulateur matériel. En revanche, l'isolation des modèles atomiques en composants est préservée, et ce pour deux raisons. Tout d'abord, cela nous permet d'optimiser légèrement la mise en œuvre, en mutualisant au niveau du composant certains signaux qui peuvent apparaître à de multiples reprises au sein de celui-ci. Ensuite, dans une perspective d'évolution future consistant à permettre la modification des composants de la simulation dynamiquement, et qui nécessitera de conserver cette séparation.

Par ailleurs, aux côtés de ces composants matériels représentant les composants atomiques DEVS, on trouve différents autres composants faisant partie du simulateur. Ainsi du coordinateur, composant supplémentaire destiné à rythmer la simulation en s'assurant que tous les composants ont terminé leur cycle de simulation avant de commencer le suivant, également décrit dans (Albert et Foucher, 2017). En plus de ces éléments, nous ajoutons deux composants principaux : le gestionnaire de temps, et le gestionnaire d'exécution.

Dans notre mise en œuvre, les composants évoluent de manière indépendante les uns des autres au cours d'un cycle de simulation, ceux-ci étant synchronisés entre deux cycles par le coordinateur. Ainsi, plusieurs transitions du RdP appartenant à des composants différents peuvent être tirées au cours d'un cycle d'horloge. Du point de vue du RdP global (comprenant tous les RdP partiels), cela implique

que plusieurs transitions sont tirées simultanément. Cela ne pose pas de problème car, dans notre cas, les transitions appartenant à des composants différents dépendent de ressources (places, variables) indépendantes. Néanmoins, ce comportement ne pourrait pas être généralisé à n'importe quel RdP.

4.2.1 Les composants atomiques

Les composants atomiques mettent en œuvre des Réseaux de Petri temporisés partiels, dans le sens où certains arcs proviennent de l'extérieur du composant, et d'autres en sortent. Un composant atomique est donc constitué de places et de transitions reliés par des arcs. Afin de s'approcher au plus près de la structure des RdP, les places sont représentées par des mémoires contenant l'état de celles-ci, les transitions par des composants asynchrones et les arcs par des équations logiques reliant les places aux transitions. Comme notre modèle de RdP n'utilise que des places binaires, un simple bascule à un seul bit suffit à mémoriser l'état. Les bascules sont synchrones à une horloge, marquées et démarquées par le jeu des équations booléennes reliant celles-ci.

Pour la gestion du temps, notre RdP temporisé utilise des intervalles réduits à un seul point ($[t_a : t_a]$) associés aux transitions. Une transition doit être activée depuis un temps t_a pour être tirable. Le temps pris en compte est bien entendu le temps simulé, géré donc par le simulateur.

Afin de simplifier la coordination entre les composants, nous avons fait le choix d'externaliser la gestion du temps dans un composant spécialisé. En effet, dans un RdP temporisé, le temps avance au gré du tir des simulation temporisées, la transition ayant le plus faible t_a étant tirée lorsque aucune transition non temporisée n'est sensibilisée. Ainsi, dans notre mise en œuvre, les transitions temporisées fournissent leur t_a initial et leur état (sensibilisée ou non), et attendent une instruction extérieure leur indiquant qu'elles peuvent être tirées. Le choix du tir n'est donc pas fait dans le composant.

Chaque composant atomique comportant des transitions temporisées dispose donc de sorties indiquant si une transition temporisée est sensibilisée, et le t_a associé le cas échéant. On notera que, par construction, au maximum une transition temporisée peut être sensibilisée à la fois au sein d'un composant atomique. Il n'est donc nécessaire de prévoir qu'un seul port de ce type, même si plusieurs transitions temporisées existent au sein du composant.

4.2.2 Le gestionnaire de temps

Le gestionnaire de temps est en charge de deux éléments majeurs de la simulation : l'avance du temps

de simulation, et le tir des transitions temporisées. Ce composant dispose d'autant d'entrées qu'il y a de composants comportant des transitions temporisées. À chaque pas de simulation, le gestionnaire de temps vérifie si aucune transition non temporisée n'est sensibilisée. Si c'est le cas, celui-ci démarre, et calcule le t_a minimal parmi toutes les transitions temporisées sensibilisées. On notera que, pendant cette période, le RdP est figé, car il n'y a aucune transition non temporisée sensibilisée, et toutes les transitions temporisées sont en attente d'une autorisation du gestionnaire de temps pour être tirées.

Lorsque le t_a minimal $t_{a_{min}}$ a été déterminé, le gestionnaire de temps incrémente le temps de simulation de ce même $t_{a_{min}}$, diminuant du même coup tous les t_a de toutes les transitions temporisées sensibilisées de cette valeur. Ensuite, il autorise la ou les transitions pour lesquelles $t_a = t_{a_{min}}$ à être tirées, débloquent du même coup le RdP qui peut alors continuer la simulation.

4.2.3 Le gestionnaire d'exécution

Le gestionnaire d'exécution est a pour but de gérer l'avancement de la simulation : démarrage et arrêt de la simulation. Afin de permettre de gérer la simulation, celui-ci délivre une autorisation d'exécution qui est utilisée par le réseau places-transitions pour inhiber ou libérer son évolution. C'est lui qui réagit aux ordres en provenance du bloc logiciel via les registres du composant matériel.

Celui-ci effectue deux tâches principales : réagir aux ordres de démarrage et de suspension de l'exécution du simulateur en provenance de l'interface utilisateur via le réseau et la partie logicielle, et arrêter le simulateur lorsque le temps maximal est atteint. En effet, lorsque l'on génère un composant matériel, la liberté est donnée de définir le nombre de bits des signaux, là où la taille de ceux-ci dans un code logiciel est fixée par l'architecture à un multiple d'octets. Le choix de ce nombre de bits est donc défini en fonction du temps maximal de simulation défini par l'utilisateur. Néanmoins, cela entraînera une limite fixe, physique, au temps qu'il sera possible d'atteindre dans la simulation, et il convient de s'assurer que celle-ci ne sera jamais dépassée au risque d'entraîner des résultats imprédictibles. Le gestionnaire d'exécution se charge donc de bloquer l'exécution du RdP lorsque le temps maximal est atteint afin d'éviter ce comportement.

4.3 Limites de la génération

La génération automatique de code permet de construire automatiquement le réseau places-transitions ainsi que la gestion du temps. Les conditions supplémentaires, ou gardes, du réseau de Petri sont également prises en compte lors de la génération au-

tomatique. En revanche, une autre partie cruciale du code, à savoir les actions, n'est pour l'instant pas générée automatiquement sauf pour certaines actions simples du type incrément d'une variable. Mais des actions plus complexes requièrent pour l'instant que l'utilisateur tape à la main le code correspondant. Cette partie fait actuellement l'objet d'études de notre part, notamment au sujet de l'utilisation d'outils de *High Level Synthesis* (HLS), qui permettent la génération automatisée de code matériel depuis du code logiciel.

5 EXEMPLES DE DÉPLOIEMENT ET RÉSULTATS

Nous avons testé cette génération automatisée sur deux exemples représentant des cas d'utilisation très différents. Dans les deux cas, les actions des modèles restent simples afin de permettre leur génération automatisée.

Le premier exemple est un exemple classique DEVS : le générateur-buffer-processeur (Gen-Buf-Proc), composé de trois modèles très simples. Le générateur émet des événements à intervalle fixe, le buffer stocke le nombre d'événements, et le processeur les consomme à intervalles fixes. Selon les durées définies du générateur et du processeur, le nombre d'événements dans le buffer croît à l'infini ou oscille entre 0 et 1. Dans notre cas, les durées ont été définies respectivement à 2 et à 3 unités de temps.

Un autre modèle plus complexe met en œuvre le jeu de la vie, en 8×8 cellules toriques. Chaque cellule est connectée à ses 8 voisins afin de mettre à jour son état à chaque cycle de simulation. Ce modèle est donc très connexe, car il comporte de très nombreuses connexions entre les composants, et très parallélisable, car toutes les cellules peuvent s'enquérir de l'état de leurs voisins simultanément.

Le déploiement des modèles en VHDL comporte quatre étapes : la génération du code, la synthèse (ou compilation du code), la programmation de la cible et l'exécution à proprement parler. L'utilisation de code logiciel est similaire mais ne comporte que trois étapes, car l'unité d'exécution du modèle est la même que celle qui génère le code : le processeur. On n'a donc pas de phase de programmation de la cible. Il est à noter que les modèles logiciels sont actuellement générés en Java, et pourraient probablement gagner à utiliser un autre langage plus rapide, mais les ordres de grandeur obtenus restent néanmoins pertinents. Les exécutions ont été réalisées sur une durée de 100 000 unités de temps.

Le Tableau 1 présente les résultats obtenus sur le Gen-Buf-Proc, tandis que le Tableau 2 présente ceux obtenus sur le jeu de la vie. La colonne *sw* présente les

résultats obtenus sur une exécution logicielle sur un PC équipé d'un processeur à 3.2 GHz. La colonne *hw1* présente les résultats sur une exécution matérielle libre, c'est-à-dire sans récupération des résultats intermédiaires. Enfin, la colonne *hw2* présente les résultats obtenus avec la même exécution matérielle, mais en récupérant toutes les valeurs intermédiaires des variables internes : taux de remplissage du buffer pour le Gen-Buf-Proc, et état de toutes les cellules pour le jeu de la vie. On note ainsi un très fort ralentissement entre les deux exécutions matérielles dues au temps de communication avec le GUI par le réseau.

Néanmoins, quelle que soit l'exécution matérielle, on note la chose suivante : Sur un modèle très simple de type Gen-Buf-Proc, aucun gain n'est réalisé au passage en matériel, en raison du très fort impact de la durée de compilation du code matériel. Par contre, la durée d'exécution à proprement parler est diminuée d'un ordre 100.

Sur un modèle très connexe et fortement parallélisable en revanche, on note un gain en termes de durée d'exécution d'un facteur 10 lors du passage en matériel. Plus encore, on remarque que les durées totales du matériel sont constituées presque exclusivement du temps de compilation du code, temps qui ne change pas lorsque l'on augmente les durées d'exécution. Un accroissement de la durée de simulation augmenterait donc encore le gain du passage en matériel.

6 CONCLUSION ET PERSPECTIVES

En conclusion, on peut dire que selon le type de modèle que l'on simule, un passage à une exécution matérielle peut être envisagée. Les critères qui justifient ce déploiement sur FPGA sont : une grande connectivité du modèle, une forte faculté à la parallélisation (évolutions simultanées et indépendantes au sein du modèle), et un temps simulé important. L'exécution matérielle est handicapée par une durée de synthèse du code très longue, mais qui ne s'accroît pas avec la durée simulée.

Le travail n'est pas terminé, car une grosse partie du travail reste à faire concernant la génération automatisée des actions associées au modèle DEVS. Pour cela, nous étudions actuellement la génération de code utilisant les outils de HLS.

On note aussi que notre mise en œuvre n'est utilisable que sur les Réseaux de Petri représentant des composants DEVS issus de Project DEVS. Ainsi, un certain nombre de restrictions sont appliquées, telles les places binaires ou les intervalles de temps ponctuels. De la même manière, certaines libertés sont prises avec le modèle de calcul des RdP comme la possibilité de tirer des transitions simultanément dans certains

Étape	Durée (sw)	Durée (hw1)	Durée (hw2)
Génération du code	4.3 ms	4.7 ms	4.8 ms
Compilation du code	844.0 ms	102 s	102 s
Programmation de la cible	N/A	3 s	3 s
Durée de la simulation	624.0 ms	6.0 ms	1.0 s
Total	1.5 s	105.0 s	106.0 s

Tableau 1 – Durées de génération et d’exécution pour le modèle du générateur-buffer-processeur.

Étape	Durée (sw)	Durée (hw1)	Durée (hw2)
Génération du code	85.3 ms	1718.6 ms	1717.5 ms
Compilation du code	2.8 s	655 s	663 s
Programmation de la cible	N/A	3 s	3 s
Durée de la simulation	5194.8 s	113.0 ms	8.58 s
Total	5197.7 s	659.8 s	676.3 s

Tableau 2 – Durées de génération et d’exécution pour le modèle du jeu de la vie.

cas. Tout ceci interdit d’utiliser notre mise en œuvre pour des Réseaux de Petri temporisés génériques, certaines précautions devant être prises.

Un autre travail à réaliser concernera le bornage des durées d’exécution. En effet, dans une perspective d’utilisation des modèles dans un contexte de Rapid Control Prototyping, les modèles doivent être capables de s’exécuter en temps réel. Cette partie reste délicate à réaliser, car nos modèles sont à événements discrets. Ainsi, contrairement à des modèles à temps discret dans lesquels il suffit de borner une itération de la boucle d’exécution pour borner la totalité de la simulation, les modèles à événements discrets utilisent des pas de temps variables. Borner une itération de la boucle d’exécution délimite ainsi un temps réel variable, en fonction de la durée représentée dans cette boucle.

Ce travail reste donc ouvert, et ne représente que la première étape de l’extension de notre simulateur Project DEVS à la simulation matérielle.

RÉFÉRENCES

- Albert, V. et Foucher, C. (2017). Formal framework for discrete-event simulation, *The 20th World Congress of the International Federation of Automatic Control*.
- Albert, V. et Foucher, C. (2018). Site web prodevs.
URL: <https://www.laas.fr/projects/prodevs/>
- Avnet (2018). Site officiel zedboard.
URL: <http://www.zedboard.org/>
- Bukowiec, A. (2013). Distributed control systems design as petri nets for fpgas, *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems (IJDATICS)* 4: 1–9.
- Dunkels, A. (2018). lwip - a lightweight tcp/ip stack.
URL: <https://savannah.nongnu.org/projects/lwip/>
- Gomes, L., Costa, A., Barros, J. et Lima, P. (2007). From petri net models to vhdl implementation of digital controllers, *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*, pp. 94–99.
- Real Time Engineers Ltd. (2018). Site officiel freertos.
URL: <https://www.freertos.org/>
- Soto, E. et Pereira, M. (2005). Implementing a petri net specification in a fpga using vhdl, *Design of Embedded Control Systems*, Springer US, pp. 167–174.
- Węgrzyn, M. et Węgrzyn, A. (2011). Penlogic – system for concurrent logic controllers design, in M. Adamski, A. Barkalov et M. Węgrzyn (eds), *Design of Digital Systems and Devices*, Vol. 79 of *Lecture Notes in Electrical Engineering*, Springer Berlin Heidelberg, pp. 215–228.
- Xilinx (2017). Zynq-7000 all programmable soc data sheet: Overview.
URL: https://www.xilinx.com/support/documentation/data_sheets/ds190-Zynq-7000-Overview.pdf
- Zeigler, B. P., Kim, T. G. et Praehofer, H. (2000). *Theory of Modeling and Simulation*, 2nd edn, Academic Press, Inc., Orlando, FL, USA.

Mixed-integer and constraint programming for a multi-skill project scheduling problem with partial preemption

Oliver POLO MEJIA, Marie-Christine ANSELMET Christian ARTIGUES, Pierre LOPEZ

DEC/SETC, CEA Cadarache
St Paul lez Durance, France

oliver.polomejia@cea.fr, marie-christine.anselmet@cea.fr

LAAS-CNRS, Université de Toulouse, CNRS
Toulouse, France

lopez@laas.fr, artigues@laas.fr

ABSTRACT: *In this paper, we consider the weekly scheduling problem of activities within one of the research facilities of the French Alternative Energies and Atomic Energy Commission (CEA in short for French). To better represent this problem we propose a new variant of the multi-skill project scheduling problem (MSPSP) involving partial preemption. We describe the new MSPSP variant and we present two formulations for the problem: one using mixed-integer linear programming (MILP) and a second one using constraint programming (CP). Computational experiments on realistic data are carried out and discussed.*

KEYWORDS: *RCPSP, MSPSP, Partial preemption, Scheduling, Nuclear laboratory.*

1 INTRODUCTION

There are not much research works studying the application of optimization techniques for scheduling activities within research facilities. At operational level, scheduling research activities becomes a very complex problem and the literature on this subject is almost non-existent. In this paper we work on the weekly scheduling of the activities within one of the research facilities of the French Alternative Energies and Atomic Energy Commission (CEA in short for French). After analyzing the operations and characteristics of the studied laboratory, we conclude that the problem under consideration amounts to a new extension of the classical Resource-Constrained Project Scheduling Problem (RCPSP).

The RCPSP is a classical scheduling problem that allows the modeling of a broad spectrum of real-life situations. The problem consists in scheduling non-preemptive tasks on limited renewable resources. These tasks are linked together by precedence relationships (task i cannot start until task l is finished). Usually, the objective is to find a solution that minimizes the makespan of the project, while complying both the precedence constraints and the resource constraints.

Formally, the RCPSP can be defined by a 7-tuple (I, d, E, R, B, b, T) where I is a set of activities, d is a vector of activity durations, E is a set of precedence relationships, R is a set of resources, B is a vector of resource availability, b is a matrix of resource demands or consumptions per activity, and T is the set of scheduling periods (Artigues *et al.*, 2013).

Even if the classical version of the RCPSP is very expressive, it cannot cover all the situations that happen in real-life problems. That is why researchers have developed more general or extended versions of the RCPSP using the classical version as starting point. Surveys on this topic are proposed for example by (Hartmann and Briskorn, 2010) and (Orji and Wei, 2013). Among all these variants, we distinguish one that is of great interest for the modeling of the scheduling problem at hand: the Multi-Skill Project Scheduling Problem (MSPSP).

The MSPSP, presented by the first time in (Néron, 2002), combines characteristics of both the classical RCPSP for the project description, and the Multi-Purpose Machine model with the addition of new resource constraints. In this variant a resource is therefore characterized by the set of skills it possesses; and a task is now defined by the number of required resources with a specific competence.

This problem consists in determining a feasible schedule, respecting the precedence constraints between activities and the resource constraints: a resource cannot execute a skill it does not master, cannot be assigned to more than one competence requirement at a given time, and must be assigned to the corresponding activity during its whole processing time. The aim is to minimize the total duration of the project (Bellenguez-Morineau, 2008). Activities in the MSPSP are supposed to be non-preemptive, that means, once started an activity must run continuously until its completeness. The main characteristics of the MSPSP are showed in Figure 1.

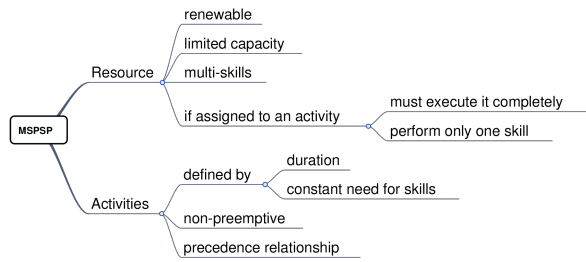


Figure 1 – Characteristics of the MSPSP

With the MSPSP as starting point, adapting some of its characteristics and relaxing the non-preemption constraint, we have developed a new variant that should better represent the real-life problem we are trying to solve: an MSPSP with partial preemption.

The remainder of the paper is as follows. In the next section, we describe the problem under consideration. In Section 3, we present the mixed-integer linear programming model representing the partially preemptive MSPSP. As a modeling alternative, a Constraint Programming model is also presented in Section 4. In Section 5, we discuss the computational experiments carried out. Finally, in Section 6 we conclude and discuss future research.

2 PROBLEM DESCRIPTION

When scheduling research or engineering activities, it may be interesting to allow the preemption, in order to reduce the makespan of the project, especially when resource availability is very limited. In our case, due to some safety and operational constraints, proper to nuclear regulation, we can only allow the preemption of a subset of activities. In (Polo Mejia *et al.*, 2017), we proposed several MILP formulations, for a first model of the considered nuclear research facility, where some activities are non-preemptive while some other activities are fully preemptive (but with a penalty every time the activity was preempted). This model only fulfilled partially the operational requirements. Indeed, when working with preemptive scheduling problem, commonly we assume that all resources are released during the preemption periods. However, for some research activities, we are interested in avoiding the release of some equipment or resource (for specific activities) having an important setup time.

That is why we propose in this paper to work with a variant allowing the partial release (partial preemption) of resources according to the characteristics of the activities. We must indicate for each activity what resource can be released during the preemption periods. Preemption is now handled in three levels

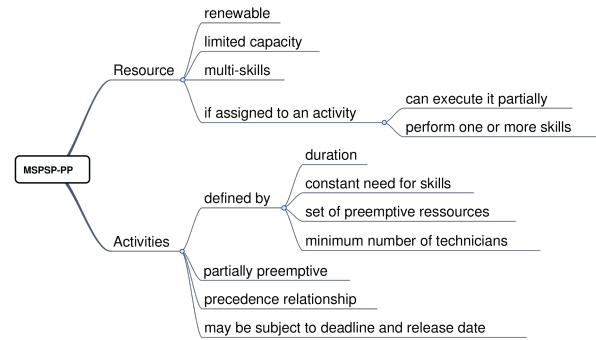


Figure 2 – Characteristics of the MSPSP-PP

according to the activities characteristics: 1) Non-preemption, for activities where none of the resources can be preempted; 2) Partial preemption, for activities where a subset of resources can be preempted; and 3) Full preemption, for activities where all resources can be preempted.

In order to have a better representation of reality, we need to do some changes over the technicians behavior. Unlike the traditional MSPSP, in our practical case, technicians may respond to more than one skill requirement per activity. Also, due to operational and safety reasons, we need to guarantee a minimum number of technicians present during the execution of the activity. Finally, due to the durations of some activities (larger than technicians' work shifts), we need to relax the constraint stating that the same technician must execute the totality of the activity.

Additionally, we must include some other characteristics to our problem concerning the time windows for scheduling. In the laboratory, the regulatory test must be executed before a restrictive date (deadline). Moreover, some of the activities are in collaboration with other nuclear facilities, such activities are then restricted by a release date fixed by external partners. A recap of our MSPSP with partial preemption (MSPSP-PP) is presented in Figure 2.

The complexity of the MSPSP with partial preemption can be established using the classical RCPSP as starting point. For each instance of the RCPSP we can match an instance of the MSPSP with partial preemption, where all resources are mono-skilled and none of the resources can be preempted. Thus, we can define the RCPSP as a particular case of the MSPSP with partial preemption. Since the RCPSP has been proved to be strongly NP-hard (Blazewicz *et al.*, 1983) we can therefore infer that the MSPSP with partial preemption is also strongly NP-hard.

Once defined the characteristics and the complexity of the proposed problem, we proceed to formalize

the problem using the two more common approaches in the literature: Integer Linear Programming (ILP) and Constraint Programming (CP). These models are presented in the following sections.

3 ILP MODEL

Usually, the variants of the RCPSP and the MSPSP can be modeled using different approaches such as: continuous time-based models based on flows, discrete-time mixed integer linear programming (MILP) formulations, or event-based MILP formulations. An analysis of these approaches allows us to identify the so-called on/off formulation as the most suitable for the preemptive case we are working on. This time-indexed formulation uses binary variables $Y_{i,t}$, where $Y_{i,t} = 1$ if activity i is in progress at time t and $Y_{i,t} = 0$ otherwise.

Using the on/off formulation as basis, we tested two models having as only difference the way in which we modeled the preemption periods. After preliminary computational experiments, one of the models showed significantly better results, and it is presented below:

3.1 Variables

- $Y_{i,t} \in \{0,1\}$, $Y_{i,t} = 1 \iff$ activity i is in progress at time t
- $O_{j,i,t} \in \{0,1\}$, $O_{j,i,t} = 1 \iff$ technician j is allocated to activity i at time t
- $Z_{i,t} \in \{0,1\}$, $Z_{i,t} = 1 \iff$ activity i starts at time t or before
- $W_{i,t} \in \{0,1\}$, $W_{i,t} = 1 \iff$ activity i ends at time t or after
- $Pp_{i,t} \in \{0,1\}$, $Pp_{i,t} = 1 \iff$ activity i is preempted at time t
- $End_i \in \mathbb{Z}_+$: Completion time of activity i
- $Cmax \in \mathbb{Z}_+$: Project makespan

3.2 Objective function

The most common objective function found in the literature for the RCPSP and the MSPSP is to minimize the project makespan:

$$\min(Cmax) \quad (1)$$

However, as we are working in a research facility, we have the interest to assure that all activities are completed as soon as possible. We can translate this as the minimization of the average completion time for all activities:

$$\min \sum_{i \in I} End_i \quad (2)$$

Using an aggregate objective function may have a negative impact on the solving time of the problem. That is why, we decide to test both objective functions and evaluate the impact of this change.

3.3 Constraints

$$\sum_i O_{j,i,t} \leq DO_{j,t} \quad \forall j, \forall t \quad (3)$$

$$\sum_i ((Y_{i,t} + PR_{i,k} * Pp_{i,t}) * Br_{i,k}) \leq DR_{i,k} \quad \forall t, \forall k \quad (4)$$

$$(Y_{i,t} + Pc_i * Pp_{i,t}) * Bc_{i,c} \leq \sum_j (O_{j,i,t} * CO_{j,c}) \quad \forall i, \forall t, \forall c \quad (5)$$

$$\sum_j O_{j,i,t} \geq (Y_{i,t} + Pc_i * Pp_{i,t}) * Nt_i \quad \forall t, \forall i \quad (6)$$

$$\sum_t Y_{i,t} \geq D_i \quad \forall i \quad (7)$$

$$D_l * (1 - Y_{i,t}) \geq \sum_{t'=t}^T Y_{i,t'} \quad \forall (i,l) \in E, \forall t \quad (8)$$

$$\sum_{t=dl+1}^T Y_{i,t} \leq 0 \quad \forall i \quad (9)$$

$$\sum_{t=1}^{r_i-1} Y_{i,t} \leq 0 \quad \forall i \quad (10)$$

$$Pp_{i,t} = Z_{i,t} + W_{i,t} - Y_{i,t} - 1 \quad \forall i, \forall t \quad (11)$$

$$Z_{i,t} \geq Y_{i,t'} \quad \forall i, \forall t, \forall t' \leq t \quad (12)$$

$$W_{i,t} \geq Y_{i,t'} \quad \forall i, \forall t, \forall t' \geq t \quad (13)$$

$$Z_{i,t} \leq \sum_{t'=1}^t Y_{i,t'} \quad \forall t \quad (14)$$

$$W_{i,t} \leq \sum_{t'=t}^T Y_{i,t'} \quad \forall t \quad (15)$$

$$End_i \geq t * Y_{i,t} \quad \forall t, \forall i \quad (16)$$

$$Cmax \geq End_i \quad \forall i \quad (17)$$

Equations (3) ensure that operator's capacities ($DO_{j,t}$) are satisfied. In equations (4), we ensure that all resource requirements ($Br_{i,k}$) are satisfied respecting the resource capacities ($DR_{k,t}$). Parameter $PR_{i,k}$ indicates whether the resource k can be preempted ($PR_{i,k}=0$) or not ($PR_{i,k}=1$). Equations (5) ensure the respect of skill requirements ($Bc_{i,c}$) taking into account the set of skills of technicians ($CO_{j,c}$; $CO_{j,c} = 1$ if technician j has the competence c , 0 otherwise). Parameter Pc_i indicates whether technicians can be preempted ($Pc_i=0$) or not ($Pc_i=1$). The constraints given in (6) and (7) ensure the respect of the minimum number of technicians (Nt_i)

and duration of activities (D_i), respectively. Precedence constraints are given in (8). Inequalities (9) and (10) are the constraints for deadlines (dl_i) and release dates (r_i). Equations (11) determine whether an activity is preempted or not. Inequalities (12) to (15) are constraints for getting the values of variables $Z_{i,t}$ and $W_{i,t}$. Equations (16) allows to calculate the completion time of each activity. Finally, inequalities (16) calculate the makespan of the project. Note that it is not necessary to declare variables $Y_{i,t}$, $Z_{i,t}$, End_i and $Cmax$ as integer variables since an integer optimal solution for the other variables enforces integrity on these variables. However, the modern branch-and-bound procedures inside MILP solvers may be able to exploit the knowledge of the integrity of these variables via cuts and preprocessing techniques to converge faster.

4 CP MODEL

In the last decades constraint programming has attracted high attention among experts from many areas of computer science due to its potential for solving hard real life problems. The main idea of constraint programming is to solve problems by stating constraints (requirements) about the problem area and, consequently, finding solutions satisfying all the constraints. The increasing interest for this technique led us to use it and evaluate its performance over the studied problem. Our first CP model for the MSPSP with partial preemption is presented below; further research needs to be done in order to improve its performance. We use the IBM CP Optimizer (CPO) as software to model and solve this problem. The presented model refers to the concept of interval variables, a constrained object tailored to scheduling problem, and also to other specific scheduling constraints in CPO.

4.1 Variables

- $itvs_i$: Interval variable between the start and the end of activity i
- $par_{i,p}$: Optional ¹ interval variable indicating the start and the end of every possible part p , $p \in \{1, 2, \dots, D_i\}$, of activity i
- $InTech_{j,i,p}$: Optional interval variable indicating the periods when technician j is working in the part p of activity i

4.2 Objective function

As mentioned before, we want to study the impact of using an aggregate objective function. In our CP

¹Optional interval variables may or may not be present in the solution, so as to satisfy the constraints.

model the objective functions are calculated as follows:

Minimize the project makespan:

$$\min(\max_{i \in I} itvs_i.end) \quad (18)$$

Minimize the average completion time of activities:

$$\min \sum_{i \in I} itvs_i.end \quad (19)$$

4.3 Constraints

Span(a, {b1, ..., bn}) constraint: states that the interval variable a (if present) spans over all present interval variables from the set $\{b1, \dots, bn\}$. In other words, interval variable a starts together with the first present interval from $\{b1, \dots, bn\}$ and ends together with the last present interval (Laborie, 2009). We use this kind of constraint to span the $par_{i,p}$ and $InTech_{j,i,p}$ variable within the $itvs_i$ variables.

Span all the parts p of the activity i within the execution interval of variable i :

$$span(itvs_i, par_{i,p} : \forall p) \forall i \quad (20)$$

Span allocation intervals of technicians ($InTech_{j,i,p}$) within the variables $itvs_i$. With constraints (20) this will allow only the necessary technician assignment interval variables to be present in the solution. These constraints stand over the hypotheses that every activity requires at least 1 technician for its processing:

$$span(itvs_i, InTech_{j,i,p} : \forall j, \forall p) \forall i \quad (21)$$

noOverlap({b1, ..., bn}) constraint: states that none of the interval variables within the set $\{b1, \dots, bn\}$ overlap over the time. We use this variable to ensure a disjunctive constraint over the technicians:

$$noOverlap(InTech_{j,i,p} : \forall i, \forall p) \forall j \quad (22)$$

Ensure the activity duration: the *sizeOf* expression is used to access the size (duration) of an interval variable:

$$D_i = \sum_p sizeOf(par_{i,p}) \forall i \quad (23)$$

Respect the deadlines and release dates:

$$dl_i \leq itvs_i.end \quad (24)$$

$$r_i \leq itvs_i.start \quad (25)$$

Precedence relationships:

$$itvs_i.end < itvs_l.start \forall (i, l) \in E \quad (26)$$

Let us define $rUsage_k$ as a cumulative function indicating the usage of resource k over the time. Also let $pulse(F, h)$ be an elementary pulse function taking the value of h over interval F . We can state the resource constraint as follows:

$$rUsage_k = \sum_{i \in I: PR_{i,k}=0} \sum_p pulse(par_{p,i}, Br_{i,k}) + \sum_{i \in I: PR_{i,k}=1} pulse(itvs_i, Br_{i,k}) \quad \forall k$$

$$rUsage_k \leq DR_k \quad \forall k \quad (27)$$

For ensuring the skill requirements, we define $SkAll_{i,c}$ as a cumulative function indicating the allocation of skill c for the activity i over the time.

$$SkAll_{i,c} = \sum_j \sum_p pulse(InTech_{j,i,p}, CO_{j,c}) \quad \forall i, \forall c$$

The $alwaysIn(F, B, min, max)$ constraint is used to confine the values of a cumulative function F during an interval B inside interval $[min, max]$. We can then define the skill constraints for preemptive technicians as follows:

$$alwaysIn(SkAll_{i,c}, par_{i,p}, Bc_{i,c}, \infty) \quad \forall i : Pc_i = 0, \forall c, \forall p \quad (28)$$

If technicians are declared as non-preemptive the constraint is applied to the interval variable of the activity rather than on its parts:

$$alwaysIn(SkAll_{i,c}, itvs_i, Bc_{i,c}, \infty) \quad \forall i : Pc_i = 1, \forall c \quad (29)$$

We must add the constraints on the minimum number of technicians required for executing an activity. In order to do that, we use a cumulative function $AllTech_i$ indicating the number of technicians allocated for an activity over the time. This function is calculated as follows:

$$AllTech_i = \sum_j \sum_p pulse(InTech_{j,i,p}, 1) \quad \forall i$$

The constraints are then defined for preemptive technicians as follows:

$$alwaysIn(AllTech_i, par_{i,p}, Nt_i, \infty) \quad \forall i : Pc_i = 0, \forall p \quad (30)$$

For non-preemptive technicians we have:

$$alwaysIn(AllTech_i, itvs_i, Nt_i, \infty) \quad \forall i : Pc_i = 0 \quad (31)$$

We also need to ensure that there is not overlapping of every activities part variable. However, in order to improve our model we decided not to use the $noOverlap$ expression we presented before and state these constraints as follows, so as to break symmetries:

$$par_{i,p}.end < par_{i,s}.start \quad \forall i, \forall p, \forall s \in Parts : s > p$$

(32)

$$presenceOf(par_{i,p}) \implies presenceOf(par_{i,s}) \quad (33)$$

$$\forall i, \forall p, \forall s \in Parts : s > p$$

$presenceOf()$ is a boolean constraint, which is true when the interval variable is present. Together with Constraints (20) and (23), these constraints ensure that only the necessary part interval variables are present.

Technicians cannot be assigned during their absence periods. To model these constraints, we define a step function describing the present and absent periods of each technician ($PreTech_j$). We must also use the predefined constraint $forbidExtent(a, F)$ that states that whenever the interval variable a is present, it cannot overlap a point t where the step function $F(t) = 0$. We can define these constraints as follows:

$$forbidExtent(InTech_{j,i,p}, PreTech_j) \quad \forall j, \forall i, \forall p \quad (34)$$

Finally, in order to improve the model we added a synchronization constraint between the $InTech_{j,i,p}$ and the $par_{i,p}$ variables for all the activities where technicians can be preempted. The constraint $synchronize(a, \{b_1, \dots, b_n\})$ makes intervals $\{b_1, \dots, b_n\}$ start and end together with interval variable a (if a is present).

$$synchronize(par_{i,p}, InTech_{j,i,p} : \forall j) \quad \forall i : Pc_i = 0, \forall p \quad (35)$$

5 EXPERIMENTAL RESULTS AND DISCUSSION

Computational tests have been carried out using the solver CPLEX 12.7 for the ILP model and CP Optimizer 12.7 for the CP model. We generated our sets of instances using a basic instance generation algorithm that allows the control of certain aspects such as: proportions of preemption type, percentage of activities with deadline and release date, number of precedence relationships, skill number per technicians, etc.

The proposed ILP model requires that the user indicates the initial scheduling horizon (an estimation of project makespan, T) in order to initialize the decision variables. Since we are using a “time-indexed” formulation, this estimation may have an important role in the performance of the model. That is why, the first thing to want to analyze is the behavior of the solving time when the makespan estimation (T) change. We tested the ILP model using the objective function (1) (ILP1). A small set of 7 dummy instances have been generated with a makespan lower than 30 time units (this instance have 20 activities with duration between 1 and 5 time units, 4 precedence relationship, 13 skills, all other characteristics are random).

	ILP1 T=30	ILP1 T=50
Instance 1	15.05 sec	60.77 sec
Instance 2	2.05 sec	36.59 sec
Instance 3	15.23 sec	69.93 sec
Instance 4	23.07 sec	138.61 sec
Instance 5	27.23 sec	105.32 sec
Instance 6	36.71 sec	171.65 sec
Instance 7	1403.31 sec	228.98 sec

Table 1 – Time to solve to optimality

As shown in Table 1, when the initial estimation gets far from the real value, the time needed to solve the problem increases significantly. This can be explained by the fact that when T increases, the ILP model will need to handle more decision variables to solve to optimality the same instance.

	ILP2 T=30	ILP2 T=50
Instance 1	15.00%	36.00%
Instance 2	13.35%	25.68%
Instance 3	13.93%	55.96%
Instance 4	17.85%	58.39%
Instance 5	11.00%	51.42%
Instance 6	13.56%	33.38%
Instance 7	13.84%	50.35%

Table 2 – Gap after 30 min of computation

We did the same test using the objective function (2) (ILP2). Results are shown in Table 2. Since the time required to get the optimal solution is bigger than the fixed time limit, we present the final gap (percentage of deviation between the best feasible found solution and the optimal solution) after 30 minutes of computing. As it can be seen from Table 2, this gap increases with T . This shows us the importance of having a good estimation of the project makespan in order to improve the performance of the proposed ILP model. In the following of this research project, we must develop heuristic methods allowing us to have a good estimation of the project makespan in order to reduce the required time to solve to optimality the instances.

Another observation we can do from the results in Table 1 and Table 2, is that the use of an aggregate objective function has a negative impact in the computation time needed to solve the instances. This may be due to the fact that they can exist a big number of feasible solutions having different values for the aggregate function but with the same makespan. So, when branching the solver cannot cut these branches as fast as it would do it for a makespan optimization.

Another possible reason is the quality of upper/lower bounds we can get for an aggregate objective function.

In order to compare the performance of the CP model against the ILP model, we generate a bigger set of 95 heterogeneous instances (15 activities, duration between 1 to 10 time units, 15 skills, 8 resources, 20% of activities with release date and deadline, all other characteristics are random). We solved these instances for the two configurations of the ILP and CP models (CP1 for objective function (18), CP2 for objective function (19)).

		Solved to opt.	Solved not opt.	No able to find a first solution
I L P 1	Number	58	32	5
	Mean GAP	-	45.86%	-
	Standard deviation	-	28.93%	-
C P 1	Number	1	61	33
	Mean GAP	-	18.63%	-
	Standard deviation	-	23.41%	-

Table 3 – Results configuration 1 (Time lim: 15 min)

Results in Table 3 show that the proposed ILP model outperforms the CP model for configuration 1 (minimization of makespan). The ILP1 was able to solve to optimality a large number of instances (58 out of 95) within the limited time while the CP1 only was able to demonstrate the optimality of 1 instance. Additionally, the number of not solved instances is bigger for the CP. In the following of the project, we must try to improve the performance of the CP model adding or modifying some constraints to break symmetries and reduce the search space.

	Solved to opt.	Solved not opt.	Better value	No first solution
Number (ILP2)	0	78	40	17
Number (CP2)	0	59	19	36

Table 4 – Results configuration 2 (Time lim:15 min)

Table 4 presents the obtained results for configuration 2 (minimization of end times). These results allow us to confirm what we said before, the use of an aggregate objective function has a negative impact in the computation time required to solve the problem (a lower number of optimally solved instances within

the limit time). The ILP2 seems to outperform the CP2 being able to get values for a bigger number of instances within the limited time and a lower number of not solved instances. Additionally, ILP2 gets more better solution than the CP2 for the solved instances. Again, we must remember that the proposed CP model is our first version and it should require some improvements.

A deep analysis of individual results of this test suggested us a difference of performance of the models when the portion of preemptive, semi-preemptive and non-preemptive activities changes. In order to corroborate this, we have generated 4 sets (A, B, C and D) of 40 instances. For each instance in a set, there is a similar instance in the others having as only difference the distribution of the preemption type of activities (this distribution its presented in Table 5). These instances have 15 activities with duration between 1 to 10 time units, 15 skills, 8 resources, 20% of activities with release date and deadline, all other characteristics are random.

	A	B	C	D
Non-preemptive	10%	10%	80%	33.3%
Partially preemptive	10%	80%	10%	33.3%
Preemptive	80%	10%	10%	33.3%

Table 5 – Distribution of preemption type

We solved the 4 sets of instances to optimality using the configuration 1 of the ILP model. Using a t-test for two paired samples with significance level of 0.05 (Derrick *et al.*, 2017), we try to identify if there is enough statistical evidence to claim a difference of performance (time to solve) according to the distribution of the preemption type. The test results, showed in Table 6, led us to conclude that the ILP model give faster answer when the fraction of preemptive activities increases. For the other kind of preemption, we do not evidence any special impact.

	B	C	D
A	time A < time B p.val= 0.0017	time A < time C p.val=0.0002	Statistically equal p.val=0.0595
B	-	Statistically equal p.val=0.1636	Statistically equal p.val=0.1491
C	-	-	time C > time D p.val=0.0073

Table 6 – T-test for difference in solving time - ILP1

In order to test the impact of the type of preemption over the CP model, we try to solve the same sets of instances using the CP1 configuration. However, the CP1 model was not able to solve to optimality any of the instance within the time limit (30 min). It was not even able to found initial solutions for a large number of instances. This fact made impossible for us to do an analysis similar to the one done for the ILP1. However, we decided to analyze the number of instances for which the model CP1 was able to found at least an initial solution. Results in Table 7 may confirm the same behavior observed for the ILP1: the CP1 model goes faster when it works over highly preemptive instances.

	A	B	C	D
Number of Instances	50	21	7	26

Table 7 – Solved instances per preemption type - CP1

One of the bigger advantages of the MSPSP-PP is the possibility of better handle the partial preemption we can find in real-life problems. When using a traditional preemptive/non-preemptive approach, semi-preemptive activities are usually modeled as non-preemptive. In order to prove the convenience of using a partial preemption approach for reducing the makespan, we decide to analyze the optimal values obtained for the 4 sets of instances.

As expected, the set with most of preemptive activities (set A) allows us to get lower optimal values than the other sets. When comparing the set having a big portion of non-preemptive activities (C) against the one having a big portion of semi-preemptive activities (B), we appreciate that for set B we get lower values than for set C. This confirms the fact that using a semi-preemptive approach may reduce the makespan of the project. With these results we can establish the following relation for makespan value: Preemptive ≤ Semi-preemptive ≤ Non-preemptive. This simple relation may help us to calculate the lower/upper bounds we need to develop exact solving algorithms.

The main objective of this project is to schedule the weekly activities of the nuclear laboratory. The first approach is to generate scheduling using the hour as time unit. In a normal week the number of working hours is 108, what give us an idea of the expected makespan. Usually, we must schedule between 40 and 50 activities with a duration between 4 and 20 hours. There is at least 20 technicians, more than 25 resources and at least 20 skills. When trying to solve instances having these characteristics, the models were not able to find optimal solutions after 2 hours of computation, getting an optimality gap go-

ing from 20% to 50% for the ILP1. These results lead us to start the development of heuristics methods for solving large instances in reasonable times.

6 CONCLUSIONS

The literature about scheduling research activities is very sparse. That is why, in this paper we presented a new variant of the Multi-skill Project Scheduling Problem (MSPSP) developed to schedule the research activities in a nuclear laboratory: MSPSP with partial preemption. The proposed problem can be easily adapted to a huge number of real-life situations for scheduling activities within very restrictive environments. After describing all the characteristics of the proposed problem, we modeled it using two different techniques: Integer Linear Programming and Constraint Programming.

The first experimental results showed the importance of the initial estimation of the project makespan for the ILP model performance. We also evidenced that using aggregate objective functions has a negative impact on the computation time required to solve the ILP and CP models. Finally, we observed that the ILP model outperformed the CP model for almost all instances and all model configurations, at least for our current CP modeling, which could certainly be improved.

We can conclude that the use of a partially-preemptive approach may lead to reduction of the makespan of the project when compared again a traditional preemptive/non-preemptive approach. Additionally, the proportion of different kinds of preemptions seems to have an impact in the performance of the ILP and CP models. Indeed, the models are faster when the portion of preemptive activities is high.

In practice, the MSPSP with partial preemption may be easily adapted to many real-life situations, especially for scheduling activities within very controlled environments such as pharmaceutical, chemical and aeronautical. For the studied laboratory, the application of the proposed model gives to the engineers the ability to have a more performant schedule in a matter of minutes, at the same time that it eliminates the risk of forgetting constraints, which can happen easily in the manual scheduling process.

As future work, we must study the ways of improvements of the proposed CP model such as adding or modifying some constraints to break symmetries and reduce the search space. We also need to develop heuristics allowing us to have good solutions in reasonable times. The solving times obtained in the experimental tests lead us to develop dedicated algorithms for exact solving. In order to do that we will need to study different approaches for calculat-

ing good lower bounds.

REFERENCES

- Artigues C., Demassey S. and Néron E., 2013. The Resource-Constrained Project Scheduling Problem. *Resource-constrained project scheduling: models, algorithms, extensions and applications*, John Wiley & Sons.
- Bellenguez-Morineau O., 2008. Methods to solve multi-skill project scheduling problem. *4OR*, 6(1), pp. 85-88.
- Blazewicz J., Lenstra J.K. and Rinnooy Kan A.H.G., 1983. Scheduling projects subject to resource constraints: classification and complexity. *Discrete Applied Mathematics*, 5, pp. 11-24.
- Derrick B., Toher D and White P., 2017. How to compare the means of two samples that include paired observations and independent observations. *Tutorials in Quantitative Methods for Psychology, Vol 13, Iss 2, Pp 120-126 (2017)*, 13(2), pp. 120-126.
- Hartmann S. and Briskorn D., 2010. A survey of variants and extensions of the resource-constrained project scheduling problem. *European Journal of Operational Research*, 207(1), pp. 1-14.
- Laborie P., 2009. IBM ILOG CP Optimizer for detailed scheduling illustrated on three problems. *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, Springer, Berlin, Heidelberg, pp. 148-162.
- Montoya C., Bellenguez-Morineau O., Pinson E. and Rivreau D., 2015. Integrated column generation and lagrangian relaxation approach for the multi-skill project scheduling problem. *Handbook on Project Management and Scheduling*, Springer International, pp. 565-586.
- Néron E., 2002. Lower bounds for the multi-skill project scheduling problem. *Proceedings of the Eighth International Workshop on Project Management and Scheduling*, pp. 274-277, Valencia, Spain.
- Orji I. M. J. and Wei S., 2013. Project scheduling under resource constraints: A recent survey. *International Journal of Engineering Research and Technology*, 2(2), pp. 1-20.
- Polo Mejia O., Anselmet M.-C., Artigues C. and Lopez P., 2017. A new RCPSP variant for scheduling research activities in a nuclear laboratory. *47th International Conference on Computers & Industrial Engineering (CIE47)*, Lisbon, Portugal.

MODÉLISATION D'UNE CHAÎNE LOGISTIQUE HYBRIDE PAR LA SIMULATION À ÉVÉNEMENTS DISCRETS

Najat Bara¹, Frédéric Gautier^{1,2}

Vincent Giard^{1,3}

¹EMINES School of Industrial Management, University Mohammed VI – Polytechnique, Ben Guerir, Morocco

²IAE de Paris, Université Paris I Panthéon-Sorbonne Paris, France

najat.bara@emines.um6p.ma
frederic.gautier@emines.um6p.ma,
frederic.gautier@univ-paris1.fr

³Université Paris-Dauphine, PSL Research University, Paris, France

Vincent.Giard@emines.um6p.ma,
Vincent.Giard@dauphine.fr

RÉSUMÉ : *La simulation permet d'évaluer les conséquences des décisions opérationnelles dans une chaîne logistique complexe. Une chaîne logistique hybride (CLH) enchaînée, sans restriction, des processus continus (PC) et/ou discrets (PD). Aucun logiciel de simulation à événements discrets (SED) ou de la dynamique des systèmes (DS) ne permet à lui seul de simuler le fonctionnement d'une CLH au niveau de granularité voulu, restreignant l'usage de la simulation ou conduisant à des mécanismes complexes de différentes approches. La prise en charge de PC par la SED implique la création d'items par une discrétisation périodique de flux émis par certains PC de la CLH. Ces items « continus » (IC) peuvent constituer des inputs d'autres PC et être stockés, ce qui pose plusieurs problèmes : les caractéristiques d'IC sortant d'un PC diffèrent souvent de celles requises par le PC qui l'utilise comme input ; la capacité d'un stock ne peut être définie en un nombre maximal d'IC si ceux-ci ne sont pas homogènes ; il faut permettre la circulation, régulée par des vannes, d'IC entre stocks. Nous proposons ici les bases d'un ensemble de six primitives obtenues par combinaison de composants de base d'un SED permettant de modéliser une CLH. Un cas réel complexe d'utilisation de ces primitives est présenté.*

MOTS-CLÉS : *chaîne logistique hybride, simulation à événements discrets, flux poussés, flux tirés.*

1 INTRODUCTION

La définition des processus élémentaires discrets et continus (§1.1) est un préalable à celle des chaînes logistiques discrètes, continues et hybrides (§1.2). On s'intéresse ici à l'évaluation des conséquences de décisions de pilotage opérationnel d'une chaîne logistique (§1.3). L'approche retenue est celle de la modélisation/ simulation de chaînes logistiques hybrides (§1.4), ce qui nous conduira à présenter l'organisation de notre article (§1.5).

1.1 Processus élémentaires discrets et continus

Un *processeur* discret a pour vocation de traiter (transformations physiques ou spatiale (transport)) des items correspondant à des objets (produits, composants...), que l'on qualifie ici d'*items*. Un processus élémentaire discret (PED) est constitué d'un processeur discret ; il est alimenté par un ou plusieurs stocks, dans des quantités régies par des nomenclatures et, après un traitement, dont la durée est déterminée par une gamme opératoire, il expédie un ou plusieurs items dans un ou plusieurs stocks. La définition du processeur dépend du niveau d'observation retenu et de son usage dans une modélisation.

De la même façon, un processus élémentaire continu (PEC) a pour vocation de transformer des flux (liquides, granulés...) dans un processeur continu qui prélève ces

flux d'un ou plusieurs stocks et émet des flux en sortie, vers un ou plusieurs stocks. Les débits de ces flux d'entrée et de sortie sont liés par des proportions correspondant à une nomenclature, qui dépendent des performances du processeur. Sous certaines conditions, on peut calculer analytiquement le temps moyen de séjour des flux entrants dans le processus de transformation. Cette durée s'apparente à un temps opératoire.

1.2 Chaînes logistiques discrètes, continues et hybrides

On note {PED} un enchaînement de PED, le lien s'opérant habituellement par des stocks partagés (un stock de sortie d'un PED constituant un stock d'entrée du PED suivant). De la même façon, on note {PEC} un enchaînement de PEC, liés par des stocks. Le graphe correspondant à ces enchaînements peut être quelconque. Il s'ensuit qu'une chaîne logistique discrète (CLD) relève de la configuration {PED} en y ajoutant un ou plusieurs points d'entrée {PE} pour faire arriver des items dans des stocks qui seront traités dans la CLD, ainsi que un ou plusieurs points de sortie {PS} par lesquels sortent les items après leur dernier traitement. La CLD se définit donc par la séquence {PE} → {PED} → {PS}.

De la même façon, une chaîne logistique continue (CLC) relève de la configuration {PEC} en y adjoignant des

points d'entrée {PE} et de sortie {PS}. Une CLC se définit donc par la séquence {PE} → {PEC} → {PS}.

Une chaîne logistique hybride (CLH) combine {PED}, {PEC}, {PE} et {PC}. Par la suite, on sera amené à distinguer deux types de CLH :

CLH Type 1 {PE} → {PEC} → {PED} → {PS}

CLH Type 2 {PE} → {PEC} → {PED} → {PS}

1.3 Décisions de pilotage des chaînes logistiques

La typologie classique (Anthony, 1965) distingue les décisions stratégiques tactiques et opérationnelles ; par la suite, on y adjoint les décisions de type temps réel. On s'intéresse ici aux décisions opérationnelles. Ceci implique, d'une part, l'usage d'une maille temporelle allant de plusieurs minutes à quelques heures et un horizon pouvant aller jusqu'à quelques semaines et, d'autre part, de travailler avec une maille spatiale privilégiant les échanges entre des machines (ou groupe de machines) reliés par des stocks, sans se préoccuper de ce qui se passe dans ces machines.

Cette orientation décisionnelle implique de tirer les conséquences de la distinction classique entre productions à flux tirés et flux poussés dans la modélisation / simulation d'une CLH.

- Dans la production à flux tirés, la production se fait à la commande (carnet de commandes, kanbans...). Les flux d'informations (état du système, décisions programmées...) pilotent les flux physiques. La modélisation / simulation doit synchroniser des flux informationnels et des flux physiques et prendre en compte la disponibilité prévisionnelle des ressources (décision initiale).

- Dans la production à flux poussés, les processeurs produisent tant qu'ils disposent de l'ensemble requis des matières (composants ou fluides ou granulés) et que les stocks de destination ne sont pas saturés. Les flux circulent alors dans le système productif sans que cela ne relève de décisions explicites. Dans ce cadre, les décisions portent principalement sur la disponibilité prévisionnelle des ressources productives. Certaines décisions programmables peuvent être intégrées dans le modèle de simulation mais elles relèvent d'une logique cybernétique.

Les CLH auxquelles on s'intéresse ici combinent pilotages opérationnels par flux poussés, en amont de la CLH, et par flux tirés, en aval. L'un des objectifs de la simulation est de vérifier la cohérence d'un ensemble de décisions prises à différents endroits de la CLH et de faciliter la comparaison de scénarios décisionnels en termes d'efficacité et d'efficacité.

1.4 Modélisation/ simulation des chaînes logistiques discrètes, continues et hybrides et éclairage de la prise de décisions opérationnelles

Le problème auquel on s'intéresse ici est celui de la modélisation simulation d'une CLH utilisée pour éclairer la prise de décisions opérationnelles ou tactiques.

La modélisation / simulation des CLD et CLC dispose d'instrumentations éprouvées : en particulier, la simulation à événements discrets (SED) pour une CLD et la dynamique des systèmes (DS) pour une CLC. Plusieurs approches sont envisageables pour la modélisation / simulation de la CLH (on reviendra sur ce point en section 2).

- La première repose sur un mécano d'instrumentations prenant chacune en charge un {PED} ou un {PEC}, dans le cadre d'une approche intégrée gérant les échanges d'informations entre ces instrumentations.

- La seconde, privilégiée ici, consiste à mobiliser, en l'adaptant si nécessaire, l'une des instrumentations dédiées (SED ou DS). On peut souligner dès à présent que la granularité spatio-temporelle qu'implique la modélisation / simulation du pilotage d'une CLH par un ensemble de décisions opérationnelles rend très difficile l'utilisation de la DS pour décrire des processus discrets.

L'utilisation directe de la SED pour une modélisation/simulation d'une CLH de type 2 n'est pas possible pour deux séries de raisons.

- La capacité d'un stock dans une SED est définie en nombre d'items, ce qui n'est utilisable pour un fluide que si les caractéristiques volumiques ou pondérales de tous les IC qu'il contient sont identiques, ce qui n'est le cas que dans des exemples simplistes. Ensuite, dans les processus continus, les stocks peuvent être directement connectés et s'alimenter mutuellement, les échanges étant contrôlés par des vannes. Les primitives de base des SED ne permettent pas ce type d'échange.

- Dans l'enchaînement de deux PEC (séparés par un stock), les caractéristiques volumiques de l'IC produite par le PEC-amont a peu de raison de correspondre à celles requises par le PEC-aval. Il faut alors disposer d'un mécanisme effectuant la transformation requise, tout en garantissant la conservation de la matière

1.5 Organisation de la communication

On commencera par une revue de la littérature (section 2) qui permettra de montrer qu'il n'existe pas de travaux significatifs proposant de modéliser une CLH de type 2, par la SED, afin d'évaluer des décisions opérationnelles. Dans une section 3, nous proposerons un ensemble de *primitives* combinant les composants de base des SED, pour être en mesure d'effectuer une modélisation simulation d'une CLH, au niveau de granularité pertinent pour l'analyse de décisions opérationnelles. Enfin dans une section 4, nous présenterons un exemple réel d'une simulation d'un CLH complexe, mobilisant la SED et les composants proposés ici. Nous terminerons par une rapide conclusion en section 5.

2 REVUE DE LA LITTÉRATURE

Notre objectif de cette revue de littérature est de présenter les raisons du choix de modéliser une CLH par la SED, et de faire un état des lieux des techniques déjà utilisées ; permettant la modélisation des processus hybrides par la SED. Dans plusieurs travaux de recherche, le choix du type de la simulation est souvent basé sur le niveau déci-

sionnel étudié (niveau opérationnel, tactique ou stratégique) (Tako et Robinson, 2012), (Jahangirian *et al.*, 2010). Dans ce présent article, en plus du niveau de décision, nous nous intéressons à étudier le choix de la technique de simulation (SED, DS, SH) vis-à-vis la nature ; discrète ou continue ; des processus composant la CL étudiée.

2.1 Méthodologie de la revue de littérature

Aucune restriction n'a été mise dans la recherche des articles sur l'année, la conférence ou la revue. Nous avons essayé les combinaisons suivantes de mots clés : « Continuous process », « discrete process », « discretization », « operational management », « supply chain », « System dynamics » et « Discrete event simulation ». Les questions que nous nous posons sont les suivantes :

- La DS et les SED ne sont-ils utilisés respectivement que pour une CLC et une CLD ?
- Quel lien existe-t-il entre le modèle de simulation et le niveau décisionnel à éclairer ?
- Si la SED est utilisée pour une CLH, quelles sont les techniques de discrétisation mobilisées ?

2.2 Etat des lieux

J. Forrester, fondateur de la DS, définit cette méthodologie comme une approche de modélisation des politiques de l'entreprise et comme un outil aidant à résoudre des problématiques du top management. Elle modélise le pilotage de l'évolution des systèmes dans lesquels des flux circulent et leurs évolutions sont représentées par des équations différentielles. La DS modélise la CL à un niveau agrégé et permettent d'avoir une vision macroscopique du fonctionnement du système étudié. (Scholz-Reiter *et al.*, 2005), (Kleijnen, 2005), (Venkateswaran et Son, 2004). Elle a été utilisée pour étudier l'impact du phénomène du *Bullwhip Effect*. (Pierreval *et al.*, 2007), (Higuchi *et al.*, 2004), (Rozman *et al.*, 2014)... La DS est également appropriée pour modéliser l'évolution en temps réel des processus continus et d'en faciliter le pilotage ; dans ce cadre, elle permet de comprendre l'évolution de paramètres physiques (température, pression...) à garder sous contrôle.

La SED est utilisée pour étudier le comportement des systèmes en réponse à certains événements (franchissement de seuils, décisions de pilotage...) survenant à des moments précis au cours de la simulation (Venkateswaran et Son, 2004). La SED est plus appropriée pour l'étude des décisions opérationnelles et tactiques (Andersson *et al.*, 1998). Elle permet une représentation du système productif à une granularité (spatiale, temporelle...) pertinente pour évaluer l'impact de décisions de pilotage.

Nous distinguons trois conceptions possibles de la simulation hybride :

- Deux modèles différents de simulation sont mobilisés pour la modélisation de deux niveaux décisionnels différents. Généralement un premier modèle basé sur

la DS est développé pour répondre à des problématiques stratégiques à un niveau agrégé (planning annuel de production, politique d'externalisation ...). Par la suite un modèle de SED est utilisé pour simuler le comportement des décisions opérationnelles intégrées dans la solution obtenue par le premier modèle (Jamalnia *et al.*, 2013), (Rabelo *et al.*, 2007), (Venkateswaran and Son, 2004)... Cette approche se retrouve dans d'autres combinaisons d'instrumentations couplant des décisions stratégiques et opérationnelles ou tactiques. Il est très fréquent d'utiliser la programmation mathématique au niveau stratégique pour définir des niveaux de ressources permanentes à affecter à des besoins ; la SED est ensuite utilisée pour vérifier la viabilité de ces décisions stratégiques à un niveau fin, pouvant prendre en compte des perturbations aléatoires. La rapidité de l'obtention de la solution finale dépend du couplage réalisé entre les deux modèles. Dans un premier cas où les résultats du premier modèle sont automatiquement utilisés par le deuxième modèle par un interfaçage direct, le couplage est fort. Il est faible dans le cas contraire, ce qui implique plus de temps pour prendre une décision satisfaisante.

- La seconde conception est souvent qualifiée de simulation combinée. Elle correspond à une modélisation intégrée incorporant simultanément des variables discrètes et continues, dans un même modèle (Lee *et al.*, 2002). Certains auteurs (Barbosa et Azevedo, 2017) qualifient d'enrichissement ce type de SH qui implique l'usage d'un logiciel fondé sur l'une des deux approches de simulation, avec une mobilisation de méthodes caractérisant l'autre approche. Par exemple, dans (Music et Matko, 1999) et (Fabre *et al.*, 2011), la DS est utilisée pour le suivi des paramètres physicochimiques des réactions chimiques se faisant par batch : ces processus sont de nature hybride, des événements discrets déclenchent le lancement ou l'arrêt d'une transformation chimique contrôlée ensuite par un système d'équations différentielles. Dans ce contexte, la partie discrète est représentée par des contrôleurs logiques modélisés par exemple, à l'aide des réseaux de Petri. Symétriquement, un exemple d'intégration d'un module continu dans un SED peut être trouvé dans (de Ugarte, 2008) ; dans ce cas, un PED qui produit à la commande des lots différents, de manière continue, c'est-à-dire sans purge entre deux lots, ce qui implique un certain mélange en début et fin de production d'une commande. Un module continu (basé sur des équations différentielles) est encapsulé dans le premier PED d'un {PED} ; ce programme sélectionne le meilleur ordonnancement des commandes et détermine les temps opératoires.
- L'utilisation de la SH à un même niveau décisionnel pour une chaîne logistique de type 2, modéliser les PEC continus par la DS et les PED par la SED et opérer un couplage fort entre ces modèles. Nous n'avons pas trouvé d'articles qui défendent cette approche. La raison en est sans doute que cette approche implique

l'usage des hypothèses restrictives de modélisation et augmente la complexité de la programmation logicielle et les temps de calcul (Chen *et al.*, 2002).

Dans le reste de cette revue de littérature nous regarderons comment la SED est utilisée pour modéliser des chaînes logistiques hybrides.

2.3 Simulation à événements discrets pour une CLH

Quand la technique de simulation retenue est la SED, on se retrouve face des problèmes techniques nécessitant des règles de traduction permettant la représentation d'un processus continu par un processus discret. On retrouve dans la littérature, des exemples d'utilisation de la SED pour des CLH de type 1. Le PEC situé en amont de la CLH a pour vocation de discrétiser un flux continu pour émettre, à intervalles réguliers, des IC correspondant à une quantité de liquide ou de granulé. Cette fragmentation de flux correspond généralement à un conditionnement (paquet, sac, camion...) utilisé par le transport. Par la suite, les items circulant dans la CLH ne font normalement plus l'objet de traitement par un PEC ou peuvent le faire par un traitement par lot effectué par un PED (Chen *et al.*, 2002).

(Meyer *et al.*, 2011) ont développé une approche basée sur des modèles séparés pour la chaîne logistique de la production des carburants, un modèle de simulation pour la partie gaz et un autre pour la partie liquide. Pour la production des gaz ou des liquides les inputs sont ajoutés dans le réacteur par lots. Les lots émis par le premier processus (par un PED), deviennent des inputs du second processus (PED), après transformation, les caractéristiques volumiques des lots émis ne correspondant pas à ceux demandés. On retrouve ici l'une des raisons évoquées en introduction sur l'impossibilité d'utiliser directement la SED pour modéliser une CLH.

Auteurs	Problème traité	Niveau de décision			Nature de la CL			Simulation		
		Stratégique	Tactique	Opérationnel	CLC	CLD	CLH	SD	SHD	SH
Hijachi et al (2004)	Prévision de la demande (Bullwhip effect)	X				X		X		
Rabcho et al (2007)	Pilotage opérationnel des configurations d'externalisation	X	X							X
Doloi et al (2002)	Etude de l'impact de phase développement des projets sur les coûts de leur cycle de vie	X				X		X		
Lehr et al (2013)	Simulation des stratégies de recouvrement	X				X		X		
Jamalia et al (2013)	Planification de la production	X				X				X
Rozman et al (2014)	Etude de l'impact de l'investissement dans la production	X				X	X	X		
Floutas et al (2004)	Vue générale sur les modèles discrets et continus		X		X			X	X	
Giambiasi et al (2006)	Modèle généralisé à événements discrets			X				X	X	
Hamadi et al (1994)	Etude de l'impact des microbes sur les aliments			X				X	X	
Kyamalya et al (2005)	Problème de localisation	X				X		X		
Weston et al (1999)	Planification de la capacité		X			X		X	X	
Owens et al (2002)	Management de la capacité: alternatives de design	X	X			X		X		
Lyneis (2000)	Prévisions et impact de la dynamique du marché	X				X		X		
Lin et al (2014)	Bullwhip effects	X				X		X		
Pierrevall et al (2007)	Impact des perturbations de la demande	X				X		X		
Lee et al (2002)	Gestion des stocks (du fournisseur au client)	X				X		X		
Music and Matko (1998)	Contrôle des processus chimiques			X	X			X		X
Venkateswaran and Son (2004)	Gestion des stock et management opérationnel	X	X			X				X
Chen et al (2002)	Optimisation de la configuration des processus		X			X		X	X	
Stelchoff et Cheng (2003)	Propagation de l'impact des événements sur la CL		X			X		X	X	
Meyer et al (2011)	Etude de l'impact de la variabilité des installations sur la CL des carburants	X	X		X				X	
Fabre et al (2011)	Contrôle des paramètres physicochimiques			X	X			X		X
de Ugarte et al (2009)	Ordonnement à temps réel			X	X			X		X

Tableau 1 : Grille d'analyse des articles

Dans la suite nous présentons des primitives génériques indépendante du logiciel de simulation discrète utilisé et dépassant ce que proposent (Chen *et al.*, 2002) et (Meyer *et al.*, 2011). Ces primitives permettent de modéliser, avec un seul modèle et avec un degré de détail suffisant tout

type de configuration d'une CLH, pour éclairer la prise des décisions de type opérationnel.

3 PRINCIPES DE MODÉLISATION D'UNE CHAÎNE LOGISTIQUE HYBRIDE PAR LA SIMULATION À ÉVÉNEMENTS DISCRETS

Dans cette section on s'intéresse à la description d'un ensemble de primitives génériques permettant de modéliser un {PEC}, et donc une CLH, dans le niveau de granularité induit par l'analyse du pilotage opérationnel. Les fondements logiques de certaines de ces primitives ont été présentés très sommairement dans (Degoun *et al.*, 2015); nous explicitons ici un certain nombre de principes et de règles permettant de créer ces primitives génériques dans n'importe quel SED.

Après avoir rappelé les caractéristiques des composants de base d'un SED (§3.1), nous précisons comment s'effectue le processus de discrétisation d'un flux continu (§3.2), avant de détailler la primitive *processus continu* permettant de représenter correctement un PEC dans un SED (§3.3). Ce processus continu pouvant mobiliser plusieurs inputs et produire plusieurs outputs, il convient d'adapter au cas continu, la définition des nomenclatures (§3.4). Les IC, séjourner dans des stocks; la définition de la capacité d'un stock dans un SED passe par un nombre maximum d'items qu'il peut contenir, ce qui ne convient pas si les IC ont des caractéristiques volumiques différentes, ce qui est généralement le cas; on est donc amené à créer la primitive *stock capacitaire* (§3.5). Les caractéristiques volumiques des IC produits diffèrent généralement de celles des PEC qui les utilisent, nécessitant la création de la primitive *convertisseur* (§3.6). On examinera enfin les problèmes posés par le contrôle des flux, ce qui conduit à créer une primitive *vanne* pour contrôler des échanges entre les stocks (§3.7) ou d'ordres de fabrications (OF) liés au pilotage à flux tirés, ce qui conduit à un éclatement de ces OF lorsque ces commandes portent sur des quantités de flux (§3.8).

Dans la présentation de ces primitives nous nous appuyons sur des exemples tirés de la CLH de l'OCP (cf. § 4).

3.1 Composants de base d'un SED

La SED décrit le processus de production discrète d'un ensemble de biens ou de services par un système productif (SP) dans lequel circulent des *items* qui arrivent dans ce système par des *points d'entrée* et le quittent par des *points de sortie*, après avoir transité par une séquence de *processeurs* et de *stocks*, en utilisant des *chemins* prédéfinis. Six classes d'objets sont donc mobilisées: la terminologie utilisée ici est générique et donc indépendante du logiciel utilisé; il en est de même pour les propriétés de ces classes d'objets qui peuvent être implémentées de différentes façons.

- Les **items** correspondent à des objets ou des personnes qui font l'objet d'un traitement dans le SP, ou d'informations (commandes, ordres de fabrication...) utilisées dans le pilotage des flux (principalement dans la production à flux tirés). Les caractéristiques spécifiques d'un item, utiles à son traitement par le SP sont consignées dans des

labels, qui jouent le rôle similaire à celui des variables qualitatives ou quantitatives utilisées pour caractériser les éléments d'un ensemble en statistique descriptive. Nous utiliserons ultérieurement la convention Item.Label pour accéder (en écriture ou lecture) au label d'un item.

- Les **points d'entrée** permettent l'introduction d'items dans le SP. Ils sont caractérisés par des *lois d'arrivée* de ces items (stochastiques ou déterministes) et ils définissent la *caractérisation initiale* de ces items entrants par des labels.

- Les **points de sortie** permettent l'évacuation d'items ayant séjourné dans le SP, après traitement. Ils ne possèdent pas de caractéristiques particulières et sont passifs, ce qui implique que les items devant sortir sont envoyés à ces points de sortie par des processeurs.

- Les **processeurs** ont pour fonction de réaliser une *opération* sur un ou plusieurs items, prélevés dans des stocks. Ces opérations modifient certaines caractéristiques des items qui peuvent être morphologiques (transformation de matière...), informationnelles (analyse...) ou spatiales (déplacements). Elles peuvent conduire à la modification de certains labels d'items traités. Elles peuvent également nécessiter des ressources humaines et matérielles mais ces ressources n'ont pas besoin d'être détaillées si elles sont dédiées au processeur. Un processeur est caractérisé par son temps opératoire (logique de *gamme*) qui peut être déterministe ou aléatoire et dépendre des valeurs des labels des items à traiter.

Le traitement peut être exécuté par un processeur sur un item unique. Il peut aussi nécessiter le prélèvement de plusieurs items dans un ou plusieurs stocks. Dans ce dernier cas, les items peuvent être de nature différente (définie par un label), et l'opération relève d'une logique d'assemblage, ou de nature similaire, et l'opération relève d'une logique de lotissement. Dans les deux cas, l'opération mobilise une *nomenclature*. À l'issue du traitement, le processeur expédie un (ou plusieurs items) vers un stock (ou plusieurs stocks), le processeur est alors inoccupé. Un ou plusieurs items ne peuvent être admis dans un processeur que si celui-ci est inoccupé.

Un processeur fictif (caractérisé par un temps opératoire nul) peut être mobilisé dans la description de processus complexes, en particulier pour décrire le routage en entrée ou en sortie de plusieurs items (par exemple, règles combinant des conditions logiques ET et OU).

Un processeur possède un certain nombre de labels par défaut (décompte des items traités...). Il est également possible d'y ajouter des labels spécifiques pour permettre un meilleur contrôle du pilotage du SP. Nous utiliserons ultérieurement la convention Processeur.Label pour accéder (en écriture ou lecture) au label d'un processeur.

- Un **stock** est un lieu où peuvent séjourner des items mais sans y subir de traitement, contrairement aux processeurs. Le stock est passif, c'est-à-dire qu'un item n'entre en stock que s'il est envoyé par un point d'entrée ou un processeur et ne peut en sortir que s'il est appelé par un processeur. Un stock possède un certain nombre de labels par défaut (capacité...). Il est également possible d'y ajouter des labels spécifiques pour permettre un meilleur contrôle

du pilotage du SP. Nous utiliserons ultérieurement la convention Stock.Label pour accéder (en écriture ou lecture) au label d'un stock.

- Les **routes**, représentées par des flèches reliant deux des icônes représentant des objets des quatre classes précédentes, matérialisent des déplacements possibles d'un item d'un point à un autre. L'itinéraire emprunté par un item est piloté par les valeurs de ses labels et un ensemble de règles utilisées par les processeurs. On associe parfois un temps de transport associé à ce déplacement.

L'exécution de programmes est possible lorsqu'un événement survient (date ; arrivée d'un item à un point d'entrée, dans un processeur ou dans un stock ; en cours d'exécution d'une opération par un processeur ; à la sortie d'un item à un point de sortie, d'un processeur ou d'un stock) ou à des dates. Ces programmes utilisent les labels des items, des processeurs et des stocks, plus d'autres informations stockées en mémoire, pour mieux gérer les flux dans le SP.

3.2 Discrétisation d'un flux continu

La modélisation d'un processus continu par la SED repose sur la discrétisation de flux admis en entrée du système productif, avant admission dans un stock capacitaire (voir § 3.5). Cette discrétisation crée des IC qui héritent des caractéristiques chimiques du flux et dont les caractéristiques volumiques et pondérales dépendent du **débit** du flux et de l'intervalle de temps θ qui délimite le lot de liquide ou de granulés associé aux IC. Par exemple, avec un pas de discrétisation $\theta=9'$, l'arrivée de souffre liquide dans la CLH avec un débit de 35,3 T/H conduit à la création d'un IC de 5,3 T, toutes les 9'. Ces IC sont traités ensuite par des processus discrets ou continus.

3.3 Modélisation d'un processus continu par la SED

Un processus continu se caractérise par le passage d'un (ou plusieurs, cf. § 3.4) flux dans une installation productive, où il séjourne un temps moyen μ avant d'en sortir avec des caractéristiques morphologiques et chimiques différentes. La modélisation de ce processus continu par un processeur discret de temps opératoire μ est possible, dans une CLH possédant plusieurs processeurs continus de temps opératoires μ différents, à condition d'utiliser des convertisseurs (cf. §3.6). Le problème qui se pose alors est que l'étude du comportement de la CLH, implique de recueillir périodiquement la production des processeurs, mesurée en sortie de processeurs, la plage Δ de recueil de ces productions pouvant être, par exemple, de 1 heure. La solution évoquée ci-dessus conduit à des productions horaires très irrégulières, même en régime de croisière, dès lors que $\mu > \Delta$. Il convient alors de dissocier la définition de l'intervalle θ séparant deux admissions successives dans un PC (et donc deux sorties successives de ce PC), de son temps opératoire μ .

Il est évident que plus la valeur de θ est faible, meilleur est le lissage de la production de ce PC (modélisé par un

PD). Cette définition de l'intervalle θ entre deux admissions successives dans un PC donné, doit tenir compte de trois contraintes.

- L'intervalle d'admission θ ne peut dépasser l'intervalle Δ de recueil d'information dans la simulation sans perturber les résultats de cette simulation ($\rightarrow \theta \leq \Delta$).

- Si $\theta > \mu$, le processeur est réputé ne pas produire durant la période $\theta - \mu$, d'où une diminution de production ne correspondant pas à la réalité

- Si $\theta < \mu$, on se heurte au problème de l'impossibilité, pour le processeur de base des SED, de modéliser le traitement simultané d'un ensemble d'items arrivant et partant à des dates différentes.

Pour tenir compte de ces contraintes, dans le cas logique où $\theta \leq \text{Min}(\Delta, \mu)$, nous proposons une primitive générique d'un **processus continu** (voir figure 1), consistant à le représenter par trois composants de base d'un SED dont l'usage est « dévoyé ».

- En amont, ce stock est alimenté par un **processeur d'admission et de début de traitement** (Processeur_1, dans la figure 1), doté d'un temps opératoire θ qui correspond à la fois à la cadence d'admission dans le processus continu et au début du temps de traitement du processeur continu. Ce processeur est connecté aux stocks des différents inputs du processus continu et prélève un item dans chaque stock, conformément à la nomenclature de fabrication (voir § 3.4). À titre d'exemple, on a retenu les trois stocks des produits intervenant dans la fabrication de l'acide phosphorique (voir § 3.4).

- Un **stock fictif de fin de traitement** (Stock_1), doté d'un temps de séjour minimal $\mu - \theta$ correspondant au temps résiduel de traitement par le processus continu. Il permet de prendre en compte des items se trouvant à différents stades d'avancement dans le processus continu.

- En aval, le stock fictif est connecté à un **processeur fictif d'enlèvement** (Processeur_2), de temps opératoire nul, dont la fonction est de prélever, du stock fictif, tout item ayant séjourné le temps minimal $\mu - \theta$. En général, une fois l'opération terminée, des items de natures différentes sont expédiés dans autant de stocks, conformément à la nomenclature utilisée (cf. 3.4), ces stocks étant généralement capacitaires (cf. § 3.5).

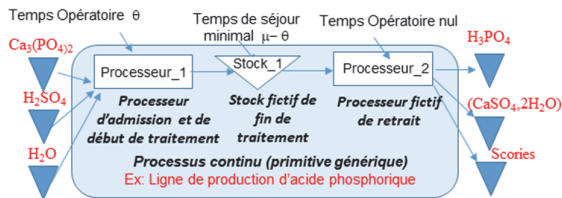
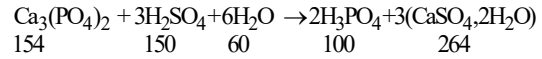


Figure 1 : Modélisation de la primitive processus continu

3.4 Nomenclature d'un processus continu

D'une manière générale, un processus continu prélève un certain nombre d'inputs différents et produit des outputs différents. Appuyons-nous sur l'exemple de la nomenclature d'acide phosphorique. Sa production fait appel à

l'équation molaire suivante, complétée par les nombres atomiques de ces molécules, qui vérifie le principe de conservation de la matière dans la transformation et dans laquelle, le phosphogypse ($\text{CaSO}_4, 2\text{H}_2\text{O}$) est un solide qui contient de l'eau.



En toute rigueur, les débits entrants et sortants d'un processus continu sont des dérivées, dq/dt , où q est un poids. Dans le cadre retenu, ces débits sont définis sur une courte période de temps ($\rightarrow \Delta q / \Delta t$) ; on conviendra ici de retenir la minute ($\rightarrow \Delta t = 1$) et de définir les débits en kilogrammes/minute. On définit ensuite l'un des flux comme étant le *flux-maître* dont le débit sert de référence pour décrire les débits des autres flux. Il est préférable de retenir le flux du produit fabriqué (ici acide phosphorique) comme flux-maître, mais n'importe quel autre intrant ou extrant aurait pu être retenu. On peut alors écrire, en indiquant q par les produits chimiques :

$$\begin{aligned} \frac{\Delta q_{\text{Ca}_3(\text{PO}_4)_2}}{\Delta t} &= \frac{154}{100} \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} = 1,54 \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} \\ \frac{\Delta q_{\text{H}_2\text{SO}_4}}{\Delta t} &= \frac{150}{100} \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} = 1,5 \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} \\ \frac{\Delta q_{\text{H}_2\text{O}}}{\Delta t} &= \frac{60}{100} \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} = 0,6 \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} \\ \frac{\Delta q_{(\text{CaSO}_4, 2\text{H}_2\text{O})}}{\Delta t} &= \frac{264}{100} \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} = 2,64 \cdot \frac{\Delta q_{\text{H}_3\text{PO}_4}}{\Delta t} \end{aligned}$$

Dans la simulation du processus de fabrication d'acide phosphorique, la fixation du débit de production d'acide phosphorique détermine automatiquement ceux des arrivées de $\text{Ca}_3(\text{PO}_4)_2$, de l'acide sulfurique et de l'eau. Avec un pas d'admission $\theta = 60'$ et un débit horaire de production d'acide phosphorique de la ligne de production de 144 T/heure, ce processus chimique admet toutes les 60 minutes un item correspondant à 221,76 T de $\text{Ca}_3(\text{PO}_4)_2$, un item correspondant à 216 T d'acide sulfurique et un item correspondant de 86,4 T d'eau. Toutes les 60 minutes, sortent un item correspondant à 144 T d'acide phosphorique et un item correspondant à 380,16 T de phosphogypse. La conservation des masses est assurée : $144 + 380,16 = 221,76 + 216 + 86,4$.

Cette nomenclature théorique ne se retrouve pas dans le contexte industriel qui ne permet pas d'atteindre ces rendements théoriques. Pour atteindre la production périodique de 144 T, il faut un peu plus d'inputs. Sur une ligne existante, par exemple, il faut 217,08 T d'acide sulfurique (au lieu de 216 T) pour produire 144 T d'acide phosphorique. Cette augmentation d'inputs consommés (qui dépend des installations industrielles utilisées) s'accompagne d'une production inévitable de scories, ce qui permet à l'équation de conservation d'être respectée. La Figure 2 représente la nomenclature industrielle utilisée par une ligne de production de Jorf (cf. § 4).

Enfin, pour effectuer cette production des inputs additionnels n'intervenant pas dans la réaction chimique sont nécessaires (énergie, liquides de refroidissement et, éventuellement, catalyseurs). Ces inputs additionnels peuvent être importants à prendre en compte pour faciliter la gestion de ces ressources, qui peuvent être onéreuses, et dont la consommation a souvent un fort impact dans l'écosystème dans lequel ces unités de production sont insérées.

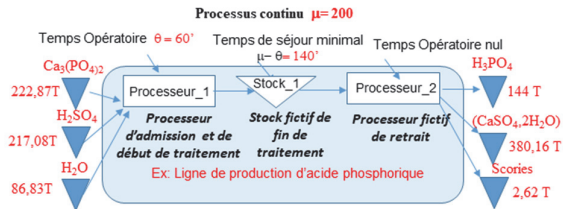


Figure 2 : Nomenclature d'un processus continu

3.5 Gestion des stocks capacitaires contenant des liquides ou des granulés

Dans les SED la capacité d'un stock est exprimée en nombre d'items qu'il peut contenir. Dans les processus continus la capacité est exprimée en poids ou en volume. Un stock capacitair (à capacité limitée) est généralement alimenté par des processus continus différents, ce qui implique que les IC stockés n'ont pas de caractéristiques homogènes et, donc, qu'il n'est pas possible de définir la capacité de ce type de stock par un nombre maximal d'IT. Nous avons créé une primitive générique du **stock capacitair** (voir figure 3), consistant à le représenter par trois composants, un processeur fictif et deux stocks. Le stock capacitair est caractérisé par les labels `Stock_3.Capacite_Offerte` et `Stock_3.Capacité_Utilisée`, dont les valeurs, exprimées en poids (ou en volume), sont initialisées en début de simulation. Le premier label est statique et garde la valeur définie en début d'initialisation. Le second label varie dynamiquement en fonction de l'utilisation instantanée de ce stock. Les items entrants sont caractérisés par le label `Item.Poids`.

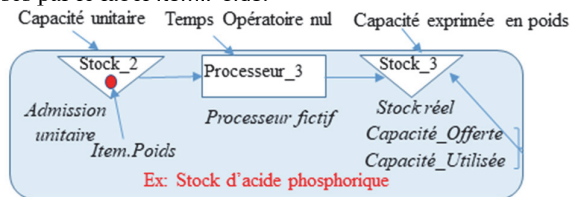


Figure 3 : Modélisation d'un stock capacitair

- Le stock **Admission unitaire** (Stock_2), accueille un item caractérisé par son label `Item.Poids`. Pour éviter certains problèmes de désamorçage en amont, la capacité maximale de ce stock, qui est normalement de 1, devient égale au nombre de processus continus qui l'alimentent.
- Un **processeur fictif** (Processeur_3), avec un temps opératoire nul n'accepte d'items que si la capacité résiduelle du stock réel est suffisante ce qui est réalisé par le programme suivant exécuté avant prélèvement de l'item
`Si Stock_3.Capacité_Utilisée+Item.Poids < Stock_3.Capacite_Offerte`
`alors admettre l'item dans Processeur`
`Sinon Bloquer le prélèvement de l'item`

- À l'entrée d'un item dans **stock réel capacitair** (Stock_3), la mise à jour de la capacité utilisée est effectuée par le programme suivant:
`Stock_3.Capacité_Utilisée ← Stock_3.Capacité_Utilisée + Item.Poids`

- À la sortie d'un item de ce stock, la mise à jour de la capacité utilisée est effectuée par le programme suivant:
`Stock_3.Capacité_Utilisée ← Stock_3.Capacité_Utilisée - Item.Poids`

Un stock capacitair peut être alimenté par d'autres stocks (cf. § 3.7) ou par des processus continus. Un processus continu qui émet des outputs alimentant des stocks capacitaires, doit s'arrêter dès que l'un de ces stocks capacitaires est saturé. La solution consiste alors à ajouter un programme s'inspirant du test décrit ci-dessus pour bloquer l'alimentation du processus continu ; ce test est à effectuer à l'entrée du Processeur_1 de la figure 1.

3.6 Couplage de deux processus continus par un convertisseur

Le couplage entre deux processus continus, caractérisés par les cadences d'admission θ_1 et θ_2 , se fait nécessairement par l'intermédiaire d'un stock (capacitair, le plus souvent). Le problème auquel on s'intéresse ici est celui posé par la non-coïncidence du poids des items produits par l'amont et demandés par l'aval, ce qui conduit à créer une primitive générique appelée **convertisseur** qui mobilise quatre composants (deux processeurs fictifs et deux stocks fictifs) permettant de créer un item ayant le poids requis en aval à partir d'un ou plusieurs items produits en amont.

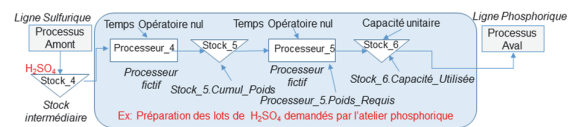


Figure 4 : Modélisation d'un convertisseur

- Le Stock_6 est utilisé pour recevoir le lot ayant la valeur du poids demandée. Ce **stock** a une capacité unitaire pour assurer une conversion au fur et à mesure des besoins de la production. Ce stock possède le label `Capacité_Utilisée`.
- Le Stock_5 est utilisé pour recevoir les items qui vont permettre la création d'un item ayant le poids requis. Ce **stock** possède les labels `Cumul_Poids` et `Nombre_items`. À l'arrivée d'un item dans ce stock le programme suivant est exécuté :
`Stock_5.Cumul_Poids ← Stock_5.Cumul_Poids+ Item.Poids`
- Un premier **processeur fictif**, Processeur_4, de temps opératoire nul, a pour fonction de prélever, du stock intermédiaire (Stock_4), des items pour les pousser dans le stock Stock_5 à condition que l'on ne dispose pas déjà d'un item ayant le poids voulu, et que l'on n'ait pas déjà dépassé le poids requis par le processus continu aval, valeur consignée dans le label `Processeur_5.Poids_requis` (valeur mise à jour dynamiquement en fonction de l'évolution des productions). Ces conditions sont vérifiées par le programme suivant, exécuté avant l'admission par le Processeur_4.
`Si Stock_6.Capacité_Utilisée = 1 alors Bloquer le prélèvement de l'item`
`Si Stock_5.Cumul_Poids ≥ Processeur_5.Poids_requis`
`alors Bloquer le prélèvement de l'item`
- Un second **processeur fictif**, Processeur_5, de temps opératoire nul, est utilisé pour prélever et fusionner tous

les items séjournant dans Stock_5. Il ne peut rien faire si le Stock_6 a un item de poids requis en stock ou si le cumul du poids des items dans le stock Stock_5 (valeur du label Stock_5.Cumul_Poids) n'est pas suffisant. Ces conditions sont vérifiées par le programme suivant exécuté avant l'admission par le Processeur_5 ; ce programme utilise un paramètre k pour gérer l'utilisation de ce processeur

```

k ← 0
Si Stock_6.Capacité_Utilisée = 1 alors k ← 1
Si Stock_5.Cumul_Poids < Processeur_5.Poids_requis alors k ← 1
Si k = 1 alors Bloquer le prélèvement des items
Sinon :
Prélever tous les items du Stock_5 et les fusionner en un item unique
avec Item.Poids ← Stock_5.Cumul_Poids - Processeur_5.Poids_requis
Mettre à jour la capacité utilisée du Stock_5 : Stock_5.Cumul_Poids ←
Stock_5.Cumul_Poids - Processeur_5.Poids_requis

```

Le Processeur_5 émet ensuite deux items identiques, le premier vers le Stock_5 et le second vers Stock_6. À l'entrée de l'item dans Stock_6, le programme suivant est exécuté.

```
Item.Poids ← Processeur_5.Poids_requis
```

3.7 Le contrôle des flux

On abordera ici le cas du pilotage en flux tirés, qui suppose la synchronisation d'un flux d'informations (commandes...). Le transfert d'un stock de fluide à un autre stock relève de cette logique décisionnelle et fait intervenir une vanne, puisque les stocks sont passifs et ne peuvent émettre d'items vers un autre stock. La gestion des flux se fait généralement suivant un programme de transferts, chaque transfert est planifié pour une date bien définie, pour une quantité et un débit bien définis. Ceci fait que l'intervalle de temps séparant deux transferts successifs dépend d'un planning des transferts prédéfini. Le démarrage d'un transfert et le réglage des débits se font via le réglage des vannes.

La primitive générique proposée, appelé **vanne** et présentée dans la figure 5, mobilise un Convertisseur (cf. § 3.6), un point d'entrée, un **stock** (Stock_8) et deux **processeurs fictifs** (Processeur_6 et Processeur_7).

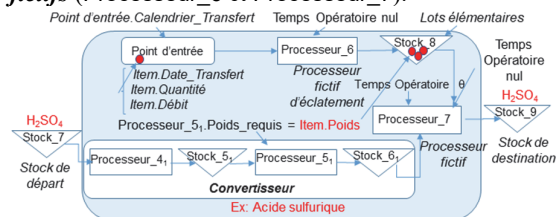


Figure 5 : Modélisation de la primitive Vanne

Le **Point d'entrée** est utilisé pour recevoir les items correspondant à des ordres de transfert d'un stock (Stock_7) à un autre stock (Stock_9). Les labels Item.quantité et item.Débit, sont lus dans un tableau et affectés à l'item entrant.

Le Processeur_6 est utilisé pour dupliquer l'item entrant en m items en sortie de ce processeur. Ce nombre m est calculé par l'exécution du programme suivant lors de l'admission de l'item dans le Processeur_6 :

```
m ← arrondi[Item.Quantité/Item.Débit]
```

L'item sortant se voit attribuer la valeur

```
Item.Poids ← Item.Quantité/m
```

Chacun de ces m items correspond à la commande d'un lot à expédier. La préparation des lots est assurée par le convertisseur reliant le stock de départ (Stock_7) au processeur_7. Ce dernier fusionne l'item prélevé dans le Stock_7 (information) à celui prélevé dans le Stock_6 (matière), ce qui correspond à un processus d'assemblage classique. Le temps opératoire θ du Processeur_7 correspond à l'intervalle séparant deux expéditions successives par la vanne.

3.8 Modélisation du pilotage d'un processus continu par des ordres de fabrication

Dans la modélisation d'un atelier de production, travaillant à la commande, le flux physique est piloté par le flux informationnel. On suppose un atelier comportant plusieurs lignes de production parallèles (on prendra l'exemple de lignes de production d'engrais). Le carnet des ordres de fabrication (OF) contenant l'affectation des commandes sur les lignes est un input du modèle de la simulation. Chaque ligne traite les commandes qui lui sont affectées suivant la logique FIFO selon l'ordre indiqué sur le carnet des OF. Le début du traitement d'une commande n'est donc pas défini par une date bien définie (comme dans le cas des transferts, cf. § 3.7) mais il dépend de la date de fin de la commande précédente et du temps de lancement qui dépend quant à lui du séquençement des commandes.

La modélisation que nous proposons pour la représentation d'une ligne de production d'atelier géré à flux tirés est présentée dans la figure 6.

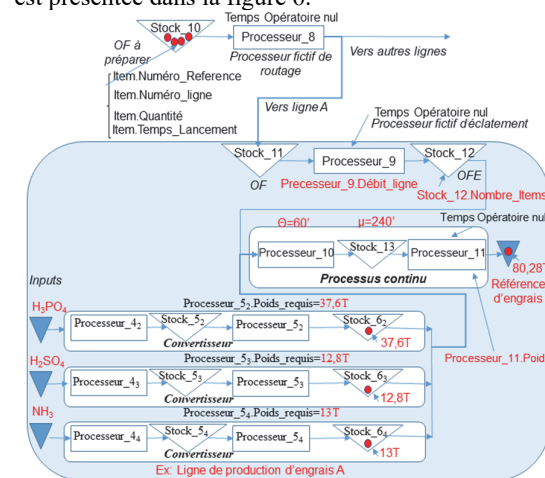


Figure 6 : Modélisation d'un atelier à flux tirés

- Le Stock_10 est initialisé en début de simulation à partir d'une liste d'OF à exécuter. Chaque OF est caractérisé par la référence à produire (Item.Numéro_Reference) et un label contenant le numéro de la ligne pour laquelle elle est affectée (Item.Numéro_ligne), la quantité à fabriquer (Item.Quantité) et le temps de lancement (Item.Temps_Lancement).

- Le **processeur fictif** Processeur_8 assure, suivant le label item.Numéro_ligne, le routage des OF vers les lignes de production. Les OF impliquant la ligne de production étudiée ici sont reçues dans le Stock_11.

- Le *processeur fictif* Processeur_9 éclate ensuite chaque OF en n OFE (OF Élémentaires), ce nombre étant calculé comme le quotient de la quantité à produire (par exemple 11 640T), par le débit de production de la ligne (par exemple 1,347T/minute), multiplié par l'intervalle d'admission θ (par exemple 60 minutes). Dans notre exemple, l'OF de 11 640T est éclaté en 145 OFE caractérisés par un poids à produire de 80,28 T. Le Processeur_9 ne peut pas accepter nouveau item que si le stock_12 est vide (stock_12.Nombre_item = 0). Ces conditions sont vérifiées par le programme ci-dessous exécuté à l'entrée du Processeur_9 :

```
Si Stock_11.Nombre_Items > 0 alors
  Si Stock_12.Nombre_Items > 0 alors
    Bloquer prélèvement de cet item
```

Le processus continu, piloté par le Processeur_10, prélève ensuite, à la cadence d'admission retenue pour ce processus ($\theta=60'$), un OFE (information) et les items (matière) correspondants aux inputs nécessaires (acides...) et les assembler en un item unique dont le poids 80,28 T. Dans cet exemple, le poids des composants utilisés est de 37,6 T pour l'acide phosphorique, 54,13 T pour l'ammoniac et 12,8 T, pour l'acide sulfurique.

Nous ne tenons pas en compte ici de la mise à jour du temps d'admission θ et du temps de traitement μ qui pourraient dépendre de la gamme en cours, mais elle facile et passe par une paramétrisation de θ et de μ en fonction de numéro de la ligne et la référence de l'OF en cours.

Les nomenclatures et gammes de deux commandes successives, sur une même ligne, n'ont aucune raison d'être les mêmes. Une actualisation des poids des inputs, à préparer par les convertisseurs et le poids des lots de l'OFE, est réalisée juste après l'admission d'un nouvel item dans le Processeur_9, en exécutant le programme suivant :

```
Lire Item.Numéro_Reference
Si référence à produire différente de la référence en cours de production
alors mettre à jour :
  Processeur_5s.Poids_requis
  Processeur_5s.Poids_requis
  Processeur_5s.Poids_requis
```

4 ETUDE DE CAS DE LA CL DE L'OCP

L'Office Chérifien des Phosphates (OCP) est la plus grande entreprise Marocaine. Il est intégrée dans toute la chaîne de valeur des phosphates. La partie de la CL de l'OCP qui se trouve à Jorf (voir figure 7), est caractérisée par des processus de production continus (production d'acide phosphorique) et discrets (production d'engrais), ce qui lui confère les caractéristiques d'une CLH.

Ce site industriel comporte quatre ateliers successifs. L'atelier sulfurique, l'atelier de production de l'acide phosphorique concentré à 29, l'atelier de concentration de cet acide phosphorique à 54% et l'atelier des engrais.

La CL de l'OCP est caractérisée par deux modes de pilotage de la production. Les ateliers de production d'acide sulfurique et phosphorique fonctionnent à flux poussés tandis que l'atelier des engrais travaille à flux tirés en suivant un carnet des OF qui décrit les références et les quantités à produire, ainsi que leurs affectations sur les lignes. Le site industriel de Jorf est en interaction permanente avec d'autres entités : en amont avec l'approvisionnement du soufre et l'arrivée de la pulpe de phosphate, en aval par

un plan de chargement des navires et latéralement avec des joint-ventures (JV).

4.1 Système décisionnel de Jorf et objectif du simulateur

Le comportement du système productif de Jorf est déterminé (sauf cas d'aléas) par l'ensemble des informations représentatives des différentes décisions prises sur l'horizon retenu. On s'intéresse ici aux décisions opérationnelles correspondant : au carnet des OF d'engrais à produire, à l'ouverture ou la fermeture des lignes de production ainsi que leurs régimes de marche, au programme de maintenance, et au programme d'échanges d'acide avec les JV.

Le simulateur a été développé dans l'optique d'amélioration des processus décisionnels de Jorf. Il permet d'anticiper l'impact des décisions prises, de détecter d'éventuelles incohérences décisionnelles, ainsi que d'évaluer la propagation des conséquences d'une décision prise au niveau d'un maillon sur les autres maillons de la CL.

Outre la détection d'éventuelles incohérences décisionnelles, le simulateur a pour vocation d'aider à la prise de décision, en facilitant l'exploration rapide de décisions alternatives permettant de stabiliser le comportement du système productif.

4.2 Développement du simulateur

Dans le développement du simulateur, nous sommes partis de l'ensemble des règles de modélisations décrites en section 3 et nous les avons dupliquées autant de fois qu'elles sont appelées dans le modèle.

- La connexion entre les entités du SP est assurée par la primitive vanne que nous avons créée et dupliquée onze fois.

- Pour la modélisation des différents ateliers nous avons mobilisé la primitive de discrétisation du processus continu en nous l'avons dupliquée 39 fois (nombre total des lignes de production dans le site)

- Les différents ateliers sont connectés par des stocks capacitaires intermédiaires. La primitive stock capacitaire est utilisée 37 fois.

- Chaque processus, continu correspondant à une ligne, prélève du stock qui lui est lié moyennant un convertisseur. Pour la modélisation de ces connexions et la modélisation de l'ensemble des vannes nous avons mobilisé la primitive convertisseur et nous l'avons dupliquée 133 fois.

La granularité temporelle retenue dans la modélisation est de type opérationnelle (de l'ordre de l'heure ou de la dizaine de minutes). La simulation est effectuée sur un horizon de deux à trois semaines.

4.3 Benchmark du simulateur

Nous avons effectué plusieurs benchmarks du simulateur pour le valider en partant de données réelles. Ces tests ont permis de détecter quelques incohérences décisionnelles, en particulier des arrêts non programmés des lignes de production induits par une rupture de stock en amont ou, surtout, par une saturation des stocks en aval (avec propagation de l'impact de l'arrêt sur l'amont). L'anticipation

de ces incohérences est difficile à faire par les méthodes classiques.

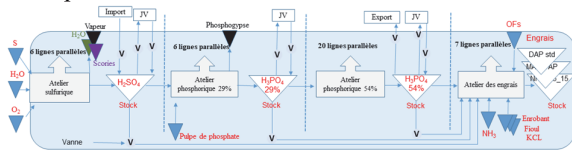


Figure 7 : chaîne logistique de Jorf

5 CONCLUSION

Dans cet article nous avons établi un ensemble de règles permettant d'utiliser les SED pour piloter une CLH, ce qu'aucun logiciel commercial, à notre connaissance, ne sait faire. Cette approche nous a permis de créer un outil d'analyse et d'aide à la prise de décision pour l'un des plus grands sites de production chimique d'une CLH phosphorique ; nous en avons rapidement montré l'architecture. L'usage de ces règles est en cours d'utilisation pour aider au pilotage d'autres activités de l'OCP.

RÉFÉRENCES

- Andersson, M. and G. Olsson, 1998. A simulation based decision support approach for operational capacity planning in a customer order driven assembly line. *Proceedings of the 30th conference on Winter simulation.*, p. 935–942.
- Anthony, R.N., 1965. *Planning and Control Systems: a framework for analysis.* Harvard University Press
- Barbosa, C. and A. Azevedo, 2017. Hybrid simulation for complex manufacturing value-chain environments. *Procedia Manufacturing*, vol. 11, p. 1404–1412, 2017.
- Chen, E.J., Y.M. Lee, and P. L. Selikson, 2002. A simulation study of logistics activities in a chemical plant, *Simulation Modelling Practice and Theory*, vol. 10, n° 3-4, p. 235–245.
- De Ugarte, B.S., A. Hajji., R. Pellerin, and A. Artiba, 2009. Development and integration of a reactive real-time decision support system in the aluminum industry. *Engineering Applications of Artificial Intelligence*, vol. 22, n° 6, p. 897–905.
- Degoun, M., P. Fénies., V. Giard., K. Retmi and J. Saadi, 2015. Propositions de règles de modélisation pour une simulation discrète d'une chaîne logistique hybride. *Congrès international du génie industriel (CIGI2015)*. Québec, Canada.
- Doloi, H. K. and A. Jaafari, 2002. Toward a dynamic simulation model for strategic decision-making in life-cycle project management, *Project Management Journal*, vol. 33, n° 4, p. 23–23.
- Fabre, F., G. Hétreux., J.-M. Le Lann, and P. Zaraté, 2011. Dynamic hybrid simulation of batch processes driven by a scheduling module. *Computers & Chemical Engineering*, vol. 35, n° 10, p. 2098–2112.
- Floudas, C.A. and X. Lin, 2004. Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review. *Computers & Chemical Engineering*, vol. 28, n° 11, p. 2109–2129.
- Forrester, J.W, 1961. *Industry dynamics.* Cambridge, Massachusetts
- Giambiasi, N. and J. C. Carmona, 2006. Generalized discrete event abstraction of continuous systems: GDEVs formalism. *Simulation Modelling Practice and Theory*, vol. 14, n° 1, p. 47–70, 2.
- Hamaidi, L., P. Bourseau., G. Muratet., O. Cerf, and A. K. Sadoudi, 1994. A mixed discrete-continuous approach for the simulation of bacterial contamination and growth in a food process. *Computers & chemical engineering*, vol. 18, p. S633–S637.
- Higuchi, T. and M. D. Troutt, 2004. Dynamic simulation of the supply chain for a short life cycle product—Lessons from the Tamagotchi case. *Computers & Operations Research*, vol. 31, n° 7, p. 1097–1114.
- Jamalnia, A. and A. Feili, 2013. A simulation testing and analysis of aggregate production planning strategies. *Production Planning & Control*, vol. 24, no 6, p. 423–448.
- Jahangirian, M., T. Eldabi., A. Naseer., Stergioulas, and T. Young,

2010. Simulation in manufacturing and business: A review. *European Journal of Operational Research*, vol. 203, no 1, p. 1–13

- Kleijnen J. P., 2005. Supply chain simulation tools and techniques: a survey. *International Journal of Simulation and Process Modelling*, vol. 1, n° 1-2, p. 82–89.
- Kyamakya, K. and K. Jobmann, 2005. Location management in cellular networks: classification of the most important paradigms, realistic simulation framework, and relative performance analysis. *IEEE transactions on vehicular technology*, vol. 54, n° 2, p. 687–708.
- Lee, Y. H., M. K. Cho, and Y. B. Kim, 2002. A discrete-continuous combined modeling approach for supply chain simulation. *Simulation*, vol. 78, n° 5, p. 321–329.
- Lehr, C. B., J.-H. Thun, and P. M. Milling, 2013. From waste to value—a system dynamics model for strategic decision-making in closed-loop supply chains. *International Journal of Production Research*, vol. 51, n° 13, p. 4105–4116.
- Lin, W.-J., Z.-B. Jiang., R. Liu, and L. Wang, 2014. The bullwhip effect in hybrid supply chain. *International Journal of Production Research*, vol. 52, n° 7, p. 2062–2084.
- Lyneis, J. M., 2000. System dynamics for market forecasting and structural analysis. *System Dynamics Review*, vol. 16, n° 1, p. 3.
- Meyer, M., H. Robinson., M. Fisher., A. Van der Merwe., G. Streicher., J.J. Van Rensburg., H. Van den Berg., E. Dreyer., J. Joubert, and G. Bonthuis., 2011. Innovative decision support in a petrochemical production environment », *Interfaces*, vol. 41, n° 1, p. 79–92.
- Mušič, G. and D. Matko, 1999. Combined simulation for process control: extension of a general purpose simulation tool. *Computers in Industry*, vol. 38, n° 2, p. 79–92.
- Owens, S.F. and R. R. Levary, 2002. Evaluating design alternatives of an extruded food production line using simulation. *Simulation*, vol. 78, n° 10, p. 626–632.
- Pierrelval, H., R. Bruniaux, and C. Caux, 2007. A continuous simulation approach for supply chains in the automotive industry. *Simulation Modelling Practice and Theory*, vol. 15, n° 2, p. 185–198.
- Rabelo, L., H. Eskandari., T. Shaalan, and M. Helal, 2007. Value chain analysis using hybrid simulation and AHP », *International Journal of Production Economics*, vol. 105, n° 2, p. 536–547.
- Rozman, Č., A. Škraba., K. Pažek, and M. Kljajić, 2014. The development of sugar beet production and processing simulation model—A system dynamics approach to support decision-making processes. *Organizacija*, vol. 47, n° 2, p. 99–105.
- Scholz-Reiter, B., M. Freitag., C. De Beer, and T. Jagalski, 2005. Modelling dynamics of autonomous logistic processes: Discrete-event versus continuous approaches. *CIRP Annals-Manufacturing Technology*, vol. 54, n° 1, p. 413–416.
- Stchedroff, N. and R. C. Cheng, 2003. Supply chain management simulation: modelling a continuous process with discrete simulation techniques and its application to LNG supply chains », in *Proceedings of the 35th conference on Winter simulation: driving innovation*, p. 1607–1611.
- Tako A. A. and S. Robinson, 2012. The application of discrete event simulation and system dynamics in the logistics and supply chain context. *Decision support systems*, vol. 52, n° 4, p. 802–815.
- Venkateswaran, J. and Y.-J. Son, 2004. Distributed and hybrid simulations for manufacturing systems and integrated enterprise. in *IIE Annual Conference. Proceedings*, p. 1.
- Weston Jr, F.C., F. Herrmann, and P. H. Davidoff, 1999. Capacity planning and process analysis a simulation study of a microbrewery. *Production and Inventory Management Journal*, vol. 40, no 2, p. 48.

Modèle de planification pour les activités de recommerce de produits électroniques

Antoine Jeanjean

Nabil Absi

Recommerce Lab
R&D de Recommerce Solutions SA
54 avenue Lénine, 94250 Gentilly - France
antoine.jeanjean@recommerce.com

École des Mines de Saint-Étienne
Manufacturing Sciences and Logistics Departement
Campus G. Charpak Provence, F-13541 Gardanne - France
absi@emse.fr

RÉSUMÉ : *Dans cet article, nous nous intéressons à la modélisation et à la résolution d'un problème de planification de flux et de ressources, lié aux activités d'achat, de test, de réparation et revente de produits usagés, appelées "activités de recommerce". Nous proposons un modèle mathématique générique à variables mixtes permettant de modéliser cette problématique. Ce modèle est inspiré de l'activité de recommerce de produits électroniques de seconde vie de l'entreprise RECOMMERCER GROUP. Des expérimentations et analyses préliminaires sont conduites sur des instances industrielles.*

MOTS-CLÉS : *Modèle, planification, recommerce, remanufacturing, produits électroniques*

1 INTRODUCTION

Depuis quelques années, la notion d'économie circulaire a pris place dans les espaces publics, politiques et industriels. Cette économie a pour but de révolutionner l'économie classique dite linéaire, qui vise à extraire les matières premières, les transformer en produits finis, les utiliser et enfin recycler les produits en fin de vie. Le but d'une économie circulaire est de réduire les déchets et l'extraction des matières premières en réutilisant les matières premières et les produits semi-finis qui peuvent être récupérés sur les produits en fin de vie, ainsi que de prolonger la durée de vie des produits en prônant la réutilisation.

Redonner une seconde vie à des produits usagés est l'un des objectifs majeurs de l'économie circulaire. En effet, le cycle de vie des produits de grande consommation est de plus en plus court, particulièrement pour les produits électroniques tels que les smartphones, les ordinateurs, les tablettes, etc... En allongeant leur durée d'utilisation, cela permet d'amortir encore plus les coûts de production et de distribution de ces produits.

Plusieurs filières permettant de redonner une seconde vie aux produits ont vu le jour naturellement pour développer cet objectif. Nous pouvons citer à titre d'exemple, les filières automobile, électroménager, textile et téléphonie. C'est une tendance d'abord encouragée par les directions développement durable qui ont vu le jour dans de nombreux groupes, et qui aujourd'hui s'inscrit aussi dans la politique marketing

de ces sociétés. Le recommerce devient notamment un moyen d'aider le consommateur à réduire le coût d'acquisition d'un produit neuf en le finançant grâce à la revente du bien précédent. Cet acte de reprise ayant désormais lieu au sein même du point de vente où le client vient acquérir son nouveau produit.

De plus, la réglementation se structure elle-aussi sur ces sujets, en obligeant désormais les distributeurs à reprendre tous les produits en fin de vie. Pour la filière automobile, depuis le 1er janvier 2017, les réparateurs de voitures doivent informer leurs clients de la possibilité d'utiliser des pièces de rechange usagées issues de l'économie circulaire à la place de pièces neuves. Les opérateurs téléphoniques proposent tous désormais à leurs clients la reprise de leur téléphone usagé moyennant une remise financière sur le nouvel appareil qu'ils souhaitent acquérir, si celui-ci a encore une valeur, ou un recyclage du produit dans le cas contraire. Ces téléphones sont par la suite revendus à des entreprises qui ont pour but de leur redonner une seconde vie ou de les désassembler afin de réutiliser leurs composants.

Les entreprises qui rachètent ou récupèrent ces produits usagés doivent avoir un modèle économique viable afin de pérenniser leur activité. Par exemple, une entreprise qui rachète des smartphones usagés doit pouvoir dégager des bénéfices en proposant des prix de rachat sur chaque produit tout en s'assurant que les coûts de logistiques, tests et éventuellement réparations pourront lui permettre de revendre ensuite ce produit reconditionné à un prix intéressant

pour le consommateur. Pour ce faire, de telles activités doivent utiliser les ressources d'une manière optimisée et écouler les stocks rapidement afin de limiter la dépréciation des produits remanufacturés. Toutes ces activités peuvent être regroupés sous le terme *recommerce* introduit par l'entreprise RECOMMERCE GROUP ¹.

Dans cet article, nous nous intéressons à la modélisation et à la résolution d'un problème de planification de flux et de ressources, lié aux activités d'achat, de réparation et revente de produits usagés. La suite de l'article est articulée autour de 5 sections. Dans la Section 2 nous définissons notre problème et nous positionnons nos travaux par rapport à l'état de l'art. Dans les Sections 3 et 4, nous proposons une modélisation mathématique du problème étudié ainsi qu'une étude de complexité. Dans la Section 5, nous proposons une analyse des premiers résultats expérimentaux et quelques discussions. La Section 6 conclut notre travail et ouvre à diverses perspectives.

2 Définition du problème et état de l'art

Le problème que nous étudions traite de la planification du traitement des produits électroniques usagés dans l'optique de les tester et de les réparer si besoin, afin de leur donner une seconde vie ou de les recycler si la réparation s'avère non profitable. Notre modèle est inspiré de l'activité de l'entreprise RECOMMERCE GROUP qui rachète, test, répare et revend des smartphones usagés. Le modèle de recommerce que nous proposons reste générique et peut s'appliquer à n'importe quelle activité de rachat, test, réparation et de revente de produits en fin de vie.

Les activités de RECOMMERCE GROUP peuvent être résumées comme suit. Les produits usagés, rachetés à des tiers, sont collectés à travers divers canaux qui peuvent être : des réseaux de boutiques, des sites internet ou des centres de collecte. Ces produits peuvent être de différentes qualités. Par exemple, ils peuvent être caractérisés comme étant : fonctionnel comme neuf, fonctionnel usagé, fonctionnel abîmé, écran cassé, composants cassés, ne s'allume pas, ou à recycler. Les produits collectés sont tout d'abord envoyés vers des centres de tests afin de les enregistrer, de les tester et d'apporter diverses réparations simples. Ces activités sont contraintes par les ressources disponibles aux niveaux des centres de tests. En fonction de leurs qualités, les produits peuvent être acheminés vers des centres de remise à neufs distants afin de les reconditionner à neuf et de les revendre à des prix plus importants. A cause de la complexité et du délai de l'opération, la remise à neuf ne peut se faire que par lots de produits et allonge les

temps de traitement par rapport à une vente directe en l'état. Il y a donc un arbitrage à opérer dans cette décision de remettre à neuf les produits, la condition étant de générer au final une marge incrémentale plus importante. En effet, l'objectif final est la revente des produits, qui sont revendus à travers différents canaux et avec différents prix : sites internet, boutiques spécialisées, négociants... Dans la section suivante, nous décrivons tous les paramètres nécessaires à notre modélisation.

Le problème auquel nous nous intéressons dans ce travail est principalement modélisé en utilisant des contraintes de conservation de flots. En effet, le flot entrant (produits à remanufacturer) doit parcourir un ensemble d'états avant d'être orienté vers un flot de sortie (produits remanufacturés) dans le but de satisfaire au mieux les demandes prévisionnelles tout en minimisant les coûts de stockage et de réparation. Ces flots sont contraints par les ressources disponibles. Les contraintes qui rendent le problème étudié difficile sont les contraintes de batch (taille de lots) et les contraintes de FIFO (First In First Out). En effet, ces deux contraintes font intervenir des variables de décision binaires ce qui rend notre modèle linéaire à variables mixtes. Les contraintes de batch permettent de modéliser le processus de remise à neuf qui nécessite une taille de lot minimale et une taille de lot maximale. Les contraintes de FIFO permettent de modéliser le processus de test qui traite les produits dans l'ordre des arrivées (un produit arrivé en période t doit être traité avant tout autre produit arrivant à t' tel que $t' > t$).

Notre problème fait intervenir deux concepts assez récents qui sont le remanufacturing et le buy-back. En effet, nous nous intéressons à la planification du remanufacturing et le rachat (buy-back) de produits en fin de vie.

La notion de remanufacturing en planification de production est assez récente. Le remanufacturing est l'ensemble d'opérations de réparation nécessaires afin que le produit remanufacturé ait les mêmes caractéristiques qu'un produit neuf. La principale question que le décideur se pose est : comment faire correspondre les produits à remanufacturer (flux entrant) et les demandes clients (flux sortant). Guide et al. 1999 sont parmi les premiers à présenter un schéma général pour le remanufacturing incluant l'utilisation de composants issus du désassemblage des produits retournés.

Plusieurs travaux se sont intéressés par la suite à intégrer la notion de remanufacturing à des modèles de lot-sizing classiques. A titre d'exemple, Teunter et al. 2006 ont proposé deux modèles de lot-sizing à un seul produit intégrant la notion de remanufacturing. Dans le premier modèle, les auteurs ont considéré des

¹RECOMMERCE GROUP est une entreprise experte dans la revente, du reconditionnement et de la reprise de smartphones.

coûts de setup de remanufacturing disjoints des coûts de setup de production. Dans le second modèle, ces coûts de setups sont supposés communs. Ces deux problèmes ont été prouvés NP-Difficiles respectivement par Retel Helmrich et al. 2014 et Fazle Baki et al. 2014. Pour une littérature plus détaillée, le lecteur peut se référer à la Section « Remanufacturing » de l'état de l'art traitant des problèmes de lot-sizing à un seul produit (Brahimi et al. 2017).

Plusieurs travaux se sont intéressés au remanufacturing sous différentes facettes. Guide et Van Wassenhove 2001 ont proposé un schéma permettant l'analyse de la rentabilité des activités de réutilisation, et ont montré que la manière dont sont gérés les retours de produits peut fortement influencer la rentabilité de l'entreprise. Plus tard, Guide et al. 2003 se sont intéressés au problème de correspondance de la demande et des produits retournés dans le but de maximiser le profit. Savaskan et al. 2004 se sont intéressés au remanufacturing dans une chaîne logistique en boucle fermée. Ils ont étudié des modèles avec plusieurs options de collecte. Ils ont montré que l'acteur le plus proche du client est le plus à même de collecter les produits à remanufacturer. Dans notre étude, le remanufacturing a deux options, une simple réparation ou une remise à neuf opérée par taille de lots (batch).

La notion de batch est bien connu de la littérature car différents processus industriels nécessitent une taille de lot afin d'assurer la production. Ceci est le cas par exemple pour l'industrie de la chimie, des produits laitiers ou des semi-conducteurs. Dans notre cas, au vu des sites distants et des contraintes fixées par les sous-traitants, une taille de lot minimale et une taille de lot maximale sont nécessaires pour la réalisation des opérations de remise à neuf. On retrouve cette notion de batch dans les problèmes de lot-sizing avec quantités minimales de commande (Anderson et Cheah 1993) ou production par batch (van Vyve 2007, Akbalik et Rapine 2012). Les problèmes de lot-sizing intégrant les batch (Akbalik et Rapine 2013) ou les quantités minimales de commande (Absi et al. 2016) sont NP-Difficiles.

Les travaux traitant du rachat (buy-back) traitent souvent du retour des invendus. Dans le cas de rachat de produits en fin de vie dans le but de les remanufacturer et de les commercialiser, la littérature est également assez récente. Jing et Huang 2013 se sont intéressés aux politiques d'établissement de prix (pricing) dans le cas de produits rachetés dans le cadre de l'achat de nouveaux produits par les clients. Cole et al. 2017 se sont intéressés à la comparaison de deux modèles d'acquisition de produits en fin de vie : le rachat ou l'échange. Kleber et al. 2012 se sont intéressés à l'option de rachat de produits en fin de vie afin d'utiliser leurs composants pour des

produits remanufacturés. La majorité de ces travaux s'intéressent au pricing et aux décisions liées à l'achat de produits en fin de vie. Dans notre étude, le flux entrant est supposé connu et déterministe, et les prix sont fixés à priori.

L'originalité de notre étude est double : (1) nous proposons tout d'abord un modèle représentant une activité de commerce générique (2) nous analysons ce modèle sur des données réelles issues de l'activité de rachat et de réparation de smartphones.

3 Modélisation

Dans cette section, nous présentons un modèle de planification d'une activité de commerce pour laquelle nous utiliserons les notations suivantes :

- Un ensemble de canaux de collecte $\mathcal{I} = \{1, \dots, I\}$. Ces canaux de collecte sont des réseaux de magasin ou des sites Internet ;
- Un ensemble de type de produits $\mathcal{M} = \{1, \dots, N\}$. Un type de produit peut être défini comme une famille de produits collectés qui ont les mêmes caractéristiques et des similarités dans les processus de tests, par exemple : un produit Premium d'une certaine marque, un produit standard d'une autre marque ;
- Un ensemble de centres de traitement $\mathcal{R} = \{1, \dots, R\}$. Ces centres de traitement sont soit des centres de diagnostics, de réparations ou d'expédition où les produits sont reçus, enregistrés, testés, réparés et expédiés au client final ;
- Un horizon de temps discret $\mathcal{T} = \{1, \dots, T\}$ de T périodes. Généralement, une période est fixée à une semaine, puisque les prévisions de collecte, ainsi que les capacités des centres de traitement et les demandes de produits sont fournies avec cette granularité ;
- Un ensemble de qualités $\mathcal{J} = \{1, \dots, J\}$. Une qualité peut être une qualité espérée (après un rachat en boutique par exemple) ou une qualité réelle (après un diagnostic en ateliers). Cela définit l'état du produit. Cela peut être par exemple "Fonctionnel comme neuf" qui signifie un produit passant l'ensemble des tests et dont la qualité cosmétique est comme neuve ;
- $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2$. \mathcal{J}_2 est l'ensemble des qualités candidates au processus de remise à neuf. \mathcal{J}_1 représente les qualités de produits qui ne peuvent pas être remis à neuf ;
- Un ensemble de classes de produits revendus $\mathcal{C} = \{1, \dots, C\}$. La demande en produits est exprimée en classe de produits revendus plutôt

qu'en type de produits. Par exemple : Premium, Standard, etc.;

- Chaque type de produit m appartient à une seule classe de produits \mathcal{C}_c . L'ensemble des types de produits \mathcal{M} est donc l'union des classes de types de produits $\mathcal{M} = \bigcup_{c \in \mathcal{C}} \mathcal{C}_c$

Dans ce qui suit, nous décrivons tous les paramètres qui sont utilisés dans notre modèle. Une quantité q_{mit} de produits de type m est collectée à travers un canal i à une période de temps t . Les produits collectés ont des qualités réelles différentes de celles annoncées, elles sont découvertes une fois les produits arrivés dans les centres de diagnostic. Le taux de produits avec une qualité donnée dépend du canal de collecte ainsi que de la période de temps. Nous définissons τ_{mijt} , le taux de produits de type m avec la qualité j qui sont collectés via un canal de collecte i à la période de temps t . A chaque produit de type m avec une qualité j et qui a été recueilli par un canal de collecte i à la période t est associée un coût d'achat unitaire p_{mijt} . Un coût unitaire b_{ir} est engagé pour transférer des produits à partir d'un canal de collecte i à un centre de diagnostic r . Un nombre de périodes l_{ir} est nécessaire pour transférer les produits d'un canal de collecte i vers un centre de diagnostic r .

Une fois collectés, ces produits sont enregistrés, testés et réparés. Après ces étapes, ces produits peuvent être soit vendus directement avec leur qualités ou remis à neuf dans des unités de remise à neuf et vendus avec un prix de vente plus haut.

Chaque centre de diagnostic r a une capacité en ressources C_{rt} à chaque période t . Une consommation unitaire en ressources w_{mjrt} et un coût unitaire c_{mjrt} sont nécessaires pour traiter un produit de type m avec la qualité j dans un centre de diagnostic r à la période t .

Le coût d'exploitation unitaire des produits non testés de type m de qualité j dans le centre d'essai r à la période t est noté h_{mjrt}^- et le coût unitaire de détention des produits testés de type m avec la qualité j à la période t est noté h_{mjt}^+ .

Les produits de type m d'une qualité donnée $j \in \mathcal{J}_2$ peuvent être remis à neuf pour atteindre une meilleure qualité. Le taux de produits de type m de qualité j pouvant être rénovés en qualité j' est noté $\tau'_{mjj'}$. Un coût unitaire r_{mj} est nécessaire pour rénover un produit de type m avec une qualité j . Un certain nombre de période l est nécessaire pour remettre à neuf les produits. Le processus de rénovation est effectué par lots. Une taille de lot doit être comprise entre une quantité minimale Q^{\min} et une quantité maximale Q^{\max} .

La demande finale des clients est exprimée par classe de produits. Par exemple, les clients finaux

achèteront un produit premium, standard ou basique (pas un type de produit donné avec une marque donnée). Pour chaque classe de produits c , qualité m et période t , on définit les prévisions déterministes de demandes d_{cjt} . Chaque produit de type m avec une qualité donnée j a un prix de vente s_{mjt} qui dépend de la période t .

Au niveau du centre de tests, les produits sont traités selon la règle FIFO (First In First Out). Donc, les produits qui arrivent à un centre de tests à une période donnée ne peuvent pas être traités avant les produits qui sont arrivés au même centre d'essai à une période antérieure. Pour modéliser cette règle, les niveaux de stocks de produits non testés devraient être horodatés. Nous définissons les variables de décision $I_{mirtt'}^-$ indiquant les niveaux de stock de produits non testés de type m collectés à partir du centre de collecte i à la période t dans le centre d'essai r à la fin de la période t' . Nous introduisons aussi une variable de décision binaire $B_{rtt'}$ qui indique si on est en mesure ou pas de tester en période t' dans le centre r les produits reçus à la période t . Le niveau de stock de produits transformés de type m avec la qualité j à la fin de période t est noté I_{mjt}^+ .

Le but de notre modèle est de décider d'abord les quantités X_{mirt} de produits de type m collectés à travers le canal de collecte i à transférer au centre de tests r à la période t . Deuxièmement, de décider des quantités à traiter par centre de tests. Puisque les processus des centres de tests suivent la règle FIFO, les quantités traitées doivent être horodatées. Nous définissons une variable de décision $Y_{mirtt'}$ qui donne la quantité de produits de type m livrés à la période t depuis le canal i qui doit être traités en période t' par le centre de traitement r .

Les produits testés peuvent être vendus avec leur qualité actuelle ou remis à neuf dans d'autres unités de rénovation pour atteindre une qualité supérieure, mieux valorisées. Nous définissons W_{mjt} , la quantité de produits de type m avec la qualité j à envoyer pour rénovation à la période t . Cette quantité remise à neuf sera disponible à la période $t+l$, avec l le nombre de pas de temps nécessaire à l'ensemble du processus complet : aller, remis à neuf puis retour en stock. Nous définissons une variable binaire V_t pour décider si la remise à neuf est effectuée à la période t ou pas. Nous définissons également Z_{mjt} , la quantité de produits de type m avec une qualité j à vendre à la période t .

L'objectif est de maximiser le profit en minimisant les coûts de transfert, de test, de stockage et de traitement tout en satisfaisant les contraintes de capacité.

Les données du problème et les variables de décision sont résumées dans les tableaux 1 et 2

Paramètres	
q_{mit}	Quantité de produits de type m collectés via le canal de collecte i à la période t
p_{mijt}	Prix unitaire d'achat moyen par type de produits m en qualité j collecté via le canal de collecte i à la période t
τ_{mijt}	Le taux de produits de type m en qualité j qui sont collectés via le canal de collecte i à la période t
$\tau'_{mjj'}$	Le taux de produits de type m en qualité j qui peuvent être reconditionnés en qualité j'
d_{cjt}	La demande des clients en produits de classe c en qualité j à la période t
s_{mjt}	Le prix moyen de vente des produits de type m en qualité j à la période t
l	Le nombre de périodes nécessaire pour reconditionner à neuf un produit
l_{ir}	Le nombre de périodes nécessaire pour transférer un produit d'un canal de collecte i à un centre de tests r .
b_{ir}	Le coût unitaire pour transférer un produit depuis un canal de collecte i à un centre de tests r
C_{rt}	La capacité totale en ressources du centre de tests r à la période t
w_{mjrt}	Consommation unitaire en ressources pour traiter les produits de type m en qualité j dans le centre de tests r à la période t
Q^{\min}	La taille minimale d'un lot de remise à neuf
Q^{\max}	La taille maximale d'un lot de remise à neuf
c_{mjrt}	Coût unitaire pour traiter un produit de type m et de qualité j dans un centre de tests r à la période t
h_{mjrt}^-	Coût unitaire de stockage des produits non testés de type m et de qualité j dans le centre de tests r à la période t
h_{mjrt}^+	Coût unitaire de stockage des produits testés de type m et qualité j à la période t
r_{mj}	Coût unitaire de remise à neuf des produits de type m en qualité j

Tableau 1 – Notations : Paramètres

En utilisant ces notations, le problème peut être modélisé par le programme linéaire mixte suivant :

$$\begin{aligned} & \text{maximiser } \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} s_{mjt} Z_{mjt} \\ & - \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} \left(b_{ir} X_{mirt} \right. \\ & \left. + \sum_{t' \in \{1, \dots, t\}} \sum_{j \in \mathcal{J}} (c_{mjrt} \tau_{mijt}) Y_{imrt't} \right) \end{aligned}$$

Variables de Décisions	
X_{mirt}	La quantité de produits de type m collectée via le canal de collecte i à la période t à transférer au centre de tests r
$Y_{mirtt'}$	La quantité de produits de type m arrivée à la période t via le canal de collecte i à traiter à la période t' au centre de tests r .
L_{cjt}	La demande non satisfaite pour les produits de classe c en qualité j à la période t .
$I_{mirtt'}^-$	Le niveau d'inventaire des produits non traités de type m collectés via le canal de collecte i à la période t dans le centre de tests r à la fin de la période t'
I_{mjt}^+	Le niveau d'inventaire des produits traités de type m en qualité j à la fin de la période t
$B_{rtt'}$	$= 1$ si on est en mesure de tester en période t' dans le centre r les produits reçus à la période t , 0 sinon
V_t	$= 1$ Si un batch de remise à neuf est prévu à la période t , 0 sinon
W_{mjt}	La quantité de produits de type m en qualité j à envoyer en remise à neuf à la période t .
Z_{mjt}	La quantité de produits de type m en qualité j vendue à la période t

Tableau 2 – Notations : Variables de décisions

$$\begin{aligned} & - \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} \left(h_{mjrt}^+ I_{mjt}^+ \right. \\ & \left. + \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}} \sum_{t' \in \{1, \dots, t\}} h_{mjrt}^- \tau_{mijt'} I_{mirt't}^- \right) \\ & - \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{t \in \mathcal{T}} (p_{mijt} \tau_{mijt} q_{mit}) \\ & - \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_2} \sum_{t \in \mathcal{T}} (r_{mj} W_{mjt}) \end{aligned} \tag{1}$$

Sous contraintes :

$$\sum_{r \in \mathcal{R}} X_{mirt} = q_{mit} \quad \forall i \in \mathcal{I}, m \in \mathcal{M}, t \in \mathcal{T} \tag{2}$$

$$\begin{aligned} I_{mirtt'}^- &= X_{mir(t-l_{ir})} - \sum_{k=t}^{t'} Y_{mirtt'k} \\ \forall m \in \mathcal{M}, i \in \mathcal{I}, r \in \mathcal{R}, t \in \mathcal{T}, t' \in \mathcal{T}/t' \geq t \end{aligned} \tag{3}$$

$$\begin{aligned} & \sum_{j \in \mathcal{J}_2} \tau'_{mj'j} W_{mj'(t-l)} + I_{mjt-1}^+ \\ & \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}} \sum_{t' \in \{1, \dots, t\}} \tau_{mijt'} Y_{mirt't} = Z_{mjt} + I_{mjt}^+ \\ & \forall m \in \mathcal{M}, j \in \mathcal{J}_1, t \in \mathcal{T} \quad (4) \end{aligned}$$

$$\begin{aligned} & I_{mjt-1}^+ + \sum_{r \in \mathcal{R}} \sum_{i \in \mathcal{I}} \sum_{t' \in \{1, \dots, t\}} \tau_{mijt'} Y_{mirt't} \\ & = W_{mjt} + Z_{mjt} + I_{mjt}^+ \quad \forall m \in \mathcal{M}, j \in \mathcal{J}_2, t \in \mathcal{T} \quad (5) \end{aligned}$$

$$\begin{aligned} & \sum_{m \in \mathcal{C}_c} Z_{mjt} + L_{cjt} = d_{cjt} \\ & \forall c \in \mathcal{C}, j \in \mathcal{J}_1, t \in \mathcal{T} \quad (6) \end{aligned}$$

$$\begin{aligned} & \sum_{j' \in \mathcal{J}_2} \tau_{mj'j} W_{mj'(t-l)} + \sum_{m \in \mathcal{C}_c} Z_{mjt} + L_{cjt} = d_{cjt} \\ & \forall c \in \mathcal{C}, j \in \mathcal{J}_2, t \in \mathcal{T} \quad (7) \end{aligned}$$

$$\begin{aligned} & \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}} Y_{mirt't} \leq M(B_{rtt'}) \\ & \forall r \in \mathcal{R}, t \in \mathcal{T}, t' \in \mathcal{T}/t' \geq t \quad (8) \end{aligned}$$

$$\begin{aligned} & \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}} I_{mir(t-1)t'}^- \leq M(1 - B_{rtt'}) \\ & \forall r \in \mathcal{R}, t \in \mathcal{T}, t' \in \mathcal{T}/t' \geq t \quad (9) \end{aligned}$$

$$\begin{aligned} & \sum_{m \in \mathcal{M}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{t' \in \{1, \dots, t\}} w_{mjrt} \tau_{mijt'} Y_{mirt't} \leq C_{rt} \\ & \forall r \in \mathcal{R}, t \in \mathcal{T} \quad (10) \end{aligned}$$

$$\sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_2} W_{mjt} \leq Q^{\max} V_t \quad \forall t \in \mathcal{T} \quad (11)$$

$$\sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{J}_2} W_{mjt} \geq Q^{\min} V_t \quad \forall t \in \mathcal{T} \quad (12)$$

$$L_{cjt} \leq d_{cjt} \quad \forall c \in \mathcal{C}, j \in \mathcal{J}, t \in \mathcal{T} \quad (13)$$

$$\begin{aligned} & B_{rtt'} \in \{0, 1\} \quad \forall r \in \mathcal{R}, t \in \mathcal{T}, \\ & t' \in \mathcal{T}/t' \geq t \quad (14) \end{aligned}$$

$$V_t \in \{0, 1\} \quad \forall t \in \mathcal{T} \quad (15)$$

$$\begin{aligned} & I_{mirtt'}^- \geq 0, Y_{mirtt'} \geq 0 \quad \forall m \in \mathcal{M}, \forall i \in \mathcal{I}, \\ & \forall r \in \mathcal{R}, t \in \mathcal{T}, t' \in \mathcal{T}/t' \geq t \quad (16) \end{aligned}$$

$$X_{mirt} \geq 0 \quad \forall m \in \mathcal{M}, \forall i \in \mathcal{I}, \forall r \in \mathcal{R}, t \in \mathcal{T} \quad (17)$$

$$\begin{aligned} & I_{mjt}^+ \geq 0, Z_{mjt} \geq 0, W_{mjt} \geq 0 \\ & \forall m \in \mathcal{M}, \forall j \in \mathcal{J}, t \in \mathcal{T} \quad (18) \end{aligned}$$

$$L_{cjt} \geq 0 \quad \forall c \in \mathcal{C}, \forall j \in \mathcal{J}, t \in \mathcal{T} \quad (19)$$

La fonction objectif (1) cherche à maximiser le profit total qui est modélisé comme la différence entre le revenu de la vente des produits et le coût total d'acquisition de ces produits à savoir leurs coûts : d'achat, de stockage, de transport, de traitement et de remise à neuf. Les contraintes (2) assurent la conservation des flux, afin que tous les produits collectés via un canal de collecte soient routés vers une usine de traitement. Les contraintes (3)-(5) sont aussi des contraintes de flux, elles assurent la conservation de flux pour les produits testés, non testés et remis à neuf. Les contraintes (6)-(7) sont des contraintes de flux assurant la satisfaction de la demande. Les contraintes (8) et (9) assurent la règle de traitement FIFO, i.e. que les produits qui arrivent à un centre de tests r à une période t ne peuvent être traités avant des produits arrivés à une période $t' < t$ au même centre de tests. Les contraintes (10) permettent de vérifier que l'activité d'une usine de traitement n'excède pas sa capacité totale. Les contraintes (11) et (12) satisfont les tailles minimale et maximale des batchs de traitement de remise à neuf. La contrainte (13) limite les ventes perdues par les demandes. Les contraintes (14) et (19) définissent les variables de décisions du problème.

4 Analyse de complexité

Le problème que nous étudions comprends principalement trois types de contraintes dont les complexités induites sont très différentes. Le premier problème est principalement modélisé par des contraintes de conservation de flots. Les problèmes de flots à coût minimum sont connus pour leurs complexités polynomiales dès lors que les coûts induits restent linéaires. Le second problème correspond principalement à la satisfaction des contraintes de taille de batch. Ces problèmes sont modélisés par des variables binaires qui

rendent les problèmes linéaires à variables mixtes. Le problème de planification de production (lot-sizing) avec contraintes de taille de batch est NP-Difficile au sens faible. Akbalik et Rapine 2013 ont montré que le problème de lot-sizing à un seul produit, sans coûts de setup, des coûts de batch constants et sans coûts d'approvisionnement est NP-Difficile. Les auteurs ont effectué une réduction polynomiale à partir du *Money change problem*. Le troisième problème correspond à la modélisation de la règle FIFO énoncée précédemment. Cette modélisation fait intervenir des variables binaires mais la complexité induite reste ouverte.

Propriété 1 *Le problème que nous étudions dans ce travail est NP-Difficile.*

Preuve: Cette preuve peut se faire en effectuant une réduction polynomiale du problème de lot-sizing à un seul produit, sans coûts de setup et coûts de batch constants (pour rappel, ce problème est prouvé NP-Difficile dans Akbalik et Rapine 2013). L'instance suivante de notre problème est équivalente au problème de lot-sizing cité ci-dessus. Un seul produit est à remanufacturer, un seul canal de collecte, un seul centre de traitement, une seule qualité, une seule classe de produits, les délais sont nuls, les capacités en ressources sont nulles et la seule alternative est le remanufacturing par batch. Les tailles maximales et minimales des batchs sont égales et sont constantes dans le temps. La règle FIFO ne s'applique pas dans ce cas car nous ne traitons qu'un seul produit et une seule qualité. Cette instance est équivalente au problème de lot-sizing à un seul produit avec des contraintes de batch. Ceci prouve que notre problème est au moins NP-Difficile au sens faible. □

5 Expérimentations et discussions

Nous avons implémenté le modèle mathématique proposé dans la Section 3 en utilisant le moteur de résolution de programmes linéaires à variables mixtes CPLEX 12.6 sous Visual Studio 2012 en utilisant la bibliothèque concert de CPLEX et le langage de programmation C++. Notre critère d'arrêt est le gap entre la meilleure solution connue et la meilleure borne inférieure trouvée par CPLEX. Il est fixé à 1%.

En se basant sur l'historique des données de l'entreprise RECOMMERCE GROUP, nous avons créé des jeux de données caractérisés par 1 seul canal de collecte, 10 types de produits, 52 semaines, 7 qualités d'achat, 9 qualités de revente, 1 seule ressource.

Les résultats préliminaires montrent que les temps de résolution sont importants (quelques heures) quand la possibilité de remanufacturing est activée. Les temps d'exécution sont réduits à quelques minutes quand cette option est désactivée. En effet, en ignorant

l'option de remanufacturing, le nombre de variables binaires est réduit considérablement.

Nous avons effectué des comparaisons entre les résultats obtenus en résolvant le modèle proposé et les décisions effectivement implémentés (historique). Ces résultats sont difficilement quantifiables à l'heure actuelle car nous ne traitons qu'un ensemble réduit de produits et qu'une conclusion plus aboutie nécessiterait la résolution du problème incluant une centaine de produits, une dizaine de canaux de collecte et plusieurs ressources, afin de pouvoir se rapprocher des scénarios réels.

6 Conclusions et perspectives

Dans ce travail, nous sommes intéressés à un problème d'actualité traitant du recommerce des produits en fin de vie. Cette problématique est l'une des préoccupations centrales de l'économie circulaire dont le but est de réduire les déchets et l'extraction des matières premières en revendant les produits semi-finis et en réutilisant les matières premières qui peuvent être récupérés sur les produits en fin de vie.

A notre connaissance, nous sommes les premiers à proposer un modèle mathématique générique permettant de tenir compte de la problématique de recommerce de produits en fin de vie. Notre étude est fortement inspirée par le recommerce des smartphones car nos expérimentations sont opérées sur des jeux de données de l'entreprise RECOMMERCE GROUP.

Le travail que nous avons proposé dans cet article n'aborde que la partie motivation et modélisation de l'activité de recommerce de produits en fin de vie. Nous n'avons pas développé de méthode de résolution à proprement parlé mais nous avons utilisé le moteur de résolution de programmes linéaires en nombre entiers CPLEX 12.6. Les temps de résolution peuvent atteindre quelques heures pour les problèmes autorisant le remanufacturing, mais sont de quelques minutes quand le remanufacturing est ignoré. Les gains obtenus sur les instances de test restent à valider sur des instances réelles incluant tous les produits et tous les canaux de collecte.

Afin d'aborder les instances réelles, le moteur de résolution de programmes linéaires à variables mixtes CPLEX a montré ses limites. Pour la résolution de problèmes de taille industrielle, d'autres approches de résolutions sont envisageables. La première méthode envisageable est une méthode de type Relax-and-Fix (Absi et Kedad-Sidhoum 2007). Elle consiste à décomposer l'horizon de planification en plusieurs fenêtres que l'on résout dans schéma itératif afin de réduire le nombre de variables binaires à chaque étape et s'assurer de l'obtention d'une solution réalisable à la fin du temps alloué à la résolution. La sec-

onde méthode est basée sur le principe de la relaxation Lagrangienne qui consiste à relâcher les contraintes qui rendent le problème difficile (contraintes de batch dans notre cas) et les pénaliser dans le fonction objectif (Trigeiro et al. 1989). Cette méthode est également itérative et permet l'obtention d'une bonne borne inférieure au problème étudié. Cette approche doit être couplée à une heuristique Lagrangienne qui consiste à obtenir une solution réalisable (borne supérieure) à partir des solutions duales qui sont souvent non réalisables.

Cette étude ouvre vers diverses perspectives permettant l'étude de nouveaux problèmes originaux. A titre d'exemple, l'entreprise qui opère le traitement des produits en fin de vie, peut avoir comme alternative à la revente des produits remanufacturés, la remanufacturing et la revente de produits semi-finis ou de matières premières issues des produits à remanufacturer. Ces alternatives peuvent par exemple s'appliquer aux produits électroniques en fin de vie, le secteur automobile, etc.

L'économie circulaire regorge de problèmes d'optimisation originaux comme par exemple : la revente et la transformation des produits dérivés, le stockage de l'énergie sous diverses formes pour des usages futurs, l'intégration des composantes sociétales pour dynamiser les économies locales, etc.

REFERENCES

- Absi, N., Kedad-Sidhoum, S. (2007). MIP-based heuristics for multi-item capacitated lot-sizing problem with setup times and shortage costs. *RAIRO-Operations Research*, 41(2), 171-192.
- Absi, N., Dauzère-Pérès, S., Kedad-Sidhoum, S., Penz, B., Rapine, C. (2016). The single-item green lot-sizing problem with fixed carbon emissions. *European Journal of Operational Research*, 248(3), 849-855.
- Akbalik, A., Rapine, C. (2012). Polynomial time algorithms for the constant capacitated single-item lot sizing problem with stepwise production cost. *Operations Research Letters*, 40(5), 390-397.
- Akbalik, A., Rapine, C. (2013). The single item uncapacitated lot-sizing problem with time-dependent batch sizes: NP-hard and polynomial cases. *European Journal of Operational Research*, 229(2), 353-363.
- Anderson, E. J., Cheah, B. S. (1993). Capacitated lot-sizing with minimum batch sizes and setup times. *International Journal of Production Economics*, 30, 137-152.
- Brahimi, N., Absi, N., Dauzère-Pérès, S., Nordli, A. (2017). Single-item dynamic lot-sizing problems: An updated survey. *European Journal of Operational Research*.
- Cole, D., Mahapatra, S., Webster, S. (2017). A Comparison of Buyback and Trade-In Policies to Acquire Used Products for Remanufacturing. *Journal of Business Logistics*.
- Fazek Baki, M., Chaouch, B. A., Abdul-Kader, W. (2014). A heuristic solution procedure for the dynamic lot sizing problem with remanufacturing and product recovery. *Computers & Operations Research*, 43, 225-236.
- Guide, V. D. R., Jayaraman, V., Srivastava, R. (1999). Production planning and control for remanufacturing: a state-of-the-art survey. *Robotics and Computer-Integrated Manufacturing*, 15(3), 221-230.
- Guide Jr, V. D. R., Teunter, R. H., Van Wassenhove, L. N. (2003). Matching demand and supply to maximize profits from remanufacturing. *Manufacturing Service Operations Management*, 5(4), 303-316.
- Guide, V. D. R., Wassenhove, L. N. (2001). Managing product returns for remanufacturing. *Production and operations management*, 10(2), 142-155.
- Jing, L., Huang, B. (2013). Jointly Consider Acquisition Price, Trade in Rebate and Selling Price in Remanufacturing. In *Re-engineering Manufacturing for Sustainability* (pp. 597-602). Springer, Singapore.
- Savaskan, R. C., Bhattacharya, S., Van Wassenhove, L. N. (2004). Closed-loop supply chain models with product remanufacturing. *Management science*, 50(2), 239-252.
- Teunter, R. H., Bayindir, Z. P., Den Heuvel, W. V. (2006). Dynamic lot sizing with product returns and remanufacturing. *International Journal of Production Research*, 44(20), 4377-4400.
- Trigeiro, W. W., Thomas, L. J., McClain, J. O. (1989). Capacitated lot sizing with setup times. *Management science*, 35(3), 353-366.
- Retel Helmrich, M. J., Jans, R., van den Heuvel, W., Wagelmans, A. P. M., 2014. Economic lot-sizing with remanufacturing: complexity and efficient formulations. *IIE Transactions* 46 (1), 67-86.
- Van Vyve, M. (2007). Algorithms for single-item lot-sizing problems with constant batch size. *Mathematics of Operations Research*, 32(3), 594-613.

MULTI LEVEL UNCAPACITATED LOT SIZING PROBLEM CONSIDERING WORKING CAPITAL REQUIREMENT

Yuan Bian, Fabrice Gayraud, David Lemoine
Thomas G. Yeung

IMT Atlantique / LS2N UMR CNRS 6004
4, rue Alfred Kastler, BP 20722
44307 Nantes - France
{yuan.bian, fabrice.gayraud, david.lemoine,
thomas.yeung}@imt-atlantique.fr

Nathalie Bostel

Université de Nantes / LS2N UMR CNRS 6004
58 rue Michel Ange, BP 420
44606 Saint-Nazaire - France
nathalie.bostel@univ-nantes.fr

ABSTRACT: *In this work, we integrate the financing cost of the working capital requirement (WCR) into the classic Multi-level Lot-sizing Problem (MLLP) model within a serial supply chain. The WCR model is adapted for this multi-site context. As resolution method, both sequential and centralized approaches are proposed with proofs of the "zero-inventory" property. A dynamic programming algorithm is proposed that solve the problem with a centralized objective in $O(NT^6)$ where N is the number of levels and T is the number of periods. Through numerical tests, the financial consequences of the operational decisions on different supply chain members are revealed.*

KEYWORDS: *MLLP, Financial flows, Working capital requirement, Net present value, Serial structure.*

1 Introduction

Since late 1960's, companies have largely implemented the Material Requirement Planning system to meet the increasing customer needs. Larger and more complex product structures must be dealt with in the production system. This evolution has magnificently revealed the significance of the Multi-Level Lot-sizing Problem (MLLP) in the MRP process. The principal concept of the MLLP is to translate the production program of a finished item into the programs of its components with demand satisfaction and cost minimization objectives. Those programs are calculated on the basis of the bill-of-material (BOM) and lead-time information. In recent years, the literature on supply chain management, following the original work of Babich & Sobel (2004), became aware of the fact that financial and operational problems are imbricated and that optimizing globally the two dimensions could improve the global performance of the company. In this work, we extend our previous work of the single-level case (see in Bian et al. (2018)) to the multi-level supply chain context with a serial structure, denoted as $MLLP_{P(WCR)}$. The objective is to jointly consider the physical and financial flows in the multi-level supply chain presented in Figure 1. To do so, we introduce the financial aspects as the financing of Working Capital Requirement (WCR) and the time value of money into the classical MLLP model. In order to efficiently solve the problem, both the sequential and centralized approaches based on the dynamic

programming technique are proposed with the proofs of zero-inventory property.

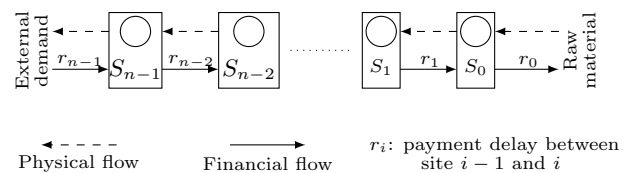


Figure 1 – Flows in a multi-level supply chain

As presented in Figure 1, we consider a supply chain with a serial structure composed of N sites. To describe this system, each site purchases and processes the material to produce intermediate goods (except for S_{n-1}) and delivers it to its successor in the supply chain. The link at the left side of the chain represents the delivery to the final client for satisfying its demand. We assume that only one type of item is manufactured in each site. The payment delay between site i and site $i-1$ is denoted r_i . More assumptions are presented in paragraphs below. Each site is characterized by:

- logistic costs: purchasing costs, setup costs, production costs and inventory holding costs. According to common sense, the purchasing unit cost of site $i+1$ should be higher than the total unit cost for its production at site i .
- terms of payment: It is the payment delay between two successive sites. For the supplier, it

should add the delivery delay to the manufacturer and the payment delay, because we assume that the manufacturer will begin to execute the payment term after receiving the intermediate goods. Noted that no particular index is given for the final customer who commands the external demand and the supplier who provides the raw material.

- financial aspects:
 - **the discount rate** is the rate for calculating the net present value,
 - **the interest rate** is used for calculating the WCR financing cost.

2 Literature review

The MLLP problem was initially investigated by Zangwill (1966) who proposed a dynamic programming (DP) algorithm for the serial BOM case after reformulating the problem as a single-source network problem with flow constraints. The complexity of this algorithm for $N = 2$ is $O(T^3)$ and for $N > 2$ is $O(NT^4)$ where N is the number of levels. Other early works are summarized in Pochet & Wolsey (2006). Recently, Melo & Wolsey (2010) considered the uncapacitated two-level problem with serial structure and developed a DP algorithm that runs in $O(T^2 \log T)$ using the approach of Eppen & Martin (1987).

In addition to those exact methods, considerable efforts have been continuously invested for developing heuristic and meta-heuristic methods over decades in order to tackle the computational difficulty of this problem. In the literature, these methods can be mainly classified into three types: sequential, simultaneous (centralized) and decentralized or distributed approach (Homberger (2010)).

The sequential approach is the one of the first methods developed for the multi-level problem in the MRP framework which decomposed the multi-level problem into several polynomial single-level problems, for instance, with the Wagner-Whitin algorithm (Veral & LaForge (1985)). A state of the art on various related methods is provided in Gupta & Keung (1990).

The simultaneous (centralized) approach has emerged much more recently and often is based on meta-heuristic algorithms that provide highly cost-efficient solutions with a reasonable computation effort. The metaheuristics that appears in the literature include simulated annealing (e.g., Jeunet & Jonard (2005)), genetic algorithms (e.g., Homberger (2008)), ant colony optimization (ACO) systems (e.g., Homberger & Gehring (2009)), particle swarm optimization algorithms (e.g., Deroussi & Lemoine (2009)) and neighborhood searches (e.g., Xiao et al. (2014)).

Traditionally, the MLLP models only focus on the physical flow through the system. However, the Net Present Value (NPV) principle has been identified as a natural approach to reveal the economic consequences in the MRP theory of the early work of Grubbström (1967). Furthermore, discrete-time models with consideration of the NPV have also been developed in the literature. Helber (1998) treats the multi-level problem in the framework of Manufacturing Resource Planning (MRP II) with multiple capacity constraints and setup times in a discrete-time horizon. Grubbström & Wang (2003) present a model of multi-level capacity-constrained system with a stochastic external demand and a constant lead time for maximizing expected the NPV of profit. A dynamic programming procedure is applied to solve the problem. Based on previous work, Grubbström & Huynh (2006) establish a similar model with deterministic demand and extend the model for any non-zero lead time. Other models using the NPV approach for multi-site problems are reviewed in Martínez-Costa et al. (2014).

In addition to the NPV principle, we also consider the financing cost of the WCR in order to integrate the financial flow into the classic MLLP model in our work. To the best of our knowledge, this financial aspect has not be addressed in the literature of the MLLP. Both sequential and centralized approaches are proposed in our research.

3 Working Capital Requirement (WCR) modeling

The generic OWCR formulation for the single-level case was previously established in Bian et al. (2018) by integrating the financial costs linked to operational decisions (i.e., the financing cost of OWCR) into a tactical planning model. Both the upstream and downstream delays in payment are taken into account in this formulation. To establish the model, we adopt the cash to cash cycle methodology presented in Theodore Farris & Hutchison (2002) where the OWCR for producing a unit of product depends on the amount invested into the related operations (e.g., purchasing, setup, production and inventory holding) and the financing duration of these investments before recovering from corresponding sales. However, we first assume that using the profit portion of the sales revenue to cover the OWCR is not allowed. Since the profit margin of the company has never been required, it does not correspond to the OWCR definition. Furthermore, this assumption supposes that the company does not prioritize its profit allocation to WCR financing. Profit can be allocated to any of a number of objectives for the firm including debt reduction, internal or external investment, or dividend payments. Since we do not indicate the allocation of profit (to WCR

financing or others objectives), our model is thus a partial model for a company. Moreover, following the Modigliani Miller Theorem, a company should not affect a specific financial resource (here profit) to a specific object (here WCR financing). In consequence, we adopt the scheme that the WCR generated by producing a product is only effectively recovered when that product is sold. Thus, we progressively and uniformly receive all production related costs from the sales revenue of products over time.

4 Proposed modeling of the WCR

4.1 Assumptions

The model proposed in this paper is built under the same assumptions as in the classic MLLP model. Furthermore, in order to model the WCR, especially in the multi-level context, the following financial assumptions are added:

- The payment from site $i + 1$ to site i for the ordered goods is executed after receipt of the products;
- Payments of all logistic and financial costs are made at the beginning of periods;
- The margin resulting from the sale of products will not be used for financing the WCR;
- Products in the same lot uniformly share the setup cost. However, the inventory holding cost is measured for each product according to its total holding time in the inventory;
- The purchasing unit cost of site $i + 1$ is higher than the unit cost for producing one unit at site i ;
- The discount rate of site i is greater than the one of site $i + 1$;
- The payment delay from site $i + 1$ is shorter or equal to the one from site i ;
- The interest rates of the two levels are assumed equal in the centralized approach.

4.2 Model

The notations are summarized in Table 1.

At each period, the WCR is composed by four terms, for covering the purchasing (*purchase*), setup (*setup*), production (*production*) and inventory holding (*inventory*) costs. The formulation of the WCR in this case is similar than the ones in the single-level case for covering the different logistic costs associated to the disaggregated production quantity, X_{itk} (where $i \in \{0, N - 1\}$, $t \in [0, T - 1]$ and $k \in [t, T - 1]$). For instance, $WCR_{itk}^{purchase}$ represents the WCR for financing the purchasing cost related to X_{itk} . In the general formulation proposed in below, we consider the begging of the

Parameters

T	Horizon length
d_{it}	Demand in period t for site i
v_i	Unit product price
h_i	Unit inventory cost per period
p_i	Unit production cost
s_i	Fixed setup cost
a_i	Unit raw material cost
L_i	Delivery delay site i to site $i - 1$
r_i	Delay in payment from site i to site $i - 1$
α_i	Discount rate per period of site i
β_i	Interest rate for financing OWCR of site i

Decision variables

Q_{it}	Total production quantity in period t of site i
X_{itk}	Production quantity in period t for satisfying (a part of) demand in period k of site i
I_{it}	Inventory level at the period t of site i
Y_{it}	setup variables

Table 1 – Parameters and decision variables for OWCR modeling

global planning horizon to be at the beginning of the supplier's production. For this reason, the delivery delay to site i will postpone its production timings in the global planning and add additional discount effects to the incoming and outgoing cash flows of site i . Therefore, for instance, the first term in the formulation of the manufacturer's WCR, $\frac{1}{(1+\alpha_i)^{L_i}}$ reflects the discount effect of this backward shifting of the planning horizon for L_i periods. The WCR of site i is formulated with the following equations, where $i \in \{0, N - 1\}$, $t \in [0, T - 1]$ and $k \in [t, T - 1]$.

$$WCR_{itk}^{purchase} = \prod_{l=0}^i \frac{1}{(1+\alpha_i)^{L_l}} \times$$

$$a_i \cdot X_{itk} \cdot \left(\sum_{j=t+r_i}^T \frac{1}{(1+\alpha_i)^j} - \sum_{j=k+L_{i+1}+r_{i+1}}^T \frac{1}{(1+\alpha_i)^j} \right)$$

$$WCR_{itk}^{production} = \prod_{l=0}^i \frac{1}{(1+\alpha_i)^{L_l}} \cdot p_i \cdot X_{itk} \times \sum_{j=t}^{k+L_{i+1}+r_{i+1}-1} \frac{1}{(1+\alpha_i)^j}$$

$$WCR_{itk}^{setup} = \prod_{l=0}^i \frac{1}{(1+\alpha_i)^{L_l}} \cdot \frac{s_i \cdot Y_{it}}{Q_{it} + 1 - Y_{it}} \times$$

$$X_{itk} \cdot \sum_{j=t}^{k+L_{i+1}+r_{i+1}-1} \frac{1}{(1+\alpha_i)^j}$$

$$WCR_{itk}^{inventory} = \prod_{l=0}^i \frac{1}{(1+\alpha_i)^{L_l}} \cdot h_i \cdot X_{itk} \times$$

$$\sum_{w=t}^{k-1} \sum_{j=w}^{k+L_{i+1}+r_{i+1}-1} \frac{1}{(1+\alpha_i)^j}$$

Combining all these terms, the total WCR for site $i \in \{0, N - 1\}$ is given by the eq. (1)

$$WCR_i = \sum_{t=0}^{T-1} \sum_{k=t}^{T-1} \left(WCR_{itk}^{purch} + WCR_{itk}^{prod} + WCR_{itk}^{setup} + WCR_{itk}^{inv} \right) \quad (1)$$

and the financing cost of WCR_i will thus be $\beta_i \times WCR_i$.

5 Proposed approaches

This part begins with the expression of the objective function. Then we present the two developed approaches:

1. a sequential resolution based on the resolution of N independent single-level sub-problems with a dynamic programming approach described by Bian et al. (2018);
2. a centralized approach that consists in solving the problem with a global profit maximization objective using a modified dynamic programming algorithm of Bian et al. (2018).

5.1 Objective function

The objective function of these approaches is to maximize the profit corresponding to the difference between the revenue and the total logistic costs. The revenue of site i is $R_i = v \times d_i$ where $d_i = Q_{i-1}$, when $i \neq 0$. Moreover, the logistic costs of purchasing, setup, production and inventory holding activities of site i are formulated as follows:

$$LC_i^{purch} = \prod_{l=0}^i \frac{1}{(1 + \alpha_i)^{L_l}} \cdot a_i \sum_{t=1}^T \frac{Q_{it}}{(1 + \alpha_i)^{t+r_i}} \quad (2)$$

$$LC_i^{prod} = \prod_{l=0}^i \frac{1}{(1 + \alpha_i)^{L_l}} \cdot p_i \cdot \sum_{t=1}^T \frac{Q_{it}}{(1 + \alpha_i)^t} \quad (3)$$

$$LC_i^{setup} = \prod_{l=0}^i \frac{1}{(1 + \alpha_i)^{L_l}} \cdot s_i \cdot \sum_{t=1}^T \frac{Y_{it}}{(1 + \alpha_i)^t} \quad (4)$$

$$LC_i^{inv} = \prod_{l=0}^i \frac{1}{(1 + \alpha_i)^{L_l}} \cdot h_i \cdot \sum_{t=1}^T \sum_{k=t}^T \sum_{m=t}^{k-1} \frac{X_{itk}}{(1 + \alpha_i)^m} \quad (5)$$

Thus, the objective of the sequential approach prioritizes the profit maximization of site $N - 1$, then the one of site $N - 2$ until site 0. In the process, the optimal production plan of the previous level is used as the demand of the following site. Furthermore, the objective of the centralized approach consist in maximizing the total profit of all sites by summing up their profits.

5.2 Sequential resolution

This approach consists in solving N single-level problems with a dynamic programming of Bian et al. (2018) beginning with the level nearest end customer (site $N - 1$). The obtained production plan of site i will be used as the demand for site $i - 1$.

5.3 Centralized resolution

In this approach, we define simultaneously the production plans of all levels. The general mathematical formulation of the centralized approach for the multi-level problem can thus be written as:

$$Max \sum_{i=0}^{N-1} [R_i - (LC_i + \beta_i WCR_i)] \quad (1.1)$$

s.t.

$$\begin{aligned} I_{i,t+L_i-1} + Q_{it} - I_{i,t+L_i} \\ = d_{i,L_i+t} + Q_{j,L_i+t} \quad \forall i \in [0, N - 1] \\ \forall t \in [T_i + 1, T_i + T] \end{aligned} \quad (1.2)$$

$$\begin{aligned} MY_{it} - Q_{it} \geq 0 \\ \forall (i, t) \\ \in [0, N - 1] \times [1, T] \end{aligned} \quad (1.3)$$

$$\begin{aligned} Q_{it}, I_{i,t+L_i} \in \mathbb{N}^{+2} \\ \forall i \in [0, N - 1] \\ \forall t \in [T_i + 1, T_i + T] \end{aligned} \quad (1.4)$$

$$\begin{aligned} Y_t = 0 \text{ or } 1 \\ \forall (i, t) \\ \in [0, N - 1] \times [1, T] \end{aligned} \quad (1.5)$$

Constraints (1.2) represent inventory balance. Constraints (1.3) determine if a setup for item i occurs at period t . Constraints (1.4) are the integrity constraints. Constraints (1.5) determine that the setup variables are binary.

After proving that the zero-inventory property remain valid with the given assumptions, to efficiently solve this problem, it is reformulated as a shortest path problem. To that purpose, a dynamic programming algorithm is developed. In this method, determining the production plan of site i requires the optimal results of corresponding subproblems of site $i + 1$. In Figure 2, an example is given to explain the arc valuation in the two-level case. The reason for setting the S_0 on top of the S_1 is that the production at S_1 can only be executed when the intermediate products are prepared by S_0 .

As shown in Figure 2, in order to calculate the exact arc value of E_{25}^0 , the corresponding subproblem at the S_1 level between the node 2 and 5 must be priorly solved. It is because that a part of costs of S_0 depend on both the plans of S_0 and S_1 . These costs, denotes as "variable" costs, includes the inventory holding cost and the WCR financing cost. We need to calculate the plan of S_1 to determine the holding periods and the timings of revenue collections from S_1 with which the financing duration can be deduced. Moreover, other costs of one site that only depend on its own plan are denoted as "independent" costs.

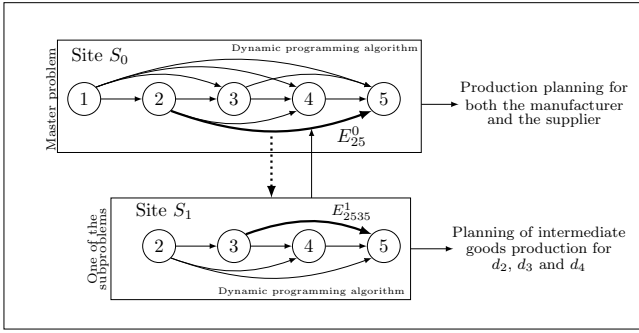


Figure 2 – Centralized approach in two level case

In a multi-level case, all intermediate levels have two adjacent levels. The arc value of an intermediate level i thus includes the "variable" costs of its lower level $i - 1$, the "independent" costs that only depend on the plan of site i and the optimal result of its corresponding subproblem at level $i + 1$. Moreover, the arc value of S_0 is the sum of its "independent" costs and the optimal result of the corresponding subproblem. At last, the arc value of S_{N-1} is the sum of the "variable" costs of S_{N-2} and all logistic and financial costs of S_{N-1} . After determining the arc value of all levels, the recurrence is similar to the one used for the single-level problem. However, we need to execute it from the S_{N-1} level to the S_0 level in order to compute the optimal results of the subproblems. The arc values of these three types of levels are formulated as follows:

- For site 0, the planning problem is considered as a master problem in which the formulation of the arc value is described as above and the recursion is similar to that of Bian et al. (2018):

$$\begin{aligned}
 - \text{Arc value: } E_{tk}^0 &= (a'_{0,t+r_f} + p'_{0t}) \sum_{l=t}^{k-1} d_l + s'_{0t} + Opt_{tkk}^1 \\
 - \text{Recursion: } Opt_t^0 &= \min_{j \in [0, t-1]} \{Opt_j^0 + E_{jt}^0\}
 \end{aligned}$$

- For an intermediate site i where $i = 1, \dots, N - 2$, the first three terms in the arc value formulation presented below are the "variable" costs, the last term is the optimal result of the corresponding subproblem and the other terms are the "independent" costs:

$$\begin{aligned}
 - \text{Arc value: } E_{tkt'k'}^i &= Inv_{tt'k'}^{i-1} + \beta_{i-1} WCR_{tkt'k'}^{i-1} - R_{t'k'}^{i-1} + (a'_{i,t'+r_1} + p'_{it'}) \sum_{l=t'}^{k'-1} d_l + s'_{it'} + Opt_{t'k'k'}^{i+1} \\
 - \text{Recursion: } Opt_{tkk'}^i &= \min_{t' \in [t, t-1]} \{Opt_{tkt'}^i + E_{tkt'k'}^i\}
 \end{aligned}$$

- For site $N - 1$, since it only responds to the known external demand, the revenue is thus fixed and is not considered in the optimization process:

$$\begin{aligned}
 - \text{Arc value: } E_{tkt'k'}^{N-1} &= Inv_{tt'k'}^{N-2} + \beta_{N-2} WCR_{tkt'k'}^{N-2} - R_{t'k'}^{N-2} + LC_{t'k'}^{N-1} + \beta_{N-1} WCR_{t'k'}^{N-1} \\
 - \text{Recursion: } Opt_{tkk'}^{N-1} &= \min_{t' \in [t, k'-1]} \{Opt_{tkt'}^{N-1} + E_{tkt'k'}^{N-1}\}
 \end{aligned}$$

In the arc value formulation presented above, thanks to the ZIO property, the production quantities are in function of the (entire) external demand of one or several periods. Therefore, it is possible to "peel off" the computation of all the shortest paths by level. At each level, all shortest paths between each pair of nodes of this level are computed and saved as well as the associated optimal value. For any level i except for level 0, these shortest paths and the optimal values for level i are also saved for further calculations of arc values at level $i - 1$. For level 0, we only need to keep one shortest path from the beginning of the planning horizon to the end. To initiate the algorithm, we start from level $N - 1$ who has only one predecessor.

The pseudo-code is presented in Algorithm 1.

Algorithm 1: Solving the $MLLP_{P(WCR)}$

Require: All parameter values

```

Level=N-1;
while Level ≥ 0 do
  for k = 1 to T do
    for t = 0 to k - 1 do
      if Level ≠ 0 then
        for l = t + 1 to k do
          for q = t to l do
            if Opt_{tll}^{Level} > Opt_{tlq}^{Level} + E_{tlq}^{Level}
            then
              Opt_{tll}^{Level} = Opt_{tlq}^{Level} + E_{tlq}^{Level}
            end if
          end for
        end for
      else {Level = 0}
        if Opt_t^0 > Opt_k^0 + E_{kt}^0 then
          Opt_t^0 = Opt_k^0 + E_{kt}^0
        end if
      end if
    end for
  end for Level = Level - 1;
end while

```

The loops t and k are for:

- generating all possible pairs of nodes for any level except for level 0;
- the recursion of dynamic programming for level 0.

Then, the loops l and q are the recursion of the dynamic programming for other levels than level 0. In total, for each of these levels (other than 0), there are $O(T^2)$ pairs of nodes and each possible shortest path between one pair of nodes is determined in $O(T^4)$ because it is based on the single-level type algorithm. Besides, the resolution at level 0 takes $O(T^2)$ for one shortest path to solve. Globally, for this N level problem, the overall complexity is $O(NT^6)$.

6 Numerical tests in two-level case

In this section, some tests results of a supplier-manufacturer two-level case are provided to show the consequences of integrating the financing cost of the WCR. Moreover, the influence of some parameter variations on the objective value is illustrated.

- Firstly, a comparison of optimal plans obtained by the classic MLLP model and the proposed approaches is provided to show the differences of the production programs;
- Secondly, differences on the manufacturer's total cost, the profit of the supplier and the objective value are calculated using the centralized approaches by varying payment delays;
- Lastly, the influence of the variation of the selling price of the supplier on the production plans with the centralized approach is tested.

For the following tests, we adopt the demand of one instance of Dellaert et al. (2000) (ph2in01st1de01, demand 1) over 24 periods. Values of other parameters are given in Table 2. These values will be used by default without further specification. The main feature of these inputs are:

- the unit selling price of the manufacturer is given in the following table. However, the manufacturer's revenue is constant with a fixed unit price and fixed external demand. Thus, this unit price does not impact the production program. However, the unit price of the supplier should be taken into account and be tested;
- the delivery delay is not considered in the following tests.

6.1 Test 1: production program comparisons

We first compare optimal programs separately calculated with the classic MLLP model, the proposed sequential approach and with the centralized approach. In order to do this comparison with the same parameters considered in the classic MLLP model, we thus set $a_i = p_i = v_i = 0$ with $i = 0, 1$ and $r_0 = 0$

Table 2 – Parameter values for the 24 periods instance

Parameter	Supplier	Manufacturer
Unit selling price	35	50
Production unit cost, p_i	3	3
Setup cost, s_i	800	600
Purchasing unit cost, a_i	3	35
Inventory holding unit cost per period, h_i	1	2
Discount rate, α	0.05	0.01
Interest rate, β	0.03	0.03
Delays in payment	$r_2 = r_1 = 2, r_0 = 1$	

to focus on the setup cost and inventory holding cost. Moreover, since the purchasing cost is not considered in test 1, the payment delay for the raw material, r_0 is thus irrelevant in this test. The number of setups in the obtained optimal plans and the total number of products held in the inventory are given in Table 3. As a result, we observe that

- through the plans obtained by the MLLP model, we prefer to hold the finished products at the manufacturer than holding the semi-finished products at the supplier. This is because the setup cost of the supplier is more expensive than the one of the manufacturer and the difference of unit holding cost between the two levels is relatively small. The optimal plans of the two level are almost synchronized, which allows to reduce the inventory level at the supplier.
- comparing the plans without and with the financial aspects that we consider in this two-level problem, the number of setups is generally increased (from 5 to 6 at the supplier level and from 6 to 10 or 8 at the manufacturer level). This confirms the observation that we found in the tests for single-level problem, that we have more interest to produce with a smaller lot size in order to suffer less from the financial consequences;
- on the one hand, using the sequential approach will significantly reduce the manufacturer's inventory level for its prioritized optimization objective. On the other hand, the centralized approach balances the inventory levels between the two levels for a global optimization objective.

6.2 Test2: program comparisons with different payment delays

In the following test, we focus on the influence of payment delays on the two levels with the centralized approach. Parameter values are adopted as in Table 2. β_0 and β_1 are changed to 0.05. A comparison of the supplier's profit ($Profit_0$), the manufacturer's total cost (TC_1) and the objective va-

Table 3 – Comparison of the total products held in inventory over time and of the number of setups

Approaches	Supplier	Manufacturer
	Number of setups	
<i>MLLP</i>	5	6
<i>Sequential</i>	6	10
<i>Centralized</i>	6	8
	Total products held (in unit/period)	
<i>MLLP</i>	670	3085
<i>Sequential</i>	1840	1230
<i>Centralized</i>	1005	2065

lue (OV) changes between $r_i = 2$ and $r_i = 18$ ($i \in \{0, 1, 2\}$) is given in Table 4. For example, $\Delta Profit_0 = Profit_0(r_i = 18) - Profit_0(r_i = 2)$. Remember that OV represents the sum of the logistic costs and the WCR financing costs of both levels minus the supplier's revenue.

Table 4 – Comparison of changes with the variations of payment delays using the centralized approach

parameter	$\Delta Profit_0$	ΔTC_1	ΔOV
r_2	762.9	60742.7	59979.8
r_1	-30869.3	-61904.4	-31035.1
r_0	4321.69	0	-4312.7

Since we attempt to minimize the OV , it is obvious that r_2 should be shortened as much as possible in order to reduce the unnecessary WCR financial cost of the manufacturer. Moreover, it is beneficial to lengthen r_1 and r_0 that may further decrease the OV . With the increase of r_1 , the decrease of the OV mainly results from the decreased total cost of the manufacturer and the decrease of the supplier's profit. However, the increase of r_0 only influences the supplier's profit in a positive way and decreases the OV . According to the amplitude of these changes, from the management point of view of each site, it should preferentially ensure the upstream payment that more significantly impacts its profit and the global profit than the changes brought by the downstream payment. It is reinforced by the fact that the WCR financing cost of all operation costs depends on the upstream payment delay that should be managed as a higher-priority task. In addition, from a global point of view, since the payment delay between sites conversely impacts the two sites, the negotiation should be carefully managed for satisfying both sites.

6.3 Production program evolution with different unit selling price of the supplier

In the single-level problem, the revenue does not change with a deterministic and fixed external de-

mand. Therefore, the unit selling price has no influence in determining the optimal production plan. However, it is no longer the case in this two-level problem because the demand of the supplier is the production plan of the manufacturer and the unit selling price of the supplier is the unit purchasing cost of the manufacturer. Consequently, the supplier's revenue depends on the manufacturer's plan considering the discount effect. In the following tests, $v_0 (= a_1)$ varies from 35 to 65 by step of 15. Other parameter values are the same as in Table 2.

- subsequently, the influence of the increase of v_0 on the optimal plans with the centralized approach is tested. Generally, an increase of the number of setups is found on both levels. The increase of setups of the manufacturer's level confirms the finding in Bian et al. (2018). As a consequence of the centralized objective, the increase of the selling price will thus decrease the production quantities of the supplier. As shown in Table 5, the increase of v_0 will obviously increase the supplier's profit but more significantly increase the total cost of the manufacturer that globally increase the objective value. Therefore, the pricing problem for the supplier should also consider the consequences at the manufacturer level with the centralized objective.

Table 5 – Comparison of the profit of the supplier, the total cost of the manufacturer and the objective value with different selling prices using the centralized approach

Varying parameter	$Profit_0$	TC_1	OV
$v_0 = 35$	32156.4	86848.2	54691.8
$v_0 = 50$	51065.9	117431	66364.6
$v_0 = 65$	68431.2	146412	77981.1

7 Conclusion and perspective

In this work, firstly, we extend the previous single-level problem presented in Bian et al. (2018) to a multi-level problem with a serial chain structure. The mathematical model and the associated WCR formulation are further extended to the multi-level case the proof of the ZIO property with some assumptions. In order to solve this multi-level problem. the sequential and the centralized approach are proposed based on the dynamic programming approach. The former is a direct application of the single-level algorithm of Bian et al. (2018). The latter tackles the problem level by level in $O(NT^6)$ where N represents the number of levels of the considered supply chain.

For future research, a generalization of the chain structure is naturally brought out. Other supply chain structures, such as the assembly and the general

structure need to be addressed. Moreover, the development of heuristic or meta-heuristic algorithms will help us to solve problems with a more realistic size and with fewer assumptions. Furthermore, in addition to the two approaches proposed in this work, the decentralized approach may also be taken into consideration for integrating the interaction between the supply chain members. The negotiation mechanism as presented in Homberger (2010) and game theory techniques as studied in the work of Hovelaque et al. (2015) could be considered.

References

- Babich, V. & Sobel, M. J. (2004). Pre-ipo operational and financial decisions, *Management Science* **50**(7): 935–948.
- Bian, Y., Lemoine, D., Yeung, T. G., Bostel, N., Hovelaque, V., Laurent Viviani, J. & Gayraud, F. (2018). A dynamic lot-sizing-based profit maximization discounted cash flow model considering working capital requirement financing cost with infinite production capacity, *International Journal of Production Economics* **196**: 319 – 332.
- Dellaert, N., Jeunet, J. & Jonard, N. (2000). A genetic algorithm to solve the general multi-level lot-sizing problem with time-varying costs, *International Journal of Production Economics* **68**(3): 241–257.
- Deroussi, L. & Lemoine, D. (2009). A particle swarm approach for the mllp, *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on*, IEEE, pp. 12–17.
- Eppen, G. D. & Martin, R. K. (1987). Solving multi-item capacitated lot-sizing problems using variable redefinition, *Operations Research* **35**(6): 832–848.
- Grubbström, R. W. (1967). On the application of the laplace transform to certain economic problems, *Management Science* **13**(7): 558–567.
- Grubbström, R. W. & Huynh, T. T. T. (2006). Multi-level, multi-stage capacity-constrained production–inventory systems in discrete time with non-zero lead times using mrrp theory, *International Journal of Production Economics* **101**(1): 53–62.
- Grubbström, R. W. & Wang, Z. (2003). A stochastic model of multi-level/multi-stage capacity-constrained production–inventory systems, *International Journal of Production Economics* **81**: 483–494.
- Gupta, Y. P. & Keung, Y. (1990). A review of multi-stage lot-sizing models, *International Journal of Operations & Production Management* **10**(9): 57–73.
- Helber, S. (1998). Cash-flow oriented lot sizing in mrrp ii systems, *Beyond Manufacturing Resource Planning (MRP II)*, Springer, pp. 147–183.
- Homberger, J. (2008). A parallel genetic algorithm for the multilevel unconstrained lot-sizing problem, *INFORMS Journal on Computing* **20**(1): 124–132.
- Homberger, J. (2010). Decentralized multi-level uncapacitated lot-sizing by automated negotiation, *4OR* **8**(2): 155–180.
- Homberger, J. & Gehring, H. (2009). An ant colony optimization approach for the multi-level unconstrained lot-sizing problem, *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, IEEE, pp. 1–7.
- Hovelaque, V., Viviani, J.-L. & Mansour, M. A. (2015). Chaîne logistique sous contrainte budgétaire, *CIGI, Quebec City, October 2015*.
- Jeunet, J. & Jonard, N. (2005). Single-point stochastic search algorithms for the multi-level lot-sizing problem, *Computers & operations research* **32**(4): 985–1006.
- Martínez-Costa, C., Mas-Machuca, M., Benedito, E. & Corominas, A. (2014). A review of mathematical programming models for strategic capacity planning in manufacturing, *International Journal of Production Economics* **153**: 66–85.
- Melo, R. A. & Wolsey, L. A. (2010). Uncapacitated two-level lot-sizing, *Operations Research Letters* **38**(4): 241–245.
- Pochet, Y. & Wolsey, L. A. (2006). *Production planning by mixed integer programming*, Springer Science & Business Media.
- Theodore Farris, M. & Hutchison, P. D. (2002). Cash-to-cash: the new supply chain management metric, *International Journal of Physical Distribution & Logistics Management* **32**(4): 288–298.
- Veral, E. A. & LaForge, R. L. (1985). The performance of a simple incremental lot-sizing rule in a multilevel inventory environment, *Decision Sciences* **16**(1): 57–72.
- Xiao, Y., Zhang, R., Zhao, Q., Kaku, I. & Xu, Y. (2014). A variable neighborhood search with an effective local search for uncapacitated multilevel lot-sizing problems, *European Journal of Operational Research* **235**(1): 102–114.
- Zangwill, W. I. (1966). A deterministic multiproduct, multi-facility production and inventory model, *Operations Research* **14**(3): 486–507.

ON THE FORECASTING OF TRAFFIC ACCIDENTS BASED ON REAL DATA

Fernando A. Crespo

Elyn L. Solano-Charris

Escuela de Ingeniería Industrial
Facultad de Ciencias, Universidad Mayor
Manuel Montt 367
Providencia, Santiago, Chile
fernando.crespo@umayor.cl

EICEA
Universidad de La Sabana
Km. 7, Autopista Norte de Bogotá
Chía, Cundinamarca, Colombia
erlyn.solano@unisabana.edu.co

Jairo R. Montoya-Torres

Grupo de investigación en Sistemas Logísticos
Facultad de Ingeniería, Universidad de La Sabana
Km. 7, Autopista Norte de Bogotá D.C.
Chía, Cundinamarca, Colombia
jairo.montoya@unisabana.edu.co

ABSTRACT: *In this paper, a preliminary method is proposed to forecast traffic accidents based on Big data analytics. The study is conducted using the traffic accident records from Bogotá, Colombia from December 21, 2006 to April 29, 2017. The accidents are visualized considering Spatial-temporal modelling. Then, a characterization and prediction of accidents and the level of damage are done. For the problem, the computation of the Inverse Distance Weighted (IDW) Interpolation with the Radial Basis Function (RBF) in R and the Kriging model are implemented. The preliminary results pretend to support decision making for anticipating possible disruptions of the last-mile supply chain.*

KEYWORDS: *Traffic accidents, Big data analytics, Spatial-temporal modeling, Distance Weighted Interpolation, Kriging model.*

1 INTRODUCTION

Traffic accident predictions is one of the critical tasks in Intelligent Transportation Systems (ITS) due to its impact in traffic flow, cost, time and the environment. Then, different methods for traffic accident predictions have been implemented including time-series, Markov chains, neural networks and autoregressive integrated moving average, among others. However, these methods are not able to satisfy real application requirements (Chen et al., 2017). During last years, most studies have implemented data analytics to solve traffic flow predictions and the results demonstrate that such schemes are feasible and are able to improve the accuracy of prediction (Lv et al., 2015).

As summarized in Vlahogianni et al. (2014) and stated in Li et al. (2015), the main challenge for traffic flow predictions lies in how to correctly retrieve both temporal and spatial dependence of traffic flow time series collected at different locations.

The objective of this paper is use data analytics techniques to predict possible traffic accidents in urban context. Since this is an exploratory study linking data analytics and urban logistics, the focus will be on forecasting location and severity of traffic accidents. The city of Bogotá, Colombia, is taken as a research case since it allows to have a reference example of the behavior of cities in emerging economies.

It is important to note as well that the model and results presented here are part of a bigger project integrating both qualitative and quantitative research methods. The former are employed to characterize logistics distribution processes in megacities, based on semi-structured interviews of logistics / supply chain managers in companies. The latter corresponds to the work presented in this paper whose goal is to link such characterization with numerical inputs hence possible disruptions of the last-mile supply chain can be anticipated.

The remainder of this paper is organized as follows. Section 2 presents an overview about traffic predictions. Section 3 presents the methods employed to forecast the occurrence of accidents. Section 4 presents the findings. Finally, some concluding remarks are drawn in Section 5.

2 OVERVIEW OF RELATED LITERATURE

In the field of traffic engineering, several studies have been published, in particular for addressing statistical studies of accident occurrence and location, some of them being descriptive (see for example the works of Vargas et al., 2012, Arias et al., 2017), while other works employ advanced techniques based on Big data analytics. The latter has been implemented by various authors for traffic predictions (Chen et al., 2017; Li et al., 2015; Lu et al., 2015, Abdullah and Emam, 2015; Heilmann et al., 2011; Bhaskar et al., 2011; Zhang et al., 2011; Faouzi et al., 2011; Van Lint and Hoogendoorn, 2010; Vlahogianni et al., 2004; Chrobok et al., 2004; Smith et al., 2002).

However, the quantity of data generated with ITS systems makes computationally prohibitive the incorporation of all data into one prediction model (Li et al., 2015). Then, alternative solutions to this inconvenient have been studied such as parallel computing techniques and data reduction for considering the most relevant data to build predictions models (Stathopoulos et al., 2010; Kamarianakis et al., 2012; Sun et al., 2012). Here, we study data reduction for making this study tractable and for predicting traffic accidents in a big city.

3 METHODS

In order to predict the occurrence of accidents in Bogotá, Colombia, secondary data taken from the website of the Traffic Office of Bogotá, belonging to the Secretary of Mobility, was collected from <http://movilidadbogota.maps.arcgis.com/apps/webappviewer/index.html?id=d1b4eac89bc24edc95fd\3cc6e88dc96d>.

The area uses the x-coordinates between -74.17458 and -74.01296, and y-coordinates between 4.64573 and 4.716183. For the exploratory approach, the studied area includes the zones of Usaquén, Barrios Unidos and Engativá, which corresponds to a central zone of the city. The database about accidents included records from December 21, 2006 to April 29, 2017, giving a total of 51,324 cases of crash accidents, with its damage level. To keep the analysis as simple as possible, only car accidents were considered. The damage level was set to be of level 1 if the crash only presents damage in cars, level 2 with injured people and level 3 with death.

Data analyses were run using R Project software (<http://www.r-project.org>), employing libraries for graphics and geostatistical proposes. The analysis was carried out into three steps:

1. A descriptive analysis with accident characterizations, in particular to look for the patterns of accidents.
2. A computation of the Inverse Distance Weighted (IDW) Interpolation with Radial Basis Function (RBF) in R to predict the level of damage in accidents.
3. A kriging model to predict the accidents. To do so, the Fixed Rank Kriging (FRK) package is used since 51,324 cases is a big data set to calculate a normal kriging model. Although the computational complexity of the original kriging method is $O(n^3)$, our analysis employed the Fixed Rank Kriging, with complexity of $O(n)$, with n being the quantity of data (Cressie and Johannesson, 2008). This reduction is possible because the latter method uses spatial covariance function, with basis functions, which are independent on the size of data sample. Mathematical details are explained in (Cressie and Johannesson, 2008).

The advantages of using IDW and kriging is that these are two widely employed techniques because of their simplicity and hence the easiness to be expanded for analyzing very big data sets. In addition, interpretation of results can be done employing metaphors based on the knowledge of both techniques.

To model the Inverse Distance Weighted Interpolation, the Shepard (1968) Model was executed with the IDW function in R (IDW, 2017). The goal of this step is to interpolate damage level, denoted as over u the point $x \in \mathbb{R}^2$. The interpolating function is:

$$u(x) = \begin{cases} \frac{\sum_{i=1}^N w_i(x) u_i}{\sum_{i=1}^N w_i(x)}, & d(x, x_i) \neq 0 \forall i \\ u_i & d(x, x_i) = 0 \text{ for some } i \end{cases} \quad (1)$$

Where u_i is the known point and $w_i(x) = \frac{1}{d(x, x_i)^p}$. The value of $p = 5$ was set because the higher the value of p the lower the dependence of far away points. It is necessary to understand that Equation (1) is the

minimization problem from:

$$\phi(x, u) = \left(\sum_{i=0}^N \frac{(u - u_i)^2}{d(x, x_i)^p} \right)^{1/p} \quad (2)$$

IDW uses the measured values surrounding the predicted location to forecast a value for any unsampled location, based on the assumption that things that are close to one another are more alike than those that are farther apart. The predicted value is limited to the range of values used to interpolate. Because IDW is a weighted distance average, the average cannot be greater than the highest or less than the lowest input. Therefore, it cannot create ridges or valleys if these extremes have not already been sampled. IDW can produce a “bull’s-eye effect” around data locations. But unlike other interpolation methods, IDW does not make explicit assumptions about the statistical properties of the input data. IDW is often used when the input data does not meet the statistical assumptions of more advanced interpolation methods. This method is well-suited to be used with very large input datasets. The process is explained by Shepard (1968). The variables close to each other are significant, hence if the points are very close, it generates a point where the data is much stronger than in others. This method is used in a basic way to make the hotspot estimation, it is an algorithm that strongly depends on this. According to Goodchild et al. (2015), it is the most used method for geostatistical data analysis.

For the kriging problem it was not possible to use the traditional libraries in R (e.g., gstat) because of the size of the dataset. Instead, FRK package was employed (FRK, 2017), which was made public in May 2017. A domain D is needed in which the spatial variables are found. The proposed model is:

$$Y_i = t_i^T \alpha + v_i + \xi_i \quad (3)$$

where $Y_i = \frac{1}{|A_i|} \int_{A_i} Y(s) ds$ is the average over the grid of domain, the same for $t_i = \frac{1}{|A_i|} \int_{A_i} t(s) ds$ and $v_i = \frac{1}{|A_i|} \int_{A_i} v(s) ds$, being $v(s)$ and ξ_i the spatially correlated and the uncorrelated random effects, respectively. The model supposes $v(s) = \phi(s)^T \eta$, and $\phi(s)$ are functions of multiple resolutions (e.g., wavelet).

Kriging method was chosen because it is the best linear unbiased prediction of spatial data (Christensen, 1991), understanding that for the particular case under study a library of georeferenced data containing a big data FRK package (FRK, 2017) was employed. In particular, Fixed Rank Kriging is a tool

for spatial/spatio-temporal modelling and prediction with large datasets. As discussed by Cressie and Johannesson (2008), “the approach decomposes the field, and hence the covariance function, using a fixed set of n basis functions, where n is typically much smaller than the number of data points (or polygons) m . The method naturally allows for non-stationary, anisotropic covariance functions and the use of observations with varying support (with known error variance). The projected field is a key building block of the Spatial Random Effects (SRE) model, on which this package is based. The package FRK provides helper functions to model, fit, and predict using an SRE with relative ease”.

4 NUMERICAL SETTING AND ANALYSIS OF RESULTS

4.1 Case Study: City of Bogotá, Colombia

Bogotá is the capital and the largest city of Colombia. Its population is around 10 million of inhabitants. It is the fifth largest city in Latin America and twenty-fifth in the world (City Majors, 2015). The selection of Bogotá as the city under study allows us to have a complex and complete example of the behavior of cities in emerging economies.

4.2 Validation and Descriptive Pattern

The model was validated using a sample taken from the data set; Figure 1 presents the plot of randomly selected points, so the scale shows the confidence probability of the Error Type-I of model validity, which is estimated in a value lower than 10%. The kriging model was used to uncover the dependency rules and do the predictions. To do so, the method creates the variograms and covariance functions to estimate the spatial autocorrelation values that depend on the model of autocorrelation (fitting a model) and predicts the unknown values.

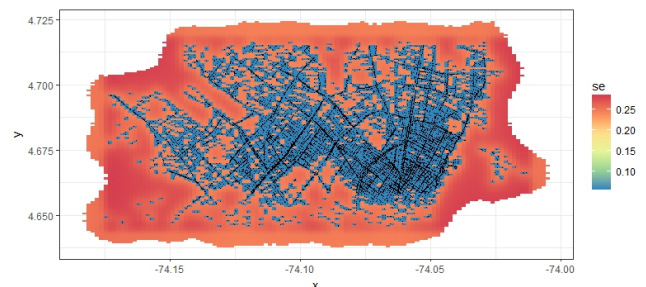


Figure 1 – Cars Accidents in a zone of Bogotá
Results of the descriptive pattern are now presented. Figure 2 shows the map of accidents. Accidents are

more frequent on Thursdays (see Figure 3) and frequently occurs the 18th day of each month (see Figure 4), day in which public salaries are paid in Bogotá. Figure 5 shows the Heat-map of levels of accidents. Red zones represents crashes with death of people. Besides, numerical outputs presented a pick of accidents in the morning between 7:30 am and 7:40 am, which corresponds to the period of time to move from home to schools and to jobs; while a second peak of accidents is about 4:00 pm corresponding to the time at which the official restriction for circulation of private cars starts. This restriction depends on the last number of the car license plate.

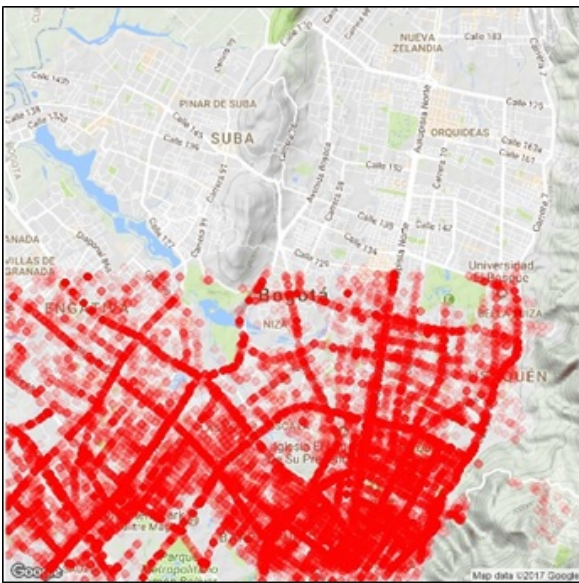


Figure 2 – Cars Accidents in a zone of Bogotá

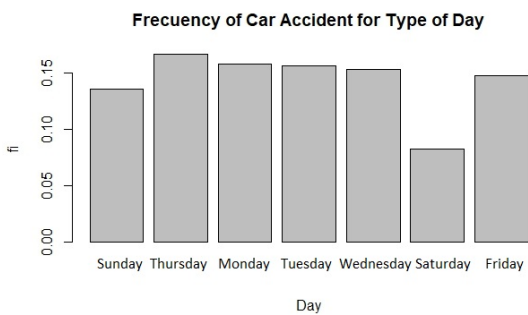


Figure 3 – Type of accidents per day

4.3 Inverse Distance Weighted Interpolation

Figure 6 shows the estimation using the Inverse Distance Weighted Interpolation with a grid with 150 cells for side. When comparing Figures 6 and 5, similar zones are observed with respect to high damage of accidents, but some other zones also appears in Figure 6.

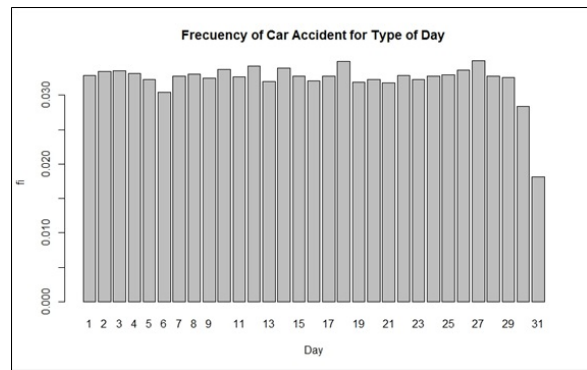


Figure 4 – Frequency of accidents per day in a month

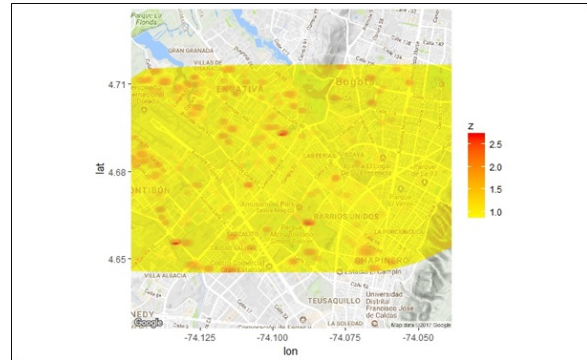


Figure 5 – Heat-map of level of accident in a zone of Bogotá

4.4 Fixed Ranking Kriging

We use the library FRK to interpolate with kriging the data. The value $\alpha_{\xi}^2 = 0.003587$. The Figure 7 shows the result of Prediction to Damage Level from traffic accidents. The results are more precise than the use of the method of IDW. The coefficient estimates is 1.17.

4.5 Discussion of Results

This study presents preliminary studies of a global project integrating qualitative and quantitative research methods. Characterization of last-mile delivery supply chains is to be done using semi-structured interviews while numerical analyses of historical data is employed in this paper to forecast critical points in the distribution network. Thus, possible delivery disruptions can be anticipated. Results presented in this paper aids the prediction of such critical points. The methodology, although applied to a specific megacity in Latin America, intends to be generic. Of course, particularities of the socio-economic context of Bogotá are to be taken into consideration when generalizing the results.

The proposed approach can predict the level of damage of traffic accidents, which is an important issue when studying mobility of both people and goods in the context of congested cities. Most critical zones were identified, as shown in heat-maps before. These

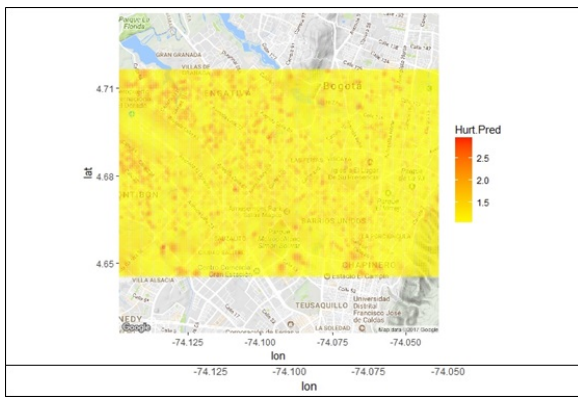


Figure 6 – Plot of Inverse Distance Weighted Interpolation

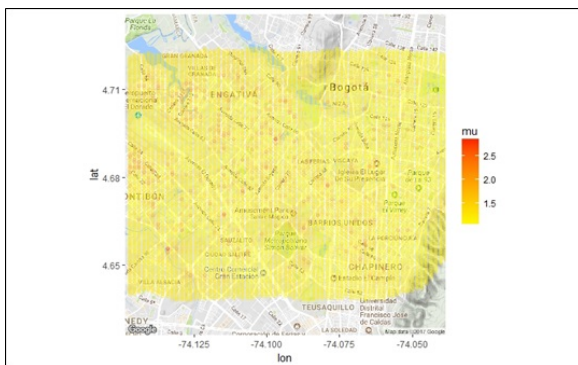


Figure 7 – Plot of FRK method to predict traffic accidents damage level

critical zones coincide with the entry points to industrial and commercial areas in the city. This is likely to disrupt goods delivery operations. Also, critical points for accidents are at the main routes to universities located within the city center, as well as to highly populated residential zones.

5 CONCLUSIONS AND FUTURE RESEARCH

This paper studied the forecasting of traffic accidents in a Latin American megacity. Big data analytics methodology was employed to analyze a large amount of data and to predict the level of damage caused by predicted accidents. Although the city of Bogotá, Colombia, was employed as a research case, the methodology can be generalized. Heat-maps can be drawn and criticality level of accidents predicted. This allowed to identify the main routes, intersections and neighborhoods of particular attention within the city when addressing the mobility of people and goods.

Future research will address the link with the characterization of last-mile goods distribution networks in Bogotá. This stage will required approaching enterprises and mapping their distribution networks. Georeferenced data will also be a key issue.

ACKNOWLEDGMENTS

This work was carried out while the first author was Visiting Scholar at Universidad de La Sabana under an Academic Mobility Fellowship from Alianza del Pacífico in 2017.

REFERENCES

- Abdullah, E., Emam, A., 2015. Traffic Accidents Analyzer Using Big Data. *Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence*, p. 392-397.
- Arias Rojas, W., Chaves Pabon, S.B., 2017. Analysis of fatalities for crash accidents in Colombia between 2011-2015. *Redes de Ingeniería*, Volumen especial, p. 226-239.
- Bhaskar, A., Chung, E., Dumont, A.-G., 2011. Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks. *Computer-Aided Civil Infrastructure Engineering*, 26(6), p. 433-450.
- Chrobok, R., Kaumann, O., Wahle, J., Schreckenberg, M., 2004. Different methods of traffic forecast based on real data. *European Journal of Operational Research*, 155(3), p. 558-568.
- Chen, Y., Guizani, M., Zhang, Y., Wang, L., Crespi, N., Myoung Lee, G., 2017. When Traffic Flow Prediction Meets Wireless Big Data Analytics. *ArXiv e-prints*, Cornell University Library.
- City Mayors. The largest cities in the world and their mayors. Available at: <http://www.citymayors.com/statistics/largest-cities-mayors-1.html>. Last accessed: 1 January 2015.
- Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society B*, 70(1), p. 209-226.
- Christensen, R. 1991. *Linear models for multivariate, time series, and spatial data*. Springer.
- FRK, 2017. FRK Project. <https://github.com/andrewzm/FRK>, Accessed: November 2017.
- Goodchild, M.F., Longley, P. A., Maguire, D.J., Rhind, D.W., 2015, *Geographic information systems and science*, 4th edition, Wiley & Sons, West Sussex, UK.
- Heilmann, B., El Faouzi, N.-E., de Mouzon, O., Hainitz, N., Koller, H., Bauer, D., Antoniou, C., 2011. Predicting motorway traffic performance by data fusion of local sensor data and electronic

- toll collection data. *Computer-Aided Civil Infrastructure Engineering*, 26(6), p. 451-463.
- IDW function, 2017. <https://www.rdocumentation.org/packages/phy\lin/versions/1.1.1/topics/idw>, Accessed: October 2017.
- Kamarianakis, Y., Shen, W., Wynter, L., 2012. Real-time road traffic forecasting using regime switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry*, 28(4), p. 297-315.
- Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., Li, Y., 2015. Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C*, 58, p. 292-307.
- Lu, H., Sun, Z., Qu, W., 2015. Big Data-Driven Based Real-Time Traffic Flow State Identification and Prediction. *Discrete Dynamics in Nature and Society*, Article ID 284906, p. 1-11 .
- Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.-Y., 2015. Traffic flow prediction with big data: a deep learning approach, *IEEE Transactions on Intelligent Transportation Systems*, 16(2), p. 865-873.
- Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 ACM National Conference*, p. 517-524.
- Smith, B.L., Williams, B.M., Oswald, R.K., 2002. Comparison of parametric and nonparametric Models for traffic flow forecasting. *Transportation Research Part C*, 10(4), p. 303-321.
- Stathopoulos, A., Karlaftis, M.G., Dimitriou, L., 2010. Fuzzy rule-based system approach to combining traffic count forecasts. *Transportation Research Record*, 2183, p. 120-128.
- Sun, S., Huang, R., Gao, Y., 2012. Network-Scale traffic modeling and forecasting with Graphical Lasso and neural networks. *Journal of Transportation Engineering*, 138(11), p. 1358-1367.
- Van Lint, J.W.C., Hoogendoorn, S.P., 2010. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil Infrastructure Engineering*, 25(8), p. 596-612.
- Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. *Transport Review*, 24, p. 533-557
- Vargas, W.E., Mozo, E., Herrera, E., 2012. Analysis of critical points more traffic accidents in Bogota, *Revista Topografica Azimut*, 4, p. 61-77.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: where we are and where we're going. *Transportation Research Part C*, 43(1), p. 3-19.
- Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., Chen, C., 2011. Data-driven intelligent transportation systems: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), p. 1624-1639.

OPTIMISATION DE L’AFFECTATION DES PATIENTS AUX MEDECINS SPECIALISTES AVEC PRISE EN COMPTE DE LA PRIORITE CLINIQUE : CAS DU CRDS MONTREALAIS

V. NAUZE

Département de Génie Mécanique, École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montréal, QC H3C 1K3, Canada

vincent.nauze@gmail.com

T. BOUKHERROUB, M. PAQUET, C. OLIVIER

Département de Génie de la Production Automatisée, École de technologie supérieure & Laboratoire Numérix, 955 rue Peel, Montréal, QC, H3C 2H8, Canada

tasseda.boukherroub@etsmtl.ca,
Marc.Paquet@etsmtl.ca, colivier@etsmtl.ca

RESUME : *Cet article propose un modèle d’optimisation multi-objectifs pour optimiser l’affectation des patients aux médecins spécialistes en tenant compte de plusieurs critères tels que la raison de consultation des patients, la priorité clinique, l’adresse du patient, la disponibilité des médecins, etc. Le modèle a été résolu sur la base d’un cas réel proposé par le Centre de Répartition des Demandes de Services (CRDS) Montréalais. La méthode Lexicographique a été utilisée. Les résultats préliminaires sont présentés dans cet article.*

MOTS-CLES : *affectation de patients, soins spécialisés, CRDS, optimisation multi-objectifs, PLNE, méthode lexicographique.*

1 INTRODUCTION

Récemment, le réseau de la santé et des services sociaux québécois a connu des changements organisationnels majeurs qui se sont notamment traduits par la création de 22 CISSS (Centres Intégrés de Santé et de Services Sociaux) et CIUSSS (Centres Intégrés Universitaires de Santé et de Services Sociaux) dans l’ensemble du Québec (MSSS, 2016). Ces CISSS/CIUSSS sont constitués de plusieurs établissements de santé et de services sociaux tels que des centres de services locaux et communautaires, des centres hospitaliers des centres d’hébergement et de soins de longue durée, des centres de protection de l’enfance et de la jeunesse, etc. La gestion de ces réseaux relève du Ministère de la Santé et des Services Sociaux (MSSS).

La mission des CISSS/CIUSSS est d’offrir des services de santé et de services sociaux généraux et spécialisés diversifiés tout en s’assurant de rendre ces services accessibles, efficaces, efficients et de qualité. Dans cette optique, le MSSS a mis en place depuis novembre 2016, des Centres de Répartition des Demandes de Services (CRDS) dans l’ensemble de la province pour faciliter l’accès aux services de soins spécialisés (i.e. première consultation). Pour ce faire, les médecins généralistes (médecins référents) acheminent les demandes de consultations de leurs patients aux CRDS qui de leur côté affectent les patients aux médecins spécialistes. Ainsi, les CRDS se situent dans la chaîne de services de soins entre les médecins généralistes et les médecins spécia-

listes en soutenant les médecins généralistes dans leur référence auprès des médecins spécialistes pour la première demande de consultation.

Dans le processus d’affectation des patients aux médecins spécialistes (processus de recherche de rendez-vous), plusieurs critères doivent être pris en compte tels que la raison médicale de consultation, la priorité clinique du patient (délai maximum pour faire une consultation), la disponibilité de médecins spécialistes et leurs horaires de consultation, le lieu de la consultation, etc. Le processus d’affectation actuel se fait selon une approche par inspection, ce qui requière beaucoup de temps et d’efforts pour des solutions sous-optimales, des risques d’erreurs élevés et des hors-délais fréquents. De plus, actuellement, les CRDS font la recherche de rendez-vous pour neuf (9) spécialités différentes : cardiologie, gastroentérologie, néphrologie, neurologie, ophtalmologie, oto-rhino-laryngologie (ORL), orthopédie, pédiatrie générale et urologie. Ces 9 spécialités seront élargies à un total de 26 lorsque l’ensemble des phases de déploiement des CRDS seront complétés (17 autres spécialités compléteront les 9 actuelles). Cette augmentation de spécialités aura pour effet de complexifier davantage le problème d’affectation.

Nous proposons de formuler ce problème multicritères comme un modèle d’optimisation multi-objectifs que nous résolvons à l’aide de la méthode Lexicographique sur un cas réel (spécialité cardiologie) proposé par le CRDS montréalais dans le cadre d’un projet pilote. En effet, en vue d’améliorer l’efficacité de ce CRDS et pré-

parer le déploiement des 17 spécialités à venir, un projet d'amélioration continue a été réalisé au sein des équipes du CRDS durant l'été 2017. Une activité Kaizen avec plusieurs employés (agents administratifs, infirmiers et gestionnaires) et partenaires du CRDS a été réalisée dans le cadre de ce projet. Lors de ce Kaizen, il a été identifié que l'étape de recherche/octroi de rendez-vous était le goulot d'étranglement du processus. Automatiser et optimiser la recherche et l'octroi de rendez-vous est une des solutions qui a été retenue pour améliorer l'efficacité du processus du CRDS. Ainsi, le modèle proposé prend compte des spécificités du problème d'affectation observé au CRDS montréalais, les modèles existants ne permettant pas de modéliser adéquatement le problème. Toutefois, le modèle peut être utilisée pour différentes spécialités et généralisé pour l'ensemble des CRDS implantés au Québec.

La suite de l'article est organisée comme suit : la section 2 décrit le cas d'étude (CRDS montréalais). La section 3 présente notre revue de la littérature sur le problème d'affectation. La section 4 introduit le modèle mathématique alors que la section 5 présente nos résultats préliminaires. Enfin, nos conclusions et perspectives de recherches sont exposées dans la section 6.

2 CAS D'ÉTUDE : CRDS MONTRÉALAIS

L'objectif premier du CRDS montréalais est de faciliter l'accès aux services de soins spécialisés en octroyant un rendez-vous aux patients pour leur première consultation auprès d'un médecin spécialiste exerçant dans la ville de Montréal. Ce CRDS est rattaché au CIUSSS du Centre-Sud de l'île de Montréal. Actuellement, le CRDS montréalais compte près de 40 employés composés essentiellement d'infirmiers et d'agents administratifs. Entre novembre 2016 et août 2017, le volume moyen des demandes de consultations reçues est estimé à 300 par semaine (toute spécialité confondue). Chacune des neuf (9) spécialités que traite actuellement le CRDS montréalais (première phase de déploiement), dispose d'un formulaire distinct élaboré par un comité directeur du MSSS, formé par des médecins généralistes et des médecins spécialistes (CRDS, 2017). Le formulaire de chacune des spécialités est standardisé pour l'ensemble des CRDS du Québec. Ces formulaires précisent la raison de consultation du patient, la date au plus tard de la consultation (appelé priorité clinique), le médecin spécialiste à consulter s'il y a une référence nominative, le numéro d'assurance maladie, les coordonnées du patient (nom, adresse et numéro de téléphone) et si requis, des renseignements supplémentaires. Les formulaires sont remplis par les médecins généralistes (i.e. médecins référents de la première ligne médicale) et envoyés (par fax) au CRDS qui octroie un rendez-vous à chaque patient pour sa première consultation auprès d'un médecin spécialiste.

Toutes les informations figurant sur les formulaires sont entrées manuellement dans un système informatique appelé SIGASS (Système Informatique de Gestion de l'Accès aux Services Spécialisés). La recherche de rendez-vous se fait à l'aide du système SIGASS où sont également entrés les noms des médecins spécialistes, leur spécialisation ainsi que leurs disponibilités (plages horaires). Le processus de traitement d'une demande de consultation au CRDS montréalais peut être résumé comme suit :

1. Saisie du formulaire dans le système SIGASS et confirmation de la réception du formulaire au médecin généraliste.
2. Recherche de rendez-vous, octroi du rendez-vous et confirmation avec le patient. L'octroi de rendez-vous doit tenir compte de trois (3) critères : la proximité du lieu de consultation, le respect de la priorité clinique (délai) et le respect de la référence nominative (si précisée dans le formulaire).
3. Confirmation du rendez-vous au médecin généraliste (référents) et au médecin spécialiste.

3 REVUE DE LA LITTÉRATURE

Dans la littérature, il existe plusieurs techniques de modélisation et de résolution pour le problème d'affectation. Chaque technique est étroitement liée aux spécificités du problème étudié. De ce fait, il est difficile de développer un modèle générique applicable à l'ensemble des problèmes rencontrés sur le terrain. Toutefois, les différents modèles proposés dans la littérature reposent sur une des formulations de base du problème classique d'affectation. L'article figure de proue sur le problème d'affectation est apparu en 1955 (Kuhn, 1955). En 2005, pour le 50^e anniversaire de l'article de Kuhn, Pentico (2007) a publié une revue de la littérature recensant les problèmes d'affectation les plus étudiés dans la littérature. L'auteur identifie des problèmes classiques d'affectation tel que le problème agent-tâche avec prise en compte de la spécialisation des tâches et des agents (Caron *et al.*, 1999), des problèmes d'affectation formulés comme des modèles du *Goal Programming* (minimisation des déviations par rapport aux objectifs souhaités) (Turgay et Taskin, 2014), ou encore comme des modèles d'optimisation lexicographiques (Burkard et Rendl, 1991). Franz et Miller (1993) ont développé un modèle pour l'affectation des ressources humaines avec ou sans prise en compte des préférences des employés et de la réglementation. Plus récemment, Boukherroub *et al.* (2017) ont proposé un modèle d'allocation des ressources naturelles aux usines de transformation en tenant compte des performances économiques, environnementales et sociales des usines.

Les techniques d'optimisation multi-objectifs sont utilisées dans plusieurs travaux qui abordent le problème d'affectation. Cet aspect est dû à la nature multicritère du problème. La modélisation multi-objectifs permet de

prendre en compte plusieurs critères et de modéliser le problème de manière plus fidèle à la réalité (Collette et Siarry, 2002). En effet, concevoir un modèle mathématique en formulant un seul objectif peut impliquer que le modèle obtenu ne représente pas les critères les plus importants sur lesquels sont basées les décisions des gestionnaires. Dans le cas de l'optimisation multi-objectifs, la notion d'optimum n'existe pas ; les solutions obtenues sont appelées solutions de Pareto (Ehrgott et Tenfelde-Podehl, 2003). Une partie de ses solutions sont dites « non-dominées », aussi appelées Optimums de Pareto, i.e., un objectif ne peut pas être amélioré sans détériorer au moins un des autres objectifs (Collette et Siarry, 2002). Les différentes techniques d'optimisation multi-objectifs sont utilisées pour traiter différentes classes de problèmes. Dans cette étude, nous utilisons la méthode Lexicographique (Collette et Siarry, 2002) pour rendre compte de la préférence des gestionnaires quant au niveau de priorité des différents critères. Cet aspect est discuté dans la section 5.

4 FORMULATION DU MODELE D'OPTIMISATION

Le modèle d'optimisation doit permettre de trouver un rendez-vous à tous les patients (dans le mesure du possible), tout en respectant les critères d'affectation que nous avons classés suivant trois (3) catégories : critères liés aux patients, critères liés aux médecins spécialistes et critères liés au CRDS.

1) Critères liés aux patients

Le patient doit consulter un médecin pour la spécialité demandée et la raison de consultation spécifiée par le médecin généraliste (réfèrent). De plus, la priorité clinique doit être respectée. Autrement dit, le rendez-vous doit être planifié à l'intérieur du délai spécifié par le médecin réfèrent. Par exemple la priorité clinique B correspond à un délai maximum de 10 jours alors que la priorité clinique C correspond à un délai maximum de 28 jours.

2) Critères liés aux médecins spécialistes

Les médecins spécialistes ne peuvent voir les patients que durant les plages horaires qu'ils ont spécifiées au préalable au CRDS (information disponible dans le système informatique SIGASSS). Chaque médecin spécialiste traite uniquement les cas (raisons de consultation) qui relèvent de sa spécialité. Par contre, certains médecins spécialistes peuvent ne pas traiter certains cas même ceux-ci relèvent de la spécialité. Bien entendu, un médecin spécialiste ne peut voir qu'un patient à la fois.

3) Critères liés au CRDS

Ces critères sont définis par le MSSS et sont communs à tous les CRDS du Québec. Par exemple, il est recommandé de minimiser la distance parcourue par un patient entre son lieu de résidence et le lieu de sa consultation.

Le modèle d'optimisation est formulé comme suit :

- Ensembles et indices :

P : ensemble des patients. $P = \{p1, p2, \dots, pm\}$

M : ensemble des médecins. $M = \{m1, m2, \dots, pm\}$

I : ensemble des jours de l'horizon de planification.

$I = \{i1, i2, \dots, in\}$, i correspond à la date de la journée

H : ensemble des plages horaires. $H = \{h1, h2, \dots, hm\}$

- Paramètres :

d_{pm} : distance entre le domicile du patient $p \in P$ et le lieu de consultation du médecin $m \in M$.

q_{pih} : paramètre binaire égale à 1 si la raison de consultation du patient $p \in P$ peut être traitée dans la plage horaire $h \in H$, du jour $i \in I$, 0 sinon.

x_{mih} : paramètre binaire égal à 1 si le médecin $m \in M$ est disponible dans la plage horaire $h \in H$, du jour $i \in I$, 0 sinon.

d^{\max} : distance maximale que peut parcourir un patient.

j_p^{\max} : jour limite d'affectation du patient $p \in P$.

- Variables de décision :

Y_{pmih} : variable binaire égale à 1 si le patient $p \in P$ est affecté au médecin $m \in M$ le jour $i \in I$ à la plage horaire $h \in H$, 0 sinon.

O_p : nombre de jours entre la date de rendez-vous du patient $p \in P$ et son jour limite d'affectation (j_p^{\max} peut être négatif en cas de hors délai).

- Fonctions-objectifs :

Le nombre d'affectations des patients doit être maximisé :

$$MAXF1 = \sum_{p \in P} \sum_{m \in M} \sum_{i \in I} \sum_{h \in H} Y_{pmih} \quad (1)$$

La distance parcourue par les patients doit être minimisée :

$$MINF2 = \sum_{p \in P} \sum_{m \in M} \sum_{i \in I} \sum_{h \in H} d_{pm} Y_{pmih} \quad (2)$$

Le nombre de rendez-vous hors délai doit être minimisé :

$$MAXF3 = \sum_{p \in P} O_p \quad (3)$$

- Contraintes :

$$\sum_{m \in M} \sum_{i \in I} \sum_{h \in H} Y_{pmih} \leq 1 \quad \forall p \in P \quad (4)$$

$$\sum_{p \in P} Y_{pmih} \leq 1 \quad \forall m \in M, i \in I, h \in H \quad (5)$$

$$Y_{pmih} \leq q_{pih} \quad \forall p \in P, m \in M, i \in I, h \in H \quad (6)$$

$$Y_{pmih} \leq x_{mih} \quad \forall p \in P, m \in M, i \in I, h \in H \quad (7)$$

$$d_{pm} Y_{pmih} \leq d^{\max} \quad \forall p \in P, m \in M, i \in I, h \in H \quad (8)$$

$$\sum_{m \in M} \sum_{i \in I} \sum_{h \in H} Y_{pmih} i = j_p^{\max} - o_p \quad \forall p \in P \quad (9)$$

Les contraintes (4) indiquent qu'un patient peut avoir au plus un rendez-vous. Les contraintes (5) indiquent qu'un rendez-vous ne peut être attribué qu'à un (1) patient, au plus. Les contraintes (6) autorisent qu'un patient soit affecté à une plage horaire seulement si sa raison de consultation est traitée dans cette plage horaire. Les contraintes (7) autorisent qu'un patient soit affecté à un médecin pendant une plage horaire donnée seulement si le médecin est disponible à cette période. Les contraintes (8) autorisent l'affectation du patient à un médecin spécialiste seulement si la distance qu'il parcourt pour visiter le médecin ne dépasse pas la distance maximale. Les équations (9) permettent de déterminer la date du rendez-vous de chaque patient affecté à une plage horaire. Maximiser les variables O_p revient à affecter un rendez-vous aux patients le plus tôt possible. Le modèle tend donc à faire respecter la priorité clinique (minimiser les hors délais).

5 RESULTATS PRELIMINAIRES

5.1 Collecte des données

Nous considérons le cas de la cardiologie du CRDS montréalais (voir Section 2). La cardiologie présente 19 raisons de consultations possibles qui figurent sur les formulaires acheminés par les médecins généralistes au CRDS (le médecin référent peut cocher une ou plusieurs raisons de consultation sur le formulaire) et dans le système SIGASS. La liste des cardiologues, 89 cardiologues sur l'île de Montréal, figure également dans le système SIGASS. Les agents administratifs peuvent consulter dans le système le lieu de consultation des cardiologues, leurs plages horaires ainsi que la raison de consultation traitées par les cardiologues. Il est à noter, que dans le cas de la cardiologie, les 19 raisons de consultations sont traitées par l'ensemble des cardiologues enregistrés dans la banque de données du CRDS montréalais. Nous disposons également d'une banque de données réelles dé-

nominalisées des patients nécessitant un rendez-vous avec un cardiologue pour une ou plusieurs raisons de consultation. Pour chacun des patients, les trois (3) premiers chiffres/lettres du code postal sont disponibles. À partir de ces données, les distances entre le lieu de résidence de chaque patient et le lieu de consultation de chaque cardiologue peut être calculée. Pour ce faire, un programme informatique écrit en VBA a été développé afin de générer la matrice des distances en km.

6000 demandes de consultations en cardiologie reçues au CRDS montréalais ont été saisies dans le système SIGASS entre novembre 2016 et septembre 2017. En moyenne, 26 demandes sont reçues quotidiennement (7j/7). Parmi ces 26 demandes, en moyenne, une (1) demande est retournée aux médecins référents (ou autres expéditeurs) pour raison de non-conformité (ex. pas le bon formulaire, ou expéditeur autre qu'un médecin généraliste), huit (8) demandes présentent des références nominatives (ex. nom de cardiologue que souhaite consulter le patient ou un établissement en particulier) et trois (3) demandes présentent des raisons de consultations autres que les 19 raisons spécifiées dans les formulaires de consultation. Nous n'avons pas tenu compte de ces trois (3) cas de figure dans notre formulation mathématique.

Par ailleurs, actuellement, il existe cinq (5) priorités cliniques A, B, C, D et E pour les demandes de consultations. La priorité A correspond à un délai maximum de trois (3) jours entre la date de consultation chez le médecin généraliste et la date du rendez-vous avec le médecin spécialiste, B correspond à un maximum de 10 jours, C à un maximum de 28 jours, D à un maximum de six (6) mois et E un maximum de 12 mois. Les demandes de consultation avec la priorité clinique A ont été exclues du cas d'étude à la demande du CRDS montréalais qui souhaitait traiter à part ce type de demandes qui sont considérées comme étant urgentes. Le nombre moyen des demandes de consultation avec la priorité clinique A est de quatre (4) par jour. Nous avons donc considéré un total de 10 demandes de consultations en cardiologie par jour en moyenne dans notre cas d'étude. Quant au nombre de plages horaires disponibles dans le système SIGASS, nous avons estimé celles-ci à une moyenne de 17 par jour. Il y a au total 21 cardiologues.

5.2 Résolution du modèle mathématique

Le modèle d'optimisation a été résolu à l'aide de la méthode Lexicographique. Le principe de cette méthode consiste à classer les fonctions-objectifs par ordre d'importance et de résoudre le modèle mathématique avec une fonction-objectif à la fois selon l'ordre d'importance de celles-ci. La valeur optimale de la fonction-objectif trouvée lors de la résolution du problème est convertie en contrainte et le modèle mathématique est résolu en considérant cette contrainte et la fonction objectif prioritaire suivante. Dans le cas du CRDS montréalais

lais, la priorisation des fonctions objectifs était comme suit :

- **Priorité 1 :** Maximiser le nombre d'affectation de rendez-vous (spécialiste et plage horaire) aux patients.
- **Priorité 2 :** Minimiser la distance parcourue par les patients entre leur lieu de résidence et le lieu de consultation.
- **Priorité 3 :** Minimiser le nombre de rendez-vous hors délais (respecter la priorité clinique).

Tableau 1 : Résultats du modèle d'optimisation obtenus à l'aide du solveur Lingo

Patients	Médecins	Jour	Heure	Y_{pmih}
Patient 1	1	1	6	1
Patient 2	12	2	4	1
Patient 3	6	3	5	1
Patient 4	6	4	15	1
Patient 5	6	3	15	1
Patient 6	6	3	16	1
Patient 7	1	1	16	1
Patient 8	6	5	13	1
Patient 9	6	4	13	1
Patient 10	6	5	14	1
Patient 11	1	1	3	1
Patient 12	6	5	19	1
Patient 13	6	1	11	1
Patient 14	6	1	13	1
Patient 15	9	1	19	1
Patient 16	6	1	12	1
Patient 17	12	2	10	1
Patient 18	6	4	16	1
Patient 19	6	5	18	1
Patient 20	3	1	8	1
Patient 21	6	3	6	1
Patient 22	8	1	18	1

Le modèle mathématique a été résolu à l'aide du solveur LINGO. Le temps de résolution a été de 30 secondes. Une solution correspond à l'affectation d'un patient à un cardiologue et à une plage horaire. Les résultats sont présentés dans le tableau 1. Pour faciliter la compréhension et l'interprétation des résultats aux agents administratifs, nous avons également généré les résultats sous forme de texte à l'aide de Lingo.

Les résultats du modèle mathématique montrent que tous les patients ont bien été affectés à un rendez-vous (tableau 1). Par ailleurs, toutes les priorités cliniques (i.e., dates limites pour l'affectation des rendez-vous) ont été respectées (tableau 2). De plus, la distance maximale de 12km est respectée pour tous les patients et la distance parcourue est minimisée (tableau 2).

Tableau 2 : Dates de rendez-vous des patients et distances parcourues

Patients	Jour d'affectation	Date limite (j^{\max})	Distances (km)	Distance max (km)
Patient 1	02/03/2017	05/03/2017	1,9	12
Patient 2	03/03/2017	06/03/2017	8,6	12
Patient 3	06/03/2017	15/03/2017	1,9	12
Patient 4	07/03/2017	15/03/2017	2	12
Patient 5	06/03/2017	07/03/2017	1,9	12
Patient 6	06/03/2017	06/03/2017	2	12
Patient 7	02/03/2017	05/03/2017	1,9	12
Patient 8	08/03/2017	08/03/2017	2	12
Patient 9	07/03/2017	07/03/2017	1,9	12
Patient 10	08/03/2017	08/03/2017	1,9	12
Patient 11	02/03/2017	02/03/2017	2	12
Patient 12	08/03/2017	09/03/2017	1,9	12

Patient 13	02/03/2017	02/03/2017	1,9	12
Patient 14	02/03/2017	03/03/2017	2	12
Patient 15	02/03/2017	04/03/2017	4,4	12
Patient 16	02/03/2017	05/03/2017	2	12
Patient 17	03/03/2017	08/03/2017	8,6	12
Patient 18	07/03/2017	07/03/2017	2	12
Patient 19	08/03/2017	08/03/2017	2	12
Patient 20	02/03/2017	02/03/2017	4,9	12
Patient 21	06/03/2017	09/03/2017	1,9	12
Patient 22	02/03/2017	04/03/2017	4,9	12

6 CONCLUSION ET PERSPECTIVES DE RECHERCHE

L'objectif de ce travail a été de proposer un modèle d'aide à la décision pour optimiser l'affectation des patients aux médecins spécialistes en tenant compte de la raison de consultation, de la priorité clinique et du lieu de résidence du patient. Un modèle multi-objectifs résolu à l'aide de la méthode Lexicographique a été proposé. Le modèle a été résolu sur la base d'un cas réel proposé par le CRDS montréalais. En termes de perspectives de recherche, plusieurs améliorations pourraient être apportées au modèle pour être au plus près de la réalité du CRDS.

Un des éléments importants du problème qui n'a pas été traité dans le modèle proposé est l'aspect dynamique des affectations (ex. arrivée de nouvelles demandes de consultation, refus ou annulation d'un rendez-vous, nouveaux médecins spécialistes et nouvelles plages horaires, retrait de certaines plages horaires, etc.). Une affectation à horizon roulant est une piste de recherche intéressante qui pourrait être investiguée. Une autre piste de recherche concerne l'intégration du principe de rotation des médecins spécialistes dans le modèle d'optimisation. Autrement dit, pour une spécialité donnée, les plages horaires de tous les médecins spécialistes devraient être utilisées par le CRDS de manière équitable dans le processus d'affectation. On pourrait par exemple, définir une cible pour chaque médecin spécialiste, en termes de pourcentage de plages horaires à utiliser et minimiser l'écart entre l'utilisation réelle et cette cible. Enfin, l'objectif de minimisation de la distance pourrait être reformulé pour considérer, non seulement les kilomètres parcourus mais également les modes de transport disponibles et la congestion routière. Par exemple, il serait intéressant de proposer une affectation qui prend en compte l'accès et la disponibilité des transports en commun. Dans la même optique, une affectation qui minimiserait le temps de trajet et/ou les émissions de carbone relatives au trajet du patient serait pertinente. Dans la métropole montréalaise, ces aspects présentent des préoccupations majeures.

REMERCIEMENTS

Cette action de recherche est soutenue par les fonds démarrage pour jeunes chercheurs de l'École de

technologie supérieure. L'équipe de recherche tient à remercier Mme Geneviève Alary, Mme Camille Bouchard-Coulombe, M. Jacques Drôlet et M. Didier Royer du CIUSSS Centre-Sud de l'île de Montréal ainsi que Marine-Sophie Chevillard du CRDS montréalais.

REFERENCES

- Benayoun R., J. De Montgolfier, J. Tergny and P. Laritchev, 1971. Linear programming with multiple objective functions: Step method (STEM). *Mathematical Programming*, 1(3), p 366-375.
- Boukherroub, T., L. Lebel and A. Ruiz, 2017. A Framework for Sustainable Forest Resource Allocation: A Canadian Case Study. *Omega*, 66(Part B) p. 224-235.
- Burkard, R.E. and F. Rendl, 1991. Lexicographic bottleneck problems. *Operations Research Letters*, 10(5), p. 303-308.
- Caron, G., P. Hansen and B. Jaumard, 1999. The assignment problem with seniority and job priority constraints. *Operations Research*, 47(3), p. 449-454.
- Collette Y and P. Siarry, 2002. *Optimisation multi-objectif*. Eyrolles, 280p.
- CRDS Montréalais : Centre de Répartition des Demandes de Services, 2017. <https://www.santemontreal.qc.ca/professionnels/medecins/crds-centre-de-repartition-des-demandes-de-services-montrealais/>
- Ehrgott M. and D. Tenfelde-Podehl, 2003. Computation of ideal and Nadir values and implications for their use in MCDM methods. *European Journal of Operational Research*, 151, p. 119-139.
- Franz, L.S. and J.L. Miller, 1993. Scheduling medical residents to rotations: Solving the large-scale multiperiod staff assignment problem. *Operations Research* 41(2), p. 269-279.
- Kelyanmoy Deb, 2001. *Multi-Objective Optimization Using Evolutionary Algorithms*. WILEY, 517p.
- Kelyanmoy Deb, K., Miettinen and S. Chaudhuri, 2010. Towards an Estimation of Nadir Objective Vector Using Hybrid Evolutionary and Local Search Approaches. *IEEE Transactions on Evolutionary Computation*, 14, p 821-841.
- Kohl N. and S.E. Karisch, 2004. Airline crew rostering: Problem types, modeling, and optimization. *Annals of Operations Research*, 127, p 223-257.

- Kuhn H.W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics*, 2, p. 83-9.
- Marler R.T. and J.S. Arora, 2003. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6), p 369-395.
- Medard C.P and N. Sawhney, 2007. Airline crew scheduling from planning to operations. *European Journal of Operational research*, 183, p 1013-1027.
- MSSS, 2016. *Le système de santé et de services sociaux au Québec en bref. Gouvernement du Québec*. 18 p. Accessible sur le lien : <http://publications.msss.gouv.qc.ca/msss/fichiers/2015/15-731-02WF.pdf>
- Pentico D.W, 2007. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176, p.774-793.
- Turgay, S. and H. Taskin, 2014. Fuzzy goal programming for health-care organization. *Computer & Industrial Engineering*, 86, p. 14-21.

OPTIMISATION D'UN PLAN DE PRODUCTION MULTI-PERIODES AVEC UNE DEMANDE DYNAMIQUE ET DES DELAIS STOCHASTIQUES

O. BEN-AMMAR¹, B. BETTAYEB², A. DOLGUI¹

¹IMT Atlantique, LS2N, UMR-CNRS 6004, La Chantrerie, 4 rue Alfred Kastler, 44300 Nantes, France

{oussama.ben-amar, alexandre.dolgui}@imt-atlantique.fr

²Groupe CESI, Laboratoire LINEACT, 1 Rue G. Marconi, Mont-Saint-Aignan, France

bbettayeb@cesi.fr

RESUME : *La planification de la production et la gestion de stocks sont d'une importance capitale pour les entreprises afin de fournir à ses clients des produits finis de grande qualité, à des prix compétitifs et juste à temps. Pour atteindre ces objectifs, une des priorités des planificateurs de la production en entreprise est de réduire le niveau moyen des stocks et déterminer les délais de sécurité lorsqu'ils opèrent dans un environnement incertain. Cette étude traite un problème de planification de la production sur plusieurs périodes avec une demande dynamique connue. Les délais d'obtention sont des variables aléatoires discrètes et indépendantes, dont les distributions de probabilité sont connues et bornées. Un modèle probabiliste, incluant une procédure récursive pour calculer l'espérance mathématique du coût total, est proposé. Un algorithme génétique classique est développé afin de déterminer les délais planifiés et le niveau de stock de sécurité qui minimisent l'espérance du coût total. Ce dernier est égal à la somme des coûts de rupture et de stockage. Afin d'illustrer sa performance, cette approche est comparée à trois autres stratégies (Newsboy, pessimiste et optimiste). Les résultats montrent qu'il est plus avantageux d'optimiser les délais planifiés plutôt que de mettre en place un stock de sécurité. Pour analyser la robustesse de la solution, différents tests ont été réalisés en faisant varier la variance de la distribution du délai.*

MOTS-CLES : *Planification de la production, Multi-périodes, Demande dynamique, Délais d'obtention stochastiques, Algorithme génétique*

1 INTRODUCTION

La gestion de l'incertitude devient l'un des défis les plus importants dans l'optimisation de la chaîne d'approvisionnement. En effet, l'incertitude entraîne plusieurs difficultés dans la planification de la production et la gestion de stocks. Les sources d'incertitude sont diverses et peuvent se situer à plusieurs niveaux de la chaîne d'approvisionnement : variabilité de la demande, pannes de machines, délais de transport, problèmes de qualité, etc. Dans cet article, nous étudions le cas de la planification de la production en une seule étape avec une politique lot-pour-lot, où les demandes sont connues et doivent être livrées à des dates d'échéance prédéfinies. Une fois qu'une demande est lancée en production, elle est prête après un délai d'obtention aléatoire. Cet aléa sur le délai entraîne des coûts de rupture et de stockage. L'objectif est de minimiser l'espérance du coût total en optimisant les délais planifiés et le stock initial de sécurité.

Sans essayer de faire un examen exhaustif de la littérature, nous soulignons et discutons les travaux les plus proches à ce travail. Pour mieux anticiper l'incertitude des délais, les planificateurs utilisent plusieurs techniques, telles que les délais de sécurité et les stocks de sécurité. Jansen et de Kok (2013) ont souligné l'importance d'anticiper ces délais, tandis que Chopra et al (2004), He (2005) et Van Kampen et al. (2010) ont expliqué dans quels cas il pourrait être avantageux de tenir compte des délais de sécurité et/ou des stocks de sécurité.

Parmi les premiers modèles considérant un système de production mono-étape avec une demande déterministe et un temps de traitement stochastique, Weeks (1981) a prouvé que le problème est équivalent au problème standard de Newsboy lorsque les délais sont indépendants et identiquement distribués. En utilisant cette dernière hypothèse, plusieurs autres approches ont été proposées pour déterminer les politiques de commande optimales (Porteus, 2002, Muthuraman et al., 2014). Pour les autres modèles prenant en compte les délais stochastiques, mais dans le cas de fournisseurs multiples, les lecteurs peuvent se référer à la revue de littérature de Minner (2003).

Dans le cas des systèmes d'assemblage, la revue de littérature peut être scindée en deux cas différents : (i) la planification d'une seule demande sur une période unique (Yano 1987, Chu et al 1993, Tang et Grubbström, 2003, Axsäter 2005 et Ben-Ammar et al. 2017) ; et (ii) la planification sur plusieurs périodes (voir par exemple le papier Ould Louly et Dolgui (2013) et les références incluses) où la demande est généralement considérée comme constante. Les lecteurs peuvent se référer à Dolgui et al. (2013) pour une revue de littérature plus exhaustive sur le sujet.

Proche de notre modèle, Deng et al. (2010) ont étudié un modèle de gestion des stocks en une étape avec à la fois une demande et un délai stochastiques. L'objectif est de déterminer une séquence optimisée de dates de lancement, tout en minimisant l'espérance du coût total sur un horizon de planification discret et fini. En supposant que la distribution des délais tend vers zéro, un algorithme de recuit simulé est introduit pour résoudre le problème. De

même, Rossi et al. (2012) ont abordé le même problème et développé des approches heuristiques rapides.

Dans cet article, nous étudions le cas où les délais sont des variables aléatoires discrètes indépendantes avec des distributions de probabilité connues et bornées, mais qui ne sont pas nécessairement identiquement distribuées. Les demandes sont supposées déterministes mais pas constantes. À notre connaissance, aucune étude de ce genre n'a été rapportée dans la littérature. Ceci est expliqué par le fait que le croisement entre les livraisons des demandes des différentes périodes est difficile à modéliser. Les travaux existants considèrent une seule distribution des délais, alors que de nos jours, les livraisons en provenance du même fournisseur peuvent suivre différentes distributions (mode de transport variable, grandes distances,...). La contribution majeure de ce travail est de combiner les aspects « demande dynamique », « délais ne suivant pas la même distribution de probabilité » et « croisement des livraisons ».

Le reste de l'article est organisé comme suit. La section 2 donne la formulation du problème et les hypothèses considérées. Les approches de résolution proposées sont détaillées dans la section 3. Les premiers résultats sont présentés et discutés dans la section 4. Enfin, la section 5 est consacrée aux conclusions et aux perspectives.

2 FORMULATION DU PROBLÈME

Les spécificités du modèle stochastique de planification multi-période sont les suivantes. Une séquence de demandes dynamiques, notées D_t , pour un seul produit fini est planifiée sur un horizon de planification composé de T périodes. Les demandes sont planifiées entre les périodes p_m et p_M (voir la figure 1). Chaque demande D_t est lancée à la période $t - X_t$ où X_t est le délai planifié. Le délai réel, noté L_t , est une variable discrète qui varie entre L_t^- et L_t^+ . Chaque demande D_t possède sa propre distribution de délai réel L_t . Tous ces délais sont des variables aléatoires indépendantes et non nécessairement identiquement distribuées. Nous notons que la capacité est supposée infinie.

Au début de chaque période t , la quantité reçue A_t correspond à la somme des demandes D_j lancées aux périodes $j - X_j$ et reçues à la période t , c'est-à-dire $j - X_j + L_j = t$. L'ensemble des indices des demandes qui peuvent vérifier cette égalité et arriver à la période t est noté par \mathcal{A}_t . Notons que les demandes peuvent arriver entre p_h la première période où le niveau de stock peut être non nul, et $p_b + 1$ où p_b la dernière période avec une rupture possible (voir la figure 1). Chaque demande est satisfaite si le niveau du stock du produit fini le permet. Si le stock est insuffisant, la quantité manquante augmente le niveau de rupture de la période et reste dans le système jusqu'à ce qu'elle soit satisfaite. Les périodes avec possibilité de rupture sont comprises entre les périodes p_b, \dots, p_B où $p_b = p_m$ puisque la rupture ne peut pas survenir tant qu'il n'y a pas de demande à satisfaire et p_B la période qui précède la dernière période avec des arrivées possibles non

nulles (voir la figure 1). Pour calculer la rupture à une période $t \in \{p_b, \dots, p_B\}$, l'ensemble \mathcal{M}_t est introduit et contient les indices des demandes D_j impliquées.

A la fin d'une période t donnée, au plus un des deux coûts suivants est non nul : (i) si le niveau de stock est positif, le coût associé est proportionnel au coût de stockage unitaire noté c^h , (ii) si il y a une rupture, un coût associé est calculé en fonction d'un coût unitaire de rupture noté c^b . Pour faire face aux effets de l'incertitude des délais et la variabilité de la demande, un stock de sécurité S_0 est considéré et est une variable de décision.

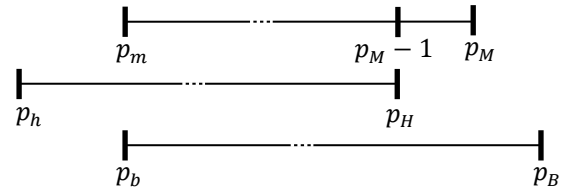


Figure 1. Intervalles de périodes avec demandes, de stockages et de ruptures possibles non nulles.

2.1 Notations

2.1.1 Paramètres et variables de décision

- t Indice de la période
- D_t Demande à la période t
- p_m Indice de la première période avec une demande non nulle
- p_M Indice de la dernière période avec une demande non nulle
- X_t Délai planifié associé à la demande de la période t , c'est une **variable de décision**
- L_t Délai réel associé à la demande de la période t lancée à la période $t - X_t$, c'est une variable aléatoire discrète bornée variant entre L_t^- and L_t^+
- A_t Somme des quantités arrivées à la période t
- \mathcal{A}_t Ensemble des indices de demandes D_j commandées aux périodes $j - X_j$ pouvant être reçues à t . Une période $j \in \mathcal{A}_t$ si $j \in \{p_m, \dots, p_M\}$ et $L_j^- \leq t - j + X_j \leq L_j^+$
- \mathcal{M}_t Ensemble des indices de demandes D_j impliquées dans le calcul de la rupture à la période t . Une période $j \in \mathcal{M}_t$ si et seulement si $j \in \{p_m, \dots, p_M\}$ et $L_j^- \leq t - j + X_j \leq L_j^+ - 1$
- m_t Premier élément dans \mathcal{M}_t
- M_t Dernier élément dans \mathcal{M}_t
- R_t Rupture à la période t
- S_t Niveau de stock à la période t
- \mathcal{R}_t^α Sous-ensemble de l'espace d'état de la variable aléatoire R_t où seulement α demandes arrivent avant la fin de la période t
- r_t^α Élément de l'ensemble \mathcal{R}_t^α , c.-à-d. une réalisation possible de la variable aléatoire R_t avec seulement α demandes arrivant avant la fin de la période t
- n_t Nombre d'éléments dans l'ensemble \mathcal{M}_t , c.-à-d $n_t = \text{card}(\mathcal{M}_t)$

- p_b Indice de la première période où il peut y avoir de la rupture, c.-à-d. $p_b = \min\{t \setminus \mathcal{A}_t \neq \emptyset\} = p_m$
- p_B Indice de la dernière période où il peut y avoir de la rupture, c.-à-d. $p_B = \max_{t=p_m, \dots, p_M} (t - X_t + L_t^+ - 1)$
- p_h Indice de la première période où il peut y avoir du stockage, c.-à-d. $p_h = \min_{t=p_m, \dots, p_M} (t - X_t + L_t^-)$
- p_H Indice de la dernière période où il peut y avoir du stockage, c.-à-d. $p_H = p_M - 1$
- c^h Coût unitaire de stockage par période
- c^b Coût unitaire de rupture par période
- S_0 Stock de sécurité rajouté à partir de la période p_b , c'est une **variable de décision**,

2.1.2 Fonctions

- $E[\cdot]$ Espérance mathématique
- $F_t(\cdot)$ Fonction de répartition de la variable aléatoire L_t
- $Q(\cdot)$ Fonction récursive
- x^+ Maximum entre x et 0, i.e. $x^+ = \max(x, 0)$
- x^- Minimum entre x et 0, i.e. $x^- = \min(x, 0)$
- $\mathbb{1}_{\{A\}}$ Fonction indicatrice, égale à 1 si l'évènement A est vrai ; 0 sinon

2.2 Modèle mathématique

L'objectif est de déterminer la séquence de délais d'obtention planifiés \vec{X} et le niveau du stock de sécurité S_0 , pour satisfaire toutes les demandes en minimisant l'espérance du coût total sur tout l'horizon de planification. Pour la suite, nous supposons que chaque délai réel L_t est une variable aléatoire discrète. L'espace d'état de cette variable varie entre L_t^- et L_t^+ . Nous notons aussi que chaque délai planifié $X_t \in \{L_t^-, \dots, L_t^+\}$. Ainsi, le coût total est une variable aléatoire dont l'expression est ci-après :

$$TC(S_0, \vec{X}, \vec{L}) = c^h \times \sum_{t=p_h}^{p_H} S_t + c^b \sum_{t=p_b}^{p_B} R_t \quad (1)$$

Afin d'évaluer l'espérance mathématique du coût total, il est nécessaire de déterminer les expressions de R_t, S_t ainsi que leurs espérances mathématiques respectives.

Propriété 1. Une forme explicite de la rupture à t est :

$$R_t = \left(\sum_{i \in \mathcal{M}_t, i \leq t} D_i - S_0 - \sum_{i \in \mathcal{M}_t} D_i \mathbb{1}_{\{i - X_i + L_i \leq t\}} \right)^+ \quad (2)$$

Preuve. La valeur de la rupture à une période $t \in \{p_b, \dots, p_B\}$ peut être exprimée par l'expression suivante :

$$R_t = \left(\sum_{i=p_m}^t D_i - S_0 - \sum_{i=p_h}^t A_i \right)^+ \quad (3)$$

D'une part, la somme des arrivées jusqu'à la période t est égale à l'ensemble des demandes commandées pouvant arriver avant la fin de la période t ($\mathbb{1}_{\{i - X_i + L_i \leq t\}}$) :

$$\sum_{i=p_h}^t A_i = \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\{i - X_i + L_i \leq t\}}$$

Sachant que pour chaque demande D_i on a $L_i \geq L_i^-$, la somme précédente peut être scindée en deux sommes :

$$\sum_{i=p_h}^t A_i = \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i - X_i + L_i^+ \leq t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i - X_i + L_i^+ > t \\ i - X_i + L_i^- \leq t \end{array} \right\}}$$

La première somme contient les demandes arrivant au plus tard à t ($i - X_i + L_i^+ \leq t$) et la seconde contient celles pouvant arriver après t ($i - X_i + L_i^- \leq t$ and $i - X_i + L_i^+ > t$). Sachant que \mathcal{M}_t contient les indices j des demandes vérifiant $L_j^- \leq t - j + X_j \leq L_j^+ - 1$, alors :

$$\sum_{i=p_h}^t A_i = \sum_{i=0}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i - X_i + L_i^+ \leq t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i \in \mathcal{M}_t} D_i \mathbb{1}_{\{i - X_i + L_i \leq t\}} \quad (4)$$

De plus, la somme des demandes jusqu'à t est égale à :

$$\sum_{i=p_m}^t D_i = \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\{i \leq t\}}$$

Parmi ces demandes quelques-unes peuvent arriver au plus tard à t ($i - X_i + L_i^+ \leq t$) et d'autres peuvent être livrées après t ($i - X_i + L_i^+ > t$), alors :

$$\begin{aligned} \sum_{i=p_m}^t D_i &= \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ \leq t \end{array} \right\}} + \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ > t \end{array} \right\}} \\ &= \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ \leq t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ > t \\ i - X_i + L_i^- > t \end{array} \right\}} \\ &+ \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ > t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ > t \\ i - X_i + L_i^- > t \end{array} \right\}} \end{aligned}$$

Comme $X_i \in \{L_i^-, \dots, L_i^+\}$, alors :

$$\sum_{i=p_m}^k D_i = \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ \leq t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ > t \\ i - X_i + L_i^- \leq t \end{array} \right\}}$$

Sachant que \mathcal{M}_t contient les indices j des demandes vérifiant $L_j^- \leq t - j + X_j \leq L_j^+ - 1$, alors :

$$\sum_{i=p_m}^k D_i = \sum_{i=p_m}^{p_M} D_i \mathbb{1}_{\left\{ \begin{array}{l} i \leq t \\ i - X_i + L_i^+ \leq t \\ i - X_i + L_i^- \leq t \end{array} \right\}} + \sum_{i \in \mathcal{M}_t} D_i \quad (5)$$

En utilisant les expressions (4) et (5), l'expression (2) peut être déduite. ■

La rupture R_t à t est une variable aléatoire discrète (parce que les délais d'obtention réels le sont aussi) avec un nombre fini de valeurs possibles. Son espérance mathématique est présentée dans la propriété suivante.

Propriété 2. La forme explicite de l'espérance mathématique de la rupture à une période t est la suivante :

$$E[R_t] = W(t) \times \left(\sum_{\alpha=0}^{n_k-1} Q(\alpha, \alpha, t, \mathcal{M}_t) \right) \quad (6)$$

Dont la fonction récursive $Q(\dots, \dots)$ est définie ci-après :

$$Q(\alpha, i, t, z_{i+1}) = \begin{cases} U(t) & i = 0, \alpha = 0 \\ \sum_{\substack{z_i = m_k + i - 1 \\ z_i \in \mathcal{M}_t}}^{z_{i+1}} Q(\alpha, i-1, t, z_i) & \forall i \in [2, \alpha], \forall \alpha \\ \sum_{\substack{z_1 = m_t \\ z_1 \in \mathcal{M}_t}}^{z_{i+1}} \left(U(k) - \sum_{k=1}^{\alpha} D_{z_k} \right)^+ \times V(\alpha, t) & i = 1, \forall \alpha \end{cases}$$

avec

$$U(t) = \sum_{\substack{i \in \mathcal{M}_t \\ i \leq t}} D_i - S_0 \quad (7)$$

$$V(\alpha, t) = \prod_{k=1}^{\alpha} \left(\frac{1}{1 - F_{z_k}(t + X_{z_k} - z_k)} - 1 \right)$$

$$W(t) = \prod_{i \in \mathcal{M}_t} (1 - F_i(t + X_i - i))$$

$$F_i(t + X_i - i) = P[L_i \leq t + X_i - i]$$

Preuve. Soit r_t^α un élément du sous-ensemble \mathcal{R}_t^α de l'espace d'états de la variable aléatoire R_t , où $\alpha \in \{0, \dots, n_t - 1\}$ est le nombre de demandes arrivant avant la fin de la période t . Par définition, l'espérance mathématique de la rupture R_t est exprimée comme suit :

$$E[R_t] = \sum_{\alpha=0}^{n_t-1} \sum_{r_t^\alpha \in \mathcal{R}_t^\alpha} r_t^\alpha \times P[R_t = r_t^\alpha]$$

$$\text{avec } \mathcal{R}_t^\alpha = \left\{ U(t) - \sum_{k=1}^{\alpha} D_{z_k} \right\}_{\substack{z_{\alpha+1} = M_t \\ z_k = m_t + t - 1, \dots, z_{k+1}}}$$

Si aucune demande lancée n'arrive avant la fin t ($\alpha = 0$), c.-à-d. toutes les demandes dont l'indices $i \in \mathcal{M}_t$ sont reçues après ($\mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} = 1$), la valeur de la rupture est calculée en utilisant les expressions (2) et (7) :

$$R_t = \prod_{i \in \mathcal{M}_t} \mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} \times (U(t))^+$$

Dans ce cas, \mathcal{R}_t^0 contient un seul élément $r_t^0 = (U(t))^+$ avec $P[R_t = r_t^0] = W(t)$, donc :

$$P[R_t = r_t^0] \times r_t^0 = W(t) \times (U(t))^+ \quad (8)$$

Si une seule demande lancée arrive avant la fin de t ($\alpha = 1$), c.-à-d. toutes les demandes $D_i \setminus i \in \mathcal{M}_t$, à l'exception de z_1 ($\mathbb{1}_{\{L_{z_1} \leq t + X_{z_1} - z_1\}} = 1$), sont reçues après t ($\mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} = 1$), la valeur de la rupture est calculée en utilisant les expressions (2) et (7) :

$$R_t = \left(\begin{array}{l} U(t) - D_{z_1} \mathbb{1}_{\{L_{z_1} \leq t + X_{z_1} - z_1\}} \\ - \sum_{\substack{i \in \mathcal{M}_t \\ i \neq z_1}} D_i (1 - \mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}}) \end{array} \right)^+$$

Or $\forall i = p_m, \dots, p_M$, les L_i sont indépendantes, donc :

$$\begin{aligned} R_t &= \mathbb{1}_{\{L_{z_1} \leq t + X_{z_1} - z_1\}} \cdot (U(t) - D_{z_1})^+ \cdot \prod_{\substack{i \in \mathcal{M}_t \\ i \neq z_1}} \mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} \\ &= \prod_{i \in \mathcal{M}_t} \mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} \cdot (U(t) - D_{z_1})^+ \cdot \frac{\mathbb{1}_{\{L_{z_1} \leq t + X_{z_1} - z_1\}}}{\mathbb{1}_{\{L_{z_1} \geq t + X_{z_1} - z_1 + 1\}}} \end{aligned}$$

Nous notons qu'il existe $n_t = \text{card}(\mathcal{M}_t)$ possibilités pour z_1 ($z_1 = m_t, \dots, M_t$ tel que $z_1 \in \mathcal{M}_t$), donc :

$$\begin{aligned} \sum_{r_t^1 \in \mathcal{R}_t^1} r_t^1 \times P[R_t = r_t^1] &= \\ W(t) \times \sum_{\substack{z_1 = m_t \\ z_1 \in \mathcal{M}_t}}^{M_t} \left((U(t) - D_{z_1})^+ \times V(1, t) \right) & \quad (9) \end{aligned}$$

Si deux demandes lancées arrivent avant la fin de la période t ($\alpha = 2$), c.-à-d. toutes les demandes $D_i \setminus i \in \mathcal{M}_t$, à l'exception de deux d'entre elles ($z_1, z_2 \in \mathcal{M}_t$), sont reçues après la période t ($\mathbb{1}_{\{L_{z_1} \leq t + X_{z_1} - z_1\}} \times \mathbb{1}_{\{L_{z_2} \leq t + X_{z_2} - z_2\}} = 1$ et $\mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}} = 1$), la valeur de la rupture est calculée en utilisant les expressions (2) et (7) :

$$R_t = \left(\begin{array}{l} U(t) - \sum_{k=1}^2 D_{z_k} \mathbb{1}_{\{L_{z_k} \leq t + X_{z_k} - z_k\}} \\ - \sum_{\substack{i \in \mathcal{M}_t \\ i \neq \{z_1, z_2\}}} D_i (1 - \mathbb{1}_{\{L_i \geq t + X_{i-1} + 1\}}) \end{array} \right)^+$$

Or $\forall i = p_m, \dots, p_M$, les variables aléatoires L_i sont indépendantes, donc :

$$R_t = \prod_{i \in \mathcal{M}_t} \mathbb{1}_{\{L_i \geq t + X_{i-1}\}} \times \left(U(t) - \sum_{k=1}^2 D_{z_k} \right)^+ \times \prod_{k=1}^2 \left(\frac{\mathbb{1}_{\{L_{z_k} \leq t + X_{z_k} - z_k\}}}{\mathbb{1}_{\{L_{z_k} \geq t + X_{z_k} - z_k + 1\}}} \right)$$

Notons qu'il existe $\binom{n_t}{2}$ possibilités pour $(z_1, z_2) \mid z_2 = m_t + 1, \dots, M_t, z_1 = m_t, \dots, z_2$ et $z_1, z_2 \in \mathcal{M}_t$, alors :

$$\sum_{r_t^2 \in \mathcal{R}_t^2} r_t^2 \times P[R_t = r_t^2] = W(t) \times \sum_{\substack{z_2 = m_t + 1 \\ z_2 \in \mathcal{M}_t}}^{M_t} \sum_{\substack{z_1 = m_t \\ z_1 \in \mathcal{M}_t}}^{z_2} \left(\left(U(t) - \sum_{k=1}^2 D_{z_k} \right)^+ \times V(2, t) \right) \quad (10)$$

Nous pouvons déduire récursivement l'expression générale de $E[R_t]$ en utilisant les expressions (8), (9) et (10). ■

Le niveau de stock S_t à une période t est aussi une variable aléatoire discrète avec un nombre fini de valeurs possibles. L'expression de son espérance mathématique est donnée dans la propriété suivante.

Propriété 3. L'expression de l'espérance mathématique de S_t à une période t est la suivante :

$$E[S_t] = S_0 + \sum_{i=p_h}^t E[A_i] - \sum_{i=p_m}^t D_i + E[R_t] \quad (11)$$

avec

$$E[A_i] = \sum_{j \in \mathcal{A}_i} D_j \times P[L_j = i - j + X_j]$$

Preuve. A une période t , le niveau de stock est donné par l'expression suivante :

$$S_t = \left(S_0 + \sum_{i=p_h}^t A_i - \sum_{i=p_m}^t D_i \right)^+$$

En utilisant l'égalité $x = x^+ + x^-$ et l'expression (3), nous obtenons :

$$S_t = R_t + S_0 + \sum_{i=p_h}^t A_i - \sum_{i=p_m}^t D_i \quad (12)$$

Et ainsi $E[S_t]$ peut être facilement déduite. ■

La demande étant dynamique entre p_m et p_M , la rupture R_t peut-être non nulle entre $p_b = p_m$ et $p_B = \max_{t=p_m, \dots, p_M} (t - X_t + L_t^+ - 1)$ et S_t peut-être non-nulle entre p_h et p_H . L'expression de l'espérance mathématique du coût total est introduite dans la propriété suivante :

Propriété 4.

$$E[TC] = c^h (p_H - p_h + 1) S_0 + (c^h + c^b) \times \sum_{t=p_b}^{p_H} E[R_t] + c^h \left(\sum_{t=0}^{p_H - p_h} (p_H - p_h + 1 - t) \times E[A_{p_h + t}] - \sum_{t=0}^{p_H - p_m} (p_H - p_m + 1 - t) \times D_{p_m + t} \right) + c^b \times \sum_{t=p_M}^{p_B} E[R_t]$$

Preuve. En utilisant l'expression (1), on en déduit que

$$E[TC] = \sum_{t=p_h}^{p_H} c^h \times E[S_t] + \sum_{t=p_b}^{p_B} c^b \times E[R_t]$$

L'expression (11) nous permet d'obtenir :

$$E[TC] = c^h S_0 \sum_{t=p_h}^{p_H} 1 + c^h \sum_{t=p_h}^{p_H} \left(\sum_{i=p_h}^t E[A_i] - \sum_{i=p_m}^t D_i \right) + c^h \sum_{t=p_h}^{p_H} E[R_t] + c^b \sum_{t=p_b}^{p_B} E[R_t]$$

Sachant que :

- $\sum_{t=p_b}^{p_B} E[R_t] = \sum_{t=p_b}^{p_H} E[R_t] + \sum_{t=p_M}^{p_B} E[R_t]$
- $\sum_{t=p_h}^{p_H} E[R_t] = \sum_{t=p_b}^{p_H} E[R_t]$
- $\sum_{t=p_h}^{p_H} \sum_{i=p_h}^t E[A_i] = \sum_{t=0}^{p_H - p_h} (p_H - p_h + 1 - t) \cdot E[A_{p_h + t}]$
- $\sum_{t=p_h}^{p_H} \sum_{i=p_m}^t D_i = \sum_{t=0}^{p_H - p_m} (p_H - p_m + 1 - t) \cdot D_{p_m + t}$

Ainsi, l'expression de l'espérance mathématique du coût total peut être déduite. ■

Nous notons qu'une demande constante ou croissante facilite la résolution du problème.

3 MÉTHODES DE RÉOLUTION

Pour déterminer les délais d'obtention planifiés et le niveau de stock de sécurité minimisant l'espérance du coût total, quatre approches ont été appliquées :

NB&SS : c'est l'approche basée sur le modèle de Newsboy utilisée par Ben Ammar et al. (2017) pour déterminer les meilleurs délais planifiés et une méthode itérative pour le dimensionnement du stock de sécurité.

Xmin&SS : c'est l'approche optimiste qui consiste à planifier la production le plus tard possible. La séquence de délais planifiés est composé des valeurs minimales que les délais réels peuvent avoir ($X_t = L_t^- \forall t$). Ensuite, la méthode itérative permet de définir le stock de sécurité.

Xmax : c'est l'approche pessimiste qui consiste à planifier la production le plus tôt possible. La séquence de délais planifiés est composé des valeurs maximales que les délais réels peuvent avoir ($X_t = L_t^+ \forall t$).

Xopt&SS : c'est l'approche que nous proposons et qui se base sur le modèle mathématique introduit dans la section précédente. Un algorithme génétique classique (AG) est utilisé pour déterminer la meilleure séquence de délais planifiés pour un niveau de stock de sécurité donné. Il est exécuté en faisant varier le niveau du stock de sécurité entre 0 et la demande totale et ainsi la meilleure solution est sélectionnée. L'AG est caractérisé par :

- Une stratégie élitiste est utilisée,
- La population initiale est générée aléatoirement,
- Chaque individu de cette population est une séquence de délais planifiés,
- Chaque gène contient un délai planifié X_t qui peut varier entre L^- et L^+ ,
- L'espérance mathématique du coût total $E[TC]$, exprimée dans la propriété 4, est la fonction fitness utilisée pour évaluer chaque solution.

4 RÉSULTATS EXPÉRIMENTAUX ET DISCUSSIONS

4.1 Jeux de données

Les approches de résolutions mentionnées ci-dessus ont été testées sur des jeux de données contenant 12 combinaisons dont chacune se caractérise par un coût unitaire de rupture et une amplitude de variabilité de la demande. Les caractéristiques des jeux de données et les différentes combinaisons sont données dans les tableaux 1 et 2. Pour chaque combinaison, 100 tests sont réalisés avec un vecteur de demandes généré aléatoirement à chaque fois entre les périodes $p_m = 11$ et $p_M = 25$. Les distributions de probabilité des délais d'obtention sont présentées dans le tableau 3.

Coût unitaire de stockage	Coût unitaire de rupture			Variabilité de la demande			
	Faible	Moyenne	Élevé	Zéro	Faible	Moyenne	Élevée
6	7	15	25	$D=100$	$D \sim U(90,110)$	$D \sim U(70,130)$	$D \sim U(0,200)$

Tableau 1. Caractéristiques des instances.

Combinaison	1	2	3	4	5	6	7	8	9	10	11	12
Variabilité de la demande	Z	Z	Z	F	F	F	M	M	M	E	E	E
Coût unitaire de rupture (c^b)	F	M	E	F	M	E	F	M	E	F	M	E

Table 2. Combinaisons de jeux de données. Z=Zéro, F=Faible, M= Moyenne, E=Élevé(e).

4.2 Résultats et discussion

La figure 2 montre la variation de l'espérance du coût total pour la combinaison 8 (variabilité de la demande : $D \sim U(70,130)$ et coût unitaire de rupture = 15) en fonction du niveau du stock de sécurité. La meilleure solution connue est trouvée par *Xopt&SS* et est égale à 6284,5. Celle trouvée par *NB&SS* correspond à une espérance égale à 7164,9 et celle proposée par *Xmax* elle est égale à 11867,1. Pour ces trois approches, l'instauration d'un stock initial de sécurité n'a pas d'intérêt économique puisqu'il ne fait qu'augmenter l'espérance du coût total. Cependant, pour l'approche optimiste *Xmin&SS* la solution optimale correspond à un niveau moyen de stock initial de 124,6. Nous notons que la méthode *Xopt&SS* trouve, pour chaque valeur de S_0 , la meilleure combinaison de délais planifiés qui minimise le coût total. Ce stock de sécurité ne fait qu'augmenter le coût total en général mais il se peut que, pour certains cas, il permet de plus baisser le coût de rupture qu'augmenter le coût de stockage.

Le tableau 4 résume les meilleures solutions connues des 4 approches pour les 12 combinaisons.

l	1	2	3	4	5	6	7
$Pr[L_{11} = l]$	0.27	0.51	0.2	0.01	0.01		
$Pr[L_{12} = l]$		0.2	0.3	0.25	0.25		
$Pr[L_{13} = l]$			0.2	0.2	0.2	0.2	0.2
$Pr[L_{14} = l]$		0.25	0.3	0.25	0.2		
$Pr[L_{15} = l]$			0.5	0.5			
$Pr[L_{16} = l]$		0.3	0.3	0.4			
$Pr[L_{17} = l]$			0.5	0.5			
$Pr[L_{18} = l]$		0.2	0.8				
$Pr[L_{19} = l]$		0.6	0.4				
$Pr[L_{20} = l]$	0.2	0.2	0.2	0.2	0.2		
$Pr[L_{21} = l]$		0.25	0.25	0.25	0.25		
$Pr[L_{22} = l]$			0.3	0.2	0.2	0.2	0.1
$Pr[L_{23} = l]$		0.25	0.25	0.25	0.25		
$Pr[L_{24} = l]$			0.5	0.5			
$Pr[L_{25} = l]$		0.3	0.3	0.4			

Tableau 3. Distributions de probabilité des délais d'obtention

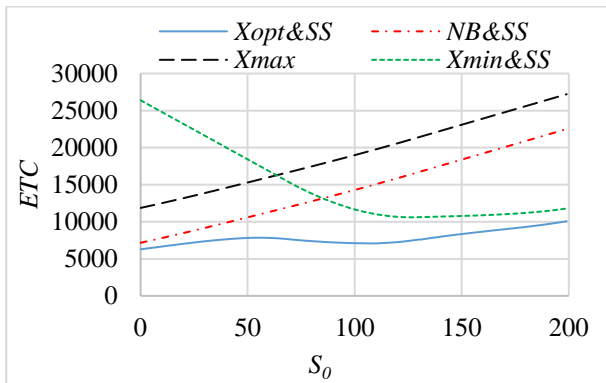


Figure 2. Variation de l'espérance mathématique du coût total en fonction du niveau moyen du stock initial.

Quelques soient la variation de la demande et le coût unitaire de rupture, l'approche *Xopt&SS* domine les trois autres. Pour *Xmin&SS*, le niveau moyen du stock initial est influencé par la variation du coût unitaire de rupture. Lorsque ce dernier est élevé, l'augmentation de la variabilité de la demande baisse légèrement le niveau moyen du stock initial.

Les performances relatives des différentes approches pour toutes les combinaisons sont présentées dans la figure 3. Nous notons que *Xopt&SS* est l'approche de référence.

Globalement, les approches *NB&SS* et *Xopt&SS* semblent dominer les deux autres (*Xmax* et *Xmin&SS*). Le gap moyen entre les deux premières approches diminue en fonction de la diminution de la variabilité de la demande. Cependant, il est petit seulement quand le coût unitaire de rupture est très proche du coût unitaire de stockage (12,42% quand il n'y a pas de variabilité de la demande et 6,44% quand la demande varie entre 0 et 200).

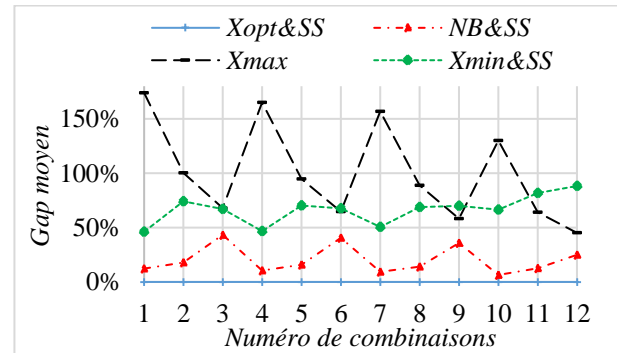


Figure 3. Performances relatives des différentes approches.

Concernant l'approche *Xmax*, même si elle paraît inefficace surtout dans le cas de coût unitaire de rupture faible, elle reste plus efficace que l'approche *Xmin&SS* dans le cas d'un coût unitaire de rupture élevé ou quand il est moyen et une grande variabilité de la demande.

Combinaison	<i>Xopt&SS</i>		<i>NB&SS</i>		<i>Xmax</i>		<i>Xmin&SS</i>	
	S_0	$E[TC]$	S_0	$E[TC]$	S_0	$E[TC]$	S_0	$E[TC]$
1	0	4287,6	0	4820,3	0	11742,0	100	6266,6
2	0	5861,6	0	6916,1	0	11742,0	100	10208,8
3	0	6995,3	0	10026,9	0	11742,0	200	11679,7
4	0	4389,6	0	4852,5	0	11638,6	100,6	6441,4
5	0	5987,8	0	6935,0	0	11658,7	107,8	10204,7
6	0	7094,5	0	9976,7	0	11673,3	195,1	11899,0
7	0	4599,2	0	5023,4	0	11810,1	104,5	6924,9
8	0	6284,5	0	7164,9	0	11867,1	124,6	10601,0
9	0	7359,3	0	10005,0	0	11651,4	187,3	12512,4
10	0	5136,5	0	5467,2	0	11823,3	104,2	8548,2
11	0	6805,3	0	7657,6	0	11164,1	157,7	12364,8
12	0	8002,1	0	9989,3	0	11612,7	184,5	15063,2

Table 4. Résultats pour les 12 combinaisons (moyenne sur 100 tests).

Pour analyser la robustesse de l'approche *Xopt&SS*, nous avons étudié l'effet de (i) la variabilité des délais d'obtention et (ii) le rapport coût unitaire de rupture/coût unitaire de stockage sur la qualité de la solution. Pour ce faire, nous faisons varier la variance des délais ainsi que le ratio c^b/c^h . Dans le cas de référence, tous les délais suivent la même loi de distribution ci-dessous : $Pr[L_t = 1] = 0,245$; $Pr[L_t = 2] = 0,48$; $Pr[L_t = 3] = 0,255$; $Pr[L_t = 4] = 0,01$ et $Pr[L_t = 5] = 0,01$.

Dans la figure 4, nous analysons l'effet de la variance des délais qui a été variée entre -75% et 75% par rapport au cas de référence.

Nous notons que la variation de l'espérance mathématique du coût total augmente quand la variance des délais augmente. Cependant, nous remarquons que : (i) les plus grandes augmentations se situent autour d'un ratio c^b/c^h proche de 1 (pour $c^b/c^h = 1,17$, la variation d'ETC varie entre -81% et 58,75%), (ii) pour $c^b/c^h \leq 0,17$ ou $2,5 \leq c^b/c^h \leq 8,33$, cette variation reste dans l'intervalle [-17%, 28%]. D'autre part, la variation de l'espérance mathématique du coût total dépend de l'estimation de la distribution de probabilité car elle augmente avec l'augmentation de la variance des délais et vice-versa. Cela signifie qu'il est préférable de surestimer la variance des

délais pour que la variation de l'espérance mathématique du coût total soit négative.

Plus généralement, nous mentionnons que, dans le cas où le coût unitaire de rupture est proche de celui du stockage, détenir une estimation fiable des distributions de probabilité des délais est primordial pour que l'approche que nous proposons soit efficace. D'autre part, la meilleure stratégie reste la prévention et l'élimination des sources d'incertitude qui perturbent les opérations et la performance d'une entreprise.

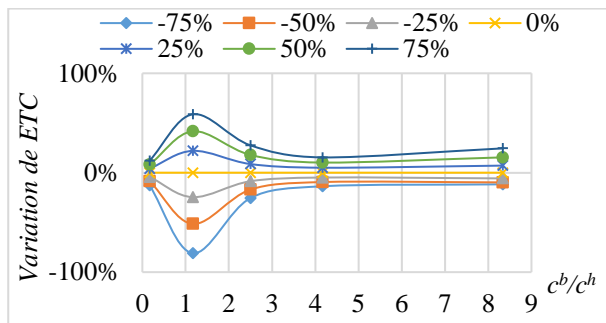


Figure 4. Effet de la variance des délais.

5 CONCLUSION ET PERSPECTIVES

Cet article étudie un problème de planification de la production multi-périodes avec une demande dynamique et des délais d'obtention stochastiques. Un modèle probabiliste est proposé et un algorithme génétique classique est développé afin de déterminer les délais planifiés et le niveau de stock de sécurité minimisant l'espérance du coût total. D'une part, l'approche proposée prouve que, il est plus avantageux d'optimiser les délais planifiés plutôt que de mettre en place un stock de sécurité. D'autre part, l'approche basée sur le modèle de Newsboy pourrait être efficace pour des problèmes de grandes tailles avec une grande variabilité de la demande et un coût de rupture relativement faible par rapport au coût de stockage.

Concernant les perspectives, nos travaux se focaliseront en premier lieu sur des tests de performances sur des horizons de planification avec plus de 15 demandes planifiées. En second lieu, il serait judicieux de tester les techniques proposées sur des cas industriels. Finalement, nous projetons étendre cette étude aux systèmes d'assemblage multi-niveaux avec des demandes incertaines. Il serait aussi intéressant d'intégrer cette approche avec des méthodes permettant de contrôler et réduire les incertitudes telles que la maintenance et le contrôle qualité.

REFERENCES

Axsäter S., 2005. Planning order releases for an assembly system with random operation times. *Operations Research Spectrum*, 27, 459–470.
 Ben-Ammar, O., Bettayeb, B., Dolgui, A. 2017. Supply planning optimization for linear product on system

with stochastic lead-times and quality control. In *International Conference on Industrial Engineering and Systems Management (IESM 2017)*.
 Ben-Ammar, O., Dolgui, A., Wu, D.D., 2017. Planned lead times optimization for multi-level assembly systems under uncertainties. *Omega*, <https://doi.org/10.1016/j.omega.2017.11.004>.
 Chopra, S., Reinhardt, G., Dada, M., 2004. The effect of lead time uncertainty on safety stocks. *Decision Sciences*, 35(1), 1-24.
 Chu, C., Proth, J.-M., Xie, X., 1993. Supply management in assembly systems. *Naval Research Logistics*, 40(7), 933-949.
 Deng, J., Song, P.J.S., Zhang, B., 2010. An inventory model with stochastic lead-time and stochastic demand. In *Control and Automation (ICCA), 8th IEEE International Conference on Control and Automation*, 1778-1782.
 Dolgui, A., Ben-Ammar, O., Hnaien, F., Louly, M.A., 2013. A state of the art on supply planning and inventory control under lead time uncertainty. *Studies in Informatics and Control*, 22(3), 255-268.
 He, X. J., Kim, J. G., Hayya, J. C., 2005. The cost of lead-time variability: The case of the exponential distribution. *International Journal of Production Economics*, 97(2), 130-142.
 Jansen, M.M., de Kok, T.G., Fransoo, J.C., 2013. Lead time anticipation in Supply Chain operations planning. *OR Spectrum*, 35(1), 251-290.
 Minner, S., 2003. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81, 265-279.
 Muthuraman, K., Seshadri, S., Wu, Q., 2014. Inventory management with stochastic lead times. *Mathematics of Operations Research*, 40(2), 302-327.
 Ould Louly, M.-A., Dolgui, A., 2013. Optimal MRP parameters for a single item inventory with random replenishment lead time, POQ policy and service level constraint. *International Journal of Production Economics*, 143(1), 35-40.
 Porteus, E. L., 2002. Foundations of stochastic inventory theory. *Stanford University Press*.
 Rossi, R., Tarim, S.A., Bollapragada, R., 2012. Constraint-based local search for inventory control under stochastic demand and lead time. *INFORMS Journal on Computing*, 24 (1), 66-80.
 Tang, O., Grubbström, R.W., 2003. The detailed coordination problem in a two-level assembly system with stochastic lead times. *International Journal of Production Economics*, 81-82, 415-429.
 Van Kampen, T.J., Van Donk, D.P., Van Der Zee, D.J., 2010. Safety stock or safety lead time: coping with unreliability in demand and supply. *International Journal of Production Research*, 48(24), 7463-7481.
 Wang, P., Hill, J. A., 2006. Recursive Behavior of Safety Stock Reduction: The Effect of Lead-Time Uncertainty. *Decision Sciences*, 37(2), 285-290.
 Yano C.A., 1987. Stochastic Leadtimes in Two-Level Assembly Systems. *IIE Transactions*, 19, 37.

OPTIMIZATION OF A HYDROGEN SUPPLY CHAIN NETWORK DESIGN: APPLICATION TO AN AIRPORT ECOSYSTEM

**Marialis GIRAUD BILLOUD, Jesus OCHOA
ROBLES, Catherine AZZARO-PANTEL**

Laboratoire de Génie Chimique, Université de
Toulouse, CNRS, Toulouse, France
4 allée Emile Monso, 31432 Toulouse Cedex 4, France.
marialis.gb@gmail.com, jesus.ochoarobles@ensiacet.fr,
catherine.azzaropantel@ensiacet.fr

Alberto Alfonso AGUILAR LASSERRE

Division of Research and Postgraduate Studies, Instituto
Tecnológico de Orizaba, Av. Oriente 9 # 852, Col.
Emiliano Zapata, 94300, Orizaba, Mexico.
aaguilar@itorizaba.edu.mx

ABSTRACT: *Hydrogen and fuel cell technologies can be one solution to address the challenges that major airports are facing today, such as upward price trends of liquid hydrocarbon fuels, greenhouse gas (GHG) emission regulations, stricter noise and air pollutant emission regulations especially for on-ground pollution. An airport can also be viewed as the centre of a hydrogen ecosystem around which multiple hydrogen users could be clustered, with cost sharing of hydrogen production and storage among users. The objective of this work within the Hyport project and in particular within the « H₂ modelling » project¹ involved is to design a hydrogen infrastructure irrigated by the airport ecosystem satisfying the airport ecosystem energy needs. This work is focused on a hydrogen airport ecosystem located in the department of Hautes-Pyrénées (France). The model is based on a multi-objective optimization framework designed to consider four echelons energy sources, hydrogen production, transportation and storage. The multi-period problem, is solved using the ϵ -constraint method. Two objective functions are considered, i.e., the Total Daily Cost of the network (TDC) and the Global Warming Potential (GWP).*

KEYWORDS: *Hydrogen Supply Chain, Airport ecosystem, MILP, optimization.*

1 INTRODUCTION

Hydrogen produced from renewable sources and used in fuel cells both for mobile and stationary applications constitutes a very promising energy carrier in the energy transition. The strategic roadmaps (ADEME, 2011; AFHYPAC, 2012; Agency for Natural Resources and Energy, 2014; Cueugnet et al., 2015; Dutta, 2014; H2ME 2, 2016, HyWays, the European Hydrogen Roadmap, 2008; IEA, 2015; Le Duigou et al., 2013; Momirlan and Veziroglu, 2002; Ozkan, 2009) that have been currently published about the potential of hydrogen at European, national and regional level as well as the analysis of the scientific publications in this field have identified that even if many of the required technologies are already available today, the deployment of hydrogen infrastructures constitutes the challenging task for the development of a «hydrogen» economy, so as to achieve competitive costs and mass market acceptance. According to these roadmaps, hydrogen demand for road transport is assumed to be a major parameter and is likely to grow in three phases according to energy trends of 2030 scenario (Ozkan, 2009):

- *Infrastructure phase I:* the demonstration phase during which a few large-scale first user centres are situated across Europe.
- *Infrastructure phase II:* the early commercialisation phase with 3-6 user centres per country, possibly including a network of transit roads connecting these centres.
- *Infrastructure phase III:* the full commercialisation phase, which encompasses existing user centres as well as newly developed regions and a dense, long-distance road network Phase III is assumed to develop in three sub-phases (see (Ozkan, 2009) for more details).

These large-scale deployment initiatives must be supported by long-term policy frameworks in countries that are early adopters. These initiatives should use current activities as platforms in order to deploy them at a national scale (Hydrogen Council, 2017). In that context, a Hydrogen Territory initiative was launched in France in 2016, which aims to demonstrate, on the scale of a particular territory, the technoeconomic feasibility as well as the environmental benefits of deploying hydrogen in energy

¹The « H₂ modelling » project (joint project with Laplace laboratory, INP Toulouse) within the Hyport project framework dedicated to the deployment of H₂ supply chains for an airport ecosystem supported by the Occitanie Region and MADELI has been labelled in November 2016 by the evaluation committee following the national call for projects « Territoires Hydrogène ». <https://www.usinenouvelle.com/article/l-occitanie-lance-son-programme-hvort-sur-deux-sites-aeronautiques.N595203>

networks or local energy applications. The Occitania region has been awarded with its meta-project HyPort, combining advanced innovation in hydrogen fuel cell applications for aeronautics, green hydrogen production and H₂ mobility deployment for Toulouse international airport, Tarbes-Lourdes-Pyrénées regional airport, and the vast urban or rural and touristic perimeters connected to them.

Due to its characteristics, the airport infrastructure is an interesting case study. An airport is a source of emissions that affect climate, including the emissions generated from activities occurring inside and outside the airport perimeter fence associated with the operation and use of an airport. The airport infrastructure accounts for 3.9% of Greenhouse Gas Emission of airplane transport and its related activities (MADEELI, 2016). Besides, urban mobility is directly impacted by airport activities: taxis, car rental, buses, tramways and subways, particular vehicles, lightweight and heavyweight utility vehicles, etc.

In that context, the hydrogen solution seems particularly interesting for the energy supply of the airport infrastructure and of the utility vehicles at non-public areas (ground support). The airport and surrounding areas are important stationary energy demanding zones, and there is an increasing interest in searching new energy solutions in order to improve environmental impact, autonomy, security and efficiency.

Several studies for Hydrogen Supply Chain (HSC) design have already been conducted (Agnolucci et al., 2013; Almansoori and Shah, 2009, 2012; De León Almaraz, 2014; Guillén-Gosálbez et al., 2010; Sabio et al., 2010) at a larger scale. The most common methodology to solve the HSC problems involves an MILP (Mixed Integer Linear Programming) approach.

For example, (Guillén-Gosálbez et al., 2010) design a HSC for vehicle use. The design task is formulated as a bi-criterion MILP problem. A case study in Great Britain is introduced to illustrate the capabilities of the proposed approach. The model optimizes an economic and environmental objective. The economic objective is given by the total discounted cost, and the environmental impact is expressed as the contribution to the climate change. Then, the problem is decomposed into two levels. The upper level refers to the so-called master problem with binary variables related to the selection of the different technologies. The lower level is represented by the original MILP model. The advantage of this methodology is the reduction of combinatorial complexity of the problem, and thus, its computational effort.

(Sabio et al., 2010) also design a HSC for vehicle use. The objective is to determine the optimal design of the production-distribution network. The model based on an MILP formulation, with uncertainty modelling introduced in the operating costs of the network has been applied to a Spain case of study.

A previous work (De León Almaraz, 2014) has been devoted to the development of a generic framework that can take into account the design of a HSC for fuel use in the time horizon 2020-2050 considering national (France) and regional (Midi-Pyrénées) scales with many energy sources and that can embed the various production and storage technologies while considering the transportation modes to link hydrogen demand to its supply. A multi-objective formulation is addressed in which cost, environmental impact and safety must be simultaneously taken into account at the earlier design stage. Even if the methodology proved to be robust enough to tackle different geographic scales, the way how the deployment can be operated from typical clusters and industrial ecosystems has not been studied so far.

To fill in this gap, this study is focused on the introduction of the hydrogen solution in an airport ecosystem and on its integration into a territory, i.e., the “department of Hautes-Pyrénées” in France. This airport is of major importance for the regional economy through its connection with tourism and industrial (aeronautics) activities.

In this work, the airport is viewed as the centre of a hydrogen ecosystem and the objective is to design a hydrogen infrastructure irrigated by the airport ecosystem satisfying the airport ecosystem energy needs. The methodological framework must be generic enough to be applied to other air(port) systems.

2 HYDROGEN IN AVIATION AND AIRPORTS

The aviation sector is increasingly facing challenges related to upward price trends of liquid hydrocarbon fuels, greenhouse gas (GHG) emission regulations, and stricter noise and air pollutant emission regulations especially for on-ground pollution at large airports. Biofuels cannot tackle all these issues, and shifting to hydrogen appears to be a promising alternative as highlighted in (Stiller and Schmidt, 2010). The replacement of fossil fuels in aviation industry has been widely studied. The most commonly used fuel for commercial aviation is kerosene which is highly pollutant, therefore, hydrogen and other fuels as methane and methanol are studied in order to replace it. In the study reported in (Yılmaz et al., 2012), a comparison between kerosene and hydrogen is presented based on their performance and environmental impacts. Hydrogen use in the aviation sector may concern an alternative fuel for future low-emission aircraft. Hydrogen-fuelled engines generate no CO₂ emissions at the point of use, may reduce NO_x emissions, and greatly diminish emissions of particulate matter. In (Janić, 2014), the GHG emissions of a conventional jet operated aircraft and one operated by cryogenic liquid H₂ for the long-term are evaluated. The author analyses different introduction rates and emphasizes that an efficient, reliable and safe supply chain is a prerequisite for hydrogen deployment. According to these studies, industry needs to overcome significant technical challenges in designing a hydrogen-powered aircraft for commercial aviation and in producing

enough hydrogen in a sustainable way, and highlight the need of a hydrogen infrastructure (Lee and Mo, 2011).

Hydrogen applications in aviation industry have been analysed for specific purposes. In (Stiller and Schmidt, 2010) an airport liquid Hydrogen infrastructure for aircraft Auxiliary Power Units (APU) is considered. The incorporation of on-site liquefaction units is considered for a long-term horizon, regarding applications and possible synergies such as hydrogen fuelled ground support equipment, apron vehicles and airport-bound vehicles.

Hydrogen is already present in airports worldwide, mostly with the incorporation of refuelling stations. Typical examples include international airports in Japan, i.e., Kansai (“Iwatani Europe: Hydrogen Station,” n.d.) and Narita (“Hydrogen - Renewable Energy - Idemitsu Kosan Global,” n.d.), Berlin Brandenburg Airport (BER) (“Green hydrogen facility opens at Berlin airport,” 2014), Munich airport in Germany “Munich Airport - Zone 3 - Cargo,” n.d.), and Oslo airport in Norway (“Prime Minister opens hydrogen filling station at Oslo Airport,” n.d.).

In this paper not only the uses of hydrogen required in the airport ecosystem are considered, but those for electromobility application of the Hautes-Pyrénées surrounding.

3 METHODS AND TOOLS

3.1 Modelling framework

For HSC deployment in Hautes-Pyrénées, the model developed by De Leon Almaraz (Sofia De León Almaraz, 2014) was adapted according to the new characteristics of the system. The model was designed in a generic way to be adapted to different scenarios, for example the addition of new energy sources or the modification in the geographic breakdown.

The original model proposes a multi-objective optimization designed to consider five stages: energy sources, production, transportation, storage and fuelling stations. It was designed to include regional and national levels in order to study the operability and evolution of the system at different scales. The problem, presented as a multi-period one, is solved using the ϵ -constraint method. There are two objective functions to be optimized, i.e. the total daily cost of the network, and the global warming potential. Finally, for the obtained Pareto front, the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) methodology is applied in order to select one of the optimal solutions.

For the sake of simplicity, the HSC base model (Sofia De León Almaraz, 2014) will be not described in detail.

The proposed approach is illustrated in Figure 1.

3.2 HSC model adaptation

As abovementioned, the mathematical model is an adaptation of the one previously developed by (Sofia De León Almaraz, 2014), which describes a Hydrogen Supply Chain Network (production, storage and distribution) for the Midi-Pyrénées region in France. The multi-period model uses a deterministic MILP (Mixed Integer Linear Programming) approach embedded in a GAMS/CPLEX environment with a multi-objective formulation implemented via the ϵ -constraint method to generate the Pareto front.

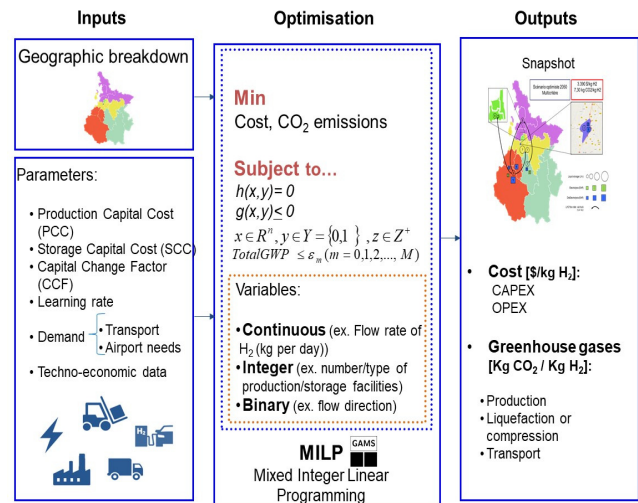


Figure 1: HSC model adapted from (Sofia De León Almaraz, 2014)

The following notations were used in the different constraints:

- g and g' : grid squares such that $g' \neq g$
- i : product physical form (LH₂)
- l : type of transportation modes (tanker truck)
- p : plant type with different production technologies (SMR, Electrolysis, DisElectrolysis)
- s : storage facility type with different storage technologies (LH₂ stock)

The following hypotheses were made:

- There are no production plants or storage units installed in the department before the first period of simulation.
- The learning rate of the system is fixed and equal to 12% per period.

The risk is not analysed in this study case, and the optimization objectives involve the Total Daily Cost of the network (TDC), and the total Global Warming potential (GWP).

3.3 Objective functions

The aim of this model is to minimize the TDC and the GWP. The total daily cost of the network in \$ per day, is defined as follows:

$$TDC = \left(\frac{FCC+TCC}{\gamma.CCF} \right) + FOC + TOC \quad (1)$$

In this expression, FCC is the facility capital cost (\$), TCC is the transportation capital cost (\$) γ is the network operating period (days per year) which is affected by the capital charge factor (CCF in years). The facility operating cost FOC (\$ per day) and the transportation operation cost TOC (\$ per day) are added to the equation in order to consider the totality of the related costs of the network.

The Global Warming Potential (in g CO₂-equivalent per day) is calculated as:

$$GWP_{TOTAL} = PGWP + SGWP + TGWP \quad (2)$$

which is given by the sum of the global warming potential due to the production facilities type p (PGWP), the storage technology (SGWP) and the daily transport (TGWP).

3.4 Geographical division

Before optimization, the geographical zone is discretized with an independent demand in each grid. A special grid is considered for the airport.

First, a study on the evolution of the municipal population of the department and on its distribution (division in canton) to predict the evolution of hydrogen demand (function of the predicted vehicles number) is performed. In the Hautes-Pyrénées, some cantons have a marked urban character: 56% of the inhabitants of the department reside in a cluster (agglomeration of more than 1 500 jobs), which is comparable to the regional average (58%).

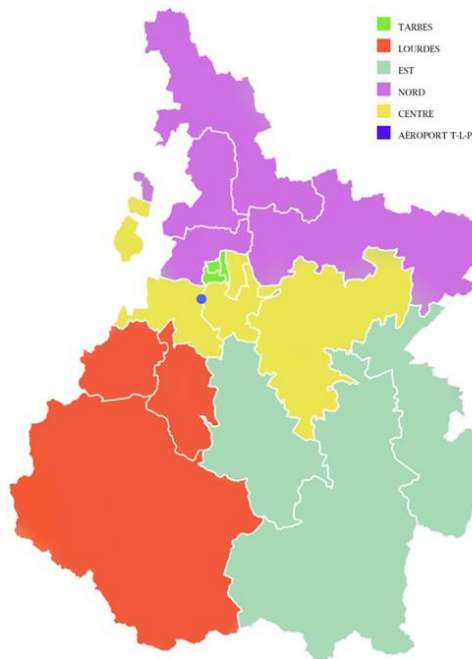


Figure 2: Definition of the grids

To simplify the problem and homogenize the identified needs, the grids have been grouped according to a geographical distribution, in order to finally obtain a model with 6 grids. Figure 2 represents the selected division.

3.5 Hydrogen needs and possible uses

The region considered is the department of Hautes-Pyrénées in France which was divided into 6 grids and each one of them is characterized by a specific demand. The grid division was made following a population density distribution based on statistic data from ("Insee - Institut national de la statistique et des études économiques," n.d.) and a geographical distribution criterion (based on proximity), assigning an independent grid for the airport.

Originally, the demand is fixed and can be satisfied either by local production, or by importation from other grids. Two base scenarios were considered for a low and a high demand case, based on the previous studies of (Salingue, 2012). The percentage of hydrogen network incorporation for mobility purposes corresponding to each scenario and period is presented in detail in Table 1.

Period	Low Demand Scenario	High Demand Scenario
2020	0.01	0.02
2030	0.075	0.15
2040	0.175	0.35
2050	0.25	0.5

Table 1: Percentages of hydrogen incorporation for transportation purposes for the scenarios considered

To identify the demand of vehicles for each grid, a study of their evolution in the last 20 years was conducted, and then a weight factor depending on population density was assigned. Finally, a 30 year-prediction following the observed trend was carried out for demand estimation. The categories considered are particular and commercial vehicles, buses, trucks and light weight vehicles (≤ 3.5 Tons) and agricultural tractors. The hydrogen demand is a function of the average distance covered in km/year and of the standard fuel economy for each category.

The particular energy needs of the airport grid have also been studied in more detail. They involve different categories: lighting, heating, generator power units, daily airplane movement and vehicle fleet. Each one of them was identified with the collaboration of the airport staff. The final aim was to cover 20% of the heating demand, and 100% demand required by vehicle fleet, generator and security power units by 2050. As the facilities are not yet fully deployed, the general demand was increased by 15% for each period, following the airport expected development scenario. The total demand was coupled with the scenarios previously defined, making the low demand half of the high demand scenario.

The assumptions based on the data provided by the airport are:

- The flight frequency ascends to 14 flights per day.
- The generator units of 90, 120 and 140 kVA, work 3 hours per day.
- The security system units of 100, 250 and 650kVA can provide energy for 48 hours if needed.
- The utility vehicles include 9 towing tractors of 400kW functioning 4 hours per day, 2 forklifts of 3kW functioning 7 hours per day and 5 light-weight vehicles with a medium distance of 10000km/year (captive fleet).
- The hydrogen consumption was calculated for each period considering the medium distances carried out by the buses and the number of planes per day.

The final amount of hydrogen needed for each grid and period, for the case of 50% more than the lower demand (the average demand between the low and high demand scenarios) is presented in Table 2.

Grid	Demand of hydrogen per period, t [kg/day]			
	2020	2030	2040	2050
1	200.5	1518	3606	5332
2	172.5	1305.5	3102	4586.5
3	223.5	1689	4012.5	5932.5
4	163.5	1236.5	2938	4344
5	155	1173.5	2788	4122.5
6	431.5	532.5	660.5	807.5

Table 2: Demand of 50% higher than the low demand used in this scenario

4 RESULTS

Several optimization runs have been carried out leading to a Pareto front. Then, a multi-criteria decision analysis method (TOPSIS: Technique for Order of Preference by Similarity to Ideal Solution) was applied in order to select the best compromise solution. A scenario of a demand 50% higher than the lower-scenario demand will be presented here in the form of a comparative table, and also through maps for the 4 periods analysed (2020, 2030, 2040, and 2050) in order to visualize the evolution of the network. The results for the chosen solutions are summarized in Table 3.

Figure 3 shows the evolution of the supply chain for the periods of 2020 until 2050, considering a 50 % demand increase. As it can be observed in the first period, there are only distributed plants, due to the low demand of hydrogen. The cost of the network for this period leads to \$19.39/kg H₂ with 4.07 kg CO₂eq/kg H₂. The investment costs of the supply chain are high in the first period, which is not surprising since there are no production plants or

storage facilities previously installed. With the development of the network, the costs decrease until reaching \$3.79/kg H₂ in 2050. Between the second and the third period, the demand increases, giving the possibility to incorporate transport units between grids, so that the costs are strongly reduced. The distribution also changes towards a centralized design, with production plants near the energy sources and hydrogen transportation to the other grids. For the airport grid, the demand is low enough to be satisfied by local production: a second production facility installed between the first and the second period.

Year	Multi-criteria Optimisation TOPSIS solution			
	2020	2030	2040	2050
Demand (t per day)	1.35	7.46	17.11	25.13
Number of total production facilities	7	14	9	14
Number of total storage facilities	6	11	25	34
Number of transport units	0	0	2	2
<i>Capital cost</i>				
Plants and storage facilities (10 ³ \$)	58184.690	79752.470	112279.700	75926.870
Transportation modes (10 ³ \$)	0.00	0.00	1000.00	1000.00
Total daily capital cost (10 ³ \$ per day)	58184.690	79752.470	113279.700	76926.870
<i>Operating Cost</i>				
Plants and storage facilities (10 ³ \$ per day)	12.821	85.802	52.628	77.418
Transportation modes (10 ³ \$ per day)	0.000	0.000	0.191	0.282
Total operating cost (10 ³ \$ per day)	12.821	85.802	52.818	77.700
Total network cost (10 ³ \$ per day)	26.105	104.010	78.681	95.263
Cost per kg H₂ (\$)	19.387	13.952	4.599	3.792
Production facilities (t CO ₂ -equiv per day)	4.539	28.688	35.377	51.959
Storage facilities (t CO ₂ -equiv per day)	0.948	5.248	12.043	17.688
Transportation modes (t CO ₂ -equiv per day)	0.000	0.000	0.373	0.551
Total GWP (t CO ₂ -equiv per day)	5.487	33.936	47.793	70.198
Kg CO₂-equiv per kg H₂	4.075	4.552	2.794	2.794

Table 3: Optimization results for the HSC

The amount of CO₂ emissions is also reduced between the second and the third period (because there are less production plants) reaching 2.79 kg CO₂ eq/kg H₂. These results depend strongly in the case of study, and for this case the energy sources are 100% renewable and mostly from hydraulic power plants, which have the lowest GWP values.

Table 3 shows the major differences between the periods that have been analysed. A different distribution network for the different periods is also observed. All the obtained results show a decrease in cost and CO₂ emissions (per kg H₂) along the periods.

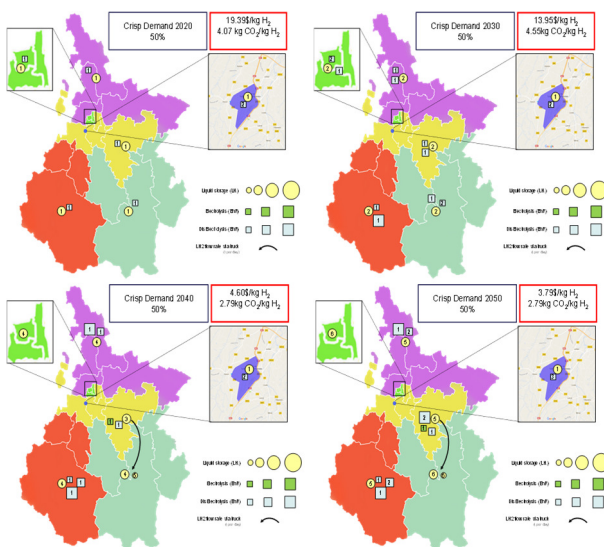


Figure 3: HSC distribution for the TOPSIS solution

5 CONCLUSIONS AND PERSPECTIVES

A HSC model based on a multi-optimization framework has been adapted in this work for an airport ecosystem. The viability of the HSC has been analysed along different time periods. In the last period (2050) the cost and the CO₂ emissions per kg of H₂ are the lowest, mostly due to the maturity of the HSC. In the first period, the cost is still prohibitive due to HSC deployment, (plants and storage units) and the subsequent low demand.

The CO₂ emissions are very low due to renewable source used, mainly hydropower, in this case. This type of energy is the less pollutant and also the cheapest, resulting in very similar solutions between periods.

The application of the airport system is a very interesting hydrogen platform as it permits to introduce hydrogen to a strategic point not only considering aircraft utilization, but also considering the activities generated by the airport (eco-mobility, tourist interest, etc.).

The conceptual project design described above may be replicated in other regions where there is a (air)port eco-

system. This modelling approach can be useful to consider various ways to cluster multiple hydrogen users around an (air)port ecosystem, with cost sharing of hydrogen production and storage among users and to develop rollout strategies.

Some perspectives can be also highlighted. For example, some parameters can be modelled under uncertainty, such as demand, costs or prices involved in the model. The TDC can be substituted by the use of discounted costs associated with each time period. The influence of other parameters such as the safety stock period, new tube trailer capacities and the use of pipelines could also be evaluated in future case studies.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank Pascal Le Houelleur, Director of Pyrenia, Mathilde Convert (MADEELI) who kindly shared their knowledge or provided data on various aspects of Tarbes-Lourdes airport.

REFERENCES

- ADEME, 2011. Roadmap on hydrogen energy and fuel cells [WWW Document]. URL <http://www.ademe.fr/sites/default/files/assets/documents/hydrogen-energy-fuel-cells-strategic-roadmap-2011-6924-summary.pdf> (accessed 12.31.17).
- AFHYPAC, 2012. Rapport d'activité de la France sur l'Hydrogène et les Piles à Combustible, période juin 2012 à novembre 2014 [WWW Document]. URL http://www.afhypac.org/documents/publications/rapports/AFHYPAC_RA_bd.pdf
- Agency for Natural Resources and Energy, 2014. Summary of the Strategic Road Map for Hydrogen and Fuel Cells.
- Agnolucci, P., Akgul, O., McDowall, W., Papageorgiou, L.G., 2013. The importance of economies of scale, transport costs and demand patterns in optimising hydrogen fuelling infrastructure: An exploration with SHIPMod (Spatial hydrogen infrastructure planning model). *Int. J. Hydrog. Energy* 38, 11189–11201. <https://doi.org/10.1016/j.ijhydene.2013.06.071>
- Almansoori, A., Shah, N., 2012. Design and operation of a stochastic hydrogen supply chain network under demand uncertainty. *Int. J. Hydrog. Energy* 37, 3965–3977.
- Almansoori, A., Shah, N., 2009. Design and operation of a future hydrogen supply chain: Multi-period model. *Int. J. Hydrog. Energy* 34, 7883–7897. <https://doi.org/10.1016/j.ijhydene.2009.07.109>
- Cueugnet, J., Legait, B., Durville, J.-L., Gazeau, J.-C., Nataf, J.-M., 2015. Filière hydrogène-énergie [WWW Document]. Portail Ministères Économiques Financ. URL <https://www.economie.gouv.fr/cge/filiere-hydrogene-energie> (accessed 12.31.17).

- De León Almaraz, S., 2014. Multi-objective optimisation of a hydrogen supply chain. Toulouse, INPT.
- De León Almaraz, S., 2014. Multi-objective optimisation of a hydrogen supply chain. *École Doctorale Mécanique, Energétique, Génie civil et Procédés (Toulouse)*; 154236012.
- Dutta, S., 2014. A review on production, storage of hydrogen and its utilization as an energy resource. *J. Ind. Eng. Chem.* 20, 1148–1156. <https://doi.org/10.1016/j.jiec.2013.07.037>
- Green hydrogen facility opens at Berlin airport, 2014. *Fuel Cells Bull.* 2014, 1. [https://doi.org/10.1016/S1464-2859\(14\)70122-1](https://doi.org/10.1016/S1464-2859(14)70122-1)
- Guillén-Gosálbez, G., Mele, F.D., Grossmann, I.E., 2010. A bi-criterion optimization approach for the design and planning of hydrogen supply chains for vehicle use. *AIChE J.* 56, 650–667. <https://doi.org/10.1002/aic.12024>
- H2ME 2 launched in Europe to grow hydrogen fueling infrastructure network and vehicle fleet, 2016. *Fuel Cells Bull.* 2016, 1. [https://doi.org/10.1016/S1464-2859\(16\)30131-6](https://doi.org/10.1016/S1464-2859(16)30131-6)
- Hydrogen - Renewable Energy - Idemitsu Kosan Global [WWW Document], n.d. URL <http://www.idemitsu.com/products/energy/battery/index.html> (accessed 7.20.17).
- HyWays, the European Hydrogen Roadmap, 2008. *European Commission*.
- IEA, 2015. Technology Roadmap. Hydrogen and Fuel Cells [WWW Document]. URL <https://www.iea.org/publications/freepublications/publication/TechnologyRoadmapHydrogenandFuelCells.pdf> (accessed 12.31.17).
- Insee - Institut national de la statistique et des études économiques [WWW Document], n.d. URL <https://www.insee.fr/fr/accueil> (accessed 6.15.17).
- Iwatani Europe: Hydrogen Station [WWW Document], n.d. URL <http://www.iwatani-europe.de/hydrogen-station.html> (accessed 7.20.17).
- Janić, M., 2014. Greening commercial air transportation by using liquid hydrogen (LH2) as a fuel. *Int. J. Hydrog. Energy* 39, 16426–16441. <https://doi.org/10.1016/j.ijhydene.2014.08.011>
- Le Duigou, A., Quéméré, M.-M., Marion, P., Menanteau, P., Decarre, S., Sinégre, L., Nadau, L., Rastetter, A., Cuni, A., Mulard, P., Antoine, L., Alleau, T., 2013. Hydrogen pathways in France: Results of the HyFrance3 Project. *Energy Policy* 62, 1562–1569. <https://doi.org/10.1016/j.enpol.2013.06.094>
- Lee, J., Mo, J., 2011. Analysis of Technological Innovation and Environmental Performance Improvement in Aviation Sector. *Int. J. Environ. Res. Public. Health* 8, 3777–3795. <https://doi.org/10.3390/ijerph8093777>
- MADEELI, 2016. Presentation du meta-projet Hyport en vue de la labellisation “Territoires Hydrogène”. Dossier Principal.
- Momirlan, M., Veziroglu, T.N., 2002. Current status of hydrogen energy. *Renew. Sustain. Energy Rev.* 6, 141–179. [https://doi.org/10.1016/S1364-0321\(02\)00004-7](https://doi.org/10.1016/S1364-0321(02)00004-7)
- Munich Airport - Zone 3 - Cargo [WWW Document], n.d. URL <https://www.munich-airport.de/en/micro/technik/zonen/fracht1/index.jsp> (accessed 7.20.17).
- Ozkan, N., 2009. A review of hydrogen demands in national roadmaps. *Policy Studies Institute*.
- Prime Minister opens hydrogen filling station at Oslo Airport, n.d.
- Sabio, N., Gadalla, M., Guillén-Gosálbez, G., Jiménez, L., 2010. Strategic planning with risk control of hydrogen supply chains for vehicle use under uncertainty in operating costs: a case study of Spain. *Int. J. Hydrog. Energy* 35, 6836–6852.
- Salingue, C., 2012. Optimisation de la chaîne logistique de l'hydrogène en région Midi-Pyrénées (Mémoire de stage). ENSEIHT and ENSIACET (INP), Toulouse, France.
- Stiller, C., Schmidt, P., 2010. Airport Liquid Hydrogen Infrastructure for Aircraft Auxiliary Power Units. Presented at the 18th World Hydrogen Energy Conference 2010 - WHEC 2010, Detlef Stolten, Thomas Grube (Eds.), Essen.
- Yılmaz, İ., İlbaş, M., Taştan, M., Tarhan, C., 2012. Investigation of hydrogen usage in aviation industry. *Energy Convers. Manag.*, 10th International Conference on Sustainable Energy Technologies (SET 2011) 63, 63–69. <https://doi.org/10.1016/j.enconman.2011.12.032>

OPTIMIZATION OF ANALYZERS CONFIGURATION IN A CLINICAL LABORATORY: A MATHEMATICAL MODEL

S. Faramarzi Oghani, E-G. Talbi, M. Bué

E. Varlet

INRIA Lille-Nord Europe - 40 Avenue Halley
59650 Villeneuve d'Ascq, France
sohrab.faramarzi-oghani@inria.fr,
el-ghazali.talbi@inria.fr,
martin.bue@inria.fr

Beckman Coulter, Normand-Info - 7 Rue Frédéric
Degeorge, CS 10451 - 62028 Arras Cedex, France
evarlet@beckman.com

ABSTRACT: *This paper presents an innovative application of operations research in clinical analysis laboratories domain. Clinical analysis laboratory is an organization gathering human and machinery resources to analyze human fluid samples such as blood and urine. Normally, analyzers require material called reagent, to be able to perform medical tests. Allocating appropriate type and amount of different reagents to the analyzers is a daily challenging problem in clinical laboratories. Analyzer configuration problem deals with the specification of the number and type of reagent bottles used in each analyzer to satisfy the daily demand with the aim of minimizing total cost of reagent bottles used in the analyzers as well as analyzers calibration costs. In this study, a mathematical model is developed to cope with this problem. The proposed model is solved optimally by Cplex solver in GAMS software for a laboratory which covers Immunology and Chemistry tests. The output results reveal the efficiency of the proposed approach to aid decision makers determining analyzers configuration effectively and intelligently. To the best of our knowledge, this study is the first effort to address and tackle analyzers configuration in clinical laboratories as an optimization problem.*

KEYWORDS: *healthcare, clinical analysis laboratory, analyzer configuration, reagent, mathematical modeling*

1 INTRODUCTION

Clinical analysis laboratory is an organization gathering human and machinery resources to analyze human fluid samples such as blood and urine. These laboratories are noticed as one of the principle and explicitly preliminary blocks in health services where most of the medical diagnoses and treatments depend on. Therefore, efficiency and effectiveness of these organizations have straight impact on the performance of other dependent health sectors. Furthermore, reducing operating costs in laboratories decreases the cost of treatment and eventually increases patient satisfaction. Surveying the current situation of clinical laboratories implies the trend of fusion among laboratories which on one side, decreases the number of clinical laboratories and on the other side, enlarges the magnitude of the emerging laboratories. This pervasive phenomenon which is mainly the consequence of financial pressures done by reimbursement organizations addresses the clinical laboratories as huge organizations which require profound and precise scientific efforts while designing and planning such complex systems. Although, a wide number of studies has been carried out to improve and optimize healthcare systems [see (Onar *et al.*, 2018) and (Brailsford and Vissers, 2011)], optimization of clinical laboratories still can be spotted as a virgin domain of research where only a few number of dispersed studies is observed in the literature of operations research.

The rest of this paper is organized as follows: section (2) reviews the research activities on modeling and optimization of clinical laboratories. In section (3), analyzer configuration problem is defined with great attention to details. Section (4) mathematically formulates the problem. Computational results are illustrated in section (5). Finally, section (6) ends the paper with conclusion and future directions.

2 LITERATURE REVIEW

Most of the research efforts concerning the application of operations research methods in improvement and optimization of clinical laboratories have been started since the advent of The OpenLabs Project (O'Moore *et al.*, 1994). This project is concerned with the use of advanced information science and telematics for the optimization of clinical laboratory services (O'Moore *et al.*, 1996). As part of this project, (Merode *et al.*, 1996) developed a simulation module for laboratory capacity planning. Introducing computerised validation systems within the clinical laboratories to automatically validate patient results is also one of the remarkable advances related to The OpenLabs program (Boran *et al.*, 1996). To optimize the structure of a clinical laboratory, (Merode *et al.*, 1998) proposed a mathematical programming model to cluster workstations to minimize

the maximum idle time of staff while assigning them to the clusters. (Maria Elena *et al.*, 2006) introduced a mathematical model to help improving the laboratory organization in relation with staff assignment and specification of samples analyzing method decisions. To efficiently schedule the personnel in laboratories, (Franses and Post, 2003) developed an algorithm to find an acceptable matching between the workers and tasks in different periods. Concerning the planning and scheduling of patient tests in laboratories, (Marinagi *et al.*, 2000) proposed a scheduling system for hospital laboratories to support the continual and dynamic nature of the problem.

Simulation modeling is the most-used tool to analyze, evaluate and improve the planning and design of clinical laboratories over the past years. Computer simulation as a powerful and versatile tool is suitable for full understanding of a complex system, bottleneck identification, change management, scenario analysis and projection. To make a wise decision on automating the central processing laboratory in Mayo clinic, (Dankbar *et al.*, 1992) developed several simulation models to evaluate the effect of different vendor proposals on system performance. (Berchtold *et al.*, 1994) endeavoured to develop a generalized flexible simulation model to assess arbitrary changes in the structure and configuration of the clinical laboratories. They expressed average turnaround time, resource utilization and average queue length as three significant performance measures in clinical laboratories studies. (Lote *et al.*, 2009) emphasized on the use of discrete-event simulation with other industrial engineering techniques to improve medical testing laboratories. As a part of the paper, optimizing courier routes and evaluating the impact of such improvement on operational performance of the laboratory as well as levelling utilization of resources to achieve optimal resource allocation were studied on a medical laboratory located in the eastern part of the United States. (Luangmul *et al.*, 2012) developed a simulation model of a hospital's clinical laboratory in Thailand focusing only on the Complete Blood Cell (CBC) test through using ARENA simulation software. The aim of the study was to verify the impact of potential changes like physical layout alternation by a verified and validated simulation model. Biochemistry and hormone test laboratories of a university hospital in Turkey were studied via simulation modeling by (Kadi *et al.*, 2016). The aims of the study were firstly, to achieve a full understanding of the current system and identify possible existing bottlenecks and secondly, to conduct scenario analysis in order to improve the current state of the system. Improvement scenarios were made based on process analysis solutions and resource alternation decisions. ARENA simulation software was used to develop the simulation model and to capture the average time in system of a patient as the main performance measure of this study. For more studies on the application of simulation modeling for clinical laboratories improvement and optimization see

[(Vogt *et al.*, 1994), (Groothuis *et al.*, 2002) and (Bodtker *et al.*, 1993)].

Investigating the literature of clinical laboratory optimization reveals the lack of coherence among the research activities in this field. There are a few number of studies demonstrating the application of operations research, mainly mathematical programming in medical laboratory decision making problems. In addition, simulation studies have been conducted on a specific problem for a given case. Since clinical laboratories are being merged, advent of giant laboratories is inevitable in near future. Therefore, new generation clinical laboratories require productive decisions concerning the planning and design processes to efficiently create and manage such a complex system on one hand and reduce the costs on the other hand.

The main aim of this paper is to firstly introduce - to the best of our knowledge - a new optimization problem in clinical laboratories, called 'Analyzer Configuration Problem' (ACP), and secondly to propose a method to deal with the introduced problem. Section (3) and (4) covers the definition and formulation of ACP in clinical laboratories.

3 ANALYZER CONFIGURATION PROBLEM

In clinical laboratories, analyzers are used to perform the tests on patient samples. Each analyzer belongs to one or more test discipline which implies the potential capability of analyzer to perform clinical tests. In better words, analyzers are able to do the tests of a discipline to which they belong. For instance, a Chemistry analyzer is potentially able to carry out only Chemistry tests. To practically equip an analyzer to a test, a matching reagent is required in the analyzer. For instance, to equip a Chemistry analyzer to Triglyceride test which is a Chemistry test, Triglyceride reagent must be available on the analyzer. Reagents are materials used by the analyzers to conduct the tests on patient samples. Normally, for each test type, a specific type of reagent is needed. Reagents are available in bottles with different sizes. It is worth noting that each analyzer has only a certain number of positions to room reagent bottles. Reagent bottles with different sizes occupy different number of positions in analyzers. In addition, Efficiency of analyzers in terms of reagent consumption to perform a test is different. Furthermore, cost of a reagent bottle rely on the bottle size, test type and analyzer in which the bottle is loaded. Each available test type on an analyzer must be calibrated which imposes a cost to the system, called calibration cost. Analyzer configuration is the problem of specification of the number and type of reagent bottles used in each analyzer to satisfy the daily demand with the aim of minimizing total cost of reagent bottles used in the analyzers as well as analyzers calibration costs.

4 PROPOSED MATHEMATICAL MODEL

In this section, assumptions considered to formulate the problem are firstly described. Then, notations used in the proposed mathematical model are introduced. Mathematical formulation and related explanation terminate this section.

4.1 Assumptions

- Analyzers only belong to one test discipline. Disciplines are Immunology, Chemistry, Hematology, etc.
- The number and type of existing analyzers in the laboratory is given.
- In order to analyze each test type, a specific reagent is required. Generally, reagents are the material used by the analyzers to perform the test.
- Capacity of each analyzer in terms of tests number relies on the analyzer technical features and the amount of reagent bottles positioned into the analyzer. Following formula is proposed to consider daily Analyzer Operational Capacity (AOC):

$$AOC_j = \min \{ \tau_j g_j, \sum_h \sum_s \delta_{hsj} x_{hsj} \}$$

According to the proposed formula, the daily operational capacity of analyzer j equals to the minimum of daily nominal capacity of the analyzer and total number of tests that can be analyzed by the analyzer regarding the number of reagent bottles assigned to that analyzer ($\sum_h \sum_s \delta_{hsj} x_{hsj}$). Nominal capacity of an analyzer is computed based on the multiplication of the analyzer capacity (g_j) provided by the manufacturer in terms of number of tests per hour to the total daily available working hours (τ_j).

- Total daily available working time for each analyzer implies the time that the analyzer is available for operating and analyzing. So, times required for calibration, maintenance and refilling reagents are reduced from this time.
- Each analyzer has a certain number of reagent positions to room reagent bottles.
- Each reagent bottle occupies certain number of positions in the analyzer depending on the reagent type and bottle size.
- Reagent bottle cost relies on the reagent type, bottle size and the analyzer in which the reagent bottle is used.

- Number of tests that can be analyzed using a bottle of reagent depends on the reagent type, bottle size and the analyzer in which reagent bottle is used.

- Calibration is performed per test type on each analyzer. So, calibration cost depends on the test and analyzer type and it is computed per test type on each analyzer neglecting the number and size of reagent bottles used for the test type in the analyzer. This cost is estimated based on the time required to perform calibration tests as well as amount of reagent required to perform these tests.

- There must be a sufficient capacity in the laboratory to meet all the requested tests within a working day.

- Configuration of the analyzers is performed on daily basis and is independent from previous days implying that the remaining reagents in the analyzers are discarded at the end of the day.

4.2 Notations

This section introduces notations used in the proposed mathematical model.

Sets

D : set of disciplines

M : set of analyzers

M^d : set of analyzers in discipline d

H : set of tests

Q : set of reagent bottle sizes

Indices

d : index of discipline; $d \in D = \{1, 2, \dots, l\}$

j : index of analyzer; $j \in M = \{1, 2, \dots, m\}$

h : index of test; $h \in H = \{1, 2, \dots, o\}$

s : index of reagent bottle size; $s \in Q = \{1, 2, \dots, q\}$

Parameters

g_j : average number of tests that can be analyzed by analyzer j per hour

τ_j : available working hours of analyzer j per day

RK_j : reagent capacity of analyzer j

F_h : average number of test h requested per day

CC_{hj} : calibration cost of test h on analyzer j

α_{hj} : average number of test h required to calibrate analyzer j

λ_{hs} : number of reagent positions occupied in an analyzer by reagent bottle h with size s

HM_{hj} : takes 1, if test h can be potentially done by analyzer j ; otherwise, takes 0

HD_{hd} : takes 1, if test h is in discipline d ; otherwise, takes 0

RC_{hsj} : cost of a reagent bottle with size s used for test h in analyzer j

δ_{hsj} : average number of test h that can be analyzed using one bottle of reagent size s by analyzer j

Decision variables

AOC_j : operational capacity of analyzer j

x_{hsj} : number of reagent bottle h with size s positioned into analyzer j

y_{hj} : takes 1, if test h is available on analyzer j ; otherwise, takes 0

4.3 Mathematical formulation

Equation (1) is the objective function which minimizes total costs of reagent bottles used in the analyzers and total calibration costs.

$$\sum_{h \in H} \sum_{s \in Q} \sum_{j \in M} RC_{hsj} x_{hsj} + \sum_{h \in H} \sum_{j \in M} CC_{hj} y_{hj} \quad (1)$$

Constraints (2) and (3) demonstrate the daily operational capacity of each analyzer which cannot be neither more than analyzer nominal capacity (Constraint (2)), nor more than the capacity created by the number of reagent bottles assigned to the analyzer denoting the total number of tests that can be processed by the analyzer (Constraint (3)).

$$AOC_j \leq \tau_j g_j \quad ; \quad \forall j \in M \quad (2)$$

$$AOC_j \leq \sum_{h \in H} \sum_{s \in Q} \delta_{hsj} x_{hsj} \quad ; \quad \forall j \in M \quad (3)$$

Constraint (4) assures that there is a sufficient capacity (capability to analyze certain number of tests) in the laboratory to handle all the daily requested tests from different disciplines.

$$\sum_{j \in M^d} AOC_j \geq \sum_{h \in H} HD_{hd} (F_h + \sum_{j \in M^d} \alpha_{hj} y_{hj}) \quad ; \quad \forall d \in D \quad (4)$$

Constraint (5) guarantees that there are sufficient reagents in the existing analyzers to calibrate the analyzers and to analyze each requested test within a day.

$$\sum_{j \in M} \sum_{s \in Q} \delta_{hsj} x_{hsj} \geq F_h + \sum_{j \in M} \alpha_{hj} y_{hj} \quad ; \quad \forall h \in H \quad (5)$$

Constraint (6) assures that the number of reagent bottles positioned into each analyzer must not exceed from the available number of reagent positions on each analyzer.

$$\sum_{h \in H} \sum_{s \in Q} \lambda_{hs} x_{hsj} \leq RK_j \quad ; \quad \forall j \in M \quad (6)$$

Constraint (7) demonstrates whether test h is available on analyzer j or not to provide useful information to compute calibration cost of each test type on the analyzers.

$$\sum_{s \in Q} \lambda_{hs} x_{hsj} \leq RK_j y_{hj} \quad ; \quad \forall h \in H; j \in M \quad (7)$$

Constraint (8) presents the potential eligibility of each analyzer to perform a test. Analyzers of each discipline are only able to analyze the tests which belong to the associated discipline.

$$y_{hj} \leq HM_{hj} \quad ; \quad \forall h \in H; j \in M \quad (8)$$

Constraints (9) to (11) imply the type of decision variables used in the model.

$$AOC_j \geq 0, \text{int} \quad ; \quad \forall j \in M \quad (9)$$

$$x_{hsj} \geq 0, \text{int} \quad ; \quad \forall h \in H; s \in Q; j \in M \quad (10)$$

$$y_{hj} \in \{0,1\} \quad ; \quad \forall h \in H; j \in M \quad (11)$$

5 MODEL VALIDATION

To verify the validity of the proposed model, analyzer configuration problem is defined and solved for a clinical laboratory where ten analyzers exist. Existing analyzers are classified into two test disciplines: five Immunology and five Chemistry analyzers. Table 1 presents analyzers discipline, daily available working time of each analyzer, analyzers nominal capacity and number of available positions for reagent bottles for each analyzer.

The laboratory purposes to analyze fifty different test types in each discipline. The average total number of daily requested Immunology and Chemistry tests is 9,702 and 36,130, respectively. Two different bottle sizes are available for each test type. Data regarding the price of reagent bottles for each type and size is extracted from the manual provided by the supplier. In addition, data concerning the efficiency of each analyzer to use a bottle of reagent for specific test and size is

adapted from laboratory database. In this laboratory, configuration of the analyzers is carried out on daily basis and is independent from previous days. In better words, the remaining reagents in the analyzers are discarded at the end of each day.

Analyzer	Discipline	Daily available working time (hour)	Nominal capacity (test/hour)	Available reagent positions
J1	Immunology	8	200	50
J2	Immunology	8	200	50
J3	Immunology	8	200	50
J4	Immunology	8	400	50
J5	Immunology	8	400	50
J6	Chemistry	8	800	54
J7	Chemistry	8	800	54
J8	Chemistry	8	800	54
J9	Chemistry	8	1200	108
J10	Chemistry	8	1200	108

Table 1. Analyzer related input data

The model is coded in GAMS 24.1.3 software and solved using Cplex solver in a PC with 2.60 GHz CPU and 4.00 GB RAM. The optimal solution is achieved in less than seven minutes with the value of 117,331 monetary unit for the objective function indicating total required cost for daily analyzer configuration and calibration. In addition, the outputs specify the optimal number and size of different reagent types used in each analyzer to meet all daily demand with minimum configuration and calibration costs.

6 CONCLUSION AND FUTURE DIRECTIONS

Reagent and calibration costs constitute a huge part of clinical laboratory operating costs. In this paper, analyzer configuration problem in clinical laboratories is introduced to allocate different type and size of reagents in different analyzers in an efficient manner. To tackle this problem, a mathematical model is developed and then solved by Cplex solver. Experimental results demonstrate the validity and usefulness of the proposed model to deal with ACP.

As the focus of the proposed model is only on the single-discipline analyzers, considering multi-discipline analyzers could be an interesting extension to this model. Multi-discipline analyzers can be spotted as a combination of single-discipline analyzers, so that they are able to handle more test disciplines and have limited number of reagent bottle position for each discipline.

Although analyzer configuration problem has relatively operational essence, costs of configuration and also calibration which mainly rely on the type of analyzers, must be taken into account while equipping the laboratory with analyzers in strategic decisions. As a

result, integration analyzer selection and configuration problem is a promising future direction.

REFERENCES

- Berchtold, G., Blaschke, H., Hanssmann, F., Liebl, F., Braun, S.L., Vogt, W., Eckert, M., Hoffmann, G. and Klose, S., 1994. Simulation modeling as a tool to evaluate alternative configurations of clinical laboratories. *Simulation*, 63(2), pp.108-120.
- Bodtker, K., Wilson, L. and Godolphin, W., 1993. Simulation modelling to assist operational management and planning in clinical laboratories. *Simulation*, 60(4), pp.247-255.
- Boran, G., Given, P. and O'Moore, R., 1996. Patient result validation services. *Computer methods and programs in biomedicine*, 50(2), pp.161-168.
- Brailsford, S. and Vissers, J., 2011. OR in healthcare: A European perspective. *European journal of operational research*, 212(2), pp.223-234.
- Dankbar, G.C., Shellum, J.L. and Bennet, K.E., 1992, December. The use of simulation to evaluate automated equipment for a clinical processing laboratory. In *Proceedings of the 24th conference on Winter simulation* (pp. 1065-1070). ACM.
- Elena, B.M., Beraldi, P. and Conforti, D., 2006. Improving the efficiency of a clinical laboratory: A mathematical approach. *IFAC Proceedings Volumes*, 39(3), pp.659-664.
- Franses, P. and Post, G., 2002, August. Personnel scheduling in laboratories. In *International Conference on the Practice and Theory of Automated Timetabling* (pp. 113-119). Springer, Berlin, Heidelberg.
- Groothuis, S., Goldschmidt, H.M., Drupsteen, E.J., Vries, J., Hasman, A. and Merode, G.G.V., 2002. Turn-around time for chemical and endocrinology analyzers studied using simulation. *Clinical chemistry and laboratory medicine*, 40(2), pp.174-181.
- Kadı, D., Kuvvetli, Y. and Çolak, S., 2016. Performance analysis of a university hospital blood laboratory via discrete event simulation. *Simulation*, 92(5), pp.473-484.
- Lote, R., Williams, E.J. and Ülgen, O.M., 2009. Simulation of Medical Laboratory Operations to Achieve Optimal Resource Allocation. In *ECMS* (pp. 249-255).
- Luangmul, K., Pichitlamken, J. and Weerawat, W., 2012. A simulation model of a hospital's clinical laboratory. *Lect Notes Manag Sci*, 4, pp.64-9.
- Marinagi, C.C., Spyropoulos, C.D., Papatheodorou, C. and Kokkotos, S., 2000. Continual planning and

scheduling for managing patient tests in hospital laboratories. *Artificial Intelligence in Medicine*, 20(2), pp.139-154.

Merode, G.G., Oosten, M., Vrieze, O.J., Derks, J. and Hasman, A., 1998. Optimisation of the structure of the clinical laboratory. *European journal of operational research*, 105(2), pp.308-316.

O'Moore, R.R., De Moor, G., Boran, G., Gaffney, P., Grimson, J., McNair, P., Groth, T., Nykänen, P., Hasman, A., Eller, J. and Yearworth, M., 1994. OpenLabs: the application of advanced informatics and telematics for optimization of clinical laboratory services. *Computer methods and programs in biomedicine*, 45(1-2), pp.137-140.

O'Moore, R.R., Groth, T., Grimson, W. and Boran, G., 1996. The OpenLabs Project. *Computer methods and programs in biomedicine*, 50, pp.85-86.

Onar, S.C., Oztaysi, B. and Kahraman, C., 2018. A Comprehensive Survey on Healthcare Management. In *Operations Research Applications in Health Care Management* (pp. 23-51). Springer, Cham.

Van Merode, G.G., Hasman, A., Derks, J., Schoenmaker, B. and Goldschmidt, H.M.J., 1996. Advanced management facilities for clinical laboratories. *Computer methods and programs in biomedicine*, 50(2), pp.195-205.

Vogt, W., Braun, S.L., Hanssmann, F., Liebl, F., Berchtold, G., Blaschke, H., Eckert, M., Hoffmann, G.E. and Klose, S., 1994. Realistic modeling of clinical laboratory operation by computer simulation. *Clinical chemistry*, 40(6), pp.922-928.

PLANIFICATION D'UN COMPLEXE MINIER, DE LA MINE A L'USINE, SOUS CONTRAINTES DE TENEUR EN ELEMENTS DU MINERAI

G. BATAIA, R. GUILLAUME, C. THIERRY

Université de Toulouse (Université Jean-Jaurès) 5 Allées A. Machado 31058 TOULOUSE, IRIT
guillaume.battaia@univ-tlse2.fr, romain.guillaume@irit.fr, thierry@univ-tlse2.fr

RESUME : L'optimisation globale de la planification à moyen terme d'un complexe minier est un problème émergent dans le domaine de l'industrie minière. La prise en compte simultanée des contraintes d'extractions, de transport et de production implique la prise en compte de la teneur en éléments (Fer, Magnesium...) dans le processus de planification. De cette prise en compte, résulte un problème de planification avec contraintes non quadratiques et non linéaires dues à la prise en compte de la teneur en éléments. De plus, le problème a une dimension multicritères : optimisation de la production du métal et des stocks au sein du complexe minier. L'objectif de cet article est de proposer un modèle linéaire et d'étudier le comportement de ce modèle sur un cas industriel en faisant varier le poids des différents objectifs d'une part et le pas de discrétisation sur la teneur en éléments utilisé pour linéariser le modèle initial d'autre part.

MOTS-CLES : *planification minière, chaîne logistique, PLNE.*

1 INTRODUCTION

Dans le domaine de l'exploitation minière, un complexe minier peut être considéré comme une chaîne logistique qui intègre des processus d'extraction du matériel minier, de stockage, de chargement, de transport et de transformation (calcination). Dans le cadre du processus de planification au sein d'une exploitation minière, plusieurs types de décisions peuvent être identifiés.

Au niveau des mines, le premier type de décision est celui du choix de la méthode d'extraction (mine sous-terrainne, à ciel ouvert, maritimes...) et de l'estimation des capacités de production. Dans le cadre de ce travail nous considérerons le cas de mines à ciel ouvert.

Parmi les décisions de planification d'une mine à ciel ouvert, on peut citer (1) la définition de la fosse ultime qui consiste à déterminer jusqu'à quelle profondeur il est souhaitable d'extraire les blocs, (2) la détermination de la séquence d'extraction d'ensemble de blocs ou de blocs, (3) la planification dans le temps de l'extraction des blocs en prenant en compte les contraintes de précédence d'extraction des différents blocs mais également les contraintes de capacités (exprimées en nombre de tonnes par périodes). En aval des mines, il s'agit ensuite de planifier l'acheminement des blocs vers les usines en vue de la production du minerai en usine qui doit prendre en compte des contraintes de fabrication (four) notamment en terme de proportion des différents éléments.

Dans le cadre de ce papier, nous considérons un complexe minier dans lequel les séquences d'extraction de chaque mine sont connues et nous nous intéressons à la planification des transports, stocks et production à la sortie de la mine. Plus précisément nous avons pour objectif de proposer un plan de stockage, chargement et transport des blocs qui respecte les contraintes de conservation de flux,

les contraintes de production de l'usine et les contraintes de teneur en éléments.

2 ETAT DE L'ART

Dans le domaine de la planification des mines et notamment des mines à ciel ouvert, un des modèles de référence dans l'état de l'art est le modèle Open Pit Mine Block Scheduling Problem (OPMBSP) [Johnson 1968]. Ce modèle a pour objectif de déterminer la séquence optimale d'extraction des blocs en prenant en compte les contraintes de précédence entre les blocs, les contraintes de sécurité, les contraintes sur les teneurs en éléments dans le mélange, les contraintes de capacité, les contraintes des processus aval. Ce modèle trop complexe pour les moyens calculatoires disponible a été simplifié puis largement étudié sous le nom de « Ultimate Pit Limit Problem (UPIT) » ou « Final Open Pit Problem » (FOP) (Limite ou contour ultime de la fosse) avec seulement les contraintes de précédence entre les blocs et avec un objectif d'optimisation de la valeur Nette de la mine ou « Net Present Value » (NPV) qui s'appuie sur la détermination d'une teneur de coupure ou « cut of grade ». Le problème dit « Open Pit Pusback and Gap problem » consiste quant à lui à définir différentes zones de la mine, les « pusbacks » [Newman et al 2010].

Le problème de séquençement des blocs ou « Block Sequencing Problem » (BSP) ou « Constrained Open Pit Limit » (CPIT) introduit quant à lui la notion de séquence d'extraction des blocs en prenant en compte les contraintes d'extraction [Liu and Kozan 2016] et dans certain cas certaines contraintes aval.

Le problème de l'optimisation aval ou « downstream optimization » s'intéresse à la planification du transport, stockage et production d'un complexe minier avec une séquence d'extraction donnée. Ce problème qui a été peu

étudié dans la littérature [Goodfellow 2014] est proche de celui que nous adressons ici. Toutefois à notre connaissance ces travaux ne prennent pas en compte les contraintes de production sur les éléments.

Le problème de planification d'un complexe minier ou « Simultaneous Optimization of a Mining Complex » (SOMC) intègre la planification de l'extraction et de la partie aval au sein d'un complexe minier [Goodfellow and Dimitrakopoulos 2017].

Ces grandes classes de problème [Newman et al 2010] [Lamghari 2017] ont été traitées avec des approches déterministes ou stochastiques lorsque les incertitudes sont prises en compte. Les approches déterministes s'appuient sur des formulations de type programmation Linéaire en Nombre entier. Pour prendre en compte les incertitudes géologiques et économiques des approches stochastiques de type programmation stochastique à deux étapes ont été utilisées. Les travaux intégrant la prise en compte des teneurs en éléments dans le processus de planification sont peu nombreux à notre connaissance même si des contraintes de bornes sur les éléments sont prises en compte pour le problème de séquençement des blocs et pour le problème d'orientation des blocs vers des stockpiles en fonction des contraintes sur les éléments [Goodfellow 2016], [Goodfellow and Dimitrakopoulos 2015] [Montiel et Dimitrakopoulos 2015] [Nogholi 2015], [Silva et al 2015].

3 PROBLEME ET MODELE

Dans le cadre de cet article nous nous intéressons au problème de planification d'un complexe minier. Plus précisément nous nous intéressons au processus de planification des activités de chargement du minerai dans des bateaux, de transport du minerai vers l'usine et de calcination effectué à l'usine. Ce problème de planification peut

être représenté par un graphe dont les nœuds représentent les quantités en stocks et les arcs les quantités extraites, chargées, transportées et traitées à l'usine.

Le modèle de planification résultant intégrant les teneurs éléments est le suivant :

Notations

\mathcal{M} : Ensemble de mines, $m \in \mathcal{M}$

\mathcal{T} : Ensemble de périodes, $t \in [0 ; T]$

\mathcal{E} : Ensemble des éléments e considérés $e \in \{E0, E1, \dots\}$

$Ex_{m,t}$: Quantité de minerai extraite d'une mine m pendant la période t . Donnée d'entrée du problème.

$pEx_{m,t,e}$: Teneur en élément e extraite d'une mine m pendant la période t . Donnée d'entrée du problème.

$Im_{m,t}$: Quantité de minerai stockée provenant d'une mine m au début de la période t .

$pIm_{m,t,e}$: Teneur en élément e du minerai stocké sur le site m au début de la période t .

$Ld_{m,t}$: Quantité de minerai en chargement au bord de mer d'une mine m pendant la période t .

$pLd_{m,t,e}$: Teneur en élément e du minerai chargé au bord de mer d'une mine m pendant la période t .

Tr_t : Quantité de minerai transportée du bord de mer vers l'usine pendant la période t .

$pTr_{t,e}$: Teneur en élément e du minerai transporté du bord de mer vers l'usine pendant la période t .

Ip_t : Quantité de minerai stockée à l'usine au début de la période t .

$pIp_{t,e}$: Teneur en élément e du minerai stocké à l'usine au début de la période t .

Cal_t : Quantité de minerai en calcination à la période t

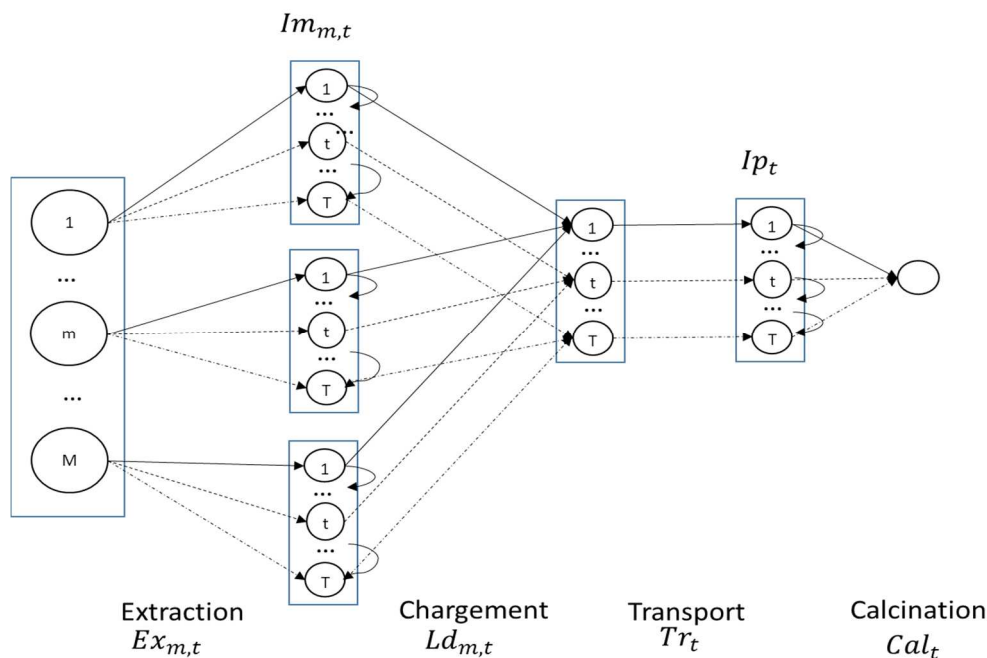


Figure 1 : Représentation graphique du problème de planification

$pCal_{t,e}$: Teneur en élément e du minerai calciné à la période t

cs_m : Coefficient correspondant au stockage à la mine m

csp : Coefficient correspondant au stockage à l'usine

α_{E0} : Coefficient sur la quantité d'élément E0 produite

a_e^s : Teneur seuil en élément e pour la calcination

$$\text{Max } \sum_t \alpha_{E0} \times pCal_{t,E0} \times Cal_t - \sum_{t,m} cs_m \times Im_{m,t} - \sum_t csp \times Ip_t$$

$$Ex_{m,t} + Im_{m,t-1} = Im_{m,t} + Ld_{m,t} \quad \forall m \in \mathcal{M}, \forall t \in [1; T] \quad (1)$$

$$\sum_{m=1}^M Ld_{m,t} = Tr_t \quad \forall m \in \mathcal{M}, \forall t \in [1; T] \quad (2)$$

$$Tr_t + Ip_{t-1} = Ip_t + Cal_t \quad \forall t \in [1; T] \quad (3)$$

$$pEx_{m,t,e} * Ex_{m,t} + pIm_{m,t-1,e} * Im_{m,t-1} = pIm_{m,t,e} * Im_{m,t} + pLd_{m,t,e} * Ld_{m,t} \quad \forall m \in \mathcal{M}, \forall e \in \mathcal{E}, \forall t \in [1; T] \quad (4'1)$$

$$pIm_{m,t,e} = pLd_{m,t,e} \quad \forall m \in \mathcal{M}, \forall e \in \mathcal{E}, \forall t \in [1; T] \quad (4'2)$$

$$\sum_{m=1}^M pLd_{m,t,e} * Ld_{m,t} = pTr_{t,e} * Tr_t \quad \forall m \in M, \forall t \in T \quad (5)$$

$$pTr_{t,e} * Tr_t + pIp_{t-1,e} * Ip_{t-1} = pIp_{t,e} * Ip_t + pCal_{t,e} * Cal_t \quad \forall e \in \mathcal{E}, \forall t \in \mathcal{T} \quad (6'1)$$

$$pIp_{t,e} = pCal_{t,e} \quad \forall e \in \mathcal{E}, \forall t \in \mathcal{T} \quad (6'2)$$

$$Cal_t^{plan.} - \varepsilon \leq Cal_t \leq Cal_t^{plan.} + \varepsilon \quad \forall t \in \mathcal{T} \quad (7)$$

$$pCal_{t,e} \leq a_e^s, \quad \forall t \in \mathcal{T} \quad \forall e \in \mathcal{E}^+ \quad (8'1)$$

$$pCal_{t,e} \geq a_e^s, \quad \forall t \in \mathcal{T} \quad \forall e \in \mathcal{E}^- \quad (8'2)$$

$$\text{où } \mathcal{E}^- \cap \mathcal{E}^+ = \emptyset \text{ et } \mathcal{E}^- \cup \mathcal{E}^+ = \mathcal{E} \setminus \{E0\}$$

L'objectif est de faire un compromis entre les quantités calcinées et les quantités stockées à la mine et à l'usine. Les contraintes (1), (2) et (3) sont les contraintes de conservation pour le minerai. Les contraintes (4), (5) et (6) sont les contraintes de conservation avec élément à chaque phase du processus. Les contraintes (7) et (8) sont des contraintes sur la production. En effet, pour la contrainte (7), en raison de contraintes techniques spécifiques aux équipements utilisés pour la calcination, la quantité calcinée est établie à un niveau constant sur la durée de l'horizon considéré (à ε près). Ce niveau est estimé en tenant compte des quantités prévisionnelles de minerai extrait. Pour la contrainte (8), on est amené à considérer deux sous-ensembles d'éléments, \mathcal{E}^+ et \mathcal{E}^- , pour lesquels l'étape de calcination ne peut être réalisée si leur teneur est supérieure, respectivement inférieure, à un certain seuil noté a_e^s .

Les contraintes de ce modèle ne sont ni quadratiques, ni linéaires. Dans le but de linéariser ces contraintes, nous supposons qu'il existe plusieurs flux selon la teneur en éléments du flux. Par ailleurs nous supposons qu'il existe plusieurs arcs pour passer d'un sommet à l'autre : chaque arc étant défini avec un intervalle de pourcentage d'éléments, un flux d'un élément e peut passer par un arc $A_{i,e}$ si et seulement si la teneur de l'élément e contenu dans ce flux est compris entre la borne inférieure $a_{i,e}^m$ et supérieure $a_{i,e}^M$ de l'arc. On considère que pour un élément e la différence $\Delta_e = a_{i,e}^M - a_{i,e}^m$ est constante. Δ_e constitue le pas de

discrétisation du modèle linéaire. On introduit alors l'ensemble \mathcal{J} défini de la façon suivante. (En toute rigueur cet ensemble doit être indicé par e car il varie d'un élément à l'autre. Toutefois, on omettra cet indice afin de ne pas surcharger les notations.) :

\mathcal{J} : Ensemble des arcs reliant deux nœuds du processus pour un élément e donné.

Cet ensemble contient d'autant plus d'élément que le pas de discrétisation est petit. Pour un pas de discrétisation donné, il est, en revanche, inchangé à tous les niveaux du processus global (chargement, transport, calcination, etc.). Par la suite, on est amené à considérer les variables suivantes :

$qIm_{m,t,e}$: Quantité de minerai stocké sur le site m au début de la période t dont la teneur en élément e est comprise entre $a_{i,e}^m$ et $a_{i,e}^M$.

$qLd_{m,t,e}$: Quantité de minerai chargé au bord de mer d'une mine m pendant la période t dont la teneur en élément e est comprise entre $a_{i,e}^m$ et $a_{i,e}^M$.

$qTr_{t,i,e}$: Quantité de minerai transporté du bord de mer vers l'usine pendant la période t dont la teneur en élément e est comprise entre $a_{i,e}^m$ et $a_{i,e}^M$.

$qIp_{t,i,e}$: Quantité de minerai stocké à l'usine au début de la période t dont la teneur en élément e est comprise entre $a_{i,e}^m$ et $a_{i,e}^M$.

$qCal_{t,i,e}$: Quantité de minerai calciné à la période t dont la teneur en élément e est comprise entre $a_{i,e}^m$ et $a_{i,e}^M$.

Et, avec la notation suivante :

$$xB_{i,e} = \begin{cases} 1 & \text{si } B \text{ passe par l'arc } A_{i,e} \\ 0 & \text{sinon} \end{cases} \quad \forall B \in \{Im_{m,t}, Ip_t, Tr_t\}$$

En fonction des exigences (contraintes ou élément de la fonction-objectif) exprimées sur les divers éléments, on peut être amené à imposer que la solution obtenue garantisse une teneur au moins égale à un certain seuil pour certains éléments ($e \in \mathcal{E}^-$) et au plus égale à un certain seuil pour d'autres ($e \in \mathcal{E}^+$). Ainsi, selon l'élément considéré, on sera amené à écrire que la teneur à la sortie d'un nœud, et donc l'arc emprunté, se déduit des arcs entrants en faisant l'hypothèse la plus pessimiste quant à leur teneur effective. On prendra donc la borne supérieure de l'arc entrant lorsqu'en bout de chaîne on souhaite garantir une teneur inférieure à un certain seuil, et la borne inférieure dans le cas contraire. Cela se justifie par le caractère monotone de l'expression de la teneur de sortie en fonction des quantités et des teneurs d'entrée. On choisira, de même, la borne inférieure de l'arc entrant lorsque l'on souhaite maximiser la teneur finale d'un élément en bout de graphe comme c'est le cas pour E0. Cela permet ainsi de garantir une teneur en élément E0 au moins égale à la valeur obtenue par la résolution numérique du problème.

Pour cela, on introduit la notation suivante $a_{i,e}^{\{m,M\}}$ qui vaut $a_{i,e}^M$ si en bout de processus on doit garantir que la teneur en élément e est inférieure à un certain seuil ou si l'on cherche à la minimiser et qui vaut $a_{i,e}^m$ dans le cas contraire. M est un nombre très grand.

$$\text{Max } \sum_{t,i} \alpha_{E0} \times a_{i,E0}^m \times q\text{Cal}_{t,i,E0} - \sum_{t,m} cs_m \times Im_{m,t} - \sum_t csp \times Ip_t$$

$$Ex_{m,t} + Im_{m,t-1} = Im_{m,t} + Ld_{m,t} \quad \forall m \in \mathcal{M}, \forall t \in [1; T] \quad (1)$$

$$\sum_{m=1}^M Ld_{m,t} = Tr_t \quad \forall m \in \mathcal{M}, \forall t \in [1; T] \quad (2)$$

$$Tr_t + Ip_{t-1} = Ip_t + Cal_t \quad \forall t \in [1; T] \quad (3)$$

$$-M(1 - xIm_{m,t,i,e}) + a_{i,e}^m \times (Im_{m,t-1} + Ex_{m,t}) \leq \sum_j a_{j,e}^{\{m,M\}} \times qIm_{m,t-1,j,e} + pEx_{m,t,e} \times Ex_{m,t} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, i \in I, e \in E \quad (4'1)$$

$$M(1 - xIm_{m,t,i,e}) + a_{i,e}^M \times (Im_{m,t-1} + Ex_{m,t}) \geq \sum_j a_{j,e}^{\{m,M\}} \times qIm_{m,t-1,j,e} + pEx_{m,t,e} \times Ex_{m,t} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, i \in J, e \in E \quad (4'2)$$

$$\sum_i qIm_{m,t,i,e} = Im_{m,t} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, e \in E \quad (4'3)$$

$$\sum_i qLd_{m,t,i,e} = Ld_{m,t} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, e \in E \quad (4'4)$$

$$\sum_i xIm_{m,t,i,e} = 1 \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, e \in E \quad (4'5)$$

$$qIm_{m,t,i,e} \leq M \times xIm_{m,t,i,e} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, i \in J, e \in E \quad (4'6)$$

$$qLd_{m,t,i,e} \leq M \times xLd_{m,t,i,e} \quad \forall m \in \mathcal{M}, t \in \mathcal{T}, i \in J, e \in E \quad (4'7)$$

$$-M(1 - xTr_{t,i,e}) + a_{i,e}^m \times \sum_m Ld_{m,t} \leq \sum_{m,j} a_{j,e}^{\{m,M\}} \times qLd_{m,t,j,e} \quad \forall t \in \mathcal{T}, \forall i \in J, \forall e \in E. \quad (5'1)$$

$$M(1 - xTr_{t,i,e}) + a_{i,e}^M \times \sum_m Ld_{m,t} \geq \sum_{m,j} a_{j,e}^{\{m,M\}} \times qLd_{m,t,j,e} \quad \forall t \in \mathcal{T}, \forall i \in J, \forall e \in E. \quad (5'2)$$

$$\sum_i qTr_{t,i,e} = Tr_t \quad \forall t \in T, e \in E \quad (5'3)$$

$$\sum_i xTr_{t,i,e} = 1 \quad \forall t \in T, e \in E \quad (5'4)$$

$$qTr_{t,i,e} \leq M \times xTr_{t,i,e} \quad \forall t \in T, i \in J, e \in E \quad (5'5)$$

$$-M(1 - xIp_{t,i,e}) + a_{i,e}^m \times (Ip_{t-1} + Tr_t) \leq \sum_j a_{j,e}^{\{m,M\}} \times (qIp_{t-1,j,e} + qTr_{t,j,e}) \quad \forall t \in \mathcal{T}, \forall i \in J, \forall e \in E. \quad (6'1)$$

$$M(1 - xIp_{t,i,e}) + a_{i,e}^M \times (Ip_{t-1} + Tr_t) \geq \sum_j a_{j,e}^{\{m,M\}} \times (qIp_{t-1,j,e} + qTr_{t,j,e}), \quad \forall t \in \mathcal{T}, \forall i \in J, \forall e \in E. \quad (6'2)$$

$$\sum_i qIp_{t,i,e} = Ip_t \quad \forall t \in \mathcal{T} \forall e \in E \quad (6'3)$$

$$\sum_i qCal_{t,i,e} = Cal_t, \quad \forall t \in \mathcal{T} \forall e \in E \quad (6'4)$$

$$\sum_i xIp_{t,i,e} = 1 \quad \forall t \in \mathcal{T} \forall e \in E \quad (6'5)$$

$$qIp_{t,i,e} \leq M \times xIp_{t,i,e} \quad \forall t \in \mathcal{T}, i \in J, e \in E \quad (6'6)$$

$$qCal_{t,i,e} \leq M \times xIp_{t,i,e}, \forall t \in \mathcal{T}, i \in J, e \in E \quad (6'7)$$

$$Cal_t^{plan.} - \varepsilon \leq Cal_t \leq Cal_t^{plan.} + \varepsilon \quad \forall t \in \mathcal{T} \quad (7)$$

$$qCal_{t,i,e} = 0, \quad \forall t \in \mathcal{T}, \forall i / a_{i,e}^M > a_e^S, \forall e \in \mathcal{E}^+ \quad (8'1)$$

$$qCal_{t,i,e} = 0, \quad \forall t \in \mathcal{T}, \forall i / a_{i,e}^m < a_e^S, \forall e \in \mathcal{E}^- \quad (8'2)$$

L'objectif est inchangé : compromis entre les quantités calcinées et les quantités stockées à la mine et à l'usine.

Les contraintes (1), (2) et (3) restent inchangées : contraintes de conservation pour le minerai. Les contraintes (4), (5) et (6) sont les contraintes de conservation avec élément qui ont été linéarisées. La contrainte (7) reste inchangée contrainte sur la production. Les contraintes (8) sont la linéarisation des contraintes sur les teneurs seuils pour l'étape de calcination.

4 ETUDE DE CAS

On considère un ensemble de 5 mines {A, B, C, D, E}, sur lesquelles on relève les teneurs de deux éléments E0 et E1.

E0 constitue la finalité du processus d'extraction. L'objectif est donc de maximiser sa teneur à l'issue de l'étape finale de calcination sur la durée de l'horizon de planification. En complément de cet objectif primaire qui conditionne la rentabilité du processus, les différentes opérations du processus global induisent potentiellement des stocks qu'il est souhaitable de minimiser. Nous tiendrons compte ici, comme cela a été présenté dans le modèle décrit plus haut, du stockage sur chacun des sites en faisant l'hypothèse, classiquement admise, qu'ils sont proportionnels à la quantité stockée lors de chaque période ($Im_{m,t}$ et Ip_t) et à un coefficient spécifique à chaque site nommé coefficient de stockage que l'on peut assimiler, par abus de langage, à un coût de stockage. Ce coefficient est plus élevé, d'environ un ordre de grandeur, sur le site dédié à la calcination que sur les sites d'extraction.

E1 est contrôlé car sa teneur fait l'objet d'une contrainte lors de l'étape de calcination. Celle-ci doit être maintenue au-dessous d'un seuil a_{E1}^S . On est donc dans le cas où $\mathcal{E}^+ = \{E1\}$ et $\mathcal{E}^- = \emptyset$.

Les tableaux 1 et 2 synthétisent les données d'entrée du modèle. Le tableau 1 présente, pour chaque site (extraction et calcination), les niveaux de stocks initiaux, les capacités maximales de stockage et les coûts de stockage par unité de masse. Il est intéressant de remarquer que sur la plupart des sites d'extraction, les stocks initiaux sont égaux aux stocks maximaux (à l'exception du site B).

Site	Stock initial	Stock max.	coût de stockage
A	48	48	5
B	30	38	8
C	44	44	6
D	28	28	7
E	38	38	10
P	15	200	50

Tableau 1 : Stocks initiaux, maximaux et coûts de stockage en chaque site d'extraction (A, ..., B) et sur le site de calcination (P).

En revanche, le site de calcination P dispose d'une large capacité de stockage excédentaire au temps initial. Toutefois, il peut être noté sur ce même tableau, comme cela a

été mentionné plus haut, que le coût de stockage sur ce site est conséquemment plus élevé que sur les sites d'extraction.

Le tableau 2 présente, pour chaque site d'extraction, l'intervalle des valeurs d'entrée du modèle relatives à la quantité de minerai extrait et les teneurs en éléments E0 et E1. Il peut être observé une certaine corrélation positive entre ces teneurs. Ainsi le problème d'optimisation vise à concilier deux exigences contradictoires à savoir maximiser la teneur en E0 lors du processus de calcination tout en maintenant celle de E1 en dessous d'un certain seuil. D'un point de vue industriel, ce seuil est de 14 %.

Mine	quantité (tonnes)	teneur E0 (%)	teneur E1 (%)
A	[11 ; 49]	[2,4 ; 2,5]	[11,8 ; 14,2]
B	[27 ; 70]	[2,7 ; 3,0]	[16,5 ; 19,4]
C	[23 ; 56]	[2,1 ; 2,3]	[11,1 ; 12,5]
D	[21 ; 52]	[2,5 ; 2,7]	[13,6 ; 14,9]
E	[27 ; 52]	[2,3 ; 2,5]	[8,2 ; 14,2]

Tableau 2 : Intervalles des quantités extraites et des teneurs sur chaque site minier.

Dans ce qui suit, nous présentons deux jeux de simulation réalisés à différents niveaux de précision. Pour le premier, les teneurs en E0 sont discrétisés en intervalle de 0,1 % d'amplitude et les teneurs en E1 en intervalle de 1 % d'amplitude. Pour le second, les amplitudes des intervalles sont ramenées respectivement à 0,05 % et 0,5 % pour E0 et E1.

Dans ce cadre, pour le premier jeu, le modèle ne permet pas d'obtenir de solution avec une teneur maximale en E1 fixée à 14 %. Nous avons donc effectué ces simulations avec une valeur de 15 %. Pour le second jeu, nous avons pu fixer la teneur maximale à 14 %.

Ces simulations ont été réalisées à l'aide du solveur Gurobi.

4.1 Simulations exploratoires

Nous présentons deux simulations exploratoires ayant pour objectif d'étudier l'influence du paramètre α_{E0} . Ce paramètre représente l'importance relative que l'on accorde à la production de E0 en comparaison des coûts de stockage induits sur chaque site. Une valeur importante signifiera que l'on met la priorité sur la production de E0, alors qu'une valeur faible signifiera que l'on met la priorité sur les coûts stockage. Deux valeurs, $\alpha_{E0} = 10000$ et $\alpha_{E0} = 1$ ont été testées pour rendre compte de ces deux cas extrêmes.

Le fait de considérer des intervalles de teneur assez important (0.1 % pour E0 et 1 % pour E1) permet de limiter le nombre d'arcs entre deux nœuds et donc le nombre de variables du problème et ainsi réduire les temps de calcul.

En revanche, la perte de précision induite peut rendre le problème infaisable. C'est le cas lorsque l'on fixe une teneur maximale de calcination α_{E1}^s à 14 %. Ainsi, ce seuil a été relevé à la valeur de 15 % pour ces simulations.

Pour ce jeu de simulation, le problème comporte 3588 variables continues, 1848 variables binaires et 7752 contraintes. Les opérations de prétraitement permettent de réduire la taille du problème pour n'avoir plus que 872 variables continues, 370 variables binaires et 1736 contraintes.

Les résultats sont synthétisés dans le tableau 3. La figure 2 compare pour chacun des scénarios envisagés l'évolution temporelle du stock présent sur le site de calcination.

α_{E0}	Coût des stocks mines	Coût des stocks usine	Coût total des stocks	E0 produit (tonnes)
1	16.3	49.0	65.3	55.4
10000	12.1	78.6	90.8	56.7

Tableau 3 : Synthèse des résultats des simulations exploratoires. Coûts en millier d'unités monétaires.

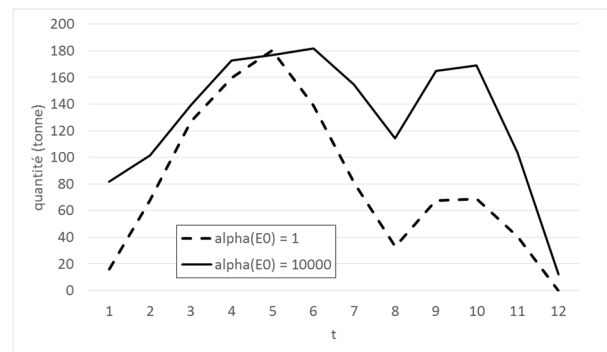


Figure 2 : Evolution du stock usine (IP_t) pour les simulations exploratoires.

Pour le premier scénario ($\alpha_{E0} = 1$), l'objectif premier étant de donner la priorité aux coûts de stockage. La solution obtenue vise donc à réduire au maximum le stock réalisé sur le site de calcination puisque c'est celui pour lequel le coût est, de loin, le plus élevé. Il n'est toutefois pas possible de s'affranchir totalement de ce stock pour deux raisons.

Premièrement, si la quantité de minerai calcinée à chaque période est constante alors que les quantités extraites observent une fluctuation saisonnière. Ainsi le fait de fixer la quantité de minerai calciné à partir des quantités extraites sur la durée de l'horizon implique que le niveau de stock total (mines + usine) revienne au niveau initial à la fin de l'horizon. Cependant les fluctuations saisonnières de l'extraction induisent des variations temporaires du stock global. Comme, par ailleurs, les stocks sur les mines sont quasiment tous saturés au temps initial, un éventuel excédent de production sur les mines induit nécessairement un besoin de stockage à l'usine.

Deuxièmement, le stock usine s'avère crucial dans la mesure où c'est le lieu où des minerais possédant des teneurs en E1 plus ou moins élevées peuvent être combinés afin de satisfaire la contrainte de teneur maximale. Il s'avère notamment, comme cela peut être remarqué à partir des données du tableau 2, que le minerai produit sur certains sites d'extraction ne peut être calciné tel quel sans avoir été préalablement combiné avec des minerais à plus faible teneur. Ainsi, il peut s'avérer nécessaire de constituer un stock usine afin de maintenir le niveau de calcination imposé et s'affranchir des fluctuations susceptibles de dépasser le seuil de 15 %.

Pour le second scénario ($\alpha_{E0} = 10000$), comme cela était anticipé, la forte valeur de α_{E0} tend à maximiser la quantité de E0 produit au détriment du coût généré par le stockage. La quantité de E0 produit est supérieure d'environ 2 % dans ce scénario par rapport au premier scénario dans lequel on ne tient compte que marginalement de la quantité de E0 produit. On peut constater que pour arriver à cela, il a été nécessaire de déporter une partie des stocks des mines vers l'usine. La façon temporelle dont s'effectue ce report peut-être vue sur la figure 2. Il apparaît sur ces courbes que la quantité maximale devant être stockée sur ce site, valeur qui correspond au pic observé au mois 5 pour le scénario 1, n'est que peu ou pas modifiée. En revanche, une plus grande quantité de minerai est transporté vers l'usine avant et surtout après ce pic. Cela permet d'incorporer une plus grande partie de minerai à forte teneur en E0 dans le processus de calcination. Celui-ci doit néanmoins être combiné à d'autres minerais, ce qui peut expliquer l'accroissement du stock usine, afin de satisfaire la contrainte de seuil sur E1.

Il peut apparaître discutable dans un tel processus de chercher, comme cela a été fait dans le premier scénario, à privilégier l'optimisation du niveau des stocks sur la production de E0. Toutefois, il est important de voir que la production de E0 est majoritairement dictée par le fait d'imposer à la fois le flux entrant du processus (extraction minière) et le flux sortant (calcination). Ainsi, la totalité du E0 extrait de la mine sera calcinée à une période plus ou moins proche. L'optimisation de la production de E0 relève, en réalité, d'un autre niveau de décision qui prend en compte un horizon plus large. Une fois décidés les flux entrant et sortant, il peut sembler légitime de focaliser les efforts sur l'optimisation du processus et notamment la gestion des stocks. Prendre à nouveau en compte l'optimisation de E0, comme cela a été fait dans le scénario 2, relève donc d'une stratégie de court terme qui, si elle a sa raison d'être, n'en demeure pas la seule stratégie envisageable.

Il est à rappeler que ces simulations ne sont valides qu'en considérant un seuil de 15 % pour E1 ce qui n'est pas acceptable industriellement. Notre levier pour assurer l'obtention d'une solution avec un seuil de 14 % est de diminuer le pas de discrétisation sur les teneurs élémentaires du modèle.

4.2 Simulation fine

Cette simulation fine a pour objectif de mesurer l'influence du pas de discrétisation sur les teneurs élémentaires pour ramener le seuil sur E1 à 14 %. Pour cela, le pas de discrétisation sur les teneurs élémentaires a été divisé par deux pour chacun des éléments par rapport aux simulations exploratoires. Par ailleurs, nous nous sommes placés dans le cas du scénario 1 c'est-à-dire avec un α_{E0} petit.

Le problème comporte, à présent, 6708 variables continues, 3528 variables binaires et 14316 contraintes. Les opérations de prétraitement permettent de réduire la taille du problème pour n'avoir plus que 1354 variables continues, 675 variables binaires et 3129 contraintes.

Les résultats de cette simulation sont synthétisés dans le tableau 4. L'évolution du stock usine est présentée sur la figure 3. En complément, la figure 4 présente l'évolution temporelle de la teneur en élément E1 du minerai chargé depuis chaque mine.

Coût des stocks mines	Coût des stocks usine	Coût total des stocks	E0 produit (tonnes)
13	69.9	83	56.4

Tableau 4 : Synthèse des résultats de la simulation fine. Coûts en millier d'unités monétaires.

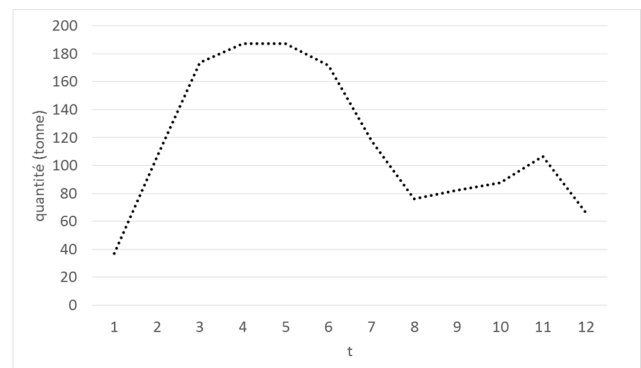


Figure 3 : Evolution du stock usine (IP_t) pour la simulation fine

Il apparaît dans ce cas que le stock usine est globalement plus approvisionné que dans le premier scénario du jeu de simulations exploratoires pour lequel la valeur de α_{E0} est identique. La raison en est évidemment l'abaissement du seuil α_{E1}^s qui requiert la combinaison d'une plus grande quantité de minerai sur ce site pour satisfaire à la fois à cette contrainte et au niveau de calcination imposé. Parallèlement, il peut être observé que la quantité de E0 produit est plus élevée par rapport à ce même scénario. Elle n'est inférieure que de 0,5 % par rapport au second scénario exploratoire dont l'objectif principal est la maximisation de E0 et qui bénéficie également d'un seuil plus élevé sur E1.

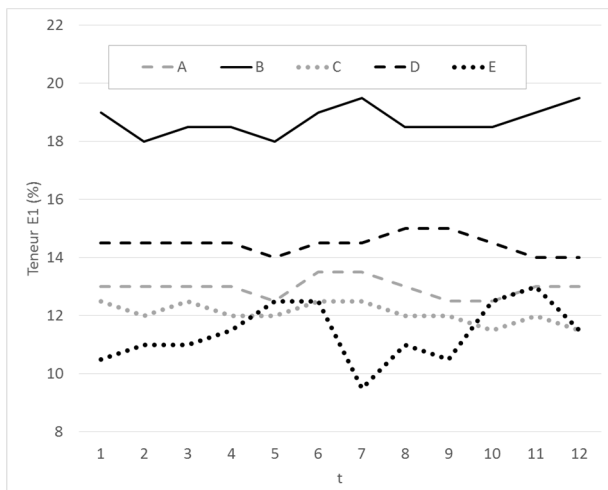


Figure 4 : Teneur maximale en élément E1 des quantités chargées ($Ld_{t,m}$) depuis chaque mine pour le transport.

Ainsi, le modèle conçu constitue un outil d'aide à la décision pour la gestion optimale des stocks permettant de satisfaire les contraintes techniques du processus. Les divers résultats fournis par le modèle permettent un pilotage du processus qui assure que les contraintes de premier ordre, comme celles exprimées ici, seront satisfaites. A titre illustratif, le graphique de la figure 3 présente les teneurs attendues des minerais chargés à chaque période depuis chaque mine. Ces teneurs ne sont pas aisément prédictibles puisqu'elles résultent, en amont, d'une combinaison entre l'extraction et le stock de la mine. Néanmoins, ces teneurs, complétées par les quantités chargées, permettent d'assurer que l'étape ultérieure de calcination pourra être effectuée dans les conditions optimales. Un écart constaté de ces valeurs peut induire des perturbations sur le processus que l'outil d'aide à la décision sera à même d'anticiper.

5 CONCLUSION

Parmi les problèmes de planification minière, la planification des opérations depuis l'extraction jusqu'au traitement dans un contexte d'extraction multi-site comporte des spécificités qui justifient le développement d'approches nouvelles et adaptées. Dans cet article, nous avons présenté une modélisation non linéaire et non quadratique d'un tel processus vu dans son ensemble et proposé une méthode permettant la linéarisation de ce modèle. Celle-ci présente notamment l'atout de pouvoir prendre en considération autant de contraintes élémentaires que souhaité.

Par la suite, nous avons utilisé cette modélisation pour simuler un cas industriel dans lequel une contrainte élémentaire est prise en compte. Cette étude de cas a été réalisée en considérant tout d'abord un problème simplifié pour

lequel des simulations exploratoires à pas de discrétisation élevé ont pu être conduites. Elles ont permis d'étudier l'influence relative des critères d'optimisation du modèle (stockage versus production) et d'en discuter la pertinence dans le cadre de la planification minière. Enfin, une simulation fine a permis de proposer une solution satisfaisant aux contraintes industrielles. Celle-ci permet d'illustrer comment le modèle développé peut être exploité dans la perspective du développement d'un outil d'aide à la décision destiné à la planification d'un complexe minier.

REFERENCES

- Goodfellow R, 2014. Unified modelling and simultaneous optimization of open pit mining complexes with supply uncertainty. PhD Thesis, McGill University, Montreal, QC
- Goodfellow, R. and R.Dimitrakopoulos, 2016. Global optimization of open pit mining complexes with uncertainty. *Applied Soft Computing*, 40, 292–304.
- Goodfellow R and R. Dimitrakopoulos, 2017. Simultaneous stochastic optimization of mining complexes and mineral value chains. *Mathematical Geosciences* 49 (3), pp 341–360
- Johnson, T.B., 1968. *Optimum open pit mine production scheduling*. Ph.D. Dissertation, University of California, Berkeley, CA.
- Lamghari A. ,2017. Mine planning and oil field development: a survey and research potentials *Mathematical Geosciences* 49 (3), 395-437
- Liu S.Q. and E. Kozan, 2016. New graph-based algorithms to efficiently solve large scale open pit mining optimisation, *Expert Systems with Applications*, Volume 43, P 59-65.
- Montiel, L. and R. Dimitrakopoulos, 2015. Optimizing mining complexes with multiple processing and transportation alternatives: An uncertainty based approach. *European Journal of Operational Research*, 247(1) : 166–178.
- Newman A. M. , E. Rubio, R. Caro, A. Weintraub, and K. Eurek, 2010. A review of operations research in mine planning. *Interfaces*, 40(3) :222–245.
- Nogholi M.A. A, 2015. *Optimisation of open pit mine block sequencing*. PhD thesis, Queensland University of Technology.
- Silva M.F, R Dimitrakopoulos, A Lamghari, 2015. Solving a large SIP model for production scheduling at a gold mine with multiple processing streams and uncertain geology, *Mining Technology* 124 (1), 24-33
- Zhang J. and R. Dimitrakopoulos, 2017. A dynamic-material-value-based decomposition method for optimizing a mineral value chain with uncertainty A dynamic-material-value-based decomposition method for optimizing a mineral value chain with uncertainty, *European Journal of Operational Research* 258 (2017) p. 617–62

PLANIFICATION SIMULTANÉE DE LA PRODUCTION, DE LA REMISE À NEUF ET DU SETUP POUR UN SYSTÈME HYBRIDE NON FIABLE

Morad ASSID, Ali GHARBI

Adnène HAJJI

Laboratoire de Conception et de Contrôle des Systèmes de Production - ÉTS - Université du Québec, 1100, rue Notre-Dame Ouest, Montréal (QC), Canada H3C 1K3
morad.assid.1@ens.etsmtl.ca, ali.gharbi@etsmtl.ca

Département des opérations et des systèmes de décision & CIRRELT - Université de Laval
adnene.hajji@fsa.ulaval.ca

RESUME : *Le secteur d'activités de la remise à neuf (ou Remanufacturing) est intéressant du point de vue économique et environnemental. Pendant plusieurs années, il a été ciblé par un nombre croissant de réglementations gouvernementales. Cette situation oblige plusieurs entreprises à adapter leurs stratégies à cette nouvelle réalité afin de rester compétitives. Cet article s'inscrit dans ce contexte en considérant un système hybride qui utilise à la fois la matière première (mode de production) et les produits retournés (mode de remise à neuf) dans son processus de transformation. Le système est composé d'une unité de transformation non fiable qui nécessite une opération de setup à chaque fois qu'elle bascule entre les deux modes. L'objectif principal est de proposer une stratégie adaptative qui aide les décideurs à gérer simultanément les activités de production, de remise à neuf et de setup. Pour résoudre le problème, une approche de résolution basée sur la simulation et des techniques d'analyse statistique et d'optimisation est envisagée. Une étude comparative impliquant une large gamme de configurations du système est également effectuée afin de démontrer l'avantage économique de la structure de politiques de commande proposée par rapport à celle trouvée dans la littérature.*

MOTS-CLES : *Planification de production, remise à neuf, logistique inverse, setup, simulation, optimisation.*

1 INTRODUCTION

Au cours des deux dernières décennies, la logistique inverse (LI) constitue désormais un nouveau domaine de recherche. Les principales forces motrices sont la croissance économique, les législations et les préoccupations concernant la durabilité environnementale. Pour de nombreuses entreprises, LI fait partie des efforts mondiaux de développement durable et une réponse à une production plus responsable. Les lecteurs peuvent se référer à certaines revues de littérature qui couvrent différents aspects de LI, i.e. Ilgin et Gupta (2010) et Govindan et al. (2015). Au niveau gouvernemental, les actions visent à résoudre les problèmes environnementaux, mais surtout à la mise en œuvre de nouvelles législations afin de stimuler le secteur manufacturier et d'augmenter la productivité des entreprises (Porter, 1990). De nos jours, les législateurs tiennent de plus en plus les fabricants d'équipement d'origine responsables de l'écologisation de leurs produits et de leurs chaînes d'approvisionnement. Parmi les solutions d'une production durable, l'option de remise à neuf (ou Remanufacturing) qui ferme la chaîne de production formant la chaîne logistique en boucle fermée (CLBF), est bien connue comme une forme particulière de LI en mettant l'accent sur la récupération de la valeur ajoutée. Cet article traite le problème de la planification et du contrôle de la production (PCP) au sein d'une CLBF. Il fait référence à la productivité, à la gestion des stocks et à l'utilisation des ressources. Par exemple, il devrait aider le gestionnaire en coordonnant efficacement les activités de production et de remise à neuf.

Dans ce sens, la littérature compte divers modèles traitant cette coordination et la prise de décision dans les systèmes hybride de production et de remise à neuf (SHPR). Notre attention sera concentrée sur les modèles d'optimisation liés à SHPR.

L'un des problèmes des SHPR est de savoir s'il faut effectuer le processus de transformation dans une installation commune ou de consacrer chaque mode (mode de production utilisant la matière première et mode de remise à neuf transformant les produits retournés) dans des installations séparées (Aras et al., 2006). L'utilisation d'une même ligne de transformation pour les activités de production et de remise à neuf est observée dans l'industrie, comme chez Hewlett Packard qui remet à neuf des serveurs haut de gamme sur la même ligne que les nouveaux serveurs (Debo et al., 2006). Teunter et al., (2006) montrent qu'un tel système peut être préférable en termes de coûts dans le cas de faibles taux de retours en analysant les données d'une étude de cas concernant un fabricant de pièces automobiles. Dans le même sens, Francas et Minner (2009) indiquent que la coordination entre les processus de transformation lorsqu'ils sont effectués dans une installation commune pourrait améliorer la flexibilité du système et permettre une plus grande économie. Or, cela a un prix, un niveau significatif d'accroissement de la complexité. Cette coordination reste peu explorée dans la littérature. Parmi les auteurs qui l'ont étudié, Teunter et al. (2006) traitent le problème de détermination de la taille des lots économiques avec des coûts de setup et fournissent un algorithme fiable pour résoudre ce problème. Flapper et al. (2014) présentent une approche

basée sur la théorie de la file d'attente et proposent un calendrier de production qui minimise le coût moyen actualisé à long terme pour un SHPR sans setups. Aucun des travaux mentionnés ci-haut n'a considéré la dynamique et la stochastique du système. Plus récemment, Polotski et al. (2017) proposent une structure générale des politiques de commande optimales combinant simultanément les opérations de setup et les activités de production et de remise à neuf. Les événements de pannes et de réparations de l'installation sont formulés à l'aide de la distribution exponentielle uniquement. Toutes les études citées ci-dessus traitent le problème du PCP au sein d'une CLBF, mais n'intègrent la dynamique et la stochastique des unités de transformation. Ces dernières peuvent faire l'objet de pannes et de réparations aléatoires suivant des distributions de probabilité générales.

Dans la littérature, les systèmes conventionnels avec des politiques de setup consistent en une installation produisant différents types de produits pour répondre aux demandes des clients correspondantes. Ainsi, les opérations de setup visent à préparer le passage de la production d'un type de produit à un autre. Cependant, les conditions de setup entre deux produits peuvent changer, de même que les taux de production. Dans ce contexte, El-hafsi et Bai (1996) développent une politique de commande optimale utilisant la capacité maximale de production tout au long le fonctionnement du système, tandis que Gharbi et al. (2006) proposent d'intégrer des taux de production variables dans le but de minimiser le coût total. La décision de choisir une politique de commande plutôt qu'une autre est discutée dans (Assid et al., 2014). La structure optimale des politiques de commande se révèle être du type seuils critiques (Hedging-Point) dans la littérature sur le contrôle de la production et prennent la forme de la politique (s, Q) dans la littérature sur la gestion des stocks. Notre article traite les systèmes pour lesquels les opérations de setup ne sont pas liées aux différents types de produits à fabriquer, mais plutôt elles sont nécessaires pour changer le mode de transformation (production ou remise à neuf). Le présent travail s'inscrit dans la continuité de celui réalisé par Polotski et al. (2017). Il propose une structure appropriée de politique de commande combinant les activités de production, de remise à neuf et de setup pour un SHPR. Ce travail est motivé par la nécessité de doter les managers de stratégies mieux adaptées au contexte industriel. Il montre comment les opérations d'installation peuvent être ajustées en prenant en compte la meilleure combinaison des processus de transformation avec précision et en temps opportun afin de minimiser le coût total encouru. Ce dernier se compose des coûts de production, de remise à neuf, d'espaces de stockage, d'inventaire des retours ainsi les coûts d'inventaire et de pénurie des produits finis. Le reste du document est structuré comme suit. La description du système et la formulation du problème de planification et de contrôle de la production sont présentées à la section 2. Dans la section 3, l'approche de résolution utilisée pour déterminer les paramètres de commande optimaux est détaillée. Elle comprend les poli-

tiques de commande étudiées, une description des modèles de simulation développés et l'exemple numérique adopté. Une étude comparative approfondie des politiques de commande considérées pour une large gamme de configurations du système est par la suite présentée dans la section 4. La section 5 conclut le document.

2 PROBLEME DE COMMANDE

Cette section définit les notations et les hypothèses utilisées dans cet article, ainsi que l'énoncé du problème

2.1 Notation

Les notations utilisées dans le reste de l'article sont listées ci-dessous. Soit $i = \{P, R\}$ avec "P" définit le mode de production des produits finis (PF) utilisant la matière première et "R" représente le mode de remise à neuf des produits retournés.

$x_F(t)$: Niveau d'inventaire (ou pénurie) des PF au temps t
$x_R(t)$: Niveau d'inventaire des retours au temps t
d	: Taux de demande des PF
$u_{pro}(t)$: Taux de production au temps t
$u_{rem}(t)$: Taux de remise à neuf au temps t
$u_R(t)$: Taux de retours au temps t
U_{pro}^{max}	: Taux maximal de production
U_{rem}^{max}	: Taux maximal de remise à neuf
Z_F	: Capacité d'espace de stockage des PF
Z_R	: Capacité d'espace de stockage des retours
c_F^+	: Coût d'inventaire des PF (\$/temps/unité)
c_F^-	: Coût de pénurie des PF (\$/temps/unité)
c_R^+	: Coût d'inventaire des retours (\$/temps/unité)
c_{pro}	: Coût de production (\$/unité)
c_{rem}	: Coût de remise à neuf (\$/unité)
c_F^e	: Coût d'espace de stockage des produits finis
c_R^e	: Coût d'espace de stockage des retours
c_S^{ij}	: Coût de Setup pour passer du mode i au mode j (\$/opération de setup)
T_S^{ij}	: Temps de setup pour passer du mode i au mode j

2.2 Description du système

Cet article considère un système hybride de production et de remise à neuf (SHPR) composé d'une installation de transformation non fiable qui fabrique un seul type de produits finis. Elle fonctionne en deux modes différents selon la source utilisée: le mode de production transforme la matière première et le mode de remise à neuf est basé sur les produits retournés. Notre système présente des caractéristiques communes rencontrées chez de nombreux fabricants d'équipement d'origine (OEM) qui peuvent remettre à neuf leurs propres marques seulement. Étant donné que les retours sont limités par rapport à la quantité de produits vendus, le processus de remise à neuf est caractérisé par une faible capacité de

transformation par rapport à la demande des clients et est utilisé principalement pour satisfaire les exigences environnementales et optimiser la performance globale du système. Un tel système est connu sous le terme anglo-saxon « Main Manufacturing System (MMS) ». Dans ce contexte, les opérations de setup deviennent inévitables pour passer d'un mode de transformation à un autre afin de répondre à la totalité de la demande clientèle. Deux stocks sont considérés : le premier contient des produits finis prêts à être livrés aux clients, tandis que le second contient les retours qui seront utilisés pour la remise à neuf. La possibilité d'éliminer certains produits retournés est également prise en compte dans notre modèle afin d'éviter des coûts d'inventaires irréalistes. Les principales hypothèses adoptées dans ce travail sont présentées comme suit :

- Les matières premières sont toujours disponibles;
- Le taux de demande des clients est constant et le taux de retour est proportionnel au taux de demande selon le facteur $\gamma_R \in]0,1[$;
- Les coûts d'élimination sont considérés négligeables.
- Les deux modes de transformation produisent le même produit fini qui peut être distribué comme neuf pour répondre à la demande des clients;

2.3 Formulation du problème

L'état du système hybride peut être décrit à l'instant t par deux composantes en temps continu. D'une part, un processus stochastique discret $\{\alpha(t), t \geq 0\}$ qui définit le mode de fonctionnement de l'installation. $\alpha(t)=1$, si l'installation est disponible pour la production ou la remise à neuf à l'instant t et $\alpha(t)=0$, si elle n'est pas opérationnelle (elle est en réparation suite une panne ou / et une opération de setup est exécutée). D'autre part, le niveau d'inventaire cumulé des produits finis et des retours $X(t)=(x_F(t), x_R(t)) \in \mathbb{R}$. Ainsi, la dynamique du système peut être décrite par les variables d'état $(X(t), \alpha(t))$. Son évolution temporelle est représentée par:

$$\begin{aligned} \dot{x}_F(t) &= u_{pro}(t) + u_{rem}(t) - d, x_F(0) = x_F^0 \\ \dot{x}_R(t) &= u_R(t) - u_{rem}(t) - u_{eli}(t), x_R(0) = x_R^0 \end{aligned} \quad (1)$$

x_F^0 et x_R^0 définissent respectivement le niveau du stock initial des produits finis et des retours. Les taux de production et de remise à neuf ainsi que la séquence temporelle des opérations de setup constituent les variables de décision. Les opérations de setup sont modélisées en utilisant le temps τ_k auquel l'opération $k^{ème}$ ($k \in \mathbb{N}$) commence et la paire $i_k j_{(k+1)}$, $i, j \in \{P, R\}$, $i \neq j$ qui représente la transition d'un mode de transformation i à un autre j. La séquence des opérations de setup est donc décrite par $\varphi = \{(\tau_0, i_0 j_1), (\tau_1, i_1 j_2), \dots\}$. La fonction de séquence de ces opérations peut également être définie par l'indicateur S_{ij} (si $i=j$ alors le système est au mode i mais si $i \neq j$ alors il est en transition vers le mode j). Sachant que l'installation ne peut pas être dans les deux

modes de transformation en même temps, l'équation suivante doit être valide: $\sum_{i,j} S_{ij}=1, i,j \in \{P, R\}$.

Soit χ l'ensemble des politiques de commande réalisables qui dépendent du processus stochastique de l'installation de transformation $\alpha(t)$ et de la disponibilité des produits retournés en stock. Il est exprimé par (2).

$$\chi = \left\{ \begin{aligned} & (u_{pro}, u_{rem}), 0 \leq u_{pro} \leq U_{pro}^{max} \cdot S_{MM}, 0 \leq u_{rem} \leq U_{rem}^{max} \cdot S_{RR} \\ & u_{pro} \cdot u_{rem} = 0 \\ & \max(U_{pro}^{max} + U_{rem}^{max} \cdot \Pr(x_R > 0)) \frac{MTTF}{MTTF + MTTR} \geq d \end{aligned} \right\} \quad (2)$$

$\Pr(x_R > 0)$ représente la probabilité que l'inventaire des retours soit strictement positif. MTTF indique le temps moyen de bon fonctionnement, tandis que MTTR définit le temps moyen de réparation. Notre objectif est de proposer une structure efficace de politiques de commande dont les taux de transformation U (incluant u_{pro} et u_{rem}) et la séquence temporelle des opérations de setup (φ) minimisent la fonction de coût actualisé $J(\cdot)$ sur un horizon infini.

$$\begin{aligned} J(X, \alpha, U, \varphi) &= E[c_F^e(Z_F) + c_R^e(Z_R)] \\ &+ E \left[\int_0^\infty e^{-\rho t} g(\cdot) dt + \sum_{k=0}^\infty e^{-\rho \tau_k} c_S^e i_k j_{k+1} \right] \end{aligned} \quad (3)$$

Où ρ est le taux d'actualisation tandis que $g(\cdot)$ est la fonction de coût instantané définie comme suit:

$$\begin{aligned} g(X, U) &= c_F^+ \cdot x_F^+ + c_F^- \cdot x_F^- + c_R^+ \cdot x_R + c_{pro} \cdot u_{pro} \\ &+ c_{rem} \cdot u_{rem} + c_{eli} \cdot u_{eli} \end{aligned} \quad (4)$$

Avec, $x_F^+ = \max(0, x_F)$ si $x_F \geq 0$ et $x_F^- = \max(-x_F, 0)$ si $x_F < 0$. Comme dans Assid et al. (2014), notre modèle prend en compte deux coûts d'espaces de stockage c_F^e et c_R^e en fonction de la capacité de stockage requise pour maintenir à la fois les produits finis et les retours respectivement. Ces coûts représentent la dépréciation de l'équipement considéré pour l'espace de stockage (ils ne représentent pas les actions de conservation des produits). En fait, il s'agit des coûts fixes qui augmentent à un niveau supérieur en fonction des changements substantiels de la capacité de stockage (Z_F et Z_R). Par exemple, si $Z_R \in [150, 350[$, $c_R^e = 280$ et peut être égal à 370 si $Z_R \in [350, 650[$. La fonction de valeur du problème de planification est définie comme suit:

$$v(\cdot) = \inf_{(U, \varphi) \in \chi} J(X, \alpha, U, \varphi) \quad (5)$$

La détermination des propriétés de la fonction de valeur (5) pour une politique intégrée en régime permanent conduit aux conditions d'optimalité décrites par les équations de Hamilton-Jacobi Bellman (HJB). Le problème d'optimisation étudié est complexe. Il est non linéaire, sous contraintes et stochastique. Les phénomènes stochastiques sont dus principalement aux pannes et aux réparations aléatoires qui suivent des distributions de probabilité générales. Cette complexité est également liée à la dynamique des deux inventaires puisqu'ils dépendent de l'état stochastique de l'installation, de la capacité maximale de transformation dans chaque mode et de la disponibilité des retours en stock. Les techniques d'optimisation classique ne peuvent pas être utilisées pour résoudre le modèle étudié, car les expressions analytiques de forme fermée pour les fonctions objectives sont difficiles à obtenir. Plusieurs travaux ont déjà tenté de résoudre les équations HJB dans différents contextes où le système de transformation comprend des opérations de setup. Pour un système hybride proche de notre contexte, Polotski et al. (2017) ont réussi à combiner les analyses analytique et numérique en développant une structure de la politique de commande de production, de remise à neuf et de setup, mais l'occurrence des pannes du système considérée dans leur modèle suivent strictement la distribution exponentielle. Cette hypothèse suppose que les événements de pannes aléatoires évoluent selon un processus de Markov qui conduit à une formulation relativement plus facile du problème mais pourrait être une hypothèse forte pour des études de cas généraux. Les auteurs ont trouvé la structure de la politique de commande sous-optimale, mais ils n'ont pas calculé la valeur des paramètres de commande, y compris la capacité de stockage des retours et des produits finis. Dans ce sens, Berthaut et al. (2010) indiquent qu'une approximation satisfaisante de (5) et des paramètres de commande associés pourrait être trop longue pour être applicable au niveau opérationnel lorsque des méthodes numériques sont appliquées. En effet, la précision résultante dépend des pas de grille discrets utilisés pour le maillage des variables d'état du système ($X(t)$, $\alpha(t)$). Dans ce travail, nous privilégions plutôt une approche alternative basée sur la simulation, les plans d'expériences et des analyses statistiques en raison de sa capacité à : calculer avec précision et détail les paramètres de commande optimaux (et le coût total associé) et assouplir des hypothèses comme la considération d'un processus non-markovien pour décrire les transitions d'états du système. Cette approche est facile à utiliser, plus flexible et bien adaptée pour étudier le comportement dynamique des décisions communes et pour comparer les politiques de commande dans une large gamme de configurations système. Cependant, avant d'optimiser les paramètres de commande, nous nous basons sur la littérature pour explorer les structures de politiques de commande qui peuvent être adaptées dans notre contexte. L'objectif est de proposer une politique de commande plus économique combinant simultanément les activités de

production, de remise à neuf et de setup. La section suivante résume le cadre de l'approche adoptée.

3 APPROCHE DE RESOLUTION

L'approche de solution adoptée a été utilisée dans plusieurs problèmes de planification de la production comme dans (Assid et al., 2014). Elle combine l'outil de la simulation et des techniques d'optimisation pour optimiser de façon heuristique le coût total encouru lié aux paramètres de commande. Les principales étapes de cette approche sont:

3.1 Formulation du problème

Cette étape vise à formuler analytiquement le problème comme indiqué dans la section 2.3. Elle décrit la dynamique du système en fonction de ses états et définit les contraintes du problème, les variables de décision et la fonction de coût total à minimiser. L'objectif du problème de contrôle des flux pour le système hybride considéré est aussi détaillé. Il vise à trouver les variables de commande (u_{pro} , u_{rem} et ϕ) qui minimisent le coût total. Sur la base de cet objectif, deux travaux sont choisis pour explorer les politiques de commande qui peuvent être adaptées dans notre contexte.

3.2 Description des politiques de commande

Deux politiques de commande conjointes de type seuils critiques inspirées de la littérature ont été sélectionnées. Des modifications ont été apportées afin d'adapter leur structure à notre contexte (p. ex. limiter les coûts d'inventaires des retours en permettant leur élimination si nécessaire).

3.2.1 Politique de commande adaptée de Polotski

La structure de la politique de commande proposée dans (Polotski et al., 2017) est représentée graphiquement dans cette section. Elle combine les régimes de transformation à des taux de production et de remise à neuf maximaux avec des régimes de production à la demande et de remise à neuf suivant le taux de retours.

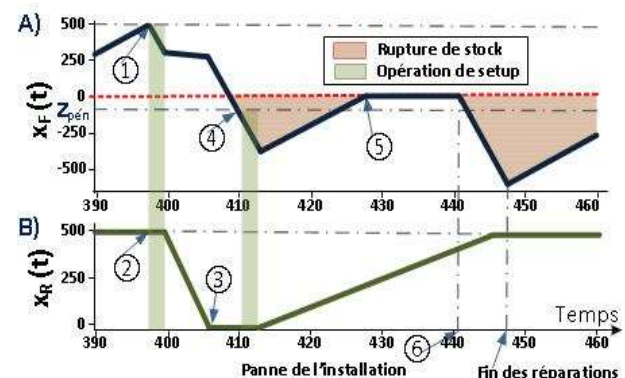


Figure 1 : Variation du niveau des inventaires lorsque la politique Polotski est appliquée.

La Figure 1 présente un exemple de la variation des niveaux d'inventaire des produits finis et des retours suivant les instructions de la politique de Polotski. Elle montre également que le système hybride restaure le stock de produits finis de taille $Z_F=500$ juste avant son passage au mode de remise à neuf (①). Pendant ce temps, le stock des produits retournés continue d'augmenter en fonction de la quantité des retours reçue (u_R). Notons que lorsque ce stock est plein ($x_R=Z_R=500$) tout produit retourné est éliminé afin de limiter les coûts d'inventaire (②). L'atteinte de Z_F a pour but de réduire le risque de rupture de stock lors de l'utilisation des retours dans le processus de remise à neuf : l'installation fonctionne à capacité maximale (i.e. au taux U_{rem}^{max}) en utilisant les retours en stock (i.e. $x_R>0$) et adapte son taux de remise à neuf au flux de retours lorsque $x_R=0$ (③). Rappelons qu'au mode de remise à neuf, la capacité de transformation est supposée inférieure à la demande des clients, d'où la baisse du niveau des stocks de produits finis. Lorsque le niveau d'inventaire des produits finis (x_F) atteint $Z_{pén}$, une opération de setup est déclenchée pour passer au mode de production (④) et restaurer leur stock. Ainsi, l'installation fonctionne à capacité maximale tant que la totalité des clients ne sont pas satisfaits ($x_F<0$). Dans le même sens, quand $x_F=0$, le système réduit son taux de production ($u_{pro}=d$) afin d'éviter des coûts d'inventaire (⑤). Cependant, des pannes aléatoires peuvent se produire causant l'arrêt des activités de transformation (⑥). Ces arrêts peuvent causer des ruptures de stock. La politique de Polotski se caractérise donc par trois paramètres de commande : Z_R , Z_F et $Z_{pén}$.

3.2.2 Politique de commande proposée

La politique proposée est inspirée de (Gharbi et al., 2006). Il s'agit de constituer le stock tampon des produits finis de taille Z_F dès que le système passe en mode de production et non jusqu'à la prochaine opération de setup comme dans la politique précédente. Un taux de production à la demande est également utilisé lorsque l'espace de stockage est plein ($x_F=Z_F$). Cela permet de stocker plus de produits finis pour faire face la non-disponibilité de l'installation qui provient principalement de pannes aléatoires. Dans le même sens, le passage au mode de production est effectué au moment où le stock de produits finis est vide ($x_F = 0$) seulement afin d'éviter des coûts de rupture de stock additionnels. La condition de passage au mode de remise à neuf se diffère également de la politique précédente en observant le niveau d'inventaire des retours. Ce passage est effectué lorsque leur espace de stockage est plein. Le but est de maximiser la quantité des produits remis à neuf. La politique proposée se caractérise alors par deux paramètres de commande: Z_R et Z_F .

La Figure 2 présente un exemple de la variation des niveaux d'inventaire des produits finis et des retours suivant les instructions de la politique de commande proposée. Elle montre que lorsque le niveau d'inventaire

atteint le seuil $Z_F=500$, le taux de production de l'installation est adapté au taux de la demande clientèle (①). Pendant ce temps, le niveau d'inventaire des retours continue d'augmenter au taux de retours (U_R) en raison du flux des produits retournés. Cependant, quand le stock des retours est plein ($Z_R=500$) (②), une opération de setup ($T_S=2$) est déclenchée afin de passer au mode de remise à neuf (③). Cette opération implique un arrêt de l'activité de transformation, d'où la diminution du niveau d'inventaire des produits finis suivant la demande des clients. Notons que pendant la période de setup, la quantité des produits retournés est rejetée à cause du manque d'espace de stockage additionnel. Dès la fin de l'opération de setup, l'installation fonctionne à capacité maximale (i.e. au taux U_{rem}^{max}) profitant ainsi de la disponibilité des produits retournés en stock (i.e. $x_R>0$). Mais, lorsque ce stock se vide, le taux de remise à neuf est adapté au flux de retours (④). Notons qu'au mode de remise à neuf, x_F diminue à taux différents en fonction du taux de remise à neuf. Cette diminution est due au fait que la capacité maximale de remise à neuf est inférieure à la demande clientèle. Comme mentionné dans le paragraphe précédent, si le niveau x_F atteint 0, une opération de setup est déclenchée mais cette fois-ci pour passer au mode de production (⑤). Dans ce mode, l'installation fonctionne à capacité maximale tant que $x_F<Z_F$ (i.e. le stock n'est pas plein). Cependant, des pannes aléatoires peuvent se produire causant l'arrêt des activités de transformation (⑥). Ces arrêts peuvent à leur tour causer des ruptures de stock.

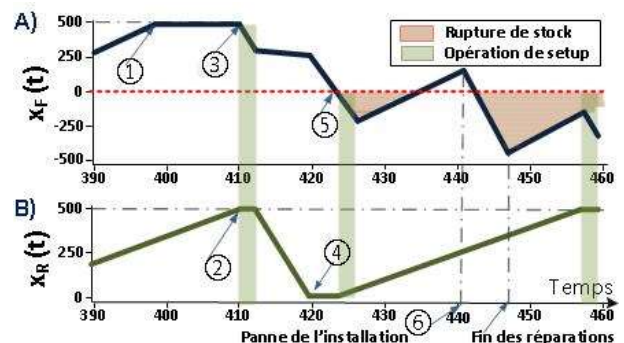


Figure 2 : Variation du niveau des inventaires lorsque la politique proposée est appliquée.

Les deux politiques de commande décrites ci-dessus sont optimisées puis prises en compte dans notre étude comparative. L'outil de simulation est utilisé pour évaluer la performance du système lorsqu'une politique de commande est appliquée.

3.3 Modèles de la simulation

Afin de quantifier la meilleure mesure de performance (i.e. coût total encouru) exprimée par (3) pour chaque politique de commande étudiée, une approche basée sur la simulation est développée en utilisant le logiciel Arena. Les modèles de simulation combinés à événements discrets-continus sont développés pour

représenter fidèlement le fonctionnement du système présenté à la section 2.2 et pour fournir une représentation précise des équations formulées à la section 2.3. Chaque modèle est contrôlé par une politique de commande et utilise à la fois des composants discrets et continus qui fonctionnent de manière synchrone comme pour le calcul des variations dans les inventaires (x_F et x_R). Par exemple, la partie continue peut déterminer le niveau des stocks instantanés en utilisant l'équation (1), tandis que la partie discrète modélise les interventions de réparations de l'installation. En effet, le modèle est composé de différents réseaux. Chacun effectue des tâches spécifiques (p. ex. pannes et dépassement d'un seuil) afin de calculer instantanément l'inventaire ou la pénurie des produits finis, de déterminer les taux de production et de remise à neuf, de vérifier l'état de l'installation, etc. Une fois l'exécution de la simulation est arrêtée, le coût total formulé par (3) est calculé. Il sera utilisé dans la procédure d'optimisation. La Figure 3 représente sous forme de diagramme à blocs nos modèles de simulation.

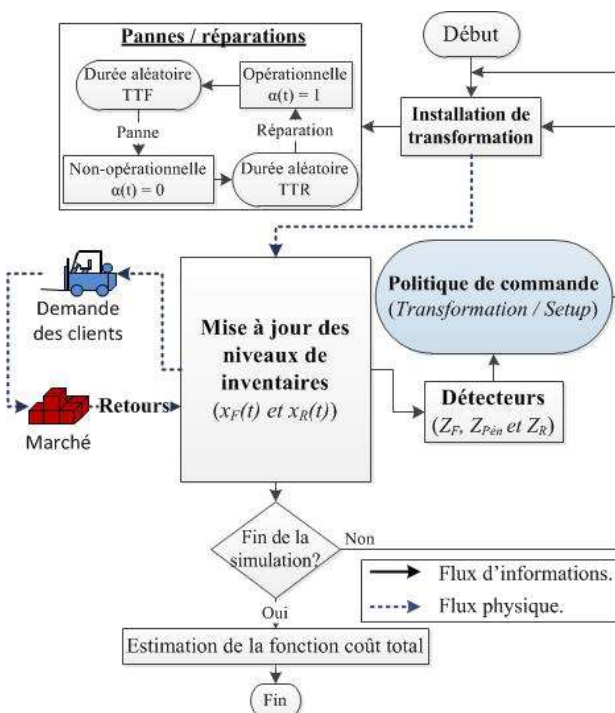


Figure 3 : Schéma-bloc des modèles de simulation.

3.4 Optimisation de la fonction de coût

Comme indiqué précédemment, notre modèle de simulation est utilisé pour effectuer une analyse statistique et optimiser les politiques de commande considérées en calculant les meilleures valeurs des paramètres de commande (Z_F , Z_R et $Z_{Pén}$) qui minimisent le coût total encouru exprimé par (3). L'optimisation est réalisée grâce à une combinaison des plans d'expériences et de la méthode de surface de réponse ainsi qu'à travers un exemple numérique. Ce dernier permet d'illustrer l'approche expérimentale adoptée et de comparer les politiques de commande étudiées. Les données du

système sont énumérées au Tableau 1. Ces données représentent le scénario de base pour lequel l'analyse comparative sera effectuée. Les paramètres de coûts sont choisis de sorte que $c_F^+ < c_F^-$ et $c_{rem} < c_{pro}$.

Différentes expériences de simulation hors ligne sont réalisées pour le scénario de base afin d'explorer le domaine d'expérimentation admissible, y compris les valeurs optimales des paramètres de commande (Z_F , Z_R et $Z_{Pén}$). Par conséquent, $Z_{Pén}$ est fixé au niveau zéro afin d'éviter de grands coûts de pénurie (voir Section 3.2.1). Le problème original est ensuite réduit pour définir deux facteurs de commande (Z_F et Z_R) pour les deux politiques considérées. Ainsi, un plan factoriel complet 3^2 sera appliqué afin d'estimer chaque interaction séparément. En utilisant cinq répliques, un total de 45 simulations est prévu pour chaque plan. La durée de simulation pour une seule réplique est définie par 1.000.000 unités de temps pour assurer l'atteinte du régime permanent. Le logiciel Statgraphics est par la suite utilisé pour produire l'analyse de variance multifactorielle et pour déterminer la fonction de surface de réponse sous forme d'un modèle de régression de second ordre reliant le coût total aux paramètres de commande (Z_F et Z_R). Les analyses de variance montrent que plus de 93% de la variabilité observée du coût total prévu est expliquée par les modèles (6) et (7):

$$C_{Polotski} = 13043,1 - 3,18Z_F - 3,22Z_R + 2,65 \cdot 10^{-4} \cdot Z_F \cdot Z_R + 8,22 \cdot 10^{-4} \cdot Z_F^2 + 1,08 \cdot 10^{-3} \cdot Z_R^2 \quad (6)$$

$$C_{Proposée} = 6815,12 - 1,46Z_F - 0,55Z_R - 6,43 \cdot 10^{-4} \cdot Z_F \cdot Z_R + 6,63 \cdot 10^{-4} \cdot Z_F^2 + 4,65 \cdot 10^{-4} \cdot Z_R^2 \quad (7)$$

Les solutions optimales des deux politiques de commande sont présentées dans le Tableau 2. Elles montrent que la politique proposée est plus avantageuse en termes de coûts par rapport à celle de Polotski avec une réduction du coût total de 42,71%. Ce résultat est dû essentiellement sa capacité à réduire significativement les ruptures de stock en construisant le stock tampon des produits finis (de taille Z_F) le plus rapidement possible contrairement à la politique de Polotski qui débute sa construction juste avant une opération de setup (voir Section 3.2.1). De plus, avoir un plus grand espace de stockage des retours ($Z_R^* = 2047$ pour la politique proposée par rapport à $Z_R^* = 1278$ pour Polotski) augmente la quantité des produits remis à neuf qui coûtent moins cher (coût de transformation plus faible). Une étude comparative est également réalisée afin de vérifier si cet avantage économique se maintient.

4 ÉTUDE COMPARATIVE

Cette section vise à étendre l'étude du scénario de base (Tableau 1) à un large éventail de paramètres du système. La Figure 4 résume les différents scénarios étudiés.

Paramètres	U_{pro}^{max}	U_{rem}^{max}	d	$T_f^{(1)}$	$T_r^{(1)}$	T_s^{ij}	c_s^{ij}	c_F^+	c_F^-	c_{pro}	c_{rem}	c_R^+	γ_R
Valeurs	130	85	100	Log-N(85,25)	Log-N(10,2)	2	500	1	30	10	5	0,5	0,15

(¹) T_f et T_r indiquent les distributions de probabilité du temps avant la panne et la durée de la maintenance corrective respectivement.

Tableau 1 : Paramètres de simulation

Politiques	Z_F^*	Z_R^*	Coût total (C^*)	Intervalle de confiné (95%)
Polotski	1730	1278	8222,27	[8217,18 ; 8231,36]
Proposée	2098	2047	4710,87	[4706,24 ; 4717,50]

Tableau 2 : Résultats d'optimisation et validation

Elle présente l'effet de plusieurs paramètres de coût (sous-figures 4.a-4.c), le temps de setup (sous-figure 4.d) et le taux de retour (sous-figure 4.e) sur la différence de coût total entre la politique de commande proposée et celle de Polotski. Cette différence de coût est exprimée par: $D.C. = (C_{Polotski} - C_{Proposée}) / C_{Polotski}$. La Figure 4 confirme la supériorité économique de la politique de commande proposée puisqu'elle génère le coût total le plus bas (la valeur de D.C. reste strictement positive), quelle que soit la valeur des paramètres du système sélectionnés. Par exemple, avec l'augmentation du coût de pénurie (c_F^-) (sous-figure 4.a), les pannes du système deviennent plus significatives puisque les coûts de rupture de stock couteront plus cher. Cela signifie que l'avantage de la politique proposée croîtra grâce à sa capacité à réduire le risque de rupture de stock (voir section 3.2.2). D'où l'augmentation de la valeur de D.C. Cette valeur augmente également lorsque le coût de production (sous-figure 4.c) ou le taux de retours augmente (sous-figure 4.e). Ce phénomène est expliqué par le fait que la politique proposée remet à neuf plus de produits retournés; qui coûtent moins cher, grâce notamment au contrôle du niveau d'inventaire des retours (voir section 3.2.2) ainsi qu'à la taille plus grande de leur espace de stockage (Z_R). Par contre, quand le temps de setup (T_S) augmente (sous-figure 4.d) D.C. diminue. En effet, la politique proposée exécute les opérations de setup pour passer au mode de remise à neuf sans tenir compte du niveau d'inventaire des produits finis (x_F), causant ainsi des ruptures de stock importantes dans certains cas. Contrairement à la politique de Polostski, x_F doit atteindre son maximum (Z_F) avant d'exécuter une telle action. Rappelant qu'une opération de setup implique un temps d'arrêt plus long des activités de transformation.

En résumé, la politique de commande proposée donne le meilleur résultat en termes de coûts pour une large gamme de configurations du système. Sa structure intègre à la fois la planification de la production et la remise à neuf et le setup pour les systèmes hybrides.

5 CONCLUSION

Cet article traite le problème de la planification des activités de production et de remise à neuf pour les systèmes hybrides. Notre travail est motivé par le besoin d'une meilleure structure de politiques de commande

dans de nombreuses applications industrielles où une unité de transformation non fiable nécessite des actions de setup à chaque fois qu'elle transite entre le mode de production et celui de remise à neuf. L'objectif principal est de proposer une structure adéquate de politiques de commande pour planifier efficacement le processus de transformation et minimiser le coût total. Ce dernier se compose des coûts de production, de remise à neuf, d'espaces de stockage, d'inventaire des retours ainsi que ceux d'inventaire et de pénurie des produits finis. En se basant sur la littérature, deux politiques de commande sont adaptées dans notre contexte. Une étude comparative approfondie est ensuite effectuée pour évaluer leurs caractéristiques et analyser l'effet d'une large gamme des paramètres du système sur leur coût total optimal. Notre étude est réalisée grâce à une approche de résolution basée sur la simulation et des techniques d'optimisation (p. ex. les plans d'expériences et la méthodologie de surface de réponse). Cette approche est capable de déterminer les paramètres de commande optimaux qui minimisent le coût total encouru. Les résultats obtenus démontrent la supériorité; en termes de coûts, de la politique de commande proposée; comparativement à celle de Polotski et al., (2017). Cela est dû notamment à sa capacité à réduire considérablement le risque de rupture de stock et à maximiser la quantité des produits remis à neuf. Plusieurs extensions possibles de ce travail peuvent être envisagées en considérant d'autres aspects liés au contexte industriel tels que les fabricants d'équipement d'origine qui vendent leurs nouveaux produits sur le marché primaire tandis que les produits remis à neuf sur le marché secondaire avec des prix différents.

REFERENCES

- Aras, N., Verter, V., & Boyaci, T. (2006). Coordination and priority decisions in hybrid manufacturing/remanufacturing systems. *Production and Operations Management*, 15(4), 528-543.
- Assid, M., Gharbi, A., & Hajji, A. (2014). Joint production and setup control policies: an extensive study addressing implementation issues via quantitative and qualitative criteria. *The International Journal of Advanced Manufacturing Technology*, 72(5-8), 809-826.
- Berthaut, F., Gharbi, A., Kenné, J. P., & Boulet, J. F. (2010). Improved joint preventive maintenance and

hedging point policy. *International Journal of Production Economics*, 127(1), 60-72.

- Debo, L. G., Toktay, L. B., & Wassenhove, L. N. V. (2006). Joint life-cycle dynamics of new and remanufactured products. *Production and Operations Management*, 15(4), 498-513.
- Elhafsi, M., & Bai, S. X. (1996). Optimal production control of a dynamic two-product manufacturing system with setup costs and setup times. *Journal of Global Optimization*, 9(2), 183-216.
- Flapper, S. D., Gayon, J. P., & Lim, L. L. (2014). On the optimal control of manufacturing and remanufacturing activities with a single shared server. *European Journal of Operational Research*, 234(1), 86-98.
- Francas, D., & Minner, S. (2009). Manufacturing network configuration in supply chains with product recovery. *Omega*, 37(4), 757-769.
- Gharbi, A., Kenné, J. P., & Hajji, A. (2006). Operational level-based policies in production rate control of unreliable manufacturing systems with set-ups. *International Journal of Production Research*, 44(3), 545-567.

- Govindan, K., Soleimani, H., & Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, 240: 603-626.
- Ilgin, M. A., & Gupta, S. M. (2010). Environmentally conscious manufacturing and product recovery (ECMPRO): a review of the state of the art. *Journal of environmental management*, 91(3), 563-591.
- Polotski, V., Kenne, J. P., & Gharbi, A. (2015). Optimal production scheduling for hybrid manufacturing–remanufacturing systems with setups. *Journal of Manufacturing Systems*, 37, 703-714.
- Polotski, V., Kenne, J. P., & Gharbi, A. (2017). Production and setup policy optimization for hybrid manufacturing–remanufacturing systems. *International Journal of Production Economics*, 183, 322-333.
- Porter, M. E. (1990). The competitive advantage of nations. *Harvard business review*, 68(2), 73-93.
- Teunter, R. H., Bayindir, Z. P., & Den Heuvel, W. V. (2006). Dynamic lot sizing with product returns and remanufacturing. *International Journal of Production Research*, 44(20), 4377-4400.

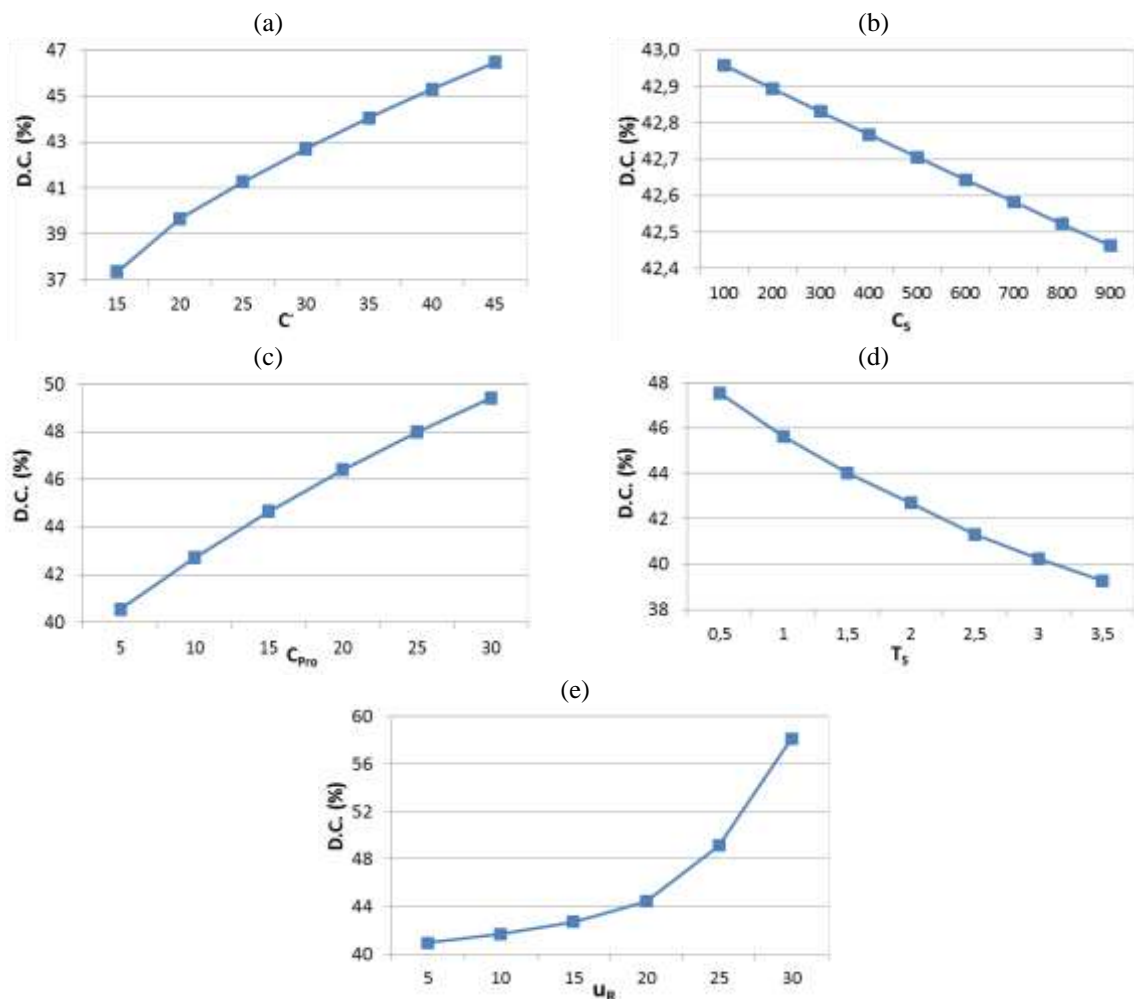


Figure 4 : Effet de la variation de paramètres du système sur la différence de coût total entre la politique proposée et celle de Polotski.

PLANIFICATION À COURT TERME DANS UN SYSTÈME DE PRODUCTION COMPLEXE GÉRÉ EN DDMRP

P. BAPTISTE, Robert PELLERIN

École Polytechnique Montréal,

P.O. Box 6079, Station Centre-ville, Montréal, Québec, Canada, H3C 3A7

pbaptiste@polymtl.ca, robert.pellerin@polymtl.ca

RÉSUMÉ : *Cet article traite des méthodes de planification dans des structures de production complexes avec des BOM multiniveaux. Le MRP, modèle traditionnel de flux poussé, utilise l'ordonnancement pour la planification à court terme. Le calcul de besoins et le lotissement créent des ensembles d'ordres de fabrication parfaitement connus sur un horizon suffisamment long. Au contraire, le Kanban, en flux tiré, offre peu de visibilité en ordonnancement puisque les ordres sont créés au dernier moment. Une nouvelle méthode, Demand Driven MRP (DDMRP) permet de gérer certains produits en « kanban généralisé » et les autres en Make-To-Order (MTO). De manière contrintuitive, les produits gérés en Kanban généralisé offrent beaucoup plus de visibilité et donc d'opportunité d'ordonnancement que les autres. Non seulement on connaît généralement à l'avance les ordres de fabrication, mais on connaît aussi certaines consommations futures. Au contraire, certains produits gérés en MTO offrent très peu de visibilité. En revanche, le fait qu'une majorité des ordres générés le soit en MTO offre au niveau opérationnel la possibilité de gérer de regroupement d'ordres portant sur le même produit et une meilleure visibilité sur les dépendances latérales. Ces conclusions induisent une vision différente de la planification en DDMRP.*

MOTS CLÉS : DDMRP, Planification, ordonnancement, nomenclatures complexes

1 CONTEXTE

Dans un système MRP, le niveau opérationnel utilise des méthodes d'ordonnancement. Le calcul des besoins a fourni un ensemble d'ordres de fabrication connus en quantités, date de début au plus tôt et au plus tard que l'on peut ordonnancer sur les ressources disponibles. Dans un système MRP, l'ordonnancement est plus difficile, car souvent, les ordres sont créés en continu on ne dispose pas d'un ensemble de Jobs connus à l'avance. Cet article se concentre sur la visibilité qu'offrira une nouvelle méthode, le DDMRP, pour l'ordonnancement, si elle est utilisée avec des nomenclatures multiniveaux complexes.

Le MRP naît dans les années 70 (Orlicky 1975) des débuts de la flexibilité et de l'apparition des ordinateurs. On y définit 3 niveaux de planification, stratégique tactique et opérationnel, chacun étant une désagrégation du précédent. Le plan stratégique utilise les prévisions commerciales pour faire un plan général de production (Master Schedule, souvent sur un an) et simuler le fonctionnement futur de l'entreprise pour ne générer des stocks qu'au moment où

l'on en a besoin. Rapidement, cet outil de flux poussé est devenu la base de la planification de toutes grandes entreprises occidentales. Son principal inconvénient est qu'il prend acte de lead times (Kanet 86) très longs et nécessite donc des prévisions fiables à long et moyen termes dont on ne dispose jamais. À l'inverse, les entreprises japonaises vont découpler la planification stratégique (dimensionnement des ressources) du niveau opérationnel ou elles utilisent la notion de cascade de boucles de stockage (les KANBANS, le flux tiré) en diminuant drastiquement les tailles de lot et en améliorant les systèmes de production (SMED, TPM, Qualité, etc.) (Womack et Al., 1990). Dans les deux cas, la génération des ordres de fabrication ne tient pas compte de l'état courant du système de production : d'où des conflits d'accès aux ressources et un besoin de stocks de sécurités. Plusieurs méthodes seront proposées dans les années 90 pour lier la génération de la demande et le taux de charge du système : le Conwip (Speraman et Al., 1990) qui permet de limiter en tout temps le nombre d'ordres de fabrication en cours de production, le Drump-Buffer-Rope issu de la théorie des

contraintes (Watson et Al., 2007) qui conditionne l'entrée d'un nouvel ordre à l'état de la file d'attente devant le goulot d'étranglement et finalement les roues de production (King 2013) qui proposent une création cyclique des ordres (méthode qui a d'ailleurs, et c'est original dans le domaine, été brevetée 0). Il est délicat de comparer ces méthodes, car certaines déclenchent les ordres (et décident des quantités), alors que d'autres ne font que libérer des ordres déjà créés. Coté MRP, la principale évolution viendra avec le Sales & Opération planning (Palmatier 2003) qui découple aussi le plan stratégique du plan opérationnel et intégrant au niveau opérationnel la prévision des ventes et la planification opérationnelle de toutes les couches du système de production. Finalement, depuis la fin des années 90, ces méthodes sont disponibles dans les logiciels (ERP), sans donner entière satisfaction, et sans évolution notable.

La parution en 2011 de la troisième édition de l'ouvrage (Orlicky 2011) qui consacre sa quatrième partie « *Looking Backward and Forward* » à la définition du *Demand Driven MRP* (DDMRP) vient introduire une nouvelle vision. Ses supporters affirment qu'il permet simultanément de diminuer les ruptures de stock et les surstocks, améliorant le niveau de service et les coûts encourus.

2 DDMRP : REVUE DE LITTÉRATURE

Globalement, le DDMRP part du principe que les stocks n'existent que pour augmenter le service client et se protéger de la variabilité de la demande. Il faut déterminer quels composants doivent être gérés à l'aide d'un KANBAN généralisé (par BUFFER) et les autres composants leur seront asservis (Make To Order).

Dans une BOM, chaque SKU a son propre *Lead Time* : *Manufacturing Lead Time* (MLT) ou *Purchasing Lead Time* (PLT), selon qu'il soit fabriqué (ou assemblé) ou acheté. Il a aussi un *Cumulative Lead Time* qui est le maximum sur toutes les branches de sa BOM des sommes de MLT. Ce Lead time cumulatif représente le délai théorique nécessaire pour fabriquer un lot de ce SKU si tous ses composants étaient fabriqués (ou achetés) à la commande pour satisfaire cet ordre de fabrication. Le choix des produits gérés sur buffer permet de découpler la production de ce SKU. Une fois positionnés les produits gérés sur buffer, les Decoupled Lead Time sont les nouveaux DLT calculés dans la sous-nomenclature limitée aux produits non gérés sur buffer.

Les Buffers ont 4 zones : une zone d'interdiction de produire, une zone où la production est autorisée, une zone de production obligatoire et une zone de sécurité. L'ensemble des composants gérés sur buffer est décomposé en 9 classes, selon que l'on estime que le lead time est long, moyen ou court et que la variabilité de la demande est forte, moyenne ou faible.

Dans chacune des 9 cases, la taille des zones est une fonction linéaire du produit de la demande moyenne (Average Dealy Usage) et du lead time.

À l'instar de la production cyclique classique, le lancement d'un ordre vise à atteindre le niveau maximum du buffer. Contrairement au Kanban, le seuil de déclenchement n'est pas seulement la projection de stock (on-hand + en attente – demande actuelle) mais aussi un mécanisme permettant d'anticiper des pics de demande.

Depuis la parution du livre définissant le DDMRP (Orlicky 2011), les auteurs ont écrit deux ouvrages susceptibles d'éclairer le chercheur sur la méthode DDMRP : l'un sur les indicateurs de performance (Smith and Smith 2016), l'autre plus étoffé sur la méthode (Ptack et Chad 2016)0. De même, à l'instar du But de E. Goldratt, Caroline Mondon a écrit un ouvrage romancé (Mondon 2016) sur le DDMRP.

Les premières recherches trouvées concernent des comparaisons entre MRP et DDMRP sur des cas-écoles ou réels. La thèse (Ihme 2015), traite d'une application concrète de DDMRP. L'auteur compare, sur un cas réel, une intégration de MRP avec les DBR de la théorie des contraintes, et une intégration DDMRP-TOC. Il constate effectivement sur ce cas une amélioration significative, mais de son aveu même, l'entreprise étudiée partait de si loin qu'il est difficile de séparer les gains liés au DDMRP de ceux liés à la simple amélioration de son système de planning. L'équipe des Mines d'Albi a cherché à valider la méthode DDMRP sur un cas-école, celui du jeu de Kanban (Miclo et Al, 2015 et 2016). Sur ce jeu simple, ils ont montré qu'effectivement, le service pouvait être amélioré. Cependant, la limite de cette étude est la simplification à outrance de l'atelier étudié avec une seule ressource. Dans ce cas, les problèmes fondamentaux d'accès aux ressources se limitent à la gestion d'une seule file d'attente. D'autre part, lorsqu'ils ont essayé d'apprendre automatiquement à paramétrer le DDMRP, des choix très simplificateurs sur les formes des demandes n'ont pas pu amener à des résultats convaincants. Finalement, les comptes rendus d'expériences relatés dans ces travaux portent sur des systèmes logistiques plus que sur des systèmes de production. Les résultats présentés par l'étudiante de Polytechnique (Velasco-Acosta, 2016, 2017), montrent aussi la difficulté de construire un scénario crédible de comparaison étant donné le grand nombre de paramètres utilisés. (Rim et Jiang 2016, se sont intéressés à déterminer la stratégie optimale de positionnement des buffers. Ils donnent dans deux cas (Make To Order ou Make to forecast) des modèles mathématiques modélisant le problème, et une optimisation à base d'algorithme génétique. Mais leur fonction de coût est exclusivement basée sur la minimisation des coûts de stockages à lead time donné. Autrement dit, ils ne tiennent pas compte de la variabilité de la demande et du taux de service qui est la justification de l'existence de ces stocks selon la méthode MRP. Bien qu'il ne s'agisse pas de

DDMRP, les travaux sur le dimensionnement des buffers en DBR (par exemple (Georgiadis et Al. 2013) et (Ye et Han 2008)) peuvent aussi inspirer le dimensionnement des buffers en DDRMP.

Finalement, la majeure partie des auteurs constatent que sur une chaîne logistique le DDMRP peut introduire des baisses importantes des niveaux de stock, en production, les résultats montrent plus un gain sur le taux de service.

Force est de constater qu'il y a peu de littérature autour du DDMRP, si ce n'est quelques travaux soutenus par des entreprises convaincues ou des sociétés développant des logiciels supportant la méthode. Rien n'a été publié sur des cas de nomenclature complexes, en particulier, si beaucoup se sont intéressés aux produits gérés sur buffer, la manière de gérer les autres produits n'a pas suscité d'effort en recherche.

3 TYPE DE PRODUCTION ÉTUDIÉE

On s'intéresse ici à des entreprises complexes disposant de plusieurs ateliers indépendants, et de nomenclatures complexes, ayant un minimum de 5 à 5 niveaux différents. Les nomenclatures sont enchevêtrées, si bien qu'un même composant peut être utilisé dans plusieurs composés, et qu'un SKU a généralement plusieurs composés.

Typiquement, dans une nomenclature comme celle de la Figure 1, le premier chiffre du produit regroupe le plus souvent tous les produits fabriqués dans un même atelier (un niveau de nomenclature correspond à 1 ou plusieurs ateliers technologiques). Ensuite, chaque produit est associé à une gamme dans cet atelier, typiquement avec plusieurs opérations sur plusieurs ressources différentes de l'atelier. Dans cette figure, le chemin en rouge et l'entourage spécial du produit 302 sont utilisés dans la suite de l'article.

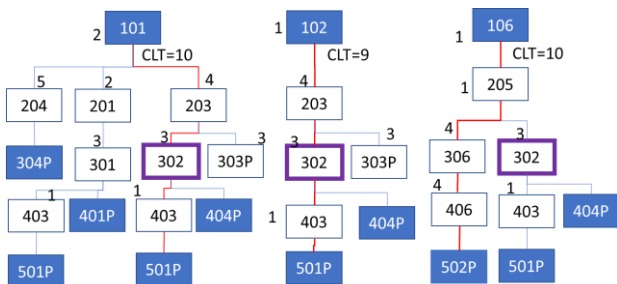


Figure 1 : exemples de nomenclatures utilisées

4 NOTATION

On utilisera les notations suivantes :

- B : ensemble des produits gérés sur buffer
- NB : ensemble des produits non gérés sur buffer
- $BOM(i)$: nomenclature de i

$LT(i)$ Lead time du produit i (acheté ou fabriqué)

$CLT(i)$ cumulative lead time de i (plus longue somme des LT sur toutes les branches de la BOM)

$DBOMD(i)$ Sous ensemble de la BOM coupé aux éléments de B

$CH(i,j), i \in B$ s'il existe, chemin de la BOM(i), allant de i à j ($j \in N \cup NB$) ne contenant que des éléments de NB

$L(CH(i,j))$ Longueur du chemin de i à j

$$L(CH(i,j)) = \sum_{k \in CH(i,j) - \{j\}} LT(k)$$

$DLT(i)$ Lead time cumulé, découplé

$$DLT(i) = \max_k (L(CH(i,k)))$$

$DES(i)$ Composants de i gérés sur buffer

$$\forall i \in B, DES(i) = \{j \in B / \exists CH(i,j)\}$$

$AS(i)$ Composés de i gérés sur buffer

$$\forall i \in B, AS(i) = \{j \in B / \exists CH(j,i)\}$$

Sur la Figure 1, les boîtes en bleu sont les produits gérés sur buffer. Chaque produit est associé à un LT (en noir, à côté de la boîte) et les CLT des produits finis sont identiques au DLT puisque parmi les produits gérés sur buffers, seuls les produits finis ont des composants. Pour les 3 produits 101, 102 et 106, le chemin en rouge est le chemin critique permettant de calculer les DLT correspondants.

Remarque : Par commodité, on représente toujours les BOM(i) ou DBOM(i) comme des arbres. Dans le cas général, ce sont des graphes orientés sans cycle, du moins si l'on souhaite ne représenter qu'une fois chaque composant. La Figure 2 donne une représentation graphique de la nomenclature donnée précédemment. Deux composants, les 301 et 302 utilisent l'un et l'autre le même composé 403. On peut indifféremment représenter la BOM comme un arbre, avec deux versions différentes du produit 403 (et de ses descendants) ou comme un graphe avec une seule version du produit 403 et de ses descendants.

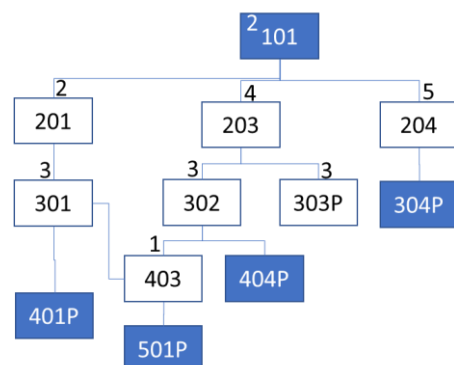


Figure 2: autre représentation de la BOM du produit 101

Par souci de simplicité, on ne considérera pas ces cas dans les notations (il suffirait de considérer les ensembles de chemins).

Le même problème se rencontre pour modéliser les nomenclatures ascendantes nécessaires pour définir les AS(i)

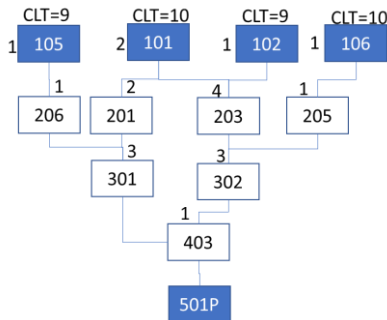


Figure 3 : nomenclature inverse

Dans cet exemple, le produit 501P rentre dans la composition de 4 produits différents. Le produit 403 est lui-même présent deux fois dans le produit 101. Les CLT des produits 105 et 106 sont plus longs que la somme des lead time des chaînes allant de 501P vers eux, ce qui signifie que les chemins critiques des produits 105 et 106 ne passent pas par 510P.

Lorsque l'on décide à la date courante t de lancer la production d'un produit $i \in B$, on crée un OF Job $J_{i,t}$ avec

- une due date $d_{i,t} = t + DLT(i)$
- une ready date $r_{i,t} = t + DLT(i) - LT(i)$

Pour chaque élément k de $DBOM(i)$, on va aussi créer un OF $J_{k,i,t}$: ce Job pour produire le produit k est créé en MTO pour satisfaire l'OF $J_{i,t}$. Les *ready* et *due dates* sont :

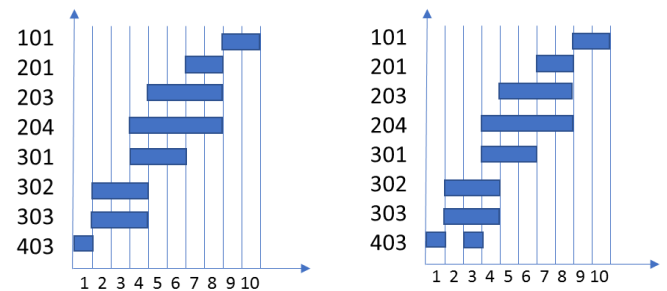
- $d_{k,i,t} = t + DLT(i) - L(CH(k, i))$
- $r_{k,i,t} = t + DLT(i) - L(CH(k, i)) - LT(k)$

Exemple :

Dans l'exemple précédent, une décision de production du produit 101 prise à la date $t=1$ déclenche la cascade présente Figure 4. Dans la seconde partie de la figure, on utilise la nomenclature en graphe et la même demande de produit 101 déclenche 2 ordres différents pour le produit 403, avec des fenêtres de temps qui n'interfèrent pas.

Remarque : Avec ce mode de fonctionnement, il est tout à fait possible de générer pour un même produit i deux ordres différents ayant la même fenêtre de temps [*ready date*; *due date*] ou avec des fenêtres qui se chevauchent. Chaque ordre correspondra à un élément de AS(i). Accessoirement, les ordres auront sans doute été générés avec des niveaux de

stock disponible différents par rapport à leurs buffers (des « priorités » différentes). La décision de grouper ou non ces ordres est donc arbitrée au niveau opérationnel.



Cas groupé

Cas non groupé

Figure 4 : cascade des ordres générés pour un OF du produit 101

5 VISIBILITÉ AU NIVEAU OPÉRATIONNEL

Cette partie explore la visibilité pour ordonnancement que la méthode DDMRP offre. Elle distinguera le cas des produits gérés en kanban généralisé avec Buffer. Bien que ce ne soit pas souvent le cas, pour simplifier l'écriture, on parlera d'atelier où tous les produits sont gérés avec Buffer ou d'atelier où les produits sont tous gérés en MTO.

5.1 Point de vue de $i \in B$

La création d'un ordre pour ces produits est longuement décrite dans les différentes références. La visibilité sera vue de points de vue différents

Ordres en attente de production.

Par construction, pour un produit i géré sur buffer, avec un Lead Time Découplé $DLT(i)$, à une date courante t , aucun nouvel ordre ne peut être créé dans l'intervalle de temps $[t; t + DLT(i) - LT(i)]$.

De la même manière, à une date t donnée, si dans un atelier (un niveau de nomenclature) tous les produits fabriqués (k) sont gérés sur buffer, la production est :

- Connue dans $\left[t; t + \min_k (DLT(k) - LT(k)) \right]$
- Inconnue après $t + \max_k (DLT(k) - LT(k))$

Entre les deux, les ordres sont partiellement connus.

Si un seul composant d'un produit i n'est pas géré sur buffer (et donc $DLT(k) > LT(k)$), contrairement au Kanban, la décision de lancer en fabrication est une décision qui n'entraîne pas la possibilité de mettre en fabrication immédiatement : elle crée un ordre placé dans une planification future.

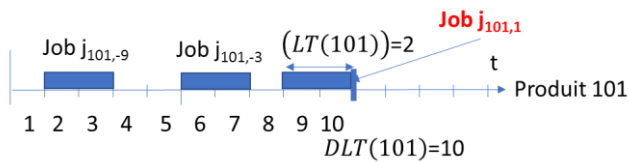


Figure 5: exemple d'ordres en attente pour le produit 101

La décision en début de période 1 de lancer une production du produit 101 entraîne un $J_{101,1}$ qui se fera dans l'intervalle [8; 10]. Dans la Figure 5, il y a aussi deux autres ordres lancés dans le passé pour ce produit-là.

5.1.1 DDMRP décentralisé

En Kanban, il est souvent possible de ne générer l'ordre de production définitif qu'au moment où on le lance en production : on ajuste la quantité au dernier moment. C'est en particulier le cas si on utilise dans l'atelier où le produit est fabriqué une boucle CONWIP. Sauf à utiliser des boucles CONWIP à cheval sur plusieurs niveaux de nomenclature (ce qui semble difficile), dans le cas d'un produit géré sur buffer en DDMRP, la quantité à produire doit être connue au moment de la création de l'ordre (puisque'il y a création d'ordres dépendants en MTO), et donc souvent bien avant le début de l'exécution.

Cela n'interdit pas de générer des ordres au fil de l'eau (à tout instant en fonction de la position du stock disponible), mais la quantité doit être fixée au moment de la création, pas de l'exécution.

De la même manière, l'usage de tableau KANBAN permettait d'anticiper ou de retarder un peu la création d'un ordre en KANBAN, en fonction du taux de charge de l'atelier (pilotage visuel). En DDMRP, il s'agit d'un taux de charge « anticipé, dans le futur », ce qui rend assez délicat un pilotage visuel.

5.1.2 Visibilité sur la future consommation de $i \in B$

De la même manière que l'on dispose d'une connaissance des productions futures d'un produit $i \in B$, on dispose aussi de certaines informations sur sa consommation future.

Pour $i \in B$, pour tous les $k \in AS(i) / L(CH(i, k)) = DLT(k)$, aucune consommation n'est connue. En revanche, pour tous les $k \in AS(i) / L(CH(i, k)) \leq DLT(k)$, k n'est pas sur le chemin critique ayant permis de calculer $DLT(k)$. Donc si t est la date courante, la création d'un ordre pour un tel k entrainera une consommation de i à la date $t + DLT(k) - L(CH(i, k))$, donc dans le futur.

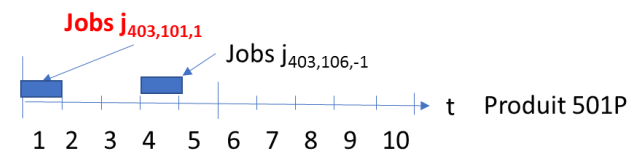


Figure 6: consommation future connue de 501P

Dans l'exemple de la Figure 6, on voit qu'un ordre créé aujourd'hui (date t) entraîne un ordre à la première période pour le produit 403, donc une consommation immédiate du produit 501. En revanche, comme 501P n'est pas sur le chemin critique ayant servi à définir $DLT(106)$, un ordre créé pour le produit 106 il y a déjà une période entraîne une consommation du produit 501P en période 4.

5.2 Point de vue de $i \in NB$

Un produit de $i \in NB$ est fabriqué en MTO lorsque l'on a décidé de produire un produit $j \in AS(i)$, donc un de ses composés. Pour chaque produit $j \in AS(i)$, à la date courante t , aucune nouvelle demande ne peut plus arriver dans l'intervalle :

$$\left[t; t + \min_{k \in AS(i)} (DLT(k) - L(CH(k, i)) - LT(k)) \right]$$

Cette durée correspond au temps minimum pour aller de l'ensemble des produits de $DES(i)$ au produit i , lorsqu'un besoin est généré par un produit d' $AS(i)$.

Si un seul composant de i n'est pas géré sur buffer, alors cet intervalle n'est pas de largeur nulle : le composant doit être fabriqué avant. En revanche, dans le cas particulier où le produit i concerné a tous ses composants gérés sur buffer et qu'il est sur le chemin critique ayant permis de définir le DLT d'un des produits de $AS(i)$, alors l'intervalle est de largeur nulle : un ordre peut arriver durant la période t avec un début immédiat imposé.

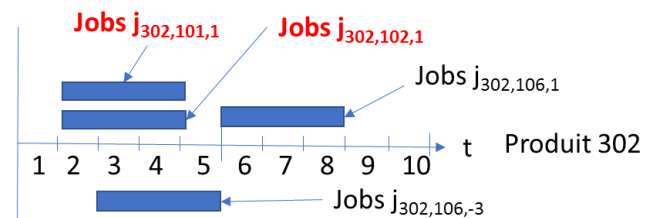


Figure 7: ordres futurs pour le produit 302

Dans la Figure 7, les 3 ordres au-dessus de l'axe sont des ordres générés à la date t courante. Les deux ordres $J_{302,101,1}$ et $J_{302,102,1}$ sont positionnés à la seconde période, car 302 est sur le chemin critique ayant permis de calculer $DLT(101)$ et $DLT(102)$, et d'autre part $LT(403)=1$ (403 est le composant de 302). En revanche, 302 n'est pas sur le chemin ayant servi à définir $DLT(106)$ et l'ordre créé aujourd'hui ne sera lancé qu'en période 6.

Dans la partie basse, un ordre pour 106 créé il y a 3 périodes de temps est positionné à la période 3.

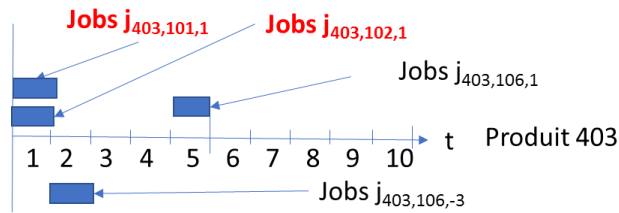


Figure 8: exemple de demandes induites sur le produit 403

Sur la Figure 8, on voit les conséquences de la création des ordres de production pour les produits 101, 102 et 106 à la date $t=1$. La conséquence est pour les produits 101 et 102 est la création de deux ordres $J_{403,101,1}$ et $J_{403,102,1}$ devant être exécuté immédiatement. Leur Lead time ne comprend pas de délais de lotissement, mais ce sont des lead time de production, Il faudrait donc que le travail commence immédiatement.

En revanche, pour ce même produit 403, il existait déjà un ordre prévisionnel pour la période 2, correspondant à un ordre $J_{106,-3}$. créé il y a déjà 3 périodes. Cette coexistence entre des productions prévues et des urgences non prévisibles est assez peu commune.

5.3 Considération sur le Lead Time

Le Lead Time est défini comme les sommes des temps de production, de mise en course et d'attente nécessaires pour faire un lot de produit de taille habituelle dans un atelier. Habituellement, les temps d'attentes sont suffisamment importants pour permettre un ordonnancement de la production respectant *ready* et *due dates*.

Dans la méthode DDMRP, comme dans la méthode MRP d'ailleurs, le *lead Time* est considéré comme préexistant. C'est une donnée. Cependant, comme cela a été vu précédemment, les articles des BOM dont un des composants est géré sur buffer et qui sont sur un chemin critique n'ont pas de visibilité sur leur ordonnancement. Un ordre peut arriver pour la date courante, sans anticipation possible. En revanche, pour des produits gérés sur buffer, on connaît généralement les ordres en avance et il est possible de faire de l'ordonnancement. On peut donc imaginer que la définition même des lead time pourrait accorder plus de temps d'attente prévisionnels pour les produits feuille des DBOM et moins pour les produits gérés en DDMRP.

5.4 Considération sur le lotissement

Si on considère la Figure 7, on constate que 3 ordres vont coexister pour le produit 302. Dans un approvisionnement de type MRP, il est certain que les trois ordres auraient été regroupés en un seul, quelle que soit la politique de

regroupement (principalement EOQ ou période fixe, éventuellement plus complexe). Plusieurs raisons peuvent conduire aux regroupements : diminuer les temps de réglage, diminuer les volumes de déplacement, etc. Mais dans une approche MRP, s'il y a un besoin urgent pour un sous-ensemble du lot, on doit impérativement prioriser le lot entier. Il serait intéressant d'étudier dans le contexte du DDMRP des règles de lotissement locales tenant compte des taux de charge, des urgences relatives des différents ordres à faire, etc. Puisqu'à priori le lotissement de type MRP est une alternative parmi les autres, on comprend qu'il doit être possible en moyenne de faire mieux. Encore faudra-t-il définir les objectifs et les critères permettant de mesurer les performances associées à différentes politiques.

5.5 Considération sur les communications latérales.

Le principe de la planification hiérarchique est de conférer une indépendance à chaque atelier. Les ordres ont des *ready* et des *due dates* qui, si elles sont respectées, rendent les décisions indépendantes. Le lotissement de MRP (regroupements des besoins bruts sous un même ordre) fait disparaître les finalités des différentes productions. Souvent, en pratique, principalement si les produits finis sont en MTO, on double le système de planification MRP par une planification « produits » dans laquelle des planificateurs « transversaux » sont responsables de synchroniser les ordres de tous les composants de la BOM (dans tous les ateliers) d'un produit fini afin de garantir sa production.

Dans le cas de DDMRP, on conserve sur tous les produits non gérés sur Buffer l'origine du besoin. Il est donc facile de visualiser tous les ordres dépendants du même ordre initial, et de tenir compte des éventuelles anticipations ou retard des autres pour ajuster sa propre planification.

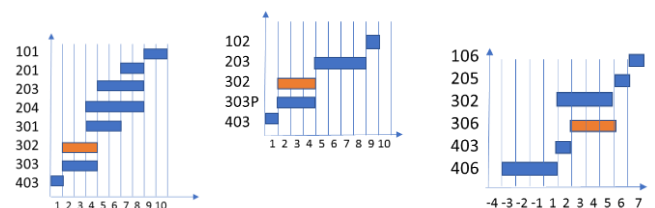


Figure 9 : historique des 3 ordres pour 302

Dans l'exemple de la Figure 9, il est facile d'extraire les contextes de chaque ordre du produit 302. Un retard constaté sur le job $J_{406,106,-3}$ ou un problème anticipé dans l'atelier susceptible de produire le $J_{301,101,10}$ auront alors une influence sur les décisions de lotissement, mais aussi sur la priorisation des lots en ordonnancement.

5.6 Considération sur l'évolution des lead times

Comme cela a été dit précédemment, tous les seuils des buffers, et par conséquent le niveau de réapprovisionnement est, pour les 9 classes d'un produit $i \in B$, une fonction

linéaire du produit $LT(i) \cdot ADU(i)$. Lors du premier congrès industriel international du DDMRP (FAPICS, Lyon 2016) de nombreuses compagnies s'interrogeaient sur la manière de tenir compte des variations de l'ADU (lissage exponentiel, largeur de la fenêtre, etc.) en temps réel. En revanche, aucune ne prévoyait d'ajuster en temps réel le $LT(i)$ dans l'équation. L'idée est cependant séduisante. En effet, une surcharge de l'atelier entraîne un allongement des LT qui, si on en tenait compte en temps réel, augmenterait alors les tailles des buffers et par conséquent les quantités à réapprovisionner en diminuant d'autant la proportion de temps accordé aux réglages, ce qui permettrait sans doute de diminuer la charge. Il s'agirait d'une première en termes de pilotage en boucle fermée de la taille des lots. Ce point mériterait certainement une étude plus poussée pour voir s'il est possible d'obtenir un système naturellement stable ou si les échelles de temps rendent cette adaptation illusoire.

5.7 Considération sur la visibilité

On s'est interrogé jusqu'ici sur la connaissance a priori des ordres futurs à réaliser pour un produit donné. Cependant, pour un produit i donné, au-delà de l'horizon ou ses ordres sont connus, pour chaque composé de $k \in AS(i)$, on connaît la date à laquelle il peut engendrer un ordre pour i et on connaît la position de son niveau actuel de stock projeté par rapport aux seuils.

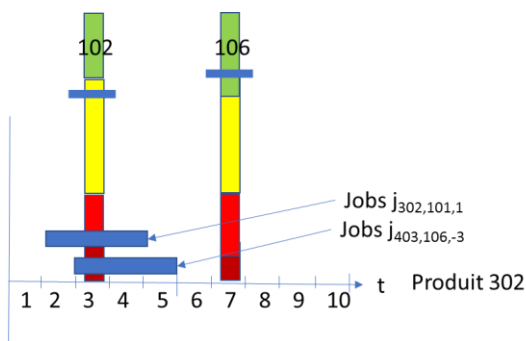


Figure 10: exemple de visibilité réelle sur le produit 302

L'exemple de la Figure 10, propose une manière de présenter sur un même graphe les ordres connus et la position des stocks projetés des composés $AS(320)$. Le produit 101 est omis puisqu'il y a déjà un ordre lancé cette période. Pour le produit 106, le niveau du stock projeté tient compte de l'ordre $J_{406,106,-3}$ qui justement doit permettre de produire l'ordre $J_{106,-3}$ qui arrive à terme en période 7.

Il semble beaucoup plus facile d'avoir une visibilité sur les ordres futurs avec un système géré en DDMRP qu'avec un système purement composé de boucle Kanban.

6 CONCLUSION

La méthode DDMRP est assez nouvelle et les principaux travaux disponibles portent soit sur les calculs des seuils des

buffers (son paramétrage), soit sur des applications en logistique ou dans des cas de production et de nomenclatures simples, voire simplifiés. Une meilleure compréhension des mécanismes de planification possibles dans les cas complexes est nécessaire.

Paradoxalement, les produits gérés sur buffer, donc en Kanban généralisé, offrent une réelle visibilité et des possibilités d'ordonnancement : on connaît habituellement tous les ordres à produire sur un horizon suffisamment long, et peu de surprises de dernière heure peuvent perturber le plan. En revanche, puisqu'ils ont un DLT souvent bien supérieur au LT , lorsqu'on décide de les produire, on place un ordre assez loin dans le futur. En conséquence, il est sans doute illusoire de penser à utiliser le taux de charge du système de production pour anticiper ou retarder la création d'un ordre.

Au contraire, pour les autres produits, gérés en MTO (par rapport aux ordres générés par les produits gérés sur buffer), la visibilité est beaucoup plus faible. Sur une chaîne allant d'un produit fini géré sur Buffer à un composant de base géré sur buffer, plus on se rapproche du composant de base, plus faible est la visibilité. Pour la dernière couche (produits dont tous les composants sont gérés sur buffer), la visibilité est même nulle : identique aux produits classiquement gérés en Kanban). Pour ces produits, des ordres peuvent arriver pour exécution immédiate.

Finalement, contrairement au MRP classique, deux opportunités pourraient permettre de mieux gérer à court terme : une visibilité latérale de la situation de tous les ordres ayant été créé pour le même besoin et la possibilité de décider ou non de regrouper les ordres pour le même produit ayant des fenêtres de temps se chevauchant, voire identiques.

7 RÉFÉRENCES

- Georgiadis, G. and A. Politou, 2013. Dynamic Drum-Buffer-Rope approach for production planning and control in capacitated flow-shop manufacturing systems, *Computers & Industrial Engineering*, 65, p.689–703.
- Hietikko, J., 2014. *Supply Chain Integration with Demand Driven Material Requirement Planning System: Wärtsilä 4-Stroke*, Master's Thesis in Industrial Management, University of Vaasa, Finland.
- Ihme, M. and R. Statton, 2015. Evaluating demand driven MRP: a case based simulated study, *International Conference of the European Operations Management Association*, Neuchatel, Switzerland, June.
- Ihme, M., 2015. *Interpreting and applying Demand Driven MRP, A case study*, PhD dissertation, Nottingham Trent university, UK, N0295945.

- Jiang, J. and S. Rim, 2016. Strategic Inventory Positioning in BOM with Multiple Parents Using ASR Lead Time, *Mathematical Problems in Engineering*, <http://dx.doi.org/10.1155/2016/9328371>, 9 pages.
- Kanet, J., 1986. Toward a better understanding of lead times in MRP systems, *Journal of Operations Management*, 6 (3–4), p. 305-315.
- King P. L. and J.S. King, 2013. *The production wheel Handbook*, CRC Press, isbn 978-1-4665-5419-1.
- Louly M. and A. Dolgui Optimal MRP parameters for a single item inventory with random replenishment lead time, POQ policy and service level constrain, *International Journal of Production Economics*, 143, p. 35–40.
- Miclo, R., F. Fontanili, M. Luras, J. Lamothe, and B. Milian, 2016. An empirical comparison of MRPII and Demand-Driven MRP, *IFAC-Papers OnLine* 49-12, p. 1725–1730.
- Miclo, R., F. Fontanili, M. Luras, J. Lamothe, and B. Milian, 2016. An empirical study of Demand-Driven MRP, *ILS2016, Bordeaux*, France.
- Miclo, R., F. Fontanili, M. Luras, J. Lamothe, and B. Milian, 2015. MRP vs. Demand-Driven MRP: Towards an Objective Comparison, *IESM2015, Séville*, Spain.
- Milco, R., 2016. *Challenging the “Demand Driven MRP” Promises: A Discrete Event Simulation Approach; thèse de doctorat*, Mines d’Albi, France.
- Mondon C., 2016. *The missing links*, Industrial Press Inc., ISBN-13: 978-0831136079.
- Orlicky, J., 1975. *Material Requirements Planning*, McGraw-Hill, New York.
- Orlicky, J., 2011. *Material Requirements Planning*, McGraw-Hill, New York, ISBN: 978-0071755634.
- Packowsky, J., S. Joswig, and T. Heckmann, 2016. Brevet “System and method for high-mix wheels for capacity planning resource planning and material resource planning”, US 20160307136 A1.
- Palmatier G., 2003. *Enterprise Sales and Operations Planning: Synchronizing Demand, Supply and Resources for Peak Performance*, J Ross Pub, ISBN: 978-1932159004.
- Ptak, C. and S. Chad, 2016. *Demand Driven Material Requirements Planning (DDMRP)*, Industrial Press Inc. ISBN-13: 978-0831135980.
- Smith, D. and C. Smith, 2016. *Demand Driven Performance using smart metrics*, McGraw-Hill Education, ISBN-13: 978-0071796095.
- Shofa, M.J. and W. O. Widyarto, 2017. Effective production control in an automotive industry: MRP vs. demand-driven MRP, *AIP Conference Proceedings* 1855, 020004 (2017); doi: 10.1063/1.4985449.
- Speraman, M.L., D. L. Woodruff, and W. Hopp, 1990. Conwip: a pull alternative to Kanban, *International Journal of Production Research*, 28 (5), p. 879-895.
- Velasco-Acosta, A.P., 2016. *L’adoption de l’approche DDMRP pour la planification de la production et de la gestion des stocks de moteurs*, Mémoire de maîtrise, Polytechnique Montréal.
- Velasco A.P., C. Acosta, C. Mascle, and P. Baptiste, 2017. L’applicabilité de l’approche demand-driven MRP dans un environnement complexe; *CIGI 2017*, Compiègne, France.
- Watson, K.J., J. H. Blackstone, and S.C. Gardiner, 2007. The evolution of a management philosophy: The theory of constraints, *Journal of Operations Management*, 25, p. 387–402.
- Womack, J. P., D. T. Jones, and D. Roos, 1990. *The machine that Changed the world*, Free press, ISBN-13: 978-0743299794.
- Ye, T. and W. Han, 2008. Determination of buffer sizes for drum–buffer–rope (DBR) controlled production systems, *International Journal of Production Research*, 46 (10), p. 2827-2844.

PREDICTIVE CROSS-MANAGEMENT OF DISASTER PLANS IN BIG DATA SUPPLY CHAINS: FUZZY COGNITIVE MAP APPROACH

**First S. CHEHBI-GAMOURA,
Second R. DERROUCHE,**

EM Strasbourg Business School,
F-67000 Strasbourg - France
samia.gamoura@gmail.com,
ridha.derrouiche@em-strasbourg.eu

Fourth M. MALHOTRA

Chandigarh University
Institute of Computing, 140413–
India
mmanishamalhotra@gmail.com

Third D. DAMAND

EM Strasbourg Business School,
F-67000 Strasbourg - France
damand@em-strasbourg.eu

ABSTRACT: *With the advent of Big Data, and the propagation of multiple business activities into supply chains, organizations are more attentive to disasters than ever. It is nearly impossible to predict their happening and the amount of the caused damages. Therefore, organizations try to cooperate in incorporating risk management in their information systems. This paper is focused on this problem by delineating understandings from the literature review. A set of fundamentals of disaster planning is also provided with qualified analysis and arguments. The proposed approach presents a risk assessment methodology that integrates Fuzzy Cognitive Maps predictive method. The approach is able to comprehend all valuation variables of the whole stakeholders that are involved in the chain. The key outcomes of this study aim to contribute in assisting enterprises in improving capability for disaster recovery. Finally, we specify a set of open challenges for further researches and a position on our future research.*

KEYWORDS: *Risk management, Big Data, Disaster Recovery Plan, Fuzzy Cognitive Map, predictive cross-management.*

1 INTRODUCTION

In today's huge digital revolution, the way we the Supply Chain (SC) is managed, all the activities are changing in Big Data context (Hashem, et al., 2015). In fact, business sustainability and risk management issues are pushed at the lead of interest for scholars and industrials (Papadopoulos, et al., 2017).

Industries today are facing plenty disordering events that could happen distinctly or concurrently (Sahebjamnia, Torabi, & Mansouri, 2015). Each issue might have different effect on resources, products, manpower, factories, etc. Typically, in order to face disasters, two main plans are settled by enterprises in their Information Systems (IS): the Disaster Recovery Plan (DRP) (Noh, Kitagawa, & Oh, 2014) and the Business Continuity Plan (BCP) (Macecevic, 2014). For example, a Swedish computer constructor should incorporate the disaster plan of Japan zone since he sub-contracts the assembly of microprocessors to an Italian constructor who purchases to factories in the Japanese islands of Ryūkyū.

To make their efforts the most beneficial, organizations should integrate all potential situations when analysing the possible risks in the whole chain (Kristo & Cingula). This leads them to consider all disaster causes of the other participants in the entire chain (Takakuwa, 2013). Though, the examination of scholar and industrial works highpoints a clear lack in collaboration practices, cross data management, and risks planning ((Brindley, 2017),

(Sarmiento, Hoberman, Jerath, & Jordao, 2016), (Chatterjee, Ismail, & Shaw, 2016)).

That is, we offer in this paper a designed methodology for predictive cross-management of disaster planning. We also try to highlight gaps in the existing researches and provide open-views of future opportunities. The proposed methodology is built on Fuzzy Cognitive Mapping (FCM) for prediction of causality relationships between disorder patterns among the chain in Big Data IS environment.

The paper is organized as follows: We begin by stating the research concern with a brief reminder of Big Data environments and risk management in SC. We then advance the proposed methodology that is based on risk assessment by FCM. Finally, we conclude by highlighting the current trends, the untaken research gaps, and the expected open views in the near future.

2 BACK GROUND AND RESEARCH GAPS

2.1 SC Sustainability in Big Data IS

The effect of natural or man-made disasters can cause vast economic damages (Brindley, 2017). The outcomes can cause colossal losses in activities and operational businesses (Mohan & Bakshi, 2017). Thus, plans and actions regarding downtimes not only disturb economic resources, but also all SC assets. This legitimizes the necessity for robust management and high ability to reestablishment and recovery (Lockamy III, 2014).

Many causes can trigger these economic stoppages. The most common of them in the context of SC organizations are: equipment (machines, laptops, projectors, etc.), facilities (electricity power, real estate, etc.), and manpower (technical and management human resources) (Wang, et al. 2015). Authors in (Li 2016) identify four sorts of seven effects (from E1 to E7) as following: E1: breakdown loss, E2: disaster recovery, E3: liability, E4: productivity decreasing, E5: customer loss, E6: idle staff and equipment, E7: damaged reputation.

The academic and industrial works that link disaster effects and negative effects are plenty; some of them are classified into the following table (Table 2).

Cost category	Samples	Reference	Effects
Cost of equipment idleness	Broken equipment, Incident in building.	(Li 2016)	E2 E6
Costs of employees idleness	Epidemic episode, Terrorist attack, Crash in transports.	(Tabikh 2014)	E2, E6, E4
Costs of revenue losses	Inaccessibility to online stores,	(Agostinho, et al. n.d.)	E5, E7, E4
Costs of repairing	Floods in factories, Earthquake in building and offices, Power-off in datacenter.	(Saltoglu, Humaira and Inalhan 2016)	E2
Liability costs	Damages of loaned cars,	(Li 2016)	E3, E7
Disaster recovery costs	Storm damages, Seismic damages in offices.	(Tabikh 2014)	E2
Customers and partners losses	Close stores, Inaccessibility to online purchasing.	(Agostinho, et al. n.d.)	E3, E5, E7

Table 1: List of costs categories related to downtime effects (Synthesized from literature review)

2.2 Disaster Recovery Planning in SC context

In a SC environment, a disaster is defined as a disturbing event related to a social alteration (Epstein and Khan 2014) that needs statement among stakeholders to manage a set of interdisciplinary challenges (Baba 2014) (Li 2016).

The main role of the Disaster Recovery Plan (DRP) is the sustenance and involvement in upper-level manage-

ment of business components (Macecevic 2014). Authors in (Brazelton et Lyons 2016) indicated that the DRP can be divided into six main processes: P1: consolidating the team, P2: evaluating risks in the organization, P3: instituting roles in divisions, P4: developing rules, strategies to perform, P5: documenting rescue procedures, P6: formulating the way to handle disasters. Each process comprises a set of sub-processes, where each one figures upon the others.

2.2.1 Challenge of the DRP in SC Context

No organization remains isolated in a SC network and businesses implies multitude partners, which can be contractors, third-party services providers, customers, suppliers, and ((Brindley 2017), (Yu, Li and Yang 2017), (Wang, et al. 2015)). In such volatile networked businesses, each edge assimilates the movement of products; the information streams; the transmission of titles; and many other flows (Papadopoulos, et al. 2017) in addition to business actions of human resources, real estate, IT systems, facilities, equipment, finance, etc. (Prasad, et al. 2015). Supposing all of these streams are operating consistently, so what occurs if something disturbs an activity?

The literature review offers two ways to face this challenge: the first way is by normalizing the communication canals in circumstances of crisis. However, efforts to standardize protocols are very restricted ((Kshetri, Fredriksson and Torres 2017), (Prasad, et al. 2015)). The second alternative approach is the cross-management of disaster plans. It seems to be the best way for enterprises to contribute and master the risk management overall the SC flows (L'Hermitte, et al. 2016). However, it is necessary to point out that the collaboration expertise is somehow daring (Wang, et al. 2015) because of the following hurdles: (1) Variation is closed to the life-cycle phase proper to each company (Noran 2014), (2) Due to the situation of crisis, enterprises may be tempted to overtake, remove or substitute dedicated teams, assuming a common control approach instead of collaborative plan (Nagurney, Masoumi and Yu 2015), (3) Interdisciplinary of the crisis depends on a extensive range of communities from economic, public institutions, healthcare, and social assets (Ojha, Salimath and D'Souza 2014).

To analyse the progress of research works in sustainability of networked companies, we reviewed widely the literature. The synthesis is shown in figure 1.

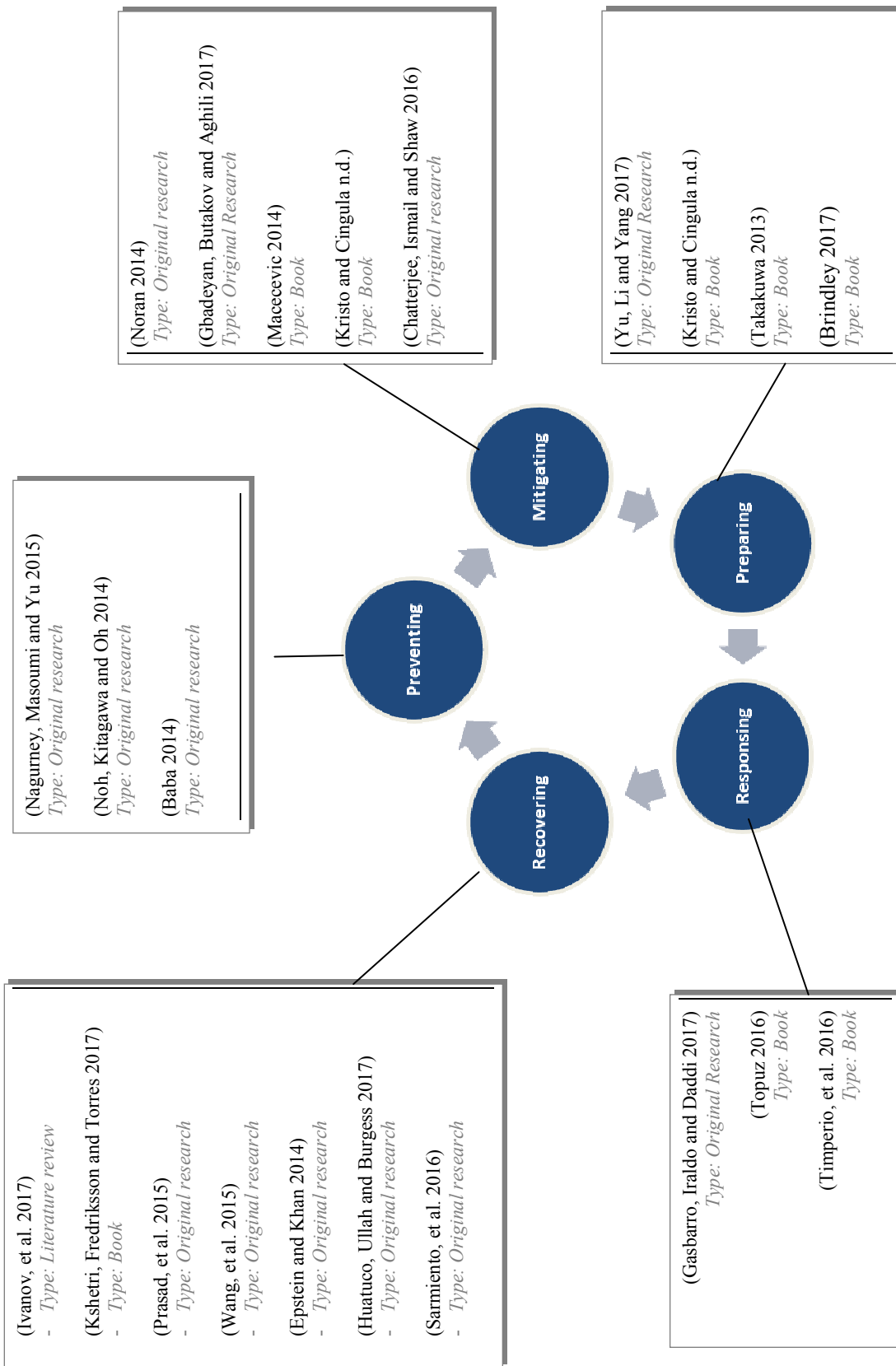


Figure 1 : Relevant Literature review by Disaster stages (with publication type)

2.2.2 Big Data to Support Disaster Planning

In sustainable development based on strong disaster infrastructure, Big Data environment can help in policy making, and risk strategies development by internalizing currently externalized environmental and societal circumstances (Kshetri, Fredriksson and Torres 2017). Big data can therefore help data scientists and risk strategy makers in: (1) developing and executing guidelines that manage man-made and natural threats (Papadopoulos, et al. 2017); (2) avoiding excess in resources and deprivation of capabilities that can offer sufficient services in situation of crisis (Lockamy III 2014), and (3) Big Data can be a good support of valuable information needed to improve prediction and then help in acting proactively to face disasters (Baba 2014).

However, academic works has been limited in giving future profits of the role of Big Data in sustainability of SC (Hashem, et al. 2015). Although there are works regarding Big Data and disaster management in enterprises, there have been no publications that use Big Data value-added to elucidate disaster recovery and preparation in enterprises networks (Papadopoulos, et al. 2017). Our work in the present paper aims to clarify some aspects related to Big Data that could facilitate deployment of our solution, but the approach is not directly based on unstructured data features. More clarifications regarding the proposed approach will be provided in section 3 below.

3 RESEARCH METHODOLOGY

3.1 Risk Assessment in SC Context

In the development of DRP, the primary function of risk valuation is to encode as many types of disasters as conceivable that a firm may occur, and then to figure out the plan of it will deal with each emergency (L'Hermitte, et al. 2016). Unfortunately, risk assessments are usually very complex to develop in advance, and very expensive in terms of tests, impact analysis and preparation (Macecevic 2014). In addition, the complexity of the collaborative aspect in the chain can make the task more tedious (Topuz 2016). State-of-the-art regarding risk assessment is plenty: In (Yu, Li and Yang 2017), authors propose a stochastic optimization model that emphasizes the ambiguity in risk preference and probability distribution through SC networks. The model is interesting but lacks in applicability and is much closed to the presented case study. Authors in (Noran 2014) proposed a new holistic model for responsivity of recovery teams in collaborative way within networks of enterprises. In (Epstein and Khan 2014), authors presented a risk-based approach named Application Impact Analysis (AIA) that is able to consider business key attributes and then provides a model for viewing data, assessing impact, and carrying out key results.

In order to create the most operative plan after a disruption, an enterprise must first study what the would-be disasters are that they could practicably meet, and how each of these might affect their business continuity (Gasbarro, Iraldo and Daddi 2017). To alleviate this task, we propose a solution to guide in risk assessment integrating prediction method for a cross-management approach. It is based on Fuzzy Cognitive Maps (FCM) and could be used taken into consideration when preparing a DRP.

3.1. FCM for Risk Management

With the advent of Big Data Analytics (BDA), application of Fuzzy Cognitive Mapping (FCM) is increasing and comes to be useful in the search for the causality between patterns and phenomena (Feyzioglu, Buyukozkan and Ersoy 2017). FCM is a graph representation of likelihood relationships between concepts (Papadopoulos, et al. 2017) and provides an understanding of complex relationships between those concepts (also called nodes or events). Connections are weighted by likelihood coefficients on directed arrows and represent causal relations between concepts (nodes). The weight value reflects the degree to which the concept is the strength of impact in the system at a specific lap of time. This value is a function of the sum of all upcoming weights multiplied and the value of the original concept at the directly previous state. A threshold function is also introduced to regulate the weighted sums. Figure 2 illustrates an example of FCM with weighted relationships.

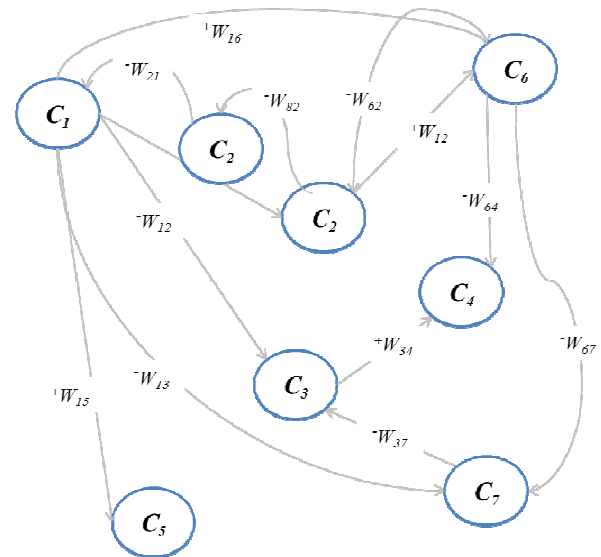


Figure 2: Example of FCM graph

Prediction by FCM can be made through the assessment of the likelihoods between concepts (nodes) (Sahebjamnia, Torabi and Mansouri 2015). The model is derived by labelling the components of the systems that are causalities values in relationships between its components (Papadopoulos, et al. 2017). As proposed in modelling process in some research works such as (Feyzioglu, Buyukozkan and Ersoy 2017), we opt to use

FCM in analysis of causality likelihood to the context of enterprises.

The DRP process-based model can be transcoded into a FCM model. In such modelling task, processes and sub-processes are represented by concepts (nodes), where likelihood relationships between them may be weighted. Administrators, and staff involved in disaster procedures can help in assessing weights of the different causal-links for each concept.

Formulation

Behind the graphical schema of the FCM, the mathematical model in context of processes/ sub-processes of a DRP is as described in the following formulation:

Suppose the set of P of N processes/sub-processes (concepts or nodes).

$$P = \{P_i\}_{1 \leq i \leq N} \tag{1}$$

For each process/sub-process p_i , there is a state vector V of N weights w_{ij} . A weight value w_{ij} represents the likelihood relationship (risk assessment) between the processes/sub-processes p_i and p_j in the interval range $[-1, +1]$.

$$\forall p_i \in P, \exists V = \{w_{ij}\}_{1 \leq j \leq N}, w_{ij} \in [-1, +1], w_{ii} = 0 \tag{2}$$

This model state can be presented using a what-if matrix W of $N \times N$ weights W_{ij} .

$$W = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n1} & \dots & W_{nn} \end{bmatrix} = \begin{bmatrix} 0 & W_{12} & \dots & W_{1n} \\ W_{21} & 0 & \dots & W_{2n} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n1} & \dots & 0 \end{bmatrix} \tag{3}$$

3.2 Aggregated FCP for Risk Cross-Management

Applying FCM to face the issue of risk management is not new. Many other research works attempted to introduce the approach to understand and model disaster management, such as (Feyzioglu, Buyukozkan and Ersoy 2017). However, as far as we know, no one tried to generalize the concept of FCM to the context of cross-management and collaboration aspect in networks of enterprises. As these research studies, we propose a method of generalization (aggregation) of the FCM to be applied on entire SC network.

Formulation

Based on FCM formulation previously detailed in the last sub-section (3.1.1), we propose the following steps, also clarified in figure 3:

Step 1: Extraction of individual FCMs

Once the stakeholders of the business line in the chain are identified, the individual FCM of processes and sub-processes are extracted from each Disaster Plan.

Step 2: Augmenting what-if matrixes

The individual maps extracted in step 1 are gathered, combined and augmented by auditioning them in superposed manner producing a single matrix.

Step 3: Normalization of values

Weights of the single matrix obtained in step 2 have to be normalized according to the number of stakeholders who supported it (suppose k stakeholders), then normalized weights (Suppose a_{ij}) is computed in the formula:

$$a_{ij} = \sum_{i=0}^k p_i \cdot a_i / k, \sum_{i=0}^k p_i = 1 \tag{4}$$

Step 4: Aggregation into one what-if matrix

The normalized values constitute a single normalized matrix (suppose A) which has to be published to all stakeholders and updated regularly.

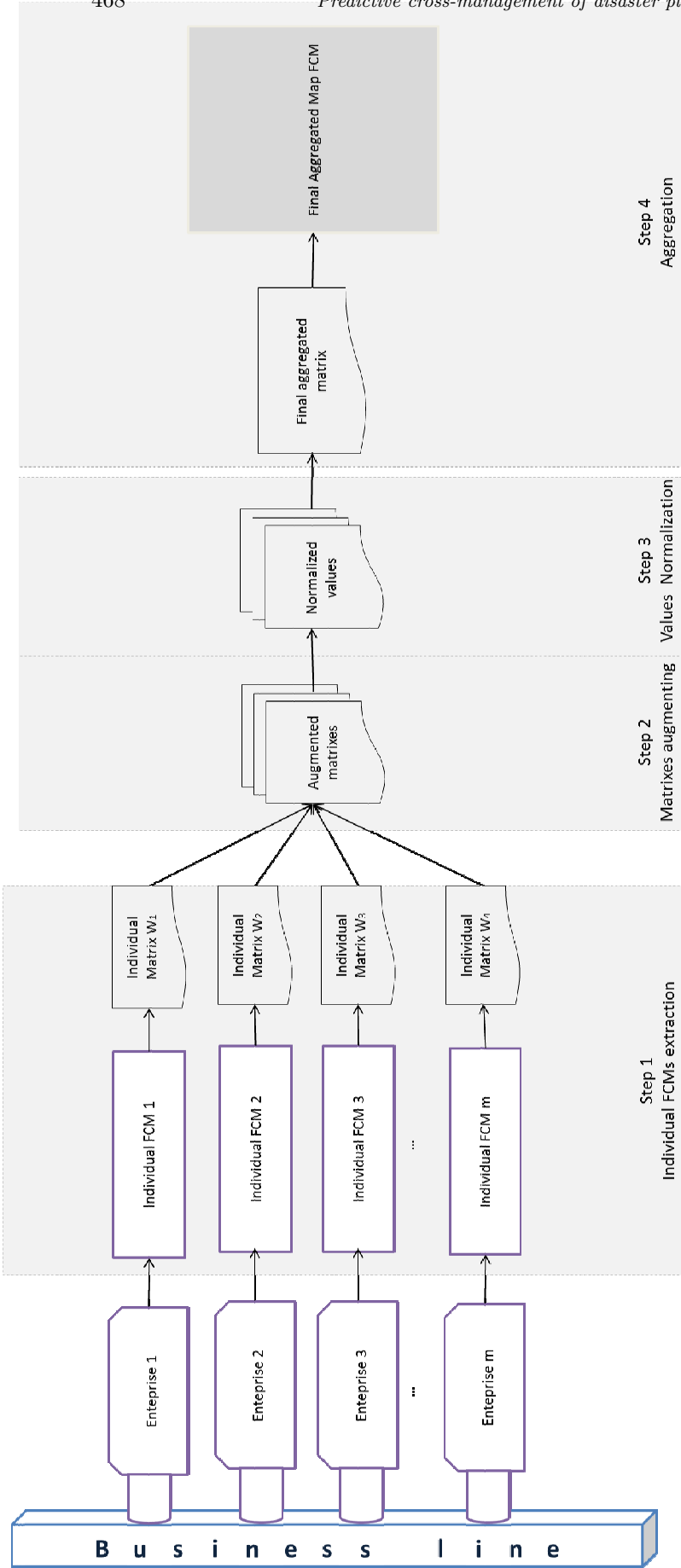


Figure 3: Steps of FCM aggregation in the proposed approach

4 CONCLUSION AND OPEN VIEWS

In large-scaled environments of Big Data, recovering Information Technology (IT) infrastructure after an outage becomes complex due to the large demand of information and the high scalability, in addition to the complexity of relationships between Supply Chain stakeholders. This also challenges organizations the ability to manage disasters. Therefore, this chapter proposed an interdependency risk assessment method that is based on a semi-quantitative approach; the Fuzzy Cognitive Map (FCM). The FCM is an adequate method that can be used to provide promises for what-if investigation scenarios of risks analysis. In addition to be a conventional predictive method, it can help in understanding the vulnerability in each part, and each division of enterprises in the chain.

The key finding of this paper is a cross-management model able to predict causality relationships between disruption patterns and risk phenomena among the Supply Chain. The outcome of this study may help in preventing enterprises that are involved in business linkages, from massive financial damages.

In the future open studies, the researchers can explore in the way that enterprises do to keep the system safe from downtimes. Further, to make the assessment model more accurate, researchers can apply the approach in their Supply Chains, gathering more details of downtime outcomes and the cost issues. An industrial case-study would be suitable in this situation. Additional research perspective can be to what scope does the organization profits from the disaster management program. Despite preventing for the potential downtimes can considerably reduce or avoid the harms, the investments in the planning programs can be colossal.

5 REFERENCES

- Agostinho, C., et al. "Towards a sustainable interoperability in networked enterprise information systems: trends of knowledge and model-driven technology." *Computers in Industry*, n.d.: 79, 64-76.
- Baba, H. "Introductory study on Disaster Risk Assessment and Area Business Continuity Planning in industry agglomerated areas in the ASEAN." *IDRiM Journal*, 2014: 3(2), 184-195.
- Brazelton, N. C., et A. M. Lyons. «Downtime and Disaster Recovery for Health Information Systems. » Dans *Health Informatics-E-Book: An Interprofessional Approach*, de R. Nelson et N. Stagers, 337-. Missouri, USA: Elsevier, 2016.
- Brindley, C. (2017). *Supply chain risk*. Routledge. New York, USA: Ashgate publishing, 2017.
- Chatterjee, R., N. Ismail, and R. Shaw. "Identifying Priorities of Asian Small-and Medium-Scale Enterprises for Building Disaster Resilience." *Urban Disasters and Resilience in Asia*, 2016: 179-194.
- Epstein, B., and D. C. Khan. "Application impact analysis: A risk-based approach to business continuity and disaster recovery." *Journal of business continuity & emergency planning*, 2014: 7(3), 230-237.
- Feyzioglu, O., G. Buyukozkan, and M. S. Ersoy. "Supply chain risk analysis with fuzzy cognitive maps." *In Industrial Engineering and Engineering Management, 2007 IEEE International Conference*. IEEE, 2017. 1447-1451.
- Gasbarro, F., F. Iraldo, and T. Daddi. "The drivers of multinational enterprises' climate change strategies: A quantitative study on climate-related risks and opportunities." *Journal of Cleaner Production*, 2017.
- Gbadeyan, A., S. Butakov, and S. Aghili. "IT governance and risk mitigation approach for private cloud adoption: case study of provincial healthcare provider." *Annals of Telecommunications*, 2017: 72(5-6), 347-357.
- Hashem, I. A. T., I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, et S. U. Khan. «The rise of "big data" on cloud computing: review and open research issues.» *Journal of Information Systems*, 2015: 47, 98-115.
- Huatuco, L. H., G. S. Ullah, and T. F. Burgess. "Supply Chain Major Disruptions and Sustainability Metrics: A Case Study." *In International Conference on Sustainable Design and Manufacturing*. Springer, Cham, 2017. 185-192.
- Ivanov, D., A. Dolgui, B. Sokolov, and M. Ivanova. "Literature review on disruption recovery in the supply chain." *International Journal of Production Research*, 2017: 1-17.
- Kristo, m., and m., Filipovic, m. Cingula. *Methodological aspects of measurnig business resilience*. 2015: Economic and social development: book of proceedings, n.d.
- Kshetri, N., T. Fredriksson, and D. C. R. Torres. *Big Data and Cloud Computing for Development: Lessons from Key Industries and Economies in the Global South*. Taylor & Francis, 2017.
- L'Hermitte, C., P. Tatham, B. Brooks, and M. Bowles. "Supply chain agility in humanitarian protracted operations." *Journal of Humanitarian Logistics and Supply Chain Management*, 2016: 6(2), 173-201.
- Li, Q. *Estimation of the costs of information system downtime in organisations: underlying threats, effects and cost factors*. Master's thesis, Twente, Netherlands: University of Twente, 2016.
- Lockamy III, A. "Assessing disaster risks in supply chains. ." *Industrial Management & Data Systems*, 2014: 114(5), 755-777.
- Macecevic, D. *Enterprise risk management-priority of export-oriented firm in emerging economy*.

- Economic and social development: book of Proceedings, 2014.
- Mohan, S., and N. Bakshi. "Supply Chain Resilience-Epidemiological Characterization." *History*, 2017.
- Nagurney, A., A. H. Masoumi, and M. Yu. "An integrated disaster relief supply chain network model with time targets and demand uncertainty." *Regional Science Matters*, 2015: 287-318.
- Noh, H. W., K. Kitagawa, and Y. S. Oh. "Concepts of Disaster Prevention Design for Safety in the Future Society." *International Journal of Contents*, 2014: 10(1), 54-61.
- Noran, O. "Collaborative disaster management: An interdisciplinary approach." *Computers in Industry*, 2014: 65(6), 1032-1040.
- Ojha, D., M. Salimath, and D. D'Souza. "Disaster immunity and performance of service firms: The influence of market acuity and supply network partnering." *International Journal of Production Economics*, 2014: 147, 385-397.
- Papadopoulos, T., A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, and S. Fosso-Wamba. "The role of Big Data in explaining disaster resilience in supply chains for sustainability." *Journal of Cleaner Production*, 2017: 142, 1108-1118.
- Prasad, S., H. C. Su, N. Altay, and J. Tata. "Building disaster-resilient micro enterprises in the developing world. ." *International Journal of Production Economics*, 2015: 39(3), 447-466.
- Sahebjamnia, N., S. A. Torabi, and S. A. Mansouri. "Integrated business continuity and disaster recovery planning: Towards organizational resilience." *European Journal of Operational Research*, 2015: 242(1), 261-273.
- Saltoglu, R., N. Humaira, and G. Inalhan. "Scheduled maintenance and downtime cost in aircraft maintenance management." *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 2016: 10(3), 580-585.
- Sarmiento, J. P., G. Hoberman, M. Jerath, and G. F. Jordao. "Enterprise engineering and management at the crossroads." *Computers in Industry*, 2016: 79, 87-102.
- Tabikh, Mohamad. *Downtime cost and Reduction analysis: Survey results*. Thesis of Master degree, Mälardalen University, School of Innovation, Design and Engineering, Innovation and Product Realisation., 2014.
- Takakuwa, S. *A perspective on manufacturing and environmental management*. Daaam international scientific book, 2013.
- Timperio, G., G. B. Panchal, R. De Souza, M. Goh, and A. Samvedi. "Decision making framework for emergency response preparedness: A supply chain resilience approach." *In Management of Innovation and Technology (ICMIT)*, 2016
- IEEE International Conference* . IEEE, 2016. 78-82.
- Topuz, ç. *Crisis management and strategies in tourism industry*. Proceedings book,, 2016.
- Wang, L., H. V Ramasamy, R. E. Harper, and M., Plattier, E. Viswanathan. "Experiences with Building Disaster Recovery for Enterprise-Class Clouds. ." *In Dependable Systems and Networks (DSN), 45th Annual IEEE/IFIP International Conference*. Rio de Janeiro, Brazil: IEEE, 2015.
- Yu, G., F. Li, and Y. Yang. "Robust supply chain networks design and ambiguous risk preferences." *International Journal of Production Research*, 2017: 55(4), 1168-1182.

PROCESSUS DE FABRICATION D'ÉLÉMENTS PRÉFABRIQUÉS : ÉTAT DE L'ART ET ANALYSE D'UN CAS D'ÉTUDES

Pierre ULTRÉ, Nadia LEHOUX

CIRCERB, Département de génie mécanique,
1065 avenue de la médecine, Université Laval,
Québec (Québec), G1V 0A6
pierre.ultré.1@ulaval.ca, Nadia.Lehoux@gmc.ulaval.ca

Sylvain Ménard

CIRCERB, DSA génie civil structure bois,
555 boulevard de l'Université, Université
du Québec à Chicoutimi,
Chicoutimi (Québec), G7H 2B1
Sylvain.Ménard@uqac.ca

RÉSUMÉ : *La construction industrialisée est actuellement en plein essor au Québec et ailleurs dans le monde. Le secteur étant toutefois fortement fragmenté, beaucoup de petits producteurs font face à un manque de capacité et ont des processus de gestion de la production peu développés. Cette étude s'intéresse aux façons d'améliorer les processus de production dans ce domaine pour gagner en efficacité. Une revue de la littérature permet tout d'abord de faire un état des lieux de la planification et de la gestion de la production dans la construction industrialisée. Elle compare ce secteur à la construction traditionnelle et au secteur manufacturier puis recense les différents problèmes auxquels les entreprises peuvent faire face ainsi que leurs solutions si elles sont connues. Une étude de cas concernant une entreprise québécoise spécialisée dans la production d'éléments préfabriqués montre par la suite les améliorations qui pourraient être effectuées afin de diminuer les temps d'opération et mieux utiliser la capacité de production tout en respectant des contraintes de financement et de disponibilité de la main d'œuvre.*

MOTS-CLÉS : *Construction industrialisée, Préfabrication, Gestion de la production, Amélioration*

1 INTRODUCTION

La demande en logements demeure une réalité et ne semble pas vouloir décroître aux cours des prochaines années puisque « le logement constitue l'une des composantes les plus cycliques de la demande globale » (Egebo *et al.*, 1990). Peu importe le pays, la zone géographique ou le niveau économique considéré, le logement est un droit et une nécessité pour tous. L'industrie du logement a cependant encore aujourd'hui des difficultés à subvenir convenablement à une telle demande. En effet, c'est un secteur d'activité très fragmenté qui utilise des méthodes ne s'avérant pas toujours des plus efficaces. C'est pourquoi l'industrie de la construction est aujourd'hui en cours de transformation. De plus en plus, les techniques de construction traditionnelles sont abandonnées ou révisées au profit de méthodes modernes de construction (*Modern Methods of Construction – MMC*)¹. La *House Builders Federation* au Royaume-Uni définit les MMC comme « des méthodes qui fournissent un processus de gestion des produits efficace pour fournir plus de produits de meilleure qualité en moins de temps » (Traduit de Sarden et Engstrom, 2010). Ces méthodes comprennent en grande majorité la « construction industrialisée » (*Industrialized Housing – IH*)² qui est définie comme « la branche de la construction qui est partiellement ou totalement réalisée dans un environnement contrôlé (i.e. une usine) utilisant des processus et des machines industrielles (Traduit de Branson *et al.*, 1990). Les MMC ne se déroulent pas forcément hors-site. Cependant, celles

qui nous intéressent ici le sont. C'est pourquoi par la suite, nous ne parlerons implicitement que des MMC hors-site. Ces pratiques, plus connues sous le nom de préfabrication, peuvent être en grande partie apparentées à celles d'autres secteurs industriels tels que le secteur manufacturier. Le principe est de réduire les gaspillages et de faire des économies tout en augmentant les capacités de production. Par ailleurs, le contexte environnemental actuel pousse les producteurs à chercher des techniques et des matériaux durables. En effet, en plus d'être plus économique, cela peut être un avantage majeur face à la concurrence puisque la clientèle se tourne, elle aussi, vers les solutions dites "vertes" ou du moins durables. Il est donc intéressant de noter que ces nouvelles méthodes permettent à l'industrie de la construction, reconnue comme étant très polluante, de faire un pas vers le développement durable.

Comme expliqué précédemment, le secteur de la construction manque encore d'efficacité pour répondre à la demande. Même si les MMC permettent de meilleures performances, l'IH fait encore face à de nombreux problèmes. Dans cette étude, il sera donc question de caractériser ces problèmes et de proposer des solutions ad hoc. D'abord, ce papier a pour objectif de répondre aux questions suivantes :

- Quelle valeur présente les MMC face à la construction traditionnelle et dans quelle mesure les MMC peuvent s'apparenter au secteur manufacturier ?
- Quels sont les problèmes majeurs que l'on peut rencontrer dans les processus de production du secteur de l'IH et comment peut-on les solutionner ?

L'article vise par ailleurs à mieux comprendre comment appliquer les bonnes pratiques dans le monde réel.

¹ Pour alléger le texte, le sigle anglais MMC sera utilisé.

² Pour alléger le texte, le sigle anglais IH sera utilisé.

Pour répondre à ces objectifs, la recherche a d'abord nécessité la réalisation d'une revue systématique de la littérature afin de relever les différents problèmes rencontrés dans le secteur de la construction industrialisée et leurs éventuelles solutions. La revue a ensuite été suivie de l'analyse d'un cas réel d'entreprise, œuvrant dans le secteur de la fabrication d'éléments préfabriqués au Québec. Une telle analyse a permis de mettre en lumière des problèmes majeurs rencontrés par le joueur industriel. Notamment, certains auteurs ont souligné que la transformation du secteur de la construction s'accompagne d'une transformation du besoin en main d'œuvre. En effet, pour l'IH, les opérateurs peuvent être moins nombreux et posséder un portfolio de compétences moins exhaustif. L'étude de cas a quant à elle montré que le problème du manque de main d'œuvre s'avère une dure réalité qui ne cesse de prendre de l'ampleur. Ce papier contribue donc à mieux faire connaître les différentes difficultés auxquelles peuvent faire face les entreprises de l'industrie du préfabriqué. Il permet également d'exposer un processus réel de production particulier, d'en relever les principaux problèmes et d'y proposer des solutions adéquates.

Ce papier est divisé de la façon suivante : La revue de la littérature conduite durant le projet est tout d'abord explicitée. L'étude de cas menée est par la suite introduite. L'article se termine par une brève conclusion.

2 REVUE DE LA LITTÉRATURE : MÉTHODOLOGIE

Comme l'explique (Tranfield *et al.*, 2003), une revue de littérature classique est souvent critiquée dans le monde scientifique. En effet, son manque de rigueur et la subjectivité du ou des chercheurs qui la mènent sont souvent pointés du doigt puisqu'il peut s'agir d'une source de biais non négligeable. Une revue systématique de la littérature apporte une solution à ce problème puisque c'est un processus défini comme reproductible, scientifique et transparent. Cette section propose donc une revue systématique de la littérature portant sur la gestion de la production dans le secteur de la construction industrialisée. Les trois principales étapes suivies sont explicitées dans les prochains paragraphes.

2.1 Planifier la revue

Le but de cette étape est de bien définir le sujet. Il faut donc poser la problématique avec une ou plusieurs questions de recherche. Il faut aussi définir une stratégie de recherche basée sur des critères d'inclusion et d'exclusion. Dans le cas présent, le sujet a été restreint aux avantages/inconvénients des activités dites hors-site ainsi qu'à la gestion de la production de la construction préfabriquée. Ainsi les études liées davantage au génie civil ont été écartées. Quatre questions de recherche ont ensuite été définies :

Q1 : Quelle valeur présente les méthodes modernes de construction face à la construction traditionnelle ?

Q2 : Quels sont les points de différenciation entre le secteur de l'IH et le secteur manufacturier ?

Q3 : Quels sont les problèmes majeurs que rencontre l'industrie du préfabriqué ? Et s'il y a lieu, quelles sont les solutions proposées ?

Q4 : Quels outils et équipements sont ou peuvent être mis en place pour supporter la production de structures préfabriquées ?

Le choix des articles via les mots-clés et l'analyse des papiers obtenus après tri ont permis de faire la lumière sur ces questions.

2.2 Conduire la revue

Il s'agit à présent de trouver des mots-clés et des termes de recherche en adéquation avec la problématique et la stratégie de recherche définies lors de l'étape précédente. Ainsi doivent ressortir uniquement les études correspondant aux critères inclusifs et n'étant pas concernées par les critères exclusifs. Ces études doivent pouvoir aider à trouver une réponse aux questions de la problématique.

En procédant par essais, de nombreux mots clés et combinaisons de mots clés ont été testés dans des bases de données spécifiques au génie industriel. Plus particulièrement, les recherches ont été faites via le moteur de recherche *Engineering Village* avec les bases de données *Compendex* et *Inspec*, puis via la base de données *Web of Science*. Ces mots clés ont été cherchés dans le titre, le résumé ou le sujet référencé de chaque étude. La littérature scientifique consultée étant principalement écrite en anglais, il fut donc nécessaire d'utiliser des mots clés en anglais. Les termes de recherche retenus sont résumés dans le Tableau 1. Chaque ensemble de termes de recherche (un par ligne du tableau 1) est concaténé par l'opérateur logique « ET ». D'autres mots-clés ont été testés puis retirés, puisqu'ils ne permettaient pas de cibler assez les articles intéressants et/ou d'éliminer les articles hors-sujet. Les recherches ont débuté au 1er mai 2017. Afin de poursuivre les recherches basées sur une banque d'articles précis et des informations déterminées, il a fallu définir une date de fin des recherches. Cette date a été fixée au 5 septembre 2017. Tous les articles de la revue de littérature font donc partie des résultats lorsque les mots clés ont été soumis à la base de données entre ces deux dates.

Mots clés	Thème ciblé
"Modern Methods of Construction" OU "Industrialized Housing" OU "Modular Building Systems"	Construction préfabriquée
Manufactur* OU Production	Production
Process* OU (Factory OU Factories OU Facility OU Facilities OU Shop-Floor* OU Work-Floor* OU Plant*)	Études centrées sur le processus ou les opérations hors-site/en usine
Manag* OU Control*	Gestion des opérations

Tableau 1 - Tableau récapitulatif des termes de recherche

Ainsi lorsque la série de termes de recherche a été soumise à *Engineering village*, 23 articles différents (28 avant retrait des doublons) sont ressortis. Sur la base des

résumés (et des introductions et conclusions en cas de doute), les articles ont été filtrés. Tout article hors-sujet a été retiré de la liste. Les articles traitant de design ou les articles dans une langue autre que l'anglais ou le français n'ont pas été traités. Il en résulte la série de 13 articles présentée dans le Tableau 2. La même série de termes de recherche a été soumise à la base de données *Web of Science*. Par souci de cohérence, la recherche a été limitée aux index *Science Citation Index Expanded (SCI-EXPANDED)*, *Conference Proceedings Citation Index-Science (CPCI-S)* et *Emerging Sources Citation Index (ESCI)*, les autres index étant consacrés aux sciences humaines et sociales. En suivant le même filtrage que celui décrit précédemment, seul 2 articles ont été retenus. Il s'agit d'articles déjà obtenus via Engineering Village.

Référence	Titre
(Hook et Stehn, 2008)	<i>Applicability of lean principles and practices in industrialized housing production</i>
(Branson et al., 1990)	<i>Automation technologies for the industrialized housing manufacturing industry</i>
(Broadway et Mullens, 2004)	<i>Shop-floor information systems for industrialized housing production</i>
(Kozlovska et Spiakova, 2011)	<i>Modern methods of construction vs. construction waste</i>
(Sarden et Engstrom, 2010)	<i>Modern methods of construction: A solution for an industry characterized by uncertainty?</i>
(Hairstans, 2010)	<i>Timber offsite Modern Methods of Construction</i>
(Johnsson et Meiling, 2009)	<i>Defects in offsite construction: Timber module prefabrication</i>
(Pan et al., 2008)	<i>Leading UK housebuilders' utilization of offsite construction methods</i>
(Bildsten et al., 2010)	<i>Applying the kraljic model to the construction sector: The case of a prefab housing factory</i>
(Noguchi, 2005)	<i>Japanese manufacturers' 'cost-performance' marketing strategy for the delivery of solar photovoltaic homes</i>
(Taylor et al., 2009)	<i>A comparison of modern methods of bathroom construction: A project case study</i>
(Nahmens et Bindroo, 2011)	<i>Is customization fruitful in industrialized homebuilding industry?</i>
(Tennant et al., 2012)	<i>Re-engineering the construction supply chain : Transferring on-site activity, offsite</i>

Tableau 2 - Liste des articles retenus pour la revue

2.3 Analyse descriptive des résultats

Afin d'avoir une meilleure compréhension des études de cette revue, il est important de savoir d'où elles proviennent et à quelle période elles ont été réalisées. Les paramètres géographiques permettent d'identifier les pôles de compétence et où se situent les besoins. En effet, les études répondent souvent à des besoins réels d'un pays ou d'une région donnée tout en s'appuyant sur des données concrètes récoltées sur le terrain. Il faut donc un secteur au minimum naissant pour avoir des données fiables. On observe sur la figure 1 que l'ensemble des études proviennent de pays riches de la triade, soit des

pays ayant une économie fortement développée. La forte demande en logements présente dans les pays industrialisés couplée avec les normes sanitaires et de sécurité élevées engendre un besoin de logements de qualité en grande quantité. Il est donc logique que ce soit des pays industrialisés qui cherchent à augmenter leur efficacité dans le secteur de la construction pour ainsi créer plus de logements en adéquation avec la demande. L'année de publication d'un papier donne une bonne indication sur la période à laquelle l'étude a été réalisée. Il est intéressant de constater ici (cf. figure 1) que la plupart des études ont été menées au cours des dix dernières années. Nous pouvons donc en déduire qu'il s'agit d'un domaine de recherche récent, ce qui peut expliquer le manque d'abondance d'études couvrant la discipline. Toutefois, leur répartition dans le temps n'est pas homogène. Cela peut s'expliquer par le caractère très appliqué du domaine. En effet, les entreprises mettent en place des actions pour évoluer, mais de telles mesures ne sont pas toujours publiées dans la littérature scientifique ni même professionnelle.

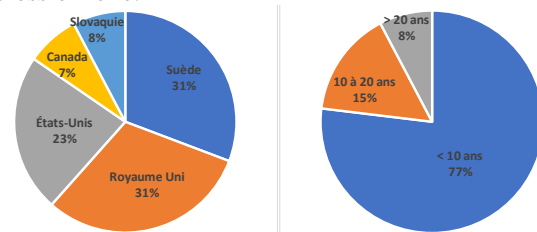


Figure 1 – Répartition des articles de la revue selon leur provenance et selon leur année de publication

3 REVUE DE LA LITTÉRATURE : ANALYSE THÉMATIQUE

Dans cette partie, il sera question d'utiliser les articles trouvés pour répondre aux questions de recherche.

3.1 Quelle valeur présente les méthodes modernes de construction face à la construction traditionnelle ?

Quasiment tous les articles explorés donnent des éléments de réponse à cette question. Comme montré dans l'analyse descriptive ci-dessus, les Méthodes Modernes de Construction (MMC) dont fait partie la construction industrialisée (IH – Industrialized Housing), constituent un sujet de recherche assez récent. Il est donc encore utile de démontrer l'intérêt que cela peut représenter pour les différentes parties prenantes. Certaines recherches (e.g., (Kozlovska et Spiakova, 2011)) sont même axées directement sur cet intérêt.

Tout d'abord, l'avancée majeure que permettent les MMC est la possibilité de standardiser une partie plus ou moins importante de la construction. D'un point de vue plus général, réaliser une partie des activités hors-site permet d'implanter un flot de production et de passer ainsi d'un fonctionnement en mode projet vers un fonctionnement en mode processus (Hairstans, 2010). (Sar-

den et Engstrom, 2010) d'ajouter que des processus bien définis permettent une coordination plus rationnelle et plus efficace entre les étapes de construction et entre les parties prenantes. (Branson *et al.*, 1990) soulignaient déjà une caractéristique importante pour permettre la standardisation, notamment via la technologie, soit la répétitivité des processus de production ramenés hors-site. La standardisation a de nombreux avantages. De manière générale, elle permet un meilleur contrôle de la production (Nahmens et Bindroo, 2011; Sarden et Engstrom, 2010), une plus grande prédictibilité (Pan *et al.*, 2008; Sarden et Engstrom, 2010) et un gain en temps (Kozlovska et Spiakova, 2011; Pan *et al.*, 2008; Sarden et Engstrom, 2010). Enfin le fonctionnement par processus apporte un avantage majeur puisque cela concentre les responsabilités sur un responsable de processus qui apportera un meilleur contrôle au service du client (Sarden et Engstrom, 2010). Par ailleurs, (Noguchi, 2005) argue que les MMC rendent possible le développement des principes de production par valeur ajoutée et l'obtention de certifications industrielles multi-secteurs telles que ISO 9000. (Hook et Stehn, 2008) montrent dans une étude dédiée au sujet que les principes du *Lean Management* peuvent bien s'appliquer à l'IH.

Principale conséquence de la standardisation et du transfert des activités hors-site, les coûts afférents aux MMC sont moindres que ceux afférents aux méthodes traditionnelles. Pointée par (Taylor *et al.*, 2009), cette réduction est expliquée par des économies d'échelle par (Noguchi, 2005; Sarden et Engstrom, 2010). En effet, la centralisation et la standardisation d'une partie des activités hors-site permet de regrouper certains produits et donc d'avoir un meilleur contrôle des coûts. (Sarden et Engstrom, 2010) ajoutent que le retour sur investissement est plus rapide. Autre atout de cette standardisation, la qualité semble être accrue (Hairstans, 2010; Pan *et al.*, 2008; Taylor *et al.*, 2009) puisqu'il y a plus d'opportunités de contrôle, ce qui mène à une diminution des rebuts en fréquence et en gravité (Johnsson et Meiling, 2009). Enfin (Sarden et Engstrom, 2010) notent une uniformisation des normes qui sont ainsi moins contradictoires.

La transformation du travail qui est opérée dans ce transfert des opérations sur-site hors-site provoque une mutation des ressources nécessaires. Sur un plan humain, la main d'œuvre connaît une importante modification. Les MMC requièrent un moins grand éventail de compétences de la part des opérateurs (Kozlovska et Spiakova, 2011; Pan *et al.*, 2008; Taylor *et al.*, 2009). En effet, le mode processus permet de découper et de simplifier les tâches de production. (Pan *et al.*, 2008) expliquent aussi que les MMC nécessitent une plus faible quantité de main d'œuvre, permettant ainsi de pallier au manque de main d'œuvre auquel plusieurs pays font face. Les ressources et méthodes de travail étant appelées à changer, il est logique que les relations clients-fournisseurs soient aussi concernées par cette transformation. (Tennant *et al.*, 2012) expliquent que la spécialisation et le développement technique des activités de chaque partie prenante

augmente l'interdépendance commerciale, ce qui constitue un incitatif pour le développement de relations à plus long-terme observé aussi par (Bildsten *et al.*, 2010). (Sarden et Engstrom, 2010) soulignent des relations moins conflictuelles et le développement d'intérêts communs, menant à une meilleure coordination.

Enfin les MMC semblent être plus durables (Noguchi, 2005) que les méthodes traditionnelles, un avantage important qui prend de plus en plus de sens de nos jours. Elles ont ainsi un impact plus modéré sur l'environnement (Pan *et al.*, 2008). (Kozlovska et Spiakova, 2011) y ont consacré une étude. Les auteurs ont pu constater que, grâce à un meilleur contrôle des quantités et à de plus bas risques matériels, les déchets de construction sont diminués de moitié par rapport à la construction traditionnelle.

3.2 Quels sont les points de différenciation entre le secteur de l'IH et le secteur manufacturier ?

L'une des motivations des MMC est de faire connaître au monde de la construction le même essor qu'a pu connaître le secteur manufacturier à la fin du XXe siècle. Ceci est bien expliqué par (Gann, 1996) dans une étude hors revue qui compare l'IH au secteur manufacturier de l'automobile au Japon et montre qu'il y a de nombreuses similarités. Il note cependant une différence fondamentale. Contrairement à la production manufacturière, l'assemblage final des produits de l'IH se font sur chantier et non en usine, donc dans un environnement moins contrôlé. De plus, le nombre de corps de métier impliqué est plus important dans l'IH que dans le secteur manufacturier. Ainsi, il est beaucoup plus difficile de contrôler le processus de production autant sur la manière de faire que sur les ressources humaines concernées. D'autres différences majeures sont relevées par d'autres auteurs. Tout d'abord, la construction présente une forte culture de projet. En effet, d'un côté le rapport entre les clients et les producteurs n'est pas le même. Dans l'IH, le client a tendance à être impliqué dès la phase initiale de design. D'un autre côté, les activités sur-site imposeront toujours cette culture projet (Hook et Stehn, 2008). Ces dernières constituent une différence majeure difficilement surmontable. De plus, chaque bâtiment ayant ses particularités, les produits de l'IH ont un plus fort degré de personnalisation, ce qui complique notamment les contrôles de qualité (Johnsson et Meiling, 2009). Par ailleurs, comme expliqué plus tôt, le secteur des MMC étant apparu plus récemment, il est donc normal de trouver des différences dans le développement des deux secteurs (Bildsten *et al.*, 2010; Johnsson et Meiling, 2009).

3.3 Quels sont les problèmes majeurs que rencontre l'industrie du préfabriqué ? Et s'il y a lieu, quelles sont les solutions proposées ?

Tout d'abord il faut noter que le changement de paradigme que représente la transformation de l'industrie de la construction peut en soi être source de problèmes.

(Hook et Stehn, 2008) pointent des éléments qui sont inhérents à la jeunesse du secteur et à traiter en priorité. Il s'agit de la forte culture de projet qui est un frein considérable, le manque de processus intégrés d'assurance qualité et le manque de programmes d'amélioration continue. (Pan *et al.*, 2008) notent la culture conservatrice des entreprises qui ont des difficultés à se défaire de la culture de projet ou encore le coût d'entrée sur le marché assez élevé. En effet, le secteur étant très fragmenté, il y a beaucoup de petits entrepreneurs qui ne parviennent pas à réaliser l'investissement nécessaire à la mise en place d'infrastructures pour pouvoir entrer en concurrence avec d'autres entreprises déjà présentes sur le marché de l'IH. (Sarden et Engstrom, 2010) expliquent que même si les MMC réduisent les incertitudes de délais, de qualité et de coûts dans les processus de production, de nouvelles incertitudes apparaissent. Elles sont principalement dues à la nouveauté des méthodes. En effet, d'un côté les clients ont moins de choix dans le design et moins de possibilités de modification et de l'autre, les constructeurs font face à de nouveaux processus, de nouveaux besoins, de nouvelles pratiques et des prises de décisions plus précoces. L'étude expose que ces incertitudes s'estomperont simplement avec le temps.

Il semble aussi y avoir un manque de documentation, notamment au niveau des processus de prise de décisions (Taylor *et al.*, 2009). (Pan *et al.*, 2008) relèvent un manque d'information et de conseil pour les entrepreneurs. Ils préconisent de développer l'apprentissage organisationnel et d'investir dans la formation des gestionnaires. (Johnsson et Meiling, 2009) montrent qu'au sein même des entreprises, les informations disponibles sont peu, voire pas, exploitées. Elles devraient l'être afin de développer l'apprentissage organisationnel au sein de chaque entreprise qui pourrait ensuite propager à l'ensemble de la chaîne logistique. Les auteurs insistent d'ailleurs sur l'importance d'investir davantage dans cette dernière pour que le rapport de force soit remplacé par des relations de confiance et que les risques soient partagés entre les différents acteurs (Tennant *et al.*, 2012).

Le contrôle de la qualité étant complexe à mettre en place dans le cadre de la construction traditionnelle, il est souvent oublié et ce défaut perdure dans la construction préfabriquée. Cela pourrait être aisément corrigé puisque (Johnsson et Meiling, 2009) estiment que la majorité des rebuts sont des rebuts d'usine. Ils préconisent aussi la mise en place de processus de traitement des défauts afin de pallier à ce problème, l'idée étant d'apprendre des erreurs commises et de créer ainsi une démarche d'amélioration continue. (Hairstans, 2010) conseille aux constructeurs de se conformer aux standards et aux certifications internationales. Cela aura aussi la conséquence de favoriser les bonnes pratiques. Par ailleurs, la capacité de production et la productivité des usines peuvent s'avérer limitées. C'est entre autres dû à des équipements et procédures peu performants. Ce problème peut être résolu en investissant dans des technologies de pointe, ce

qui pourra alors contribuer à diminuer les coûts de main d'œuvre et à améliorer la qualité (Branson *et al.*, 1990).

3.4 Quels outils et équipements sont ou peuvent être mis en place pour supporter la production de structures préfabriquées ?

Dans le cadre de la standardisation du travail et de la maintenance, les principes du *Lean Manufacturing* s'appliquent particulièrement bien, notamment dans un contexte de personnalisation de masse (Nahmens et Bindroo, 2011). Dans cette même vision, (Hook et Stehn, 2008) notent un point important qui est la responsabilisation des équipes de production. Cela permettra d'avoir des équipes plus motivées qui travailleront plus efficacement et de manière plus proactive.

Même si l'étude a plus de 25 ans aujourd'hui, (Branson *et al.*, 1990) recensent les différentes technologies qui peuvent être mises en place dans les entreprises œuvrant dans la IH. Il y a notamment les technologies de l'information constituent une grande source d'outils techniques et de gestion qui facilitent l'intégration de certaines fonctions. Ensuite il existe de nombreuses technologies d'automatisation comme des machines de coupe, des outils d'assemblage ou des appareils de maintenance. Enfin, il est possible de faire appel à la robotique qui a fait ses preuves dans d'autres domaines. Cependant, la préfabrication étant un mode de construction particulier, les possibilités d'équipements techniques sont restreintes.

4 ANALYSE D'UN CAS D'ÉTUDES

Dans cette section, l'analyse d'une entreprise est proposée afin d'explorer comment les bonnes pratiques peuvent être appliquées dans le monde réel.

4.1 Présentation du cas

L'entreprise retenue pour l'étude de cas œuvre dans le secteur de la préfabrication. Elle se veut un acteur majeur du secteur de l'IH au Québec et elle est spécialisée dans la production d'éléments préfabriqués en bois tels que des panneaux de mur, des fermes de toit et des poutrelles de plancher. Elle contrôle les étapes d'achat du bois d'œuvre, de transformation et de livraison des éléments sur les chantiers. Elle prend aussi en charge la production de plans pour les clients mais n'est pas responsable de l'assemblage des structures sur le chantier. L'étude réalisée a plus particulièrement été menée sur le processus d'assemblage des panneaux structuraux de mur. En effet, l'entreprise éprouvant une grande difficulté à recruter de la main d'œuvre dans la région où elle exerce ses activités, elle désire se pencher sur un processus pour lequel la quantité de main d'œuvre disponible limite le volume de production pouvant être produit. L'entreprise ne souhaitait cependant pas réaliser des changements majeurs, n'ayant ni les moyens financiers, ni le désir de transformer radicalement les méthodes de

travail. L'analyse devait donc mener à des solutions permettant d'augmenter la capacité de production tout en modifiant partiellement certaines étapes de production.

4.2 Présentation des processus

Afin de mieux visualiser les opérations liées au processus d'assemblage de panneaux de mur structuraux, un diagramme de processus a été tracé (cf. annexe). Il s'agit d'un processus d'assemblage classique pour ce type de produit. L'atelier d'assemblage des murs reçoit l'ensemble des pièces de bois composant le mur de la part de l'atelier de coupe (0), sauf les lisses et le revêtement. Dans un premier temps, il y a trois postes de préparation. La coupe et le marquage des lisses (1), la coupe du revêtement (2) et le pré-assemblage des éléments de cadre d'ouvertures et montants (3). Les lisses sont les pièces constituant les parties supérieur et inférieur du cadre du mur. Le marquage permet de marquer sur le bois la position des goujons, pièces perpendiculaires aux lisses. Dans un second temps a lieu l'assemblage en lui-même. À partir de l'assemblage de la structure (5), il y a deux lignes de production distinctes. Les tables pivotent pour pouvoir relever un panneau contre un plan vertical puis passer à l'étape d'isolation (8). Le panneau est ensuite couché dans l'autre sens sur une autre table pivotante. Dans certains cas, il est nécessaire d'effectuer des finitions (10) qui n'ont pu être faites à la verticale. En dernier lieu, une inspection (10) du mur est faite. Sur la gauche du diagramme, on voit apparaître la main d'œuvre nécessaire à la conduite des tâches. Les tâches 1 et 2 par exemple sont réalisées par un seul opérateur.

4.3 Problèmes observés

Plusieurs visites sur site ont permis de relever des problèmes majeurs auxquels l'entreprise fait face. Tout d'abord, l'atelier d'assemblage des murs n'est pas automatisé. En effet, aucune machine ne permet de réaliser une tâche, partiellement ou complètement, de manière automatique. Ensuite, la main d'œuvre est manquante et dans certains cas, elle est peu responsabilisée, notamment au niveau de la qualité. Aucun mode opératoire ni aucune procédure écrite sur les méthodes de travail n'existent. De tels documents sont pourtant essentiels pour endiguer les mauvaises habitudes ou les déviations à la conformité des produits. Par ailleurs, la main d'œuvre n'a pas accès à ses performances de travail via des indicateurs visibles sur le plancher. Il y a donc un manque de documentation et d'utilisation des données disponibles. De plus, les accidents ou incidents sont fréquents. Au niveau de l'approvisionnement, les relations clients-fournisseurs sont peu développées. Également, il n'y a aucun outil de gestion centralisée de l'inventaire. Enfin, l'ordonnancement de la production est fait suivant la valeur des produits et leur date de livraison. C'est une méthode biaisée car la valeur n'est pas indexée uniquement sur le temps de travail de la main d'œuvre. De plus, au sein de l'atelier de production, dès qu'une table d'assemblage est disponible,

le prochain panneau dans la liste de livraison est attribué à cette même table, peu importe le produit considéré.

4.4 Suggestions d'améliorations

Après une étude en détail du processus d'assemblage des panneaux de mur structuraux et de la demande, un ensemble de six améliorations ont pu être proposées. Elles peuvent toutes être appliquées seules, mais rien n'empêche d'en combiner quelques-unes pour obtenir de meilleurs résultats. Ces améliorations ne visent que le processus de production en lui-même et ont pour but principal d'augmenter la capacité de production. Tous les problèmes cités ci-dessus ne seront donc pas tous traités.

4.4.1 Description et impact

Pour chaque amélioration, l'idée générale et les impacts qu'elles peuvent avoir sur le processus sont ici détaillés.

Famille de produits : La demande actuelle n'est pas fixe dans le temps. En effet, selon la période de l'année, ce n'est pas le même marché qui est visé. Il y a donc une alternance entre l'été où des murs isolés sont produits et l'hiver où des murs creux sont plutôt produits. Il pourrait donc être intéressant lors des périodes transitoires d'opter pour deux lignes de production distinctes, l'une pour murs isolés, l'autre pour murs creux. L'entreprise possède déjà ces deux lignes, il suffirait de les distinguer pour permettre de garder des flux tendus lors de ces périodes transitoires et d'être ainsi plus efficace.

Guides lasers : Cette solution consiste à poser des viseurs lasers à la verticale du poste d'assemblage structure (5) pour faciliter la pose des différentes pièces et réduire la nécessité de plans papiers. Cet outil permet d'indiquer sur la table d'assemblage, à l'aide de pointeurs, les positions des pièces à assembler. Plusieurs viseurs sont toutefois nécessaires pour éviter les erreurs de lecture liées à l'incidence des faisceaux lasers. Cela permettrait aux opérateurs de travailler plus rapidement et le risque d'erreur à ce poste serait plus faible. De plus, le marquage des lisses deviendrait inutile, ce qui réduirait du même coup le temps de travail au poste de préparation des lisses (1). Ce poste pourrait même être supprimé si la coupe des lisses était confiée à l'atelier de coupe.

Écrans de guidage : Cette amélioration consiste à poser des écrans en hauteur pour permettre aux employés de visualiser le travail à effectuer étape par étape, sans avoir recours à des plans papiers aux étapes de pré-assemblage des ouvertures (3) et d'assemblage de la structure (5). Cette amélioration permettrait un gain de temps et une baisse du risque d'erreur.

Table de pré-assemblage des ouvertures : À l'heure actuelle, seuls les montants et linteaux des ouvertures sont préparés séparément sur un établi. L'installation d'une table de pré-assemblage permettrait de préparer des ouvertures de manière plus complète, c'est-à-dire des cadres complets. Cette solution représente un transfert de

la charge de travail d'un poste potentiellement goulot, l'assemblage de la structure (5), à un autre, le pré-assemblage des ouvertures (3), ce qui équilibrerait mieux la production.

Poste de coupe d'entremises : Les entremises sont coupées au fur et à mesure qu'elles sont posées. La création d'un poste de préparation supplémentaire permettrait de gagner en efficacité sur le poste d'assemblage de la structure (5) en y réduisant la charge de travail. Cette modification a pour défaut de nécessité de la main d'œuvre supplémentaire. Il serait donc intéressant de la coupler avec une autre amélioration.

Station d'assemblage : Il s'agit ici d'automatiser partiellement le poste d'assemblage de la structure (5). Un automate permettrait de réaliser à partir des plans l'assemblage de la plupart des structures. À l'aide de pinces et de cloueuses, l'automate assemble les pièces de bois positionnées par un opérateur. Il existe plusieurs modèles sur le marché. Certains plus complexes sont aussi équipés de butés pour positionner les pièces et d'écrans de contrôle pour aider l'opérateur dans sa tâche. Cela apporterait un gain de temps tout en diminuant la quantité de main d'œuvre nécessaire et le niveau de difficulté du travail pour cette dernière. Une augmentation de la qualité serait donc probablement observée. Toutefois cette amélioration peut nécessiter, selon les spécifications de la machine, d'augmenter le niveau de pré-assemblage des ouvertures (portes et fenêtres).

4.4.2 Mise en place

Une modification du processus de production n'est pas anodine. Il y a bien souvent un investissement à réaliser et une complexité de mise en œuvre. Cette complexité peut venir de différents points, de la main d'œuvre, de l'aménagement actuel, des habitudes plus ou moins ancrées, etc., mais aussi tout simplement de l'ampleur de la modification en tant que tel. Le tableau ci-après présente, pour chaque amélioration, l'investissement et la complexité de mise en œuvre attendus. Il est toutefois important de rappeler que chaque modification nécessite de former le personnel impliqué. Cela représente un investissement supplémentaire à considérer.

Amélioration	Investissement	Complexité
Famille de produits	0	Faible
Guides lasers	+ (Viseurs lasers)	Faible
Écrans de guidage	+ (Ecrans)	Faible
Table de pré-assemblage des ouvertures	+ (Table)	Moyenne
Poste de coupe d'entremises	++ (Personnel supplémentaire ou à détacher d'ailleurs)	Moyenne
Station d'assemblage	+++ (Station d'assemblage & logiciel)	Haute

Tableau 3 – Évaluation de l'investissement et de la complexité de chaque amélioration

5 CONCLUSION

Cet article a voulu faire un état des lieux de la gestion de la production dans la construction industrialisée tout en s'intéressant aux façons d'améliorer la capacité de production des entreprises du secteur. Une revue de la littérature a donc permis de présenter le secteur de l'IH en le comparant à la construction traditionnelle et au secteur manufacturier. Fort est de constater que les MMC manquent encore de développement à l'heure actuelle. En effet, les problèmes majeurs soulignés par les auteurs sont liés au manque de documentation et de communication en interne comme au sein de la chaîne logistique, à un trop faible développement des processus de contrôle de la production et à un manque d'investissement dans les technologies de fabrication. Par ailleurs, l'étude de cas a permis d'observer un processus de production du secteur de l'IH, de relever la problématique majeure à laquelle fait face l'entreprise en question et d'y répondre en amenant des améliorations potentielles au processus de production. Ces améliorations peuvent prendre plusieurs formes : ajout d'équipement (guides lasers, écrans de guidage, pont de clouage, station d'assemblage), réaménagement des lieux de travail (table de pré-assemblage des ouvertures, poste de coupe d'entremises) ou réordonnancement des tâches (Famille de produits).

Après une première étape de modélisation d'un processus de production de structures préfabriquées, cette étude se poursuivra en utilisant la simulation afin d'estimer de manière quantitative l'impact réel de ces suggestions ainsi que l'investissement que cela représente. Le but sera alors d'optimiser ce processus de production.

RÉFÉRENCES

- Bildsten, L., Rehme, J., and Brege, S., 2010, Applying the kraljic model to the construction sector: The case of a prefab housing factory. *26th Annual Conference of the Association of Researchers in Construction Management (ARCOM 2010)*, p. 1029–1037.
- Branson, T.R., Elshennawy, A.K., Swart, W.W., and Chandra, S., 1990, Automation technologies for the industrialized housing manufacturing industry. *12th Annual Conference on Computers and Industrial Engineering*, Orlando, FL, United States, vol.19, p. 587–592.
- Broadway, R.S., and Mullens, M.A., 2004. Shop-floor information systems for industrialized housing production. *IIE Annual Conference and Exhibition 2004*, p. 2203–2208.
- Egebo, T., Richardson, P., and Lienert, I., 1990. Modèles de l'investissement résidentiel pour les grands pays de l'OCDE. *Revue économique de l'OCDE*, 14, p. 165–207.
- Gann, D.M., 1996. Construction as a manufacturing process ? Similarities and differences between industrialized housing and car production in Japan. *Construction Management and Economics*, 14, p. 437–450.

Hairstans, R., 2010. Timber offsite Modern Methods of Construction. *11th World Conference on Timber Engineering 2010 (WCTE 2010)*, p. 2006–2010.

Hook, M., and Stehn, L., 2008. Applicability of lean principles and practices in industrialized housing production. *Construction Management and Economics*, 26, p. 1091–1100.

Johnsson, H., and Meiling, J.H., 2009. Defects in offsite construction: Timber module prefabrication. *Construction Management and Economics*, 27, p. 667–681.

Kozlovska, M., and Spiakova, M., 2011. Modern methods of construction vs. construction waste. 11th International Multidisciplinary Scientific Geoconference and EXPO (SGEM 2011), p. 483–490.

Nahmens, I., and Bindroo, V., 2011. Is customization fruitful in industrialized homebuilding industry? *Journal of Construction Engineering and Management*, 137, p. 1027–1035.

Noguchi, M. (2005). Japanese manufacturers' "cost-performance" marketing strategy for the delivery of solar photovoltaic homes. In *Solar World Congress 2005: Bringing Water to the World, Including 34th ASES Annual Conference and 30th National Passive Solar Conference*, p. 2341–2348.

Pan, W., Gibb, A.G.F., and Dainty, A.R.J., 2008. Leading UK housebuilders' utilization of offsite construction methods. *Building Research and Information*, 36, p. 56–67.

Sarden, Y., and Engstrom, S., 2010. Modern methods of construction: A solution for an industry characterized by uncertainty? *26th Annual Conference of the Association of Researchers in Construction Management (ARCOM 2010)*, p. 1101–1110.

Taylor, M.D., Fisher, A., and Wamuziri, S.C., 2009. A comparison of modern methods of bathroom construction: A project case study. *25th Annual Conference of the Association of Researchers in Construction Management (ARCOM 2009)*, p. 1173–1182.

Tennant, S., McCarney, M., and Tong, M.K.L., 2012. Re-engineering the construction supply chain: Transferring on-site activity, offsite. *28th Annual Conference of the Association of Researchers in Construction Management (ARCOM 2012)*, p. 739–749.

Tranfield, D., Denyer, D., and Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14, p. 207–222.

La base de données de l'OCDE sur le logement abordable - OCDE.

ANNEXES

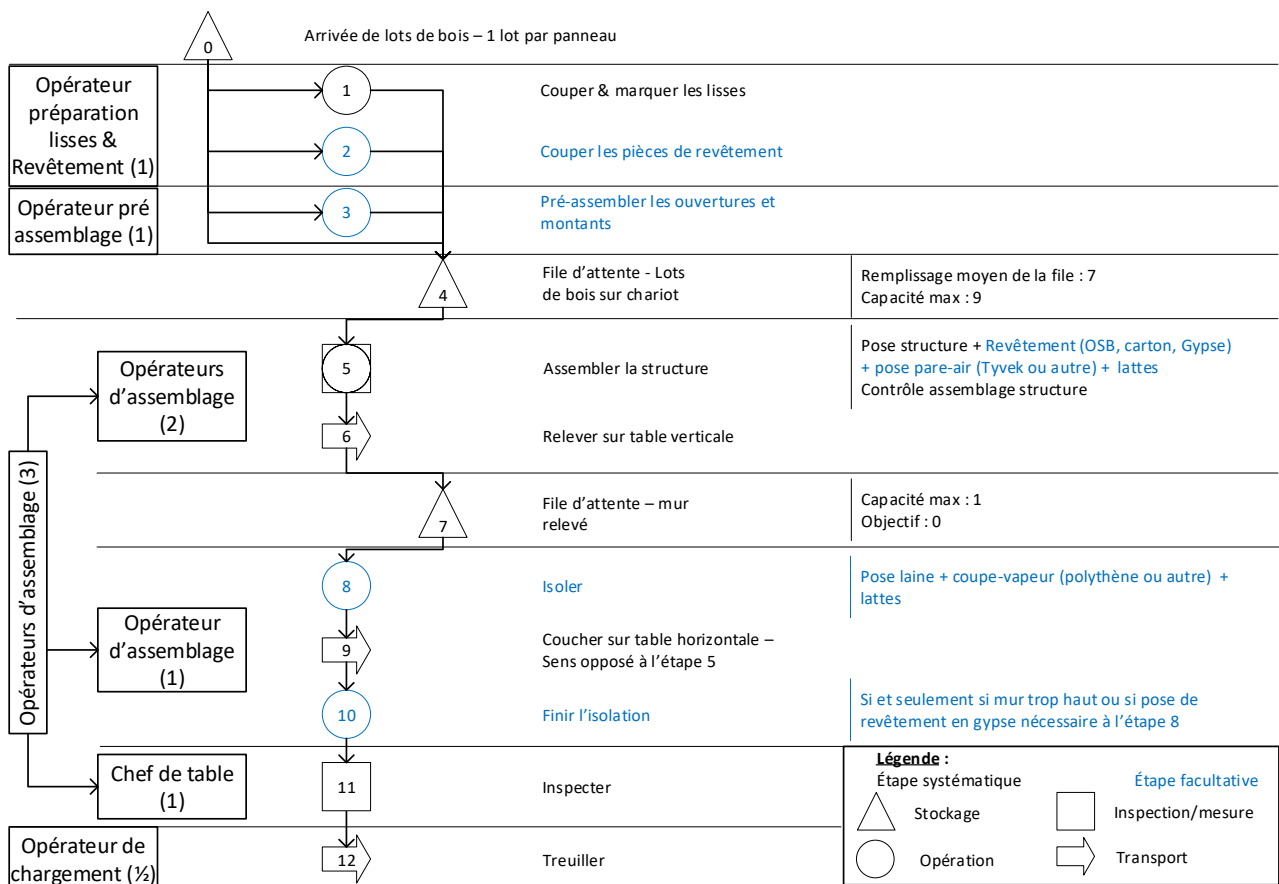


Figure 2 - Diagramme de processus de production – Atelier Murs

PROPOSITION D'UNE MODÉLISATION POUR L'AMÉLIORATION DES PERFORMANCES DU SECTEUR PUBLIC

Gautier AUBOURG

Toulouse Métropole – 31000
TOULOUSE Cedex
gautier.aubourg@toulouse-
metropole.fr

**François GALASSO, Jacques
LAMOTHE**

Université Fédérale de Toulouse,
IMT Mines Albi – 81000 ALBI
francois.galasso@mines-albi.fr,
jacques.lamothe@mines-albi.fr

Bernard GRABOT

Université Fédérale de Toulouse,
LGP/INP-ENIT – 65000 TARBES
bernard@enit.fr

RESUME : *Le secteur public, comme toute organisation, cherche à améliorer sa performance tout en conservant ou améliorant le service rendu aux usagers et citoyens. Dans une logique d'amélioration continue, les méthodes existantes issues du progrès permanent dans l'industrie peuvent être adaptées afin d'intégrer les spécificités du secteur public. Cet article présente un cadre de modélisation pour les services publics incluant la mesure de performance et l'intégration d'actions d'amélioration dans un souci d'agilité. Il montre la nécessité d'intégrer les composantes de cet environnement afin de développer des modèles de référence utiles à la modélisation globale. La cohérence de ces modèles est présentée sur un cas d'étude basé sur l'expérimentation d'une application visant, entre autres, à informer les usagers des menus proposés dans les cantines scolaires approvisionnées par la cuisine centrale de Toulouse Métropole.*

MOTS-CLES : *secteur public, modèle de référence, progrès permanent, agilité.*

1 INTRODUCTION

L'utilisation de méthodes issues de pratiques du secteur privé dans le secteur public n'est pas nouvelle. Les pratiques liées au Lean Management, englobant de multiples outils de progrès permanent et d'amélioration continue constituent l'essentiel de ces méthodes. Ainsi, de multiples tentatives ont été réalisées, dans les préfectures, au sein des collectivités territoriales (Clainche, 2008), dans le secteur hospitalier (Osorio Montoya, 2015) ou à l'échelle d'un pays (Lungisa, 2015). Ce dernier mettant en évidence que des méthodes liées au Supply Chain Management peuvent compléter l'arsenal du Lean Management. Les bénéfices attendus de telles méthodes sont liés à des gains d'efficacité visant l'amélioration des services rendus tout en maîtrisant les dépenses. Les résultats des applications de telles méthodes restent inégaux. Un des points clés pour l'obtention de bons résultats réside dans la capacité à impliquer les différents acteurs. Ces derniers devant s'approprier les méthodes et les contextualiser pour répondre aux spécificités du secteur public. Le contexte même du secteur public est en évolution permanente. Les lois sont en évolution, les engagements politiques donnent des axes stratégiques différents et dans chaque cas, les évolutions technologiques et méthodologiques viennent remettre en question les paradigmes existants. Au cœur de ces évolutions, et plus particulièrement dans le périmètre de cette étude qui est celui d'une collectivité territoriale, les concepts de Smart City et de Sustainable City sont porteurs de nouvelles opportunités.

L'objet de cet article est de s'appuyer sur une méthodologie assortie d'un cadre de modélisation pour soutenir la mutation du secteur public. Plus précisément, il s'agit de définir un modèle de référence afin d'inscrire chaque démarche dans un cadre global, adapté aux spécificités du secteur public en offrant un point d'accès vers les méthodologies issues du secteur privé. Cet article s'inscrit dans le cadre de travaux effectués au sein de Toulouse Métropole. Cette Métropole souhaite améliorer la performance de ses processus internes tout en améliorant le service rendu aux citoyens et usagers. Au-delà des méthodes industrielles usuelles centrée sur la performance des processus, elle a adopté une démarche globale intitulée « Toulouse, l'Open Métropole » visant à expérimenter l'usage de techniques issues des concepts Smart City développés ultérieurement dans cet article.

Ainsi, cet article se décompose de la façon suivante : en deuxième partie un bref état de la littérature sur l'amélioration des performances dans les secteurs privés et publics est donné et aboutit à la problématique de l'étude ; en troisième partie, un focus sur les méthodes de modélisation des processus est réalisé et introduit les compléments de modélisations nécessaires à l'adaptation des méthodes issues du secteur industriel inclus dans un cadre de modélisation ; la quatrième partie présente une application des concepts évoqués. Elle montre comment les modélisations proposées peuvent servir de support au positionnement d'actions d'améliorations des processus dans le secteur public et de leur performance d'un point de vue des citoyens.

2 ÉTAT DE L'ART SUR L'AMÉLIORATION DES PERFORMANCES

Le secteur public, à l'instar du secteur privé, est en recherche d'améliorations de performance. Cette performance s'entend d'une part, du point de vue des usagers des services et d'autre part, du point de vue du fonctionnement interne qui doit garantir un même niveau de performance tout en mobilisant moins de ressources. C'est particulièrement le cas au sein de métropoles qui doivent prendre en compte les évolutions d'une population hétérogène. Meijer et Bolívar (2015) mettent en avant l'importance du développement d'une gouvernance centrée sur l'usager et prenant en compte l'ensemble des parties prenantes. Les parties prenantes d'une métropole viennent de différents horizons avec des intérêts souvent divergents :

- société civile : patients, citoyens ou usagers,
- organisations publiques : collectivités territoriales, administrations et partenaires,
- entreprise ou acteurs privés : les industriels, les transports, les réseaux télécoms et infrastructures, les fournisseurs et les éditeurs logiciels...

Les autorités locales doivent trouver leur rôle de coordinateur au sein d'une démarche Smart City pour rendre l'environnement propice à son propre développement (Wall *et al.*, 2015).

2.1 Diverses tentatives

Le progrès permanent repose sur un système participatif qui a attiré les organismes de l'État impliqués dans divers projets. Mais, l'intégration du progrès permanent n'a souvent pas survécu à la phase de test, les principes d'évaluation du rendement toujours associés à ces méthodes ayant été mal perçus par les agents (Suarez Barraza *et al.*, 2009). Ces résistances sont, à notre avis, principalement dues à la transposition maladroite de méthodes du privé, sans un réel travail d'adaptation. Comme toute méthode, le progrès permanent nécessite une bonne compréhension de son contenu et de ses apports afin de susciter l'intérêt des hiérarchies. De plus, son adaptation doit prendre en compte le contexte particulier du secteur public et s'intégrer à ses processus spécifiques. Fin 2017, cette problématique est mise en exergue par la création de structures au niveau national français qui sont la Direction Interministérielle de la Transformation Publique (DITP) et à la Direction Interministérielle du Numérique et du Système d'Information et de Communication de l'Etat (DINSIC) chargées d'insuffler innovation au cœur des processus de l'Etat. C'est au niveau territorial que cet article est dédié.

Les concepts du Supply Chain Management (SCM) et du progrès permanent ont été testés au sein du secteur public. En 2007, la Direction Générale de la Modernisation de l'État Français (DGME) a décidé de diffuser des principes d'amélioration continue dans

divers services publics (préfectures, hôpitaux, trésorerie de la sécurité sociale, tribunaux, etc.) par la Révision Générale de la Politique Publique (RGPP) (Clainche, 2008). À notre connaissance, le secteur public hospitalier a été le premier secteur public qui a étudié et a pris en compte les approches issues des méthodes industrielles dans sa gestion quotidienne (Yoon *et al.*, 2016), en profitant de leur spécialisation sur un type de service : la santé. Des études (Landry et Beaulieu, 2001) ont montré l'avantage important découlant de l'introduction des méthodes industrielles dans le secteur hospitalier (diminution de 46% des dépenses, fournitures collaboratives, diminution des déchets et temps gaspillé, amélioration de la disponibilité des médicaments, augmentation de la disponibilité et productivité des ressources critiques).

Une étude en Angleterre a démontré l'amélioration de l'action publique par l'utilisation des indicateurs de performance au sein de l'administration (Propper et Wilson, 2003). En 2015, l'Afrique du Sud a publié un rapport (Lungisa, 2015) sur l'intégration du SCM dans ses organisations publiques. Ce rapport présente les avantages attendus du SCM : meilleure planification, transparence, diminution des coûts et des déchets, diminution de la corruption. Néanmoins, il n'explique pas précisément la manière dont les principes de SCM sont intégrés dans les services publics mais montrent un succès dans leur intégration.

La Gestion de la Relation Client (GRC) est une méthode qui n'a été implantée que récemment dans le secteur public (Hassan *et al.*, 2015). La GRC permet de personnaliser le service pour le patient ce qui apporte de la satisfaction dans le service rendu. Les hôpitaux développent les études autour de ce sujet pour accroître les interactions entre les patients et leurs organisations. L'objectif d'implanter la GRC dans les services hospitaliers réside dans l'amélioration de la transversalité de l'information et, de ce fait, de l'efficacité de la chaîne logistique hospitalière. De plus, la notion d'unicité du patient est importante dans le secteur de la santé (Vaish *et al.*, 2016).

La méthode appelée « Super-Unified Customer and Cost Evaluation Strategic System » (SUCCESS) (Melese *et al.*, 2004) soit en français « Client et Système Stratégique d'Évaluation de Coût Super-unifié » est basée sur le cycle « PDCA » et intègre quatre méthodes clés de gestion des entreprises :

- « Planning, Programming et Budgeting System » (PPBS), Planification, Programmation et Prévisions du Système budgétaire.
- « Activity Based Costing » (ABC), calcul des coûts par activité,
- « Balanced Scorecard » (BSC), le tableau de bord prospectif,
- « Total Quality Management » (TQM), gestion de la qualité totale.

Cette méthode démontre comment plusieurs méthodes d'amélioration et de suivi des performances peuvent être intégrées à une gestion de budgets publics.

L'approfondissement et l'adaptation des nouvelles méthodes de pilotage de la performance du secteur privé sont peut-être la solution pour répondre aux besoins du secteur public. Renaud, (2008) et Badja, (2010) ont proposé des cadres méthodologiques pour l'intégration des méthodes du secteur industriel dans l'évaluation de la performance du service public dans le domaine de l'enseignement supérieur. La définition d'un référentiel commun est un enjeu complexe de par l'hétérogénéité des acteurs et des activités. Osorio Montoya, (2015) a proposé un outil de pilotage permettant la création de plannings, la remontée d'information et l'élaboration de tableau de bords dans le secteur hospitalier et soulève le problème de la mise en application de l'outil et de son appropriation.

Ces différentes démarches montrent l'intérêt et la possibilité d'établir un cadre de modélisation pour le secteur public et particulièrement pour une métropole.

2.2 Un nouveau cadre pour la modélisation

Les langages et techniques de modélisation actuels présentent un caractère générique indéniable mais ne facilitent pas l'émergence de consensus autour de modèle de référence contenant des éléments liés aux métiers ou à des contextes spécifiques.

Devant le nombre important de langages de modélisation, un langage pivot baptisé UEML (« Unified Enterprise Modelling Language ») a été défini pour permettre l'intégration et le partage entre les différents modèles d'entreprises. D'autres langages sont dédiés à la modélisation des processus dans le but de les automatiser ou de se projeter vers une nouvelle organisation. Ce dernier point est régulièrement abordé par l'utilisation de méthodes comme le Value Stream Mapping (VSM).

Cet ensemble de méthodes et d'outils nécessite d'être intégré afin de proposer un support à l'analyse et à la prise de décision en incluant les spécificités du secteur public (Aubourg *et al.*, 2017). Notre méthode d'amélioration des processus du service public en Figure 2, appelée « Smart Administration Re-Engineering » (SARE), s'inspire des différentes méthodes issues du secteur industriel et plus particulièrement des travaux réalisés par Melese *et al.*, (2004) sur la méthode « SUCCESS ».

SARE intègre le SCM, le progrès permanent et le pilotage des opportunités liées à une Smart City. Albino *et al.*, (2015) présentent une revue détaillée des définitions de la Smart City sans parvenir à dégager un consensus. Trois raisons principales justifient l'intégration de la Smart City :

- elle contient un vivier important de technologies innovantes, permettant d'améliorer le vivre ensemble sur un territoire,
- elle est basée sur la communication impliquant un nombre important d'acteurs présents sur le territoire dans une perspective de collaboration autour de nouveaux besoins,
- la philosophie de la démarche Smart City est basée sur une ambition d'intérêt général en cohérence avec les missions du service public.

Notre nouvelle méthode « Smart Administration Re-engineering » se compose de cinq étapes clés :

1. définir la stratégie du service public,
2. identifier un problème inhérent au service public + favoriser l'émergence de l'écosystème,
3. sélectionner les améliorations pour le service public + identifier les opportunités liées à l'écosystème,
4. déployer des actions d'améliorations combinées aux opportunités de l'écosystème,
5. évaluer la performance du service public et de l'impact global.

La méthode est résumée en Figure 1.

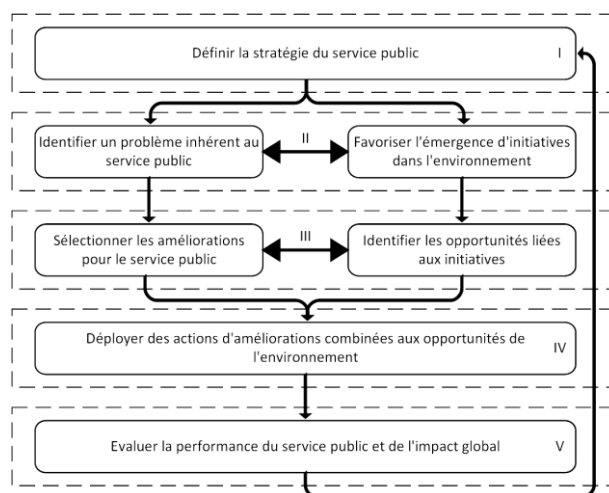


Figure 1 : Méthodologie d'amélioration du service public (Aubourg *et al.*, 2017)

Dans cet article, seuls les supports aux parties II et III sont abordés. La définition plus générale de la méthode a été publiée (cf. Aubourg *et al.*, 2017) et utilise notamment des audits, une modélisation de la performance et des définitions de processus.

Plus précisément il s'agit, dans cet article, de définir les modèles de références permettant de positionner une action d'amélioration au sein du secteur public et d'identifier, dans une logique de cartographie, quels seront les acteurs et/ou processus impactés.

3.4 Intégration de la démarche smart city à Toulouse Métropole

Concevoir l'architecture d'une ville intelligente est une tâche difficile impliquant divers domaines d'activités et parties prenantes (politique, fabricants, ingénieurs, développeurs, urbanistes, etc.). L'étude d'Abu-Matar (2016) a défini un méta modèle de la Smart City en se basant sur la norme l'ISO / CEI / IEEE 42010: 2011. Abu-Matar (2016) s'est appuyé sur des analyses qualitatives comme quantitatives. Son méta modèle est construit autour de huit vues :

- la vue capacité : c'est une représentation abstraite de toutes les capacités fournies par la ville intelligente,
- la vue des participations : cette vue représente les types de parties prenantes (citoyens, responsables politiques, entreprises, etc.),
- la vue place : elle répertorie l'ensemble des lieux pouvant être impactés par la Smart City,
- la vue service : qui propose les services spécifiques fournis par les applications de la ville intelligente,
- la vue des données : cette vue modélise tout type de données impliqué dans des applications de ville intelligente,
- la vue application : qui modélise les différents moyens nécessaires à l'utilisation du service,
- la vue infrastructure : ce modèle représente les systèmes physiques sur lesquels l'écosystème de la ville intelligente s'appuie,
- la vue processus : cette vue modélise les flux de tâches exécutées par des entreprises, des organisations et autres. Un écosystème urbain intelligent aura beaucoup de processus métier exécutés sur plusieurs frontières et réalisés par des applications et des services disparates.

La définition de notre modèle de référence de la Smart City est d'autre part conditionnée à deux hypothèses liées au terrain d'application de cette thèse :

- le secteur public joue un rôle d'animateur de la Smart City sur son territoire. Elle ne définit toutefois pas les stratégies de la Smart City, ce qui relève de l'ensemble des parties prenantes,
- la Smart City apporte des opportunités qui peuvent être utilisées pour augmenter le niveau de qualité des services publics dispensés sur le territoire.

Pour notre cas d'étude, le choix a été fait de définir un modèle axé sur les opportunités de la démarche Smart City, sur son environnement et plus principalement sur le secteur public présente sur le territoire.

Des nouvelles composantes apparaissent donc dans notre modèle de référence pour la Smart City, répertoriées dans le tableau 4. En s'appuyant sur le tableau 4, le modèle de référence lié aux opportunités de la Smart City est proposé dans la figure 5.

La Smart City peut s'intégrer à de nombreux types de processus incluant de la collaboration avec les citoyens.

Composante	Définition
Smart City (SC)	Dans le modèle de référence, la SC est représentée comme un élément de l'environnement présent au sein dans un territoire.
Dimensions de la performance d'une SC	La SC est composée de six dimensions distinctes (Gouvernance, Habitat, Citoyen, Economie, Environnement et Mobilité) qui permettent de formaliser des objectifs pour la ville intelligente.
Territoire	Il s'agit du périmètre géographique sur lequel la SC opère.
Opportunité	Une opportunité est considérée comme l'ensemble des éléments constituant une possibilité d'innovation pour la SC.
Initiative	Une ou plusieurs opportunités validées feront l'objet d'une « initiative », c'est-à-dire d'un projet expérimental. Les initiatives ont un impact sur le territoire et les services présents sur celui-ci.
Besoin	Les différents partenaires de la ville peuvent exprimer des besoins au sein de la SC. Le but est de répondre aux besoins de l'écosystème sur le territoire en sélectionnant des initiatives parmi les opportunités identifiées.
Ressource	Elles représentent l'ensemble des moyens mis à disposition pour la réalisation des initiatives sur le territoire, ainsi que les moyens pouvant être impactés par les nouvelles technologies.

Tableau 4 : Définition des nouvelles composantes pour le modèle de référence de la Smart City

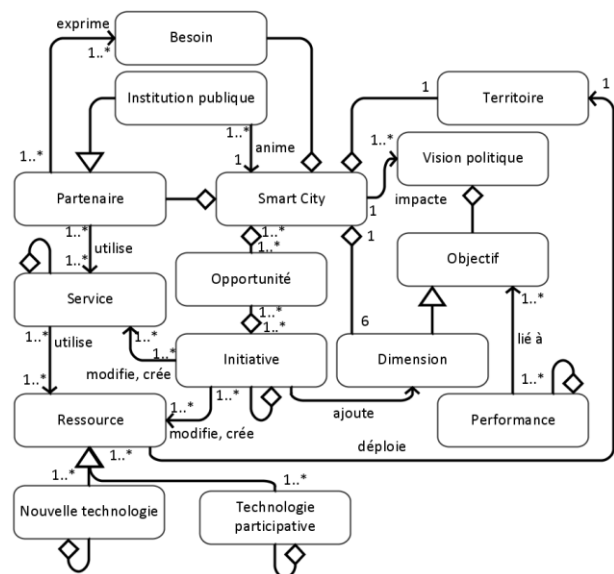


Figure 5 : Modèle de référence liant opportunités d'amélioration, services publics et performance

5 CONCLUSION

Pour répondre à des enjeux de politiques publiques dans une métropole, il s'agit d'intégrer l'ensemble des parties prenantes dans une démarche couplant d'amélioration des processus publics et développement de services dans une Smart City. Cet article présente une démarche de modélisation d'entreprise pour formaliser ces couplages. Cette démarche a été appliquée ici à l'amélioration du service de cantine scolaire. Par contre, dans le cadre de Aubourg (2017) elle a été aussi appliquée pour la justification de 17 projets portés dans la Smart City, Toulouse Métropole. L'étape suivante est d'utiliser la modélisation pour supporter une démarche d'évaluation, sélection et suivi de projets au sein d'une smart city qui couple les bénéfices de développement de la métropole avec ceux d'amélioration continue.

REFERENCES

- Abu-Matar, M., 2016. Towards a software defined reference architecture for smart city ecosystems. *IEEE International Smart Cities Conference (ISC2)*, Trento, Italie, p. 1-6.
- Albino, V. Berardi, U., and R. M. Dangelico, 2015. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), p. 3-21.
- Aubourg, G., 2017. *La démarche Smart City comme nouveau cadre d'intégration des méthodes issues du génie industriel dans les chaînes logistiques du secteur public*. Thèse de Doctorat, INP Toulouse, France.
- Aubourg, G., Galasso, F., Grabot, B., and J. Lamothe, 2017. Support to the Public Services Mutation Through Continuous Improvement in a French Metropolis. *IFIP Advances in Information and Communication Technology*, vol 514, p. 222-229, Springer, Berlin, Heidelberg.
- Badja, M., Touzi, W. and A. Thècle, 2010. « Contribution to the Definition and Modeling of Service and Service Activities ». *IFIP Advances in Information and Communication Technology*. vol 338, p. 619-626, Springer, Berlin, Heidelberg.
- Clainche, M.L., 2008. La révision générale des politiques publiques : premières annonces. *Revue française d'administration publique*, 125(1), p. 197-200.
- Hassan, R. S. ; A. Nawaz, M. N. Lashari, and F. Zafar, 2015. Effect of Customer Relationship Management on Customer Satisfaction. *Procedia Economics and Finance*, vol 23, p. 563-567.
- Landry, S. and M. Beaulieu, 2001. La logistique hospitalière : un remède aux maux du secteur de la santé ? *Gestion*, 26 (4) p. 34-41.
- Lungisa, F., 2015. Public Sector Supply Chain Management Review, *Report of the National Treasury Republic of South Africa*, online at: <http://www.treasury.gov.za/publications/other/SCMR%20REPORT%202015.pdf>.
- Meijer, A. and M. P. Rodríguez Bolívar, 2015. Governing the smart city: a review of the literature on smart urban governance. *International Review of Administrative Sciences*, 82(2), p. 392-408.
- Melese, F., J. Blandin, and S. O'Keefe, 2004. A New Management Model for Government. *International Public Management Review*, 5(2), p. 103-131.
- Osorio Montoya, G., 2015. *Conception des outils pour le suivi des activités et l'aide au pilotage dans le secteur médico-social*. Thèse de Doctorat, INSA de Lyon, France.
- Propper, C. and D. Wilson, 2003. The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy*, 19(2), p. 250-267.
- Renaud, R., 2008. *Contribution au pilotage des organismes de formation Application aux établissements de l'enseignement supérieur*. Thèse de Doctorat, Université Paul Verlaine de Metz, France.
- Suarez Barraza, M. F. Smith, T. and S. M. Dahlgaard-Park, 2009. Lean-Kaizen Public Service: An Empirical Approach in Spanish Local Governments. *The TQM Journal*, 21(2), p. 143-167.
- Vaish, A., R. Vaishya, and S. Bhawal, 2016. Customer relationship management (CRM) towards service orientation in hospitals: A review. *Apollo Medicine*, 13(4), p. 224-228.
- Wall, R., S. Stavropoulos, J. Edelenbos, and F. Pajević, 2015. Evaluating the Performance of Smart Cities in the Global Economic Network. In: *Rodríguez-Bolívar M. (eds) Transforming City Governments for Successful Smart Cities. Public Administration and Information Technology*, vol 8. p. 87-113 Springer, Cham.
- Yoon, S. N., D. Lee, and M. Schniederjans, 2016. Effects of Innovation Leadership and Supply Chain Innovation on Supply Chain Efficiency: Focusing on Hospital Size. *Technological Forecasting and Social Change*, 113(B), p. 412-421.

Proposition pour une Démarche de Modélisation des Processus

S. LEMOUSSU

J.C. CHAUDEMAR, R.A. VINGERHOEDS

ISAE - SUPAERO / DISC
sophie.lemoussu@isae-supaero.fr
STAR ENGINEERING
sophie.lemoussu@star-engineering.fr

ISAE - SUPAERO / DISC
jean-charles.chaudemar@isae-supaero.fr
rob.vingerhoeds@isae-supaero.fr

RÉSUMÉ : *Le secteur aéronautique voit actuellement arriver de nouveaux types d'entrants, des systémiers innovants, en majorité des Petites et Moyennes Entreprises (PME), souhaitant développer et commercialiser de nouveaux moyens de transport. Des projets comme les dirigeables, les drones civils de transports de marchandises, les drones dits « taxi aérien », deviennent ainsi des préoccupations sérieuses pour tous les acteurs du marché où la sécurité des personnes et des infrastructures est un enjeu majeur. Dans ce cadre, le processus de certification, cœur de l'activité aéronautique, demeure une activité critique. Le secteur aéronautique est une industrie très structurée, organisée selon de nombreuses règles, normes et standards qui ne facilitent pas l'arrivée de ces nouveaux acteurs. Les PME concernées requièrent un support méthodologique pour les accompagner dans leurs développements et faire face à la complexité du marché. Nous proposons d'aborder la problématique par une approche d'ingénierie système basée sur les modèles et une modélisation hiérarchisée des processus d'entreprise.*

MOTS-CLÉS : *Aéronautique, certification, modélisation des processus, gestion de projet, PME, ingénierie système.*

1 INTRODUCTION

Les Petites et Moyennes Entreprises (PME) du secteur aéronautique constituent une part non négligeable de la bonne croissance de l'industrie aéronautique (GIFAS 2017). Historiquement, dans ce secteur, et pas uniquement en France, les PME sont en majorité des entreprises dites sous-traitantes qui travaillent pour le compte de plus grandes entreprises qui peuvent être des systémiers (c'est-à-dire des constructeurs ensembleurs, appelés aussi intégrateurs), des motoristes ou de grands équipementiers.

Ces dernières années, le secteur aéronautique évolue avec l'émergence de nouveaux types d'aéronef et de nouvelles entreprises qui portent ces projets innovants. En effet des PME, voire parfois même de Très Petites Entreprises (TPE), s'emparent du rôle de systémier en proposant de nouveaux moyens de transport, alternatifs à ceux que nous connaissons. Plusieurs entreprises proposent par exemple de faire revivre des machines comme les dirigeables (Flywing Whales, HAV, Stratobus, ...). D'autres souhaitent exploiter le potentiel des drones et élargir ce marché avec des applications diverses le plus souvent liées au transport de marchandises (Fly Win, NATAC, ...).

Ces TPE/PME doivent comme les grands groupes répondre aux problématiques réglementaires aéri-

ennes et fournir aux instances responsables toutes les preuves suffisantes de fiabilité et de sécurité de leur système. En effet, pour tout aéronef, si son constructeur souhaite l'opérer et/ou le commercialiser, une étape cruciale est à franchir, celle de l'acquisition de son certificat de type et de tous les agréments imposés par la réglementation (EASA, FAA, CAAC, ...).

Pour les grands systémiers de la filière aéronautique cette étape de certification reste un challenge lorsqu'il s'agit d'un nouvel avion de ligne mais le risque est le plus souvent largement maîtrisé (Boeing 2011, Airbus 2017). En effet, dans ces entreprises, les équipes sont rodées, la connaissance des textes est documentée et partagée, une culture de la navigabilité est profondément intégrée au quotidien, dans toutes les tâches effectuées. La connaissance actuelle de ces organisations a été acquise au fur et à mesure, capitalisée au travers d'outils collaboratifs de gestion documentaire, de méthodes de retours sur expérience, de processus de gestion de la connaissance, etc.

Les nouveaux entrants sur le marché aujourd'hui n'ont pas cette expérience et doivent rapidement se forger la maturité organisationnelle nécessaire pour comprendre et se conformer à l'intégralité de la réglementation, en très peu de temps. Ces entreprises, de part leurs objectifs, ont des besoins très particuliers ; elles doivent pouvoir modifier leurs processus

de manière dynamique pour pouvoir se développer ; elles doivent être capables de gérer des modifications très importantes pour pouvoir s'adapter aux évolutions technologiques. Et en même temps elles doivent acquérir la confiance des organismes de certification.

Notre travail de recherche s'inscrit dans ce contexte complexe et propose une démarche d'ingénierie système dirigée par les modèles pour répondre aux objectifs de la certification aéronautique. Ce travail n'aborde pas explicitement la problématique d'architecture des systèmes d'information de l'entreprise même si nos résultats pourront contribuer à leur conception.

Cet article s'organise selon trois sections principales. Après une présentation générale de concepts issus de l'ingénierie système dans la section 2, la problématique de la modélisation des processus est introduite dans la section 3. S'appuyant sur le contenu des sections 2 et 3, la section 4 expose la démarche globale de la recherche en cours et ses premières propositions. Enfin la conclusion formule quelques axes d'orientation pour les travaux à venir.

2 APPROCHE PAR L'INGENIERIE SYSTEME

L'ingénierie système est une approche dédiée pour le développement de systèmes complexes. Elle est historiquement déployée sur des projets spatiaux, militaires et aéronautiques. Plus récemment, cette approche attire d'autres secteurs, notamment le secteur automobile, le ferroviaire et le nucléaire. L'ingénierie système est actuellement la référence dans tous les grands groupes qui conçoivent et produisent des systèmes complexes.

Les TPE/PME ne sont pas toujours sensibilisées à l'approche système et n'y trouvent pas toujours leur intérêt alors que leurs principales considérations se trouvent concentrées au niveau de la conformité du service ou du produit fourni et des processus de réalisation nécessaires pour atteindre cet objectif. Les tentatives pour s'approprier les normes, les standards et les outils disponibles sur le marché sont le plus souvent laborieuses et coûteuses. En effet, quelles normes et quels standards sont vraiment prêts en l'état à être employés par ces petites organisations? Même si l'émergence de la norme ISO 29110 pourrait apporter une aide non négligeable pour certaines TPE/PME, plusieurs travaux (Sheard & Lake 1998, Sage & Biemer 2007, Xue, Baron, Esteban & Demmou 2014, Lemoussu, Chaudemar & Vingerhoeds 2018) tendent à prouver que les moyens à disposition ne sont pas en phase avec les attentes des entreprises du secteur aéronautique et qu'une démarche d'alignement et de synthèse doit être intensivement menée au sein de chaque organisation pour

identifier ce qui est nécessaire et suffisant pour elle spécifiquement.

L'Ingénierie Système (IS) peut se définir selon plusieurs points de vue. C'est avant tout une démarche méthodologique qui englobe l'ensemble des activités pour concevoir, faire évoluer et vérifier un système apportant une solution économique et performante aux besoins d'un client tout en satisfaisant l'ensemble des parties prenantes (AFIS 2018). Le système concerné est alors appelé le système d'intérêt. La modélisation des exigences du système, de ses fonctions, de sa structure physique est l'approche de l'Ingénierie Système Dirigée par les Modèles ou ISDM (MBSE, Model-Based Systems Engineering en anglais). S'intéressant à la fois au comportement statique et dynamique du système, cette approche permet de vérifier et valider sa conception et de mieux maîtriser le système dans son ensemble. L'approche MBSE permet de mieux contrôler les interfaces et les fonctions émergentes, c'est-à-dire toutes les fonctions qui n'étaient pas identifiées dans les besoins initiaux et qu'il va pourtant falloir soit exploiter, surveiller, ou éliminer.

2.1 Le système d'intérêt: l'organisation de l'entreprise

Au sens de l'ingénierie système, l'organisation d'une entreprise peut être vue comme un système spécifique qui peut se décomposer selon différentes ressources (humaines, physiques et matérielles, immatérielles, financières, énergétiques,...). Cette organisation a pour but de rendre atteignables un ensemble d'objectifs au travers des processus mis en œuvre au sein de l'entreprise. Une organisation peut alors se décrire selon un ensemble de processus qui interagissent entre eux pour atteindre un ou plusieurs objectifs. Pour nos travaux, le système d'intérêt n'est pas le système à faire (un avion innovant) ni le système pour faire (le système d'information qui supporte les activités) mais l'organisation telle qu'elle est attendue du point de vue de la certification.

2.2 Les objets à modéliser: les processus de l'entreprise

Contrairement aux systèmes généralement traités par l'IS, le système à faire, un processus d'entreprise sera toujours un concept abstrait, qui n'aura pas de réalité physique, et qui sera donc immatériel tout au long de sa vie. Les processus sont cependant des preuves tangibles qui témoignent du niveau de complexité de l'organisation. Les formaliser puis les analyser permet de mieux les comprendre et de les améliorer. La modélisation de ces objets constitue la garantie d'une représentation simplifiée mais lisible et partageable au sein de l'entreprise.

Les activités d'identification, d'analyse et d'implémentation des processus connue sous le nom de BPM (Business Project Management) incluent généralement l'activité de modélisation des processus. Cette dernière permet d'aller plus loin dans l'analyse et permet d'identifier notamment des incohérences, des répétitions, des goulots d'étranglement, etc. Les processus, souvent décrits avec un langage naturel dans les entreprises et plus particulièrement dans les TPE/PME, méritent d'être modélisés de manière plus formelle pour déterminer leurs interdépendances et les éventuels obstacles à leur bon fonctionnement.

A l'instar d'autres travaux (Bron & Salzemann 2015), la démarche IS est utilisée ici pour la conception des processus. Les étapes suivantes font donc partie intégrante de notre travail de recherche: la définition des besoins du marché, l'analyse des exigences du secteur aéronautique, une conception de l'architecture des processus, la modélisation des processus, la vérification et la validation des modèles produits.

3 MODELISATION DES PROCESSUS

La modélisation des processus permet de mieux maîtriser les activités de l'entreprise et exerce un contrôle naturel sur leur efficacité. Toutes choses égales par ailleurs, chaque exécution du même processus, chaque instance, devrait être similaire. En réalité, chaque exécution de processus génère des instances souvent différentes. Car en plus du séquençement, les tâches sont exécutées selon des contraintes de ressources ou de temps, qui peuvent varier en fonction du contexte. Certaines contraintes peuvent constituer des règles métier. Il convient d'être capable de vérifier que les processus sont conformes à ces règles. Il s'agit donc d'être aussi capable de modéliser ces règles qui vont finalement régir les décisions prises au sein des processus. Il s'agit aussi d'être capable de modéliser les résultats issus de ces stratégies décisionnelles et in-fine la performance de l'organisation.

3.1 Les objectifs et les différents points de vue de la modélisation des processus

Les objectifs de la modélisation doivent être bien identifiés en amont de l'activité de modélisation. Ils peuvent inclure de mieux les comprendre, les documenter, les analyser, les améliorer, les optimiser, les partager, les communiquer, les surveiller, les auditer, les exécuter, les simuler, les mesurer, en sous-traiter certains, etc.

Notre démarche se concentre principalement sur les trois premiers objectifs: comprendre les processus d'intérêt, en proposer une documentation détaillée, afin de mieux pouvoir les analyser. D'autres objectifs tels que optimiser les processus pour les TPE/PME

et partager la démarche auprès de la communauté du secteur aéronautique constituent des objectifs majeurs pour valider notre approche.

La modélisation des processus nécessite aussi de faire un choix de point de vue. Il existe dans la littérature plusieurs points de vue différents (Hug, Front, Rieu & Henderson-Sellers 2009) qui sont :

- le point de vue **Activité** - le modèle se concentre sur la description des Activités et des Tâches, produites par des Rôles, assurés par des Acteurs, qui utilisent des Ressources en Entrée et génèrent des Produit de Sortie,
- le point de vue **Produit** - le modèle s'attache à décrire les différents états d'évolution d'un Produit, les phases de transition, ainsi que les processus pour atteindre ces états,
- le point de vue **Stratégie** - chaque Stratégie est définie par des Intentions qui sont soit des Sources soit des Cibles,
- le point de vue **Contexte** - le concept de Contexte résulte de la combinaison du Concept de Situation et du Concept d'Intention,
- le point de vue **Décision** - le modèle est construit selon une succession de Décisions qui produisent des Alternatives, justifiées par des Preuves.

Le choix d'un unique point de vue simplifie la modélisation. Cependant, il pourrait être intéressant de pouvoir combiner plusieurs points de vue pour proposer des modèles plus en phase avec la complexité de la réalité à modéliser. La littérature identifie ainsi trois propriétés issues des choix de combinaison (Hug et al. 2009):

- un modèle est multi-vues, si plusieurs points de vue sont disponibles,
- il est unifié, si tous les concepts nécessaires sont groupés dans un seul modèle,
- il est adapté, s'il est construit pour un type d'entreprise ou un contexte particulier; cela signifie que tous les concepts ont leur utilité et qu'aucun ne manque.

A ce stade d'avancement de nos recherches, notre choix de modélisation est orienté Activité. Cependant l'orientation Produit et l'orientation Décision pourront apporter des dimensions intéressantes à la modélisation où le cycle de vie du Produit devra être pris en compte et où les nombreuses règles métier pourront amener à identifier des chemins décisionnels différents. Un modèle multi-vues, unifié et adapté sera notre cible.

3.2 Les langages de modélisation des processus

Pour engager une activité de modélisation, le choix du langage est majeur. Tout langage de modélisation va se construire à partir de trois éléments principaux :

- une notation ou en ensemble de symboles,
- des règles sémantiques,
- des règles de syntaxe.

Ces éléments peuvent être plus ou moins riches, plus ou moins complexes selon la capacité du langage. Ils doivent cependant être embarquées dans une application logicielle, un outil capable de proposer la manipulation du langage.

De nombreux langages et de nombreux outils existent mais la plupart, surtout ceux utilisés dans l'industrie, n'ont pas de sémantique formelle. Or cette propriété est essentielle pour une analyse des processus efficace et l'assurance de la cohérence, la complétude et l'exactitude. Cet article n'a pas la volonté de présenter de manière exhaustive tous les langages disponibles (voir Lemoussu, Chaudemar, Vingerhoeds 2018).

Le BPMN ou Business Process Model Notation (OMG 2014) est le langage le plus couramment répandu dans l'industrie et pour de nombreuses références, il s'agit d'un standard "de facto" pour la modélisation des processus (Martins & Zacarias 2015, Kheldoun, Barkaoui & Ioualalen 2017). Sa force est avant tout sa grande capacité de représentation grâce à de nombreux outils graphiques (Recker, Rosemann, Indulska & Green 2009). Cependant le BPMN gère difficilement la contrainte du temps et souffre d'un manque de sémantique formelle. Cette faiblesse peut conduire à des inconsistances, des ambiguïtés et des incomplétudes dans le développement des modèles, avec de surcroît la difficulté de les identifier (Dijkman, Dumas & Ouyang 2008).

Les Réseaux de Pétri (RdP) (Recker et al. 2009) constitue un langage de modélisation semi-graphique pour le développement des systèmes à événements discrets, en particulier pour les systèmes où les événements sont parallèles et/ou distribués et où leur concurrence est une caractéristique importante. Les RdP permettent de représenter les processus d'une manière non ambiguë grâce à une sémantique extrêmement simple et formelle composée de trois types d'élément: les Places, les Transitions et les Flèches, qui relient les deux premiers. Le concept de Jetons présent dans la définition de ce langage permet à la fois d'assurer la fonctionnalité de vérification du modèle et de simulation. Ils sont Turing-Complet et en ce sens ils permettent de modéliser n'importe quel algorithme. Dans le cas de systèmes physiques, ils permettent de décrire le système d'intérêt étape par étape, en décomposant en éléments plus simples les éléments constitutifs initiaux du système. Ils permettent une analyse détaillée des propriétés du système et de sa performance. Pour le cas de la modélisation des processus, les RdP sont notamment utilisés dans les travaux de Process Mining, comparable au Data Mining mais avec pour vocation d'acquérir des connaissances sur les proces-

sus grâce aux données issues des instances (Van der Aalst 2016).

Le langage SysML ou System Modeling Language (OMG 2017) est le langage le plus couramment associé aux pratique de MBSE, car orienté système. Il permet de faire collaborer des équipes aux compétences variées autour d'un unique langage. Il s'articule autour de neuf types de diagramme dont les sémantiques sont différentes et apporte une approche multi-vues du système. SysML est semi-formel, ce qui apporte plusieurs avantages. Il offre à la fois une grande souplesse d'expression et permet de structurer le processus de modélisation. SysML permet aussi une grande liberté d'implémentation dans les outils car il n'intègre pas de méthode spécifique. La syntaxe peut être facilement adaptée pour modéliser des systèmes particuliers ou s'incorporer dans des méthodes spécifiques de développement (Saqui-Sannes, Vingerhoeds & Apvrille 2018).

Le BPMN, les RdP et le SysML sont appropriés pour formaliser des activités concourantes. Le BPMN et le SyML sont les plus visuels dans leur capacité de représentation grâce à l'usage des canaux ou piscines. Mais choisir un langage, c'est aussi faire un choix de point de vue. Il sera sans doute profitable d'avoir une approche complémentaire en utilisant plusieurs langages pour combiner différents points de vue. Aussi l'activité de modélisation est une activité critique qui nécessite de la rigueur certes mais aussi une méthode. Ces langages ne constituent pas des méthodes. Au delà du choix du langage, il est nécessaire de compléter l'approche pour améliorer les modèles produits et y adjoindre notamment des méthodes de vérification. La métamodélisation par exemple peut aider au développement de modèles consistants et complets

3.3 L'approche par la métamodélisation

La métamodélisation est l'approche que propose l'OMG (Object Management Group) qui, dans ses travaux de normalisation, décompose la démarche de modélisation en 4 niveaux où (Figure 1):

- le niveau M0 correspond au niveau le plus bas de la modélisation avec la création des instances,
- le niveau M1 correspond au niveau des modèles,
- le niveau M2 correspond au niveau du méta-modèle,
- le niveau M3 correspond au niveau le plus haut avec la définition du méta-métamodèle.

L'OMG propose un ensemble d'outils, de langage, de méthodes pour modéliser des produits et des processus. Tous sont conformes à cette approche étagée. La démarche assure ainsi que chaque niveau est conforme au niveau supérieur. En revanche l'OMG propose un seul et unique méta-métamodèle, le MOF (Meta-Object Facility), qui s'autodéfinit et permet de définir

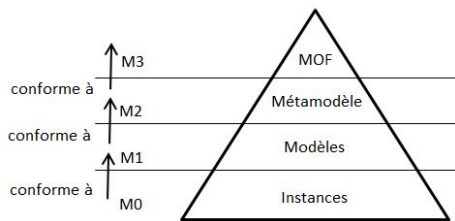


Figure 1 – Schéma de principe des 4 niveaux de modélisation issus des travaux de l'OMG

plusieurs métamodèles, notamment des métamodèles orientés Produit comme le métamodèle UML, le métamodèle SysML, des métamodèles orientés Processus comme le métamodèle SPEM (Software Process Engineering Metamodel, 2008) ou le BPDM (Business Process Definition Metamodel, 2008). L'OMG propose aussi d'autres types de métamodèle: par exemple le QVT, dédié à la transformation de langages, le XMI (XML Metadata Interchange), standard pour les échanges entre modèles et méta-modèles, l'OCL (Object Constraint Language), langage de contraintes sur modèles, ... Tous ces métamodèles de niveau M2 sont conformes au MOF. Les modèles produits au niveau M1 en se conformant au métamodèle de niveau M2 s'assurent d'être bien structurés et cohérents. De même les instances produites au niveau M0 en se conformant à un modèle de niveau M1, conforme lui-même au niveau M2, s'assurent d'être bien structurés et cohérents.

Même si le MOF de l'OMG n'est pas devenu un véritable standard en tant que tel, il fait figure de référence. Ainsi tous les autres métamodèles proposés, que ce soit du côté académique ou industriel, proposent une structure similaire qui se détermine en trois parties avec (Cadavid, Combemale & Baudry 2015):

- l'identification de concepts clés,
- la définition des relations qui relient les concepts entre eux,
- la création d'un ensemble de règles pour assurer une cohérence dans les modèles construits.

Il existe très peu de métamodèles de processus, la majorité sont plutôt orientés Activité. Le SPEM et le BPDM sont les principaux. Ce dernier construit pour supporter la sémantique du langage BPMN avait pour objectif initial de devenir un métamodèle générique pour la métamodélisation des processus. Malheureusement le BPDM n'inclut pas certains concepts qui nous semblent importants: l'objectif des processus n'est pas pris en compte, la catégorisation des processus non plus (Cruz-Cunha & Varajao 2011). Le métamodèle SPEM est construit selon une extension du profil UML. De fait, il est orienté objet et sa structure globale se présente sous forme de paquets UML (Cf. Figure 2). En plus d'être orienté objet, il

est très nettement orienté développement logiciel. Il ne prend pas en compte la particularité du développement des systèmes complexes qui présuppose la décomposition en sous-systèmes et la nécessité de collaborer entre équipes.

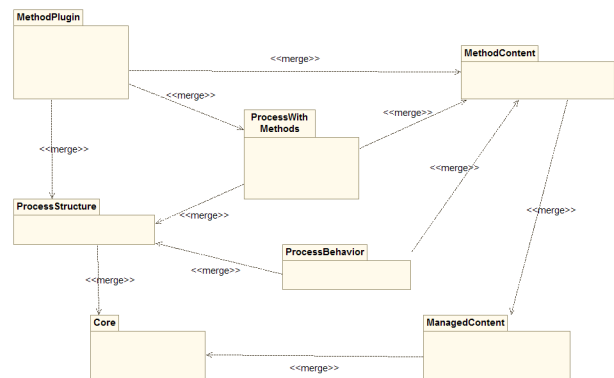


Figure 2 – La structure du SPEM V2.0 (2008)

Il existe plusieurs variantes du SPEM. Notamment le SysPem utilise le langage SysML à la place de l'UML. Il propose un nouveau métamodèle complet qui permet d'assurer une cohérence inter-modèles plus grande et une modélisation plus riche en détail et en niveau de granularité (Jakjoud, Zrikem, Baron & Ayadi 2016).

Tous les métamodèles ne sont pas fournis par l'OMG. D'autres métamodèles de processus ont été développés par d'autres organisations. Par exemple, le métamodèle OPF (Open Process Framework, 2005) proposé par le Consortium OPEN (Object-oriented Process, Environment and Notation) a ouvert un champ de recherche qui a permis de fournir d'autres propositions comme l'EPF (Eclipse Process Framework) ou OpenUp. L'OPF a l'avantage d'être compatible avec plusieurs formalismes (UML, OML, Object-Z...) mais toujours orienté objet. L'inconvénient principal est qu'il ne différencie pas toujours bien la limite entre chaque niveau de modélisation, notamment métamodèle /modèle. Enfin l'OPF ne propose pas de méta-métamodèle.

A ce stade de la recherche, nous n'avons identifié aucun métamodèle satisfaisant les besoins de la certification. Les métamodèles existants présentent des points de vue et des concepts différents qu'il est difficile de faire correspondre. Par ailleurs, ils sont le plus souvent difficilement adaptables.

4 DEMARCHE PROPOSEE

4.1 Les objectifs

S'appuyant sur une approche MBSE, notre travail de recherche a pour objectif de formaliser les processus de développement dans le cadre d'un projet de con-

ception d'un nouvel aéronef. Ce travail a donc pour vocation de proposer une cartographie des processus nécessaires et une démarche structurée mais simplifiée et adaptée au contexte des TPE/PME intégrateurs du secteur aéronautique. Ce rôle d'intégrateur signifie que ces entreprises ont un rôle de maîtrise d'ouvrage, et en ce sens elles ont nécessairement plus de responsabilités vis-à-vis de leurs sous-traitants et du processus de certification qui devient central dans la démarche. Par un ensemble de modèles, les exigences, qui incombent à une entreprise pour atteindre une certification réussie, sont formalisées. Le travail consiste dans un premier temps à modéliser le cœur organisationnel de la certification que constitue la Partie-21, point d'entrée de la réglementation aérienne en Europe. Elle permet de certifier des organismes de conception et de production. La certification est le processus qui permet de reconnaître la capacité d'une organisation à se conformer à la Partie-21, à concevoir et à produire des aéronefs, des pièces et équipements avec un niveau de sûreté suffisant, calculé en fonction de scénarios fonctionnels pré-définis. Elle comprend aussi les exigences et procédures relatives à la certification des aéronefs et de ses produits associés, des pièces et des équipements. C'est le texte pour atteindre une reconnaissance légale qu'un aéronef est apte au vol et que tous ses produits associés sont aptes à être installés dans celui-ci. En revanche, ne pas s'y conformer freine sérieusement le processus de certification et met à mal le processus de commercialisation du produit. En même temps, la Partie-21 est constituée d'un ensemble d'articles agrégés, écrits en langage naturel. Ceci force à l'interprétation et peut conduire à de possibles malentendus, à des surcoûts voire même à l'échec du processus de certification.

Nous fixons trois objectifs principaux pour notre travail de recherche:

- Comprendre les processus implicitement requis par la Partie-21, les documenter, les analyser et les formaliser grâce à un langage qui possède un pouvoir d'expression suffisant pour exprimer la structure des activités et leur coordination,
- Permettre une analyse détaillée des propriétés du système formalisé obtenu grâce à un langage suffisamment formel,
- S'assurer par un processus de vérification et de validation que les modèles produits soient conformes à la Partie-21.

4.2 La démarche générale

Notre démarche s'inspire de la hiérarchisation de modélisation de l'OMG et propose une architecture à trois couches (Cf. Figure 3) et prenant en compte:

- un métamodèle,
- deux modèles: le modèle "Business" et le modèle "Projet",

- un ensemble d'instanciations au modèle "Projet".

Le modèle "Business" constitue le modèle générique qui formalise l'intégralité des exigences de la Partie-21. Le modèle "Projet" est issu du modèle "Business" et est généré par une démarche de simplification et/ou d'optimisation qui reste à élaborer. Les instanciations sont créées à partir de cas d'utilisation: des TPE/PME nouveaux intégrateurs du tissu aéronautique européen.

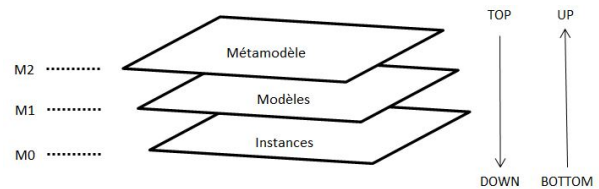


Figure 3 – La démarche générale de notre approche

De plus, l'approche de modélisation sera menée en deux étapes:

- une étape dite « Top-Down »,
- une étape dite « Bottom-Up ».

L'approche « Top-Down » sera initiée par la création du métamodèle dont une proposition est faite dans la section 4.3 suivante.

L'approche « Bottom-Up » sera initiée par l'analyse de cas d'utilisation, des TPE/PME issues du secteur aéronautique et ayant des projets innovants à faire certifier par l'EASA, l'Agence Européenne de l'Aviation Civile.

4.3 Les propositions

Une première formalisation des concepts manipulés est proposée dans cet article.

Définition 1 (Processus). Un processus est un 4-uplet (A, \rightarrow, R, a_0) , tel que:

- A représente un ensemble d'activités,
- \rightarrow est l'ensemble des interfaces entre les activités i.e. $\rightarrow \subseteq A \times A$,
- R est l'affectation univoque de ressources à des activités i.e. $R \in E \mapsto A$ (où E est un ensemble d'équipes projet, une équipe pouvant être constituée d'un seul individu),
- a_0 représente l'activité initiale.

Définition 2 (Activité). Une activité $a \in A$ est décrite par $a = \{t_1, t_2, \dots, t_n\}$ un ensemble énuméré fini de tâches t_i .

Hypothèse. En première abstraction, on fait l'hypothèse que toutes les tâches sont atomiques, i.e. leur durée d'exécution est nulle.

Réseaux de Pétri et SysML, des modèles de niveau M1 vont être construits et deux modèles majeurs seront proposés. Un modèle générique appelé modèle "Business" et un modèle simplifié pour les TPE/PME appelé modèle "Projet". L'approche par la transformation des langages et l'utilisation de patrons pourront être envisagée pour s'assurer de la convergence des résultats. Enfin l'intégration de paramètres techniques et de critères tels que la taille de l'entreprise, les types de ressources à affecter, la complexité du produit final pourrait permettre le développement d'un modèle paramétrique, qui conduirait à une optimisation des activités en adéquation avec l'entreprise et son produit.

REFERENCES

- AFIS (2018). www.afis.fr, consulté le 11-04-2018.
- Airbus (2017). Airbus A350-1000 receives EASA and FAA type certification, www.airbus.com/newsroom/press-releases/en/2017/11/airbus-a350-1000-receives-easa-and-faa-type-certification-.html, consulté le 11-04-2018 .
- Boeing (2011). Certifying Boeing's airplanes, 787updates.newairplane.com/certification-process, consulté le 11-04-2018 .
- Bron, J.-Y. & Salzemann, B. (2015). Démarche d'ingénierie système pour les systèmes de management, Tanger, Morocco.
- Browning, T. R. (2001). Applying the design structure matrix to system decomposition and integration problems: a review and new directions, *IEEE Transactions on Engineering Management* **48**(3): 292–306.
- Cadavid, J. J., Combemale, B. & Baudry, B. (2015). An analysis of metamodeling practices for MOF and OCL, *Computer Languages, Systems & Structures* **41**: 42–65.
- Cruz-Cunha, M. M. & Varajao, J. (2011). Enterprise information systems design, implementation and management: organizational applications.
- Dijkman, R. M., Dumas, M. & Ouyang, C. (2008). Semantics and analysis of business process models in BPMN, *Information and Software Technology* **50**(12): 1281–1294.
- GIFAS (2017). Rapport Annuel 2016-2017, *Technical report*.
- Hug, C., Front, A., Rieu, D. & Henderson-Sellers, B. (2009). A method to build information systems engineering process metamodels, *Journal of Systems and Software* **82**(10): 1730–1742.
- Jakjoud, A., Zrikem, M., Baron, C. & Ayadi, A. (2016). SysPEM: Toward a consistent and unified system process engineering metamodel, *Journal of Intelligent Manufacturing* **27**(1): 149–166.
- Kheldoun, A., Barkaoui, K. & Ioualalen, M. (2017). Formal verification of complex business processes based on high-level Petri nets, *Information Sciences* **385-386**: 39–54.
- Lemoussu, S., Chaudemar, J.-C. & Vingerhoeds, R. (2018). Systems engineering and project management process modeling in the aeronautics context: the smes study case, *International Journal of Mechanical and Mechatronics Engineering Vol:12, No:2* .
- Martins, P. V. & Zacarias, M. (2015). Business Process and Practice Alignment Meta-model, *Procedia Computer Science* **64**: 314–323.
- OMG (2014). Business process model and notation, version 2.0.2, *Object Management Group*, www.omg.org .
- OMG (2017). System modeling language, version 1.5, *Object Management Group*, www.omg.org .
- Recker, J., Rosemann, M., Indulska, M. & Green, P. (2009). Business Process Modeling- A Comparative Analysis, *Journal of the Association for Information Systems* **10**(4).
- Russo, A., Miller, R., Nuseibeh, B. & Kramer, J. (2000). An Abductive Approach for Handling Inconsistencies in SCR Specifications, (*ICSE2000*) *International Workshop on Intelligent Software Engineering*, Limerick, Ireland.
- Sage, A. P. & Biemer, S. M. (2007). Processes for System Family Architecting, Design, and Integration, *IEEE Systems Journal* **1**(1): 5–16.
- Saqui-Sannes, P., Vingerhoeds, R. & Apvrille, L. (2018). Early checking of sysml models applied to protocols, *MOSIM 2018, accepté pour publication* .
- Sheard, S. A. & Lake, J. G. (1998). Systems engineering standards and models compared, *Proceedings of the Eighth International Symposium on Systems Engineering, Vancouver, Canada*, pp. 589–605.
- Van der Aalst, W. M. P. (2016). *Process mining. Data Science in Action*, Springer Berlin Heidelberg.
- Xue, R., Baron, C., Esteban, P. & Demmou, H. (2014). Managing systems engineering processes: A multi-standard approach, *2014 IEEE International Systems Conference Proceedings*, pp. 103–107.

PUBLIC CONSTRUCTION PROJECTS IN TUNISIA: CURRENT PRACTICES, EXISTING PROBLEMS AND DEVELOPMENT OF A DECISION SUPPORT TOOL

Sirine TOUATI, Safa BHAR LAYEB

University of Tunis El Manar
UR-OASIS, ENIT
sirine.touati@outlook.com, safa.layeb@enit.utm.tn

Jouhaina SIALA CHAOUACHI

University of Carthage
ECSTRA, IHEC
siala.jouhaina@gmail.com

ABSTRACT: *In construction field, several interveners collaborate to ensure projects success. In this case, information exchanged during the project life cycle resides in separate and different tools (applications or papers) that could be unsynchronized and makes the monitoring process difficult and slow. Thus, the need of a unified and efficient tool becomes mandatory. The present paper describes the development of a Decision Support framework for any public construction projects in Tunisia. This framework was applied on a real case study of construction projects at the Tunisian ministry of women, family and childhood. This study is the result of a thorough diagnosis by which we have identified several problems related to lack of communication and a clear and standard strategy of work. The new tool will easily allow monitoring and evaluating the regional projects, in real time, from their beginning to their achievement. The proposed framework includes two monitoring processes for construction and maintenance projects respecting Tunisian law decrees, Tunisian civil building standards as well as Project Management fundamentals. A dashboard is also developed for better governance and visibility of the Ministry's projects.*

KEYWORDS: *Construction Projects, Project Management, Project Monitoring and Evaluation, Decision Support Framework, Public Sector, Tunisia.*

1 INTRODUCTION

Since the revolution of January 2011, the Tunisian state tries to evolve the workflows of the administration and the public service by adopting new strategies, in order to improve transparency and fight against corruption (The African Development Bank, 2012). In public administrations, management is evolving due to the large volume of activities and therefore the huge amount of information to be managed. For this purpose, Information Systems are implemented to help managers to obtain a better understanding of their work procedures (Laudon et al., 2012). Having a good management and a control of their administrations, managers support better their business decision-making for their future projects. To be more flexible, fast and collaborative, Information Technologies are used to enhance the potential of Information Systems and support the completion of daily tasks (Viborg Andersen, 2006). Thereby, Information Technologies is a core function in the public administrations.

Construction sector is taking advantages of the benefits provided by Information Technologies to support data management, since the importance of the amount of information needed to be exchanged during the project life cycle. This sector is considered as an integral component of a nation's infrastructure and industrial growth. In Tunisia, construction industry is one of the important sectors and has a key position in national economy due to

its positive impact in employment field and the population's living conditions (Btpexpo, 2016). According to the decree governing the construction of civil buildings n°2009-2617, Tunisian Ministries manage their regional projects with various interveners and ensure the monitoring during the projects achievement through regional commissioners in order to ensure their success. This success depends on a good access to a meaningful information, which is stored and exchanged in different tools (almost on Excel files). As projects become more complex and the ability to exchange information on documents within regions on a timely basis gets difficult, diverse obstacles were detected, such as: poor coordination between ministries and their regional commissioners in real time, diversity of information exchanged which make it sometimes complex and unmeaningful, lack of traceability and also the risk of document loss. In order to reduce these issues, we need to adopt standardization in construction information and documentation, to facilitate and unify the monitoring process. In this paper, we present a new decision support framework for regional projects that contains two monitoring processes for construction and maintenance projects. The proposed framework is based on Tunisian civil building standards and offers friendly interfaces that facilitate the work for commissioners because they are not necessarily experts in construction field.

The reminder of this paper is structured as follows: Section 2 reviews related research while Section 3 is devoted to explain in detail the current practices in Tun-

sian ministries. Section 4 focuses on the presentation of our proposed framework and its application on a case study of the Tunisian ministry of women, family and childhood in Section 5. Finally, we present a conclusion for this study and some guidelines for future directions.

2 RELATED RESEARCH

Project Management is an ancestral science since it allowed the construction of Egyptian pyramids (Cobb, 2015) and it has grown and be formalized under its current standards in the late 1950s and 1960s, in the United States Defence to manage complex projects (Charvat, 2003). Project Management is defined by PMBOK (Project Management Institute, 2013) as the application of knowledge, skills, tools and techniques to project activities, in order to meet the project objectives. Its major objectives include performance, cost, and time goals through optimum use of resources (Kerzner, 2013) as summarized in Figure 1.

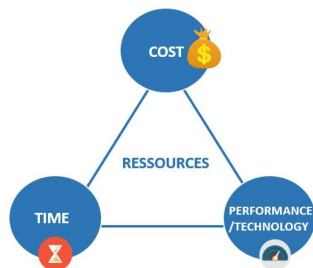


Figure 1: Overview of Project Management

Project monitoring is the process of recording, collecting and analysing actual values during project achievement that the project manager wishes to know. This task includes reporting work descriptions with the execution and the financial progress. Adopting standardization and common communication tools in construction industry play an important role for an effective communication between various interveners (Akbar, Anis Rosniza Nizam, et al., 2015) (Martínez-Rojas et al., 2016) Software tools are essential to manage projects and to bear many issues throughout projects life cycles. Nowadays, there is a wide variety of Project Management software that can fit with different needs and support all the phases of the Project Management cycle or specific part of it. Several software Project Management tools were examined in literature, including commercial as well as open source and free tools (e.g. (Pereira et al., 2013) and (Gonçalves et al., 2017)). To make the right choice and use the suitable tool, many factors should be considered such as firm size, the complexity of projects, initial year of Project Management software use and extent of Project Management software use (Liberatore et al., 2003). A comparative study was made by Mishra and Mishra (Mishra and Mishra, 2013) who presented the features of twenty popular Project Management tools. They presented the tools side by side in a table to represent and compare their features. To sum up, Primavera P6 and

MS Project are described as the most frequently used by different industries around the world world (e.g. (Fabac et al., 2010) and (AlMobarak et al., 2013)), which express the customer satisfaction due to software's performance, but they still propose other future research to improve their features (e.g. (Liberatore et al., 2001) and (Sajad et al., 2016)).

Despite their performance, Primavera and MS Project seem insufficient for our context due to the following drawbacks:

- These software tools are complex and cannot successfully work without training the local users from external consultants.
- They are usually used by Project Management professionals for planning and scheduling tasks, but the local users are not especially familiar with advanced notions of Project Management.
- Regarding their price, Primavera and MS Project are expensive tools (Fabac et al., 2010).
- According to circular focusing on the Tunisian administrative reform n°45-1999, Tunisian public administrations are required to use Arabic language in communication, exchanged documents and software tools, which is not offer by these tools.

Regarding these limitations, we conclude that it would be interesting to suggest standards in order to unify monitoring process for Tunisian ministries and to develop a customized framework that respects these standards, improves the controlling and the tracking of regional projects with dashboards for managers. Using Arabic language would be a great asset. This software should be flexible, configurable and scalable in order to offer an open source solution to future development.

3 DIAGNOSIS AND EXISTING STUDY

As previously mentioned, our framework is proposed for any public sector including Tunisian ministries to monitor their regional projects. For this, we made a diagnosis at the ministry of women, family and childhood in order to detect the current practices and to apply our proposed system.

3.1 The Tunisian Ministry of Women, Family and Childhood

By ratifying the convention on the children rights in 1992, Tunisia is committed to promote and protect their rights (Unicef, 2010). This commitment is regarded as one of the national priorities in Tunisia, which has achieved remarkable growth by ensuring children rights in the areas of health, education and wellness. According to the annual report prepared in 2017 by Kids Rights¹, in collaboration with the National Institute of Social Studies (ISS) and Erasmus School of Business (ESB), Tunisia is ranked 9th in the world and 1st in Arab level,

¹ <http://www.kidsrightsindex.org/>

which respects international commitments to promote the status of children’s right. We can therefore conclude that the child development has increased attention for a long time by Tunisian government and specifically the ministry of women, family and childhood, which held, in the last decade, implementation of specialized institutes in the protection and development of children, offering adequate and respectful environment for the child.

3.2 Current Practices

Figure 2 illustrates the management and monitoring processes of construction projects. Before embarking on construction projects, the ministry studies the feasibility of these projects and prepares, in collaboration with various interveners, a document on new projects to be launched during an annual meeting, based on Restricted Ministerial Council, regional requests and investment plan². This document will be exchanged with the ministry of finance³ to confirm and validate the final list of projects to be launched. The rejected projects will be reinstated in the original investment plan. Once the final list is signed, the ministry of women, family and childhood cooperates with their regional administrations, the governorates concerned by these projects, the ministry of equipment and the ministry of land affairs. This cooperation aims to ensure the progress of projects during all their phases, within the limits of time and budget constraints. According to the decree establishing regional commissionership of the ministry of women, family and childhood n°2013-4063, the regional commissioner is responsible for executing public procurement, monitoring the transferred credits, as well as the execution of infrastructure projects and the preparation of monthly reports. Its reports are transferred to the ministry of women, where they will be stored and organized manually in Word and Excel files.

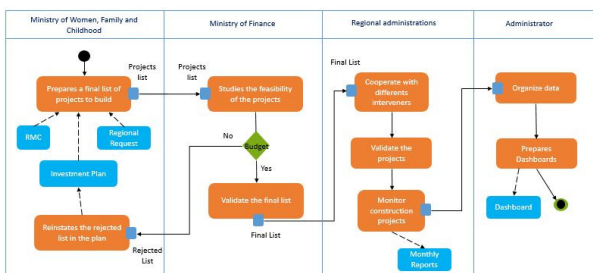


Figure 2: Management and Monitoring process of Construction projects in Tunisia

To improve the current situation of its institutions and accommodated children under the right conditions, the regional commissioner is responsible of writing detailed technical and periodic reports on its institutions current states. He could propose maintenance works if necessary [24]. Once validated, the Ministry of women, family and

² Investment plans are planned every 5 years.
³ The Tunisian Ministry of Finance supports the funding of childhood institutions.

childhood receives the funds from the ministry of finance and cooperates with the other interveners to begin maintenance works. Thus, the regional commissioner monitors the execution by providing monthly manual reports to the ministry of women, which stores the work descriptions in Excel files.

3.3 Existing Problems

After analyzing the current processes of management and monitoring of construction and maintenance projects, we have detected the following defects:

- There is a major risk of documents loss. Therefore, having a centralized and a secure database would secure all the data.
- To the best of our knowledge, there is no clear strategy of work. Every commissioner works independently and describes the work in his own way. This heterogeneity of information transferred and lack of standards make difficult the visibility of projects and their real time following up.
- There is lack of traceability and history due to the inexistence of a database.
- There is no compliance of deadlines fundamentally due to the lack of traceability. Registration of all commissioners’ actions should be mandatory.

4 PROPOSED MONITORING PROCESSES

Project monitoring is a process of collecting, recording and analyzing information about the projects state and must be continuous during all phases. To have a better view of projects progress, we have to act timely and ensure the follow-up of the project in a good way, it is important to collect the necessary data and documents throw new standards that facilitate commissioners’ work. To the best of our knowledge, there is no software tool used by the Tunisian ministries for this task. Then, we propose two new processes detailed in the next subsections. Next, a new dashboard for managers is also presented.

4.1 System Architecture

Nowadays, developers are directed towards developing complex websites and web applications using PHP. To win time without starting up a web project from scratch and to provide well-structured and maintainable codes that help detecting errors more easily, PHP Frameworks are the most used by developers. In this work, we have opted for Symfony Framework due to its powerful features (Hongkiat, 2017):

- Symfony is one of the most stable and well-documented framework.
- This framework has a growing and hyperactive community. Many tutorials are offered to facilitate its use.
- Symfony integrates preventive security measures to fight XSS, CSRF, and SQL Injection Flaws and Attacks.

Our system has been implemented on the basis of three-tier architecture as shown in Figure 4, offer a user-friendly and web-based interface coupled with secure access controls.

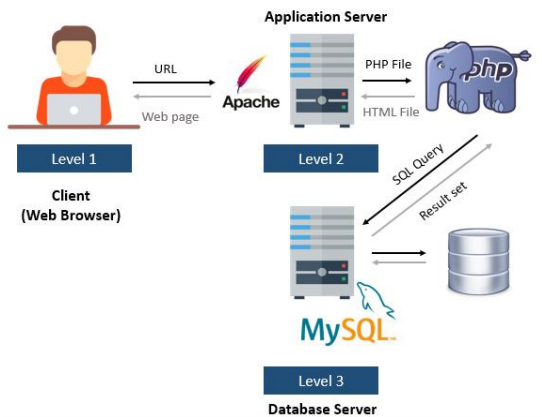


Figure 3: Three-Tier architecture of the system

4.2 Proposed monitoring process for Construction projects

As we have mentioned during the diagnosis of tasks performed by the commissioners, we identify two keywords: “building projects” and “public procurement”. Therefore, we have examined laws decree [5, 26] for each keyword and we propose to implement a new monitoring process, composed of 7 phases respecting the following characteristics:

- The proposed phases extracted from the laws decree follow a logical and chronological order respecting the standards of the Tunisian construction field.
- During these phases, we can extract and exchange the most important data necessary to monitor the project, to ensure the best view of the whole project life cycle and to make right decisions.
- The seven proposed phases are necessary and complete, containing all the critical data that managers could need.

The new proposed process respects the following criteria:

- The Tunisian law decrees,
- The Tunisian civil buildings standards,
- The fundamental concepts of Project Management.

As we respect these characteristics, the proposed framework can be applied to any Tunisian public sector that manages regional construction and maintenance projects. Two types of interveners are involved in this tool as presented in Table 1.

Intervenors	Description
Regional	24 commissioners spread over 24

commissioner	governorates
The manager	The responsible of managing childhood institutions

Table 1: The framework users

In what follows, we present the seven proposed phases for monitoring construction projects in any Tunisian public sector:

4.2.1 PHASE 1 : PRE-STUDIES PHASE

This is obviously the first phase. It begins after validating projects list to be built by the Ministry of finance and transmitting it to the regional administrations. In this case, the commissioner must declare to the ministry that the projects are validated and must send the information, presented in Table 2, concerning pre-studies.

Inputs	Intervenors	Descriptions
The governorate	The regional commissioner	In Tunisia there are 24 governorates
The municipality		Each governorate is divided into municipalities.
Institution type		Currently the Ministry of Women implements 5 types of childhood institutions: Children’s club, Children’s complex, integrated centres for youth and childhood, Children’s Vacation and Recreation Centres and public kindergartens.
The total funds		The funds committed by the Ministry of Finance for the project
The first instalment		The first instalment of the amount allocated to start the next phase.

Table 2: The inputs of Pre-Studies Phase

4.2.2 PHASE 2 : STUDIES PHASE

According to the decree governing the property deeds (article 12 n° 2009-2617), a design and achievement study should drawn up by one or several designers appointed for this purpose by the Ministry of equipment, for every civil buildings. The objective of this study is to develop the project on the architectural, functional and technical plans and to evaluate its estimated realization cost in conformity with the related functional program or the functional and technical program. Before starting this phase, the commissioner must prepare a support file for the Ministry of equipment as indicated in article 13 of the decree n° 2009-2617. The property deed, included in this support file and defined in Table 3, seems a critical

element for the Ministry. During our diagnosis phase, we have noticed that several construction projects have been started without having a property deed. This has led to the fail of projects in many cases and therefore a loss in terms of time and cost, due to the lack of control of the commissioners. To overcome this problem, we propose that the regional commissioner starts this phase, only after sending the property deed to the Ministry and then the manager will validate it.

Inputs	Interveners	Descriptions
<i>The property deed</i>	The regional commissioner	A legal document that shows the ministry possession on the ground

Table 3: The inputs before Studies Phase

Once validated, the commissioner automatically will be redirected to the study phase. In this case, the Ministry will have all the property deeds of its grounds as shown in Table 4.

Outputs	Interveners	Descriptions
<i>The property deeds list</i>	The manager	The property deed list grouped by project.

Table 4: The outputs before Studies Phase

After completing the necessary studies by architects and in order to execute the public procurement of the project, the commissioner must provide the inputs detailed in Table 5.

Inputs	Interveners	Descriptions
<i>The topographical survey of the ground</i>	The regional commissioner	Used to identify the contours and the conception of the ground
<i>The covered and uncovered area</i>		This information must be in m ²

Table 5: The inputs after Studies Phase

4.2.3 PHASE 3 : TENDER PHASE

The public procurement is defined according to decree n°2014-1039 [26] as a contract, concluded by public purchasers in order to execute works, to provide goods, services, or studies. Public procurements are awarded after competitive bidding. The notice of invitation to tender must be published in the press and on the public procurement website at least thirty days before the deadline for the receipt of tenders.

4.2.4 PHASE 4 : OPENING TENDERS PHASE

The tender opening session is held on the day set as the deadline for receipt of tenders. A permanent commission within the governorate is responsible for opening offers.

4.2.5 PHASE 5 : PUBLIC PROCUREMENT ASSIGNEMENT PHASE

According to decree regulating public procurement (article 73 n°2014-1039) and after posting the results of the call for competition and the name of the public tender holder, the commissioner must inform the Ministry. The Information of the estimated realization cost proposed by the contractor and the estimated beginning date as well as estimated end of project date must be sent as detailed in Table 6.

Inputs	Interveners	Descriptions
<i>Estimated cost</i>	The regional commissioner	Estimated cost proposed by the contractor.
<i>Estimated beginning date</i>		Estimated date proposed by the contractor to start the work.
<i>Estimated beginning date</i>		Estimated date proposed by the contractor to end the work.

Table 6: The inputs of Public Procurement Assignment Phase

The collected information will then be useful for the follow-up and for comparing the estimated with actual costs, estimated and real days of works as well as decision-making as described in Table 7.

Inputs	Interveners	Descriptions
<i>The property deed</i>	The regional commissioner	A legal document that shows the ministry possession on the ground

Table 7: The output of Public Procurement Assignment Phase

4.2.6 PHASE 6 : EXECUTION PHASE

As previously mentioned, the regional commissioner is responsible for the follow up of transferred funds as well as the project progress by providing monthly reports. Therefore, for better monitoring and budgetary control, we propose to organize this phase as follows in Table 8.

Inputs	Interveners	Descriptions
<i>Payment</i>	The regional commissioner	The Commissioner must enter the amount awarded and the date of award.
<i>Expenditure</i>		If there is an expense, the Commissioner will enter the amount and the percentage of execution.
<i>New commitment</i>		If the cost of the work has exceeded the first commitment amount, the Ministry of Finance may

	award a new commitment for the project.
Progress percentage	With every expenditure, the work progress, so the commissioner should input the progress percentage

Table 8: The inputs during the Execution Phase

The details of Table 8 must be saved and organized as shown in Figure 4. Credit management and progress percentage are grouped by year in a table with a bar chart that helps managers to have better view of the total funds management.

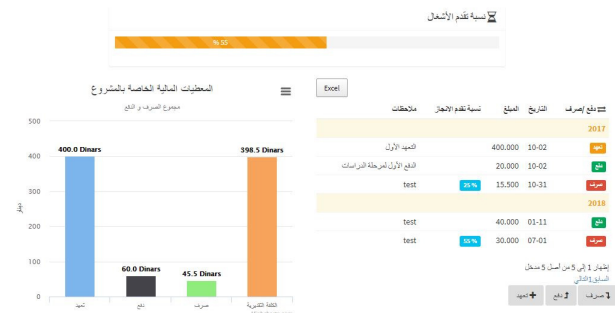


Figure 4: The output screenshot of Execution Phase

4.2.7 PHASE 7: END OF PROJECT PHASE

Once the work accomplished and the building ready to be exploited, the commissioner informs the Ministry by providing the following information presented in Table 9.

Inputs	Intervenors	Descriptions
The address	The regional commissioner	The exact address of the institute for more precision
The director name		The name of the responsible for the institute
Institute phone number		Phone Number of the director
Building reference		Every building have a unique opening reference
Opening date		The date when the institute will be available

Table 9: The inputs of the End of project Phase

With this new monitoring strategy, composed of seven sequential phases, the users will access to a list of all projects, clustered by governorates, with all the necessary details to monitor, control, intervene if

necessary and evaluate the ministry's construction projects from the beginning of their programming until their closure, in real time, as presented in Figure 5.

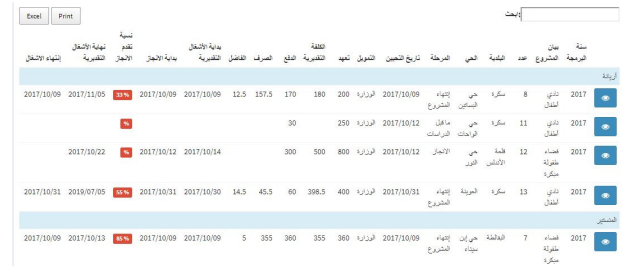


Figure 5: Construction Projects list

4.3 Proposed monitoring process for Maintenance projects

During the diagnosis phase, we have concluded that maintenance projects can be divided in three categories, presented in Table 10.

Categories	Description
Maintenance works	Includes: repairing a fault on an installation (lighting, heating, and sanitation), painting the walls and maintaining the gardens.
Planning works	Change the building design by building new rooms.
Type change	Change the institution's services, for example: a children's club will be transformed into a children's complex.

Table 10: Maintenance projects categories

As the construction projects, the monitoring process of maintenance projects follows the same strategy and we propose to divide it into the same 7 phases. The only difference is that in each category, inputs changes as shown in Table 11.

Inputs	Maintenance works	Planning Works	Type change
Phase 1: Pre-Study phase	*	*	*
Phase 2: Study phase		*	*
Phase 3: Tender phase			
Phase 4: Opening Tenders phase			
Phase 5: Public Procurement Assignment phase	*	*	*
Phase 6: Execution phase	*	*	*
Phase 7: End of Project phase	*	*	*

Table 11: Maintenance works Inputs

More precisely, Figure 6 presents a screenshot of a complete form of maintenance projects. Through this

easy-to-read and rich inference, the manager can have full visibility on the project. The progress bar that we have developed makes it possible to record the date under each phase, which allows the manager to control the commissioner and to intervene if necessary.

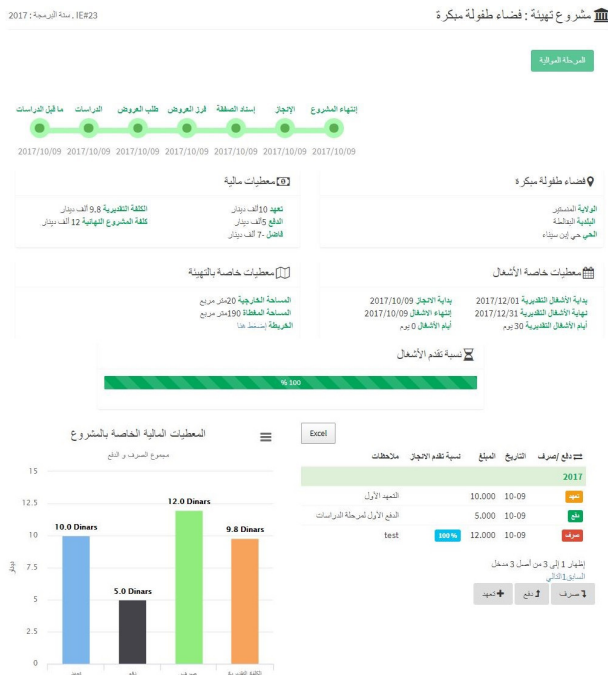


Figure 6: Complete form of Maintenance project

5 CASE STUDY

To test the proposed framework, we collect preliminary information from the Ministry’s construction and maintenance projects. As shown in Figure 7, the Tunisian ministry of women, family and childhood owns 319 childhood institutions spread over 24 Tunisian governorates. For a better and faster visibility, we have made available the map of Tunisia on which the users can consult the total number of institutions for each governorate. For more details, we have developed a table on which the users can see, for each governorate, the total number of each type of institution. All these data help managers to prepare their investment plan for the construction of new projects.

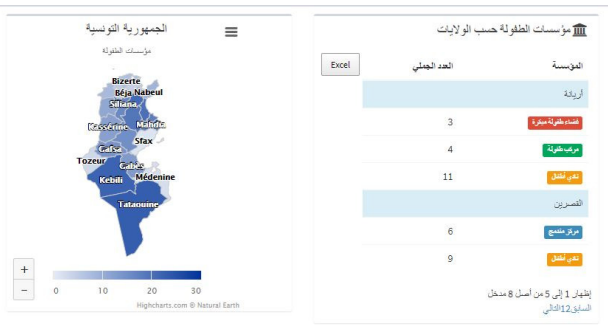


Figure 7: Complete form of Maintenance project

Figure 8 shows the result after data entry of the construction projects in 2017. Actually, the Ministry of women, family and childhood manages 36 construction projects distributed as follows: four projects in pre-study phase, six in tender phase, and two projects in procurement assignment Phase, nine under implementation and fifteen fenced projects that are ready to be exploited.

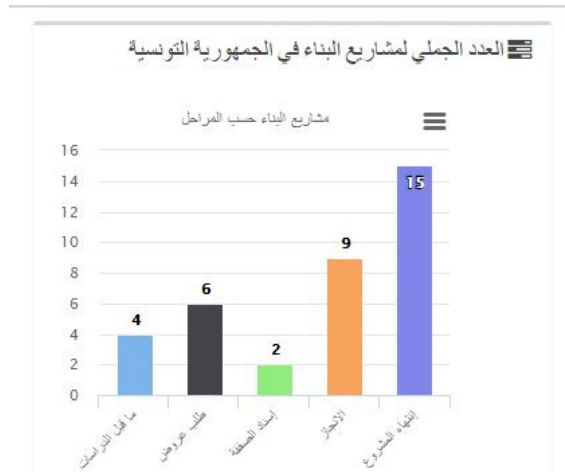


Figure 8: Construction projects grouped by different phases

To the best of our knowledge, the manager is spending approximately two weeks to collect all information about construction and maintenance projects. Thanks to our tool, few microseconds will be spent to display the project lists. Figure 9 illustrates the result of execution time to display the list of construction projects.



Figure 9: Performance metrics

6 CONCLUSION

In Tunisia, most construction and maintenance projects data are usually stored in different tools and documents. The integration of these data in a common tool is vital and mandatory for good management and for making reliable decisions as well as reducing the time spent to reach these decisions. After a diagnosis made in a Tunisian ministry, we detected several shortcomings caused by lack of clear management strategy and limited knowledge of the employees in the field of civil buildings constructions. These shortcomings have caused some projects failure, budget corruption, lack of communication between regions and absence of instantaneous information for the monitoring and evaluation of construction and maintenance projects. In this paper, a decision support tool is proposed to overcome these issues. This framework has been implemented using a web-based

application providing a user-friendly interface to handle two new proposed processes, respecting the Tunisian laws decrees, the Tunisian civil buildings and public procurement standards as well as the fundamental concept of projects management.

Our framework was applied on the Ministry of women, family and childhood, allowing a better visibility and follow-up on the construction and maintenance projects, control of the regional commissioners and time saving which facilitates communication and coordination between the ministry and its regional administrations.

REFERENCES

- African Development Bank. Tunisia, 2012. *Post - Revolution Economic and Social Challenges*. The African Development Bank.
- Akbar, Anis Rosniza Nizam, Mohamad Fadhil Mohammad, Norizan Ahmad, and Myasarah Maisyam, 2015. *Adopting standardization in construction environment: Standard method of measurement (smms)*. *Procedia-Social and Behavioral Sciences* 170: 37-48.
- AlMobarak, N., AlAbdulrahman, R., AlHarbi, S. and AlRashed, W.A., 2013. The Use of Software Project Management Tools in Saudi Arabia: An Exploratory Survey. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 4(7).
- Btpexpo, "Press Kit" 2016. [Online]. Available at: www.btpexpo.com.tn.
- Charvat, J., 2003. *Project management methodologies: selecting, implementing, and supporting methodologies and processes for projects*. John Wiley & Sons.
- Cobb, C.G., 2015. *The project manager's guide to mastering Agile: Principles and practices for an adaptive approach*. John Wiley & Sons.
- Fabac, R., Radošević, D. and Pihir, I., 2010, June. Frequency of use and importance of software tools in project management practice in Croatia. In *Information Technology Interfaces (ITI), 2010 32nd International Conference on* (pp. 465-470). IEEE.
- Gonçalves, R.Q. and von Wangenheim, C.G., 2017, May. DotProject+: open-source software for project management education. In *Proceedings of the 39th International Conference on Software Engineering Companion* (pp. 213-215). IEEE Press.
- Hongkiat, 2017. "10 PHP Frameworks for Developers – Best of". [Online]. Available at: www.hongkiat.com.
- Kerzner, Harold, 2013. *Project management: a systems approach to planning, scheduling, and controlling*. John Wiley & Sons.
- Laudon, K.C., Laudon, J.P., Brabston, M.E., Chaney, M., Hawkins, L. and Gaskin, S., 2012. *Management Information Systems: Managing the Digital Firm, Seventh Canadian Edition*.
- Liberatore, M.J., Pollack-Johnson, B. and Smith, C.A., 2001. Project management in construction: Software use and research directions. *Journal of construction engineering and management*, 127(2), pp.101-107.
- Liberatore, M.J. and Pollack-Johnson, B., 2003. Factors influencing the usage and selection of project management software. *IEEE transactions on Engineering Management*, 50(2), pp.164-174.
- Martínez-Rojas, María, Nicolás Marín, and M. Amparo Vila Miranda, 2016. *An intelligent system for the acquisition and management of information from bill of quantities in building projects*. *Expert Systems with Applications* 63: 284-294.
- Mishra, A. and Mishra, D., 2013. Software project management tools: a brief comparative view. *ACM SIGSOFT Software Engineering Notes*, 38(3), pp.1-4.
- Pereira, A.M., Gonçalves, R.Q., Von Wangenheim, C.G. and Buglione, L., 2013. Comparison of open source tools for project management. *International Journal of Software Engineering and Knowledge Engineering*, 23(02), pp.189-209.
- Rose, K.H., 2013. A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Fifth Edition. *Project management journal*, 44(3).
- Sajad, M., Sadiq, M., Naveed, K. and Iqbal, M.S., 2016. Software Project Management: Tools assessment, Comparison and suggestions for future development. *International Journal of Computer Science and Network Security (IJCSNS)*, 16(1), p.31.
- Unicef, 2010. *Rapport de la revue à mi-parcours du programme de coopération*. www.unicef.org/evaldatabase/files/TUN-2010-001-1.pdf.
- Viborg Andersen, K., 2006. Reengineering public sector organisations using information technology. In *Comparative public administration* (pp. 615-634). Emerald Group Publishing Limited.

QUESTIONING INTEGRATION OF VERIFICATION IN MODEL-BASED SYSTEMS ENGINEERING: AN INDUSTRIAL PERSPECTIVE

C. LAING, P. DAVID, E. BLANCO

Univ. Grenoble Alpes, CNRS, G-SCOP
38 000 Grenoble, France
christopher.laing@outlook.com,
pierre.david@grenoble-inp.fr, eric.blanco@grenoble-
inp.fr

X. DOREL

Schneider Electric
38000 Grenoble, France
xavier.dorel@schneider-electric.com

ABSTRACT: *To manage complexity in modern products, implementing verification using model-based methods has been identified as a key challenge for this decade (NDIA, 2011). Model-Based Systems Engineering (MBSE) has been proposed as an approach to manage such complexity through the formalized application of models throughout the systems development life cycle to support verification activities (Crisp, 2007). However, the current knowledge in MBSE does not guide systems engineers in which models to deploy nor the specific outcomes of adopting model-based approaches (Murphy & Collopy, 2012).*

The purpose of this research is to make a first step towards developing a verification strategy by performing a descriptive study on current industrial practice and the state of the art in MBSE. Through workshops of 16 engineering organizations located in the Rhone-Alpes region of France, key success criteria for integrating model-based verification with MBSE were identified allowing analysis of top verification methods presented in MBSE publications. To further validate the use of these presented methods, work studies and interviews were conducted to analyse the methods through different industrial perspectives and identify their feasibility for successful application.

KEYWORDS: *Model Based Systems Engineering, Verification, Industrial perspective*

1 INTRODUCTION

With the growth of complexity in modern products, the use of the systems engineering (SE) approach has grown across many industries to cope with increases in functionality alongside reduced product development cycle times. To manage these constraints, systems engineers need to handle considerable amounts of information regarding the requirements, interfaces, verification plans and system description throughout the system development life cycle. To overcome difficulties to maintain and synchronize this data scattered across many documents and standards, Model-Based Systems Engineering (MBSE) has been developed to enhance the ability to capture such information through the use of models (INCOSE, 2015).

This model representation offers new opportunities to verify engineering data through model analysis and execution. The potential to increase verification through model-based techniques has been a key initiative in the last decade (NDIA, 2011), challenging both systems and model-based engineering communities. Despite specialized domain or discipline techniques being developed in isolated pockets, much work is left to be done considering the virtual verification of models representing the integration of multiple domains, disciplines, and subsystems to analyse performance on a system level. This early virtual or digital verification has been recognized as a

key objective to meet the demands of current industrial product development (Maropoulos & Ceglarek, 2010) and interest has been growing towards the combined use of MBSE and digital verification to reduce errors being discovered where they are the most expensive in the integration and final verification phases of development (Frank, Broodney, Orion, & Kordova, 2016). However, the transition to MBSE is not a simple process as it involves changing over from legacy practices relying on documentation not only in the organization but also in the organization's network of suppliers and clients (Friedenthal et al., 2014). Further underscoring this difficulty, the cost of institutionalizing model-based methods may be too expensive to implement in a single project (NDIA, 2011) meaning that mixed document and model-based methods will occur for any company starting to adopt MBSE.

Despite the new verification opportunities provided by MBSE, these constraints make it infeasible to model and verify every element in the design of a complex system. A verification strategy must be drawn which incorporates various verification techniques to reduce risk in projects while controlling the resources to perform verification. As a precursor to developing such a strategy, this research seeks to identify model-based verification techniques applicable to systems engineering projects. More important than identification, it should be determined what the impacts of these techniques are and what restrictions they face when being implemented in real projects. To make this assessment, the current state of

model-based methods in systems engineering should be determined to identify possible opportunities and restrictions to developing a global verification approach based on model-based techniques. For this purpose, this study has been carried out with a group of engineering organizations in the AFIS¹ community to answer the following questions:

1. What are the desired benefits of model-based verification for companies adopting MBSE?
2. What model-based verification methods may be used to obtain these benefits in an MBSE approach?
3. What restrictions, both technical and organizational, restrict the efficiency of MBSE verification in real projects?
4. What are the differences between the published results of applying verification in an MBSE framework compared to the results of a broader systems engineering community?
5. Why do these gaps exist and what opportunities are available to improve the verification outcomes of the model-based techniques currently in use?

To approach these questions, a study composed of 16 engineering organizations located in the Rhône-Alpes region of France was conducted. Each organization, despite coming from different industries, utilize systems engineering to manage the complexity of their project development life cycles and modeling for MBSE or model-based verification. The experiences of the organizations were analysed through two workshops followed by multiple short work studies to determine their verification needs and the outcomes of their MBSE approaches to date. The analysis of the collected data on model-based verification techniques utilized in this community highlights the difficulties, impacts, and efficiency of these techniques contrasted to the experiences shown in MBSE literature and case studies.

2 MODEL-BASED VERIFICATION

In the context of this report, model-based verification refers to the verification of engineering data in the form of a model. There exist three general types of verification for such data – semantic review, testing, and formal methods (Douglass, 2015).

▪ Verification by semantic review

A semantic model is “A chart or diagram, having an underlying machine-readable representation, which shows object relationships, structure, or time sequencing of actions” (Baker et al., 1996). SysML Activity and Block Definition Diagrams may be given as examples.

Review of a semantic diagram is verification by inspection that utilizes human reasoning accompanied by simple methods of measurement to judge correctness (INCOSE, 2015). It includes solving a static analytical model.

¹ AFIS : French Association of Systems Engineering (Association Française d’Ingénierie Système)

▪ Verification by testing

Testing is verification through the execution of a model. For testing to occur, a model must be created in an executable form for a dynamic analysis study of how outputs of the system vary over time from the expected results. Testing requires a test case including a specified environment, inputs, initial conditions and all actions that must occur for the requirement under study to be verified. Testing is limited compared to formal methods in that it seeks to discover errors but does not prove that errors do not exist (Schamai, 2013). Simulation is a type of testing where the focus is on placing the model in a specific environment and replication of real system behaviour.

▪ Verification by formal methods

Formal techniques show correctness of a model through mathematical proofs. They are comprehensive verification methods as the verification result holds for any combination of inputs or system states; however, the difficulty of performing the techniques typically limits them to critical system properties (Douglass, 2015). A formal method seen in MBSE is *model checking* where numeric or symbolic computation is used to show that a system property holds for every possible system state. The technique is limited to system models with a limited number of states, due to the fact that in a complex system it may be impossible to analyse every possible execution in its entire state space (Schamai, 2013).

3 STATE OF THE ART OF VERIFICATION INTEGRATED WITH MBSE

To prepare the field study, a review was performed of the model-based verification literature focusing on the conception and development activities. To remain close to the background knowledge of the participant companies, the literature presented was pulled from INCOSE publications between 2010-2017 by searching with combinations of the key words “model”, “based”, “verification” and “virtual verification”. Papers were selected on the basis that they presented verification activities using models within an MBSE context.

The review includes relevant secondary sources from the INCOSE publications, primarily references to the journals of IEEE, ITEA, Models and work presented by groups such as NASA, JPL, the Computer and Information Sciences department of Linköping University, and the John Hopkins University Applied Physics Laboratory. This review presents the process and the methods of verification activities without prejudice to the modeling language or MBSE methodology applied.

Table 1 shows a summary of the techniques identified in the reviewed literature organized by their application in the traditional SE technical processes. The first column gives an identifier for the kind of technique. The second column shows the process in which the techniques may be applied (according to the technical processes of IEC 15288). Column 3 gives the verification type and column 4 the type of model. The last column gives the targeted element for verification. Due to space

limitations, the papers belonging to each category cannot be cited. For the complete review, the reader may consult (Laing, 2017).

Identifier	Process	Verification Type	Model Type	Verified Element
V1.1	Business or mission analysis	Semantic review	Semantic diagrams	Completeness of problem space
V2.1	Stakeholder needs and requirements definition process	Semantic review	Behaviour diagram	Stakeholder and needs identification completeness
V3.1	System requirements definition process	Semantic review	Requirement diagram	Completeness of system requirements or rational for omitting certain needs
V3.2		Testing	Executable form of semantic diagram(s)	System requirement specification
V3.3		Formal - Model Checking	Executable form of semantic diagram(s)	System requirement specification
V4.1	Architecture definition process – Functional and Logical	Semantic review	Semantic architecture	System architecture quality
V4.2		Semantic review	Requirement diagram	Requirement traceability
V4.3		Testing	Executable form of semantic diagram(s)	System requirement specification
V4.4		Formal - Model checking	Executable form of semantic diagram(s)	System requirement specification
V4.5		Formal – Ontology	Semantic architecture	System architecture correctness
V5.1	Architecture definition process – Physical	Semantic review	Requirement diagram	Requirement traceability
V5.2		Semantic review	Parametric diagram	Requirement specification
V6.1	Design Definition Process	Testing	Multiphysics system model	System requirement specification
V6.2		Testing	Multiphysics system model	System requirements
V6.3		Review (3D model)	Discipline simulation model	System requirement specification
V6.4		Testing	Discipline simulation model	System requirement specification

Table 1. Summary of verification techniques found in the literature

4 METHODOLOGY OF THE STUDY

The study was carried out with the support of the Rhône-Alpes chapter (CRR) of AFIS. The CRR consists of several organizations producing services and products similar in nature and complexity to MBSE publications and are therefore good candidates for comparison. Utilizing the exchanges within this community two stages of the Design Research Methodology (Blessing & Chakrabarti, 2009) for research in engineering design were performed. A first stage determined the success criteria for integrating verification in an MBSE approach per the AFIS systems engineering community followed by a second stage descriptive study of the state of the art literature and current practices to determine how attributes of model-based verification techniques are linked to the identified criteria. Finally, by contrasting the state of

the art to current practices in the AFIS community, an analysis is conducted of the more tangible criteria to determine how they are influenced by certain model-based verification factors and propose a basis for improving verification in MBSE. This research was conducted in this order: first, a community workshop to identify success criteria followed by two sets of data collection (through a second workshop and interviews) and a final analysis synthesis. Table 2 lists the 16 organizations consulted during the study and marks their participation in the two workshops. The heterogeneity of the domains demonstrates the broad interest in MBSE verification.

	Organization	1 st	2 nd
		WS	WS
1	Civil nuclear services firm	X	X
2	Industrial engineering research group A	X	X
3	Industrial engineering research group B	X	
4	Software and embedded system consultancy A	X	
5	Software and embedded system consultancy B	X	X
6	Electronics design and manufacturer A	X	X
7	Electronics design and manufacturer B	X	X
8	Defence systems design and manufacturer	X	
9	Medical instruments firm A	X	
10	Software solutions provider	X	
11	Engineering simulation firm	X	
12	Medical instruments firm B		X
13	System modeling software organization		X
14	Software and embedded system consultancy C		X
15	Electronics design and manufacturer C		X
16	Medical instruments firm C		X

Table 2. Attendance of Workshops (WS)

5 SUCCESS CRITERIA FOR VERIFICATION IN MBSE

The initial workshop was composed of 11 organizations and 22 participants. The industries represented may be seen in Table 2. During this workshop, a list of 10 needs concerning the deployment of verification within an MBSE framework were produced within a brainstorming and consensus definition session. Each need is a success criteria for how models must be used to improve verification, both early in design and in the final verification stages. The criteria and their explanations are listed below:

C1. Reuse of models

Reuse of models deals with the ability to reuse the modeling efforts required to perform a verification technique.

These efforts may be reused in future projects or later within the same one.

C2. Completeness of verification

Completeness of verification deals with the ability of the techniques to increase coverage above the level offered by the document based systems engineering approach.

C3. Test automation

Test automation relates to the ability to use the model-based techniques to automate verification.

C4. Collaboration and exchange tool

Collaboration and exchange refers to the ability of the technique to foster collaboration between parties using different tools or in different disciplines.

C5. Coherence between abstraction levels

Coherence between abstraction levels deals with the ability of the technique to be applied at one abstraction level effectively without the complexity of incorporating multiple system and abstraction levels.

C6. Detection & anticipation of problems

Detection and anticipation of problems deals with the technique's ability to find errors within the system due to changes or regular iterations of the system design.

C7. Verification of specification

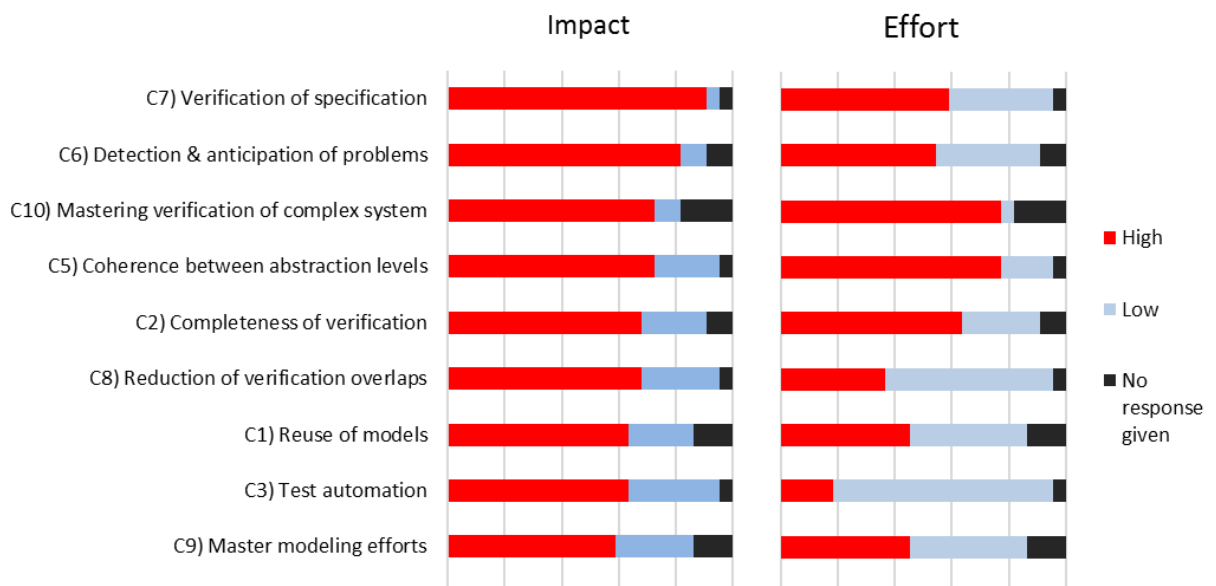
Verification of specification refers to the ability of the technique to verify a requirement specification early in the project before physical prototypes of the system are implemented.

C8. Reduction of verification overlaps

Reduce instances of work being verified multiple times at different project steps or abstraction levels.

C9. Master modeling efforts

Minimize the cost and effort of modeling while still receiving benefits from MBSE.



C10. Mastering verification of complex systems

Reducing the number of errors discovered in implementation, integration, verification and in the field.

To better understand the perceived benefits of obtaining each criteria, the AFIS community was asked to rank each by the effort required for their implementation and their possible impact on the project. In this way, each need could be labelled by the community as having either low or high effort to implement and either low or high positive impact on the verification of the project. The response of the participants is shown in Figure 1 as a percentage of 22 individuals. A first comment is that almost every criterion is viewed as having a high impact on the verification process, showing that model-based verification raises many expectations from practitioners and that its adoption would be dependent on its success in many aspects. This imposes constraints on deployed techniques to be efficient in multiple dimensions such as mastering modeling effort and reuse while supporting communication and new problem detection.

Concerning the perceived effort to reach each success criteria, a consensus is difficult to obtain. Depending on the experience of companies, the visions are often very different. The ninth criteria, “mastering modeling effort”, highlights this discrepancy as the participating companies were split equally between the two possible responses.

6 MODEL-BASED VERIFICATION IN MBSE FROM AN INDUSTRIAL PERSPECTIVE

In the second workshop, companies were asked to fill in questionnaires designed to capture any experience they might have had with the verification techniques shown in Table 1. This data along with follow up interviews of five of the organizations make up the industrial perspective in this analysis. In this section, the verification techniques classified from the state of the art (Table 1) are analysed against this industrial outlook. This involves analysing the effects of model-based verification in literature and current practice on the identified success criteria. None of the identified criteria were found to be directly measurable by the data collected; however, the influences of model-based verification were more tangible on certain criteria than others. This analysis focuses on these criteria. The results of this analysis for each technique are given in Table 3.

C1 to C7 have been considered tangible because they are direct attributes of a model-based verification technique being utilized in an MBSE framework for a given project. The other three criteria C8) Reduction of verification overlaps, C9) Mastering of modeling efforts, and C10) Mastering of verification of complex systems would require measurable data between multiple projects in the same organization; which was not collected in this research.

The industrial data validated a number of the classified state of the art techniques. Starting with the techniques for stakeholder model collaboration, V1.1 and V2.2.

Both the literature and data sources show these techniques effective for stakeholder needs elicitation and analysis and as semantic diagrams are easily saved in the system model, they are reusable. However, the AFIS community found that it was not always certain that their use for collaboration between stakeholders was efficient as there is no guarantee that two different parties understand a semantic model in the same manner. The system modeler must be careful to present something understandable to the target stakeholder without expressing a solution to their problem. This may take the form of transferring to a specialized modeling language in a certain industry standard or using a more adapted language for a large audience.

The use of verification techniques related to establishing traceability in the system model at all levels (V3.1, V4.2, and V5.1) were mirrored between the state of the art and industrial practices. In both sources, these techniques were viewed as methods to deal with complexity between modeling and abstraction levels and detect possible errors within the requirement specification when design changes occur. Techniques V4.2 and V5.1 may occur automatically; while V3.1 relates to stakeholder needs and is therefore based on human reasoning to determine if the appropriate needs have been included in the project. In the case of reused architectures and requirements, certain methodologies have suggested how packages including links between elements to perform traceability analysis may be reused to reduce the work of re-establishing traceability in each new project (Smith, 2014).

Additionally, black box testing was employed similarly between the two communities both in formal and informal versions. While its usage was not prolific within AFIS, the techniques V3.2 and V3.3 were considered useful to conduct early verification of specifications. In the case of AFIS, the focus of verification was in areas related to critical requirements before moving forward with the design. Additionally, examples were found of modeled black box specifications being reused to inform the system model design and later to build the final verification test cases. However, V3.3 formal model checking was seen to have a negative impact on C5 as the behaviour of a black box is difficult to predict accurately without specifying internal behaviour.

Generally, for the techniques related to testing architectures (V4.1 and V4.3), it was observed that the difficulty of modeling and maintaining exhaustive architectures along with a lack of know-how concerning effective testing dissuaded companies from investing in this type of verification. V4.1 was even seen to have a negative effect on C7, as one company with experience deploying quality metrics for system architecture did not find a positive influence on verification. However, a strong agreement may be seen in V4.5, the use of ontologies to enforce architecture construction rules. As a formal method, ontologies do not need specific test cases for each requirement or between projects. Therefore, they may be reused and verification may occur automatically as the model is constructed. One strongpoint of

using such ontologies is that they ensure the correct modeling rules are followed; allowing models to be imported, exported, or integrated between projects or into later stages of the current one. In this way, ontologies are expected to support certain criteria such as reusability and Organization 14 and 6 are implementing them in their practices.

For the verification techniques concerning the use of discipline models, virtual verification experience was widely seen at the industries studied. However, in these cases, communication between the system model and discipline simulation tools was not seen. Meaning that these verification steps were taken independently of the MBSE approach and therefore do not affect the success criteria.

7 DISCUSSION

In general, verification techniques were found to be applied the most frequently at high abstraction levels. This could be a reflection of the state of the industry. High abstraction levels require less investment in terms of modeling while offering better understanding of uses cases and state behaviour. Communication of these models was seen to affect all stakeholders. This included verification engineers, who were able to use the models to inform their bench testing procedures.

The companies interviewed generally welcomed verification techniques for functional and logical architectures, but capability levels rest well below those seen in the literature. The cost of building and maintaining the model seems to have discouraged work in this direction.

Finally, while verification by simulation of discipline models was widely seen as very important for the companies interviewed, there was no evidence that an MBSE approach has a direct effect on these practices. This could be for two reasons. One, the literature identifies data transfer from the central system model to the discipline models to be crucial to improving the efficiency of virtual verification. However, companies beginning to deploy MBSE have not had time to homogenize their toolchains and facilitate this transfer. Secondly, as described in the paragraph above, modeling at the functional and logical architectural levels is not mature so it is not surprising that these modeling efforts are not reused at the discipline level.

This first challenge of connecting a tool chain is mirrored in the state of the art literature. It can be viewed as a long-term goal for companies, researchers and tool vendors to improve in this area to increase the use of virtual verification in MBSE in the near future.

The second challenge of improving modeling at the functional and logical abstract levels represents a current opportunity for companies to begin adjusting their practices today. The literature reviewed suggests that verification techniques at this architectural level have a positive impact on the success criteria identified, it rests to be seen if increasing modeling investment and the use of

these techniques could lead to a positive impact on a system development project.

8 CONCLUSION

Through the participation of 16 engineering organizations located in the Rhone-Alpes region of France, key success criteria for integrating model-based verification with MBSE were identified allowing analysis of top verification methods presented in MBSE publications. To further validate the use of these techniques, analysis of the promising model-based verification techniques found in literature was conducted through different industrial perspectives to identify their feasibility and barriers to their successful application.

The findings in this study paint a picture of the current practice in integrating verification and MBSE as well as the impact of implementing such practices in companies beginning to adopt MBSE approaches. Agreement was discovered between many industrial and state of the art techniques conducted in the first processes of systems engineering to verify stakeholder and system requirements; with many techniques relating to functional and behavioural requirements leading to more consistent verification during design and in the final verification stages.

At the same time, barriers were discovered towards implementing techniques meeting the identified success criteria for verifying architectures and detail design in both the state of the art and industrial practices. The techniques for verifying architecture shows gaps as current modeling practices do not inform engineers on how to construct verifiable architecture nor direction on which requirements are verifiable with this approach.

Additionally, techniques related to verifying multiphysics and discipline designs, important for error reduction in integration, were not seen to be integrated in an MBSE framework due to technical challenges as well as missing aspects of the central system model.

While these conclusion show that techniques in the above two areas remain difficult to deploy, the industrial perspectives collected in this research suggest applying them could lead to a high positive impact on the project. Future research should focus on developing the verification techniques shown to impact the success criteria identified in this study.

ACKNOWLEDGMENTS

The authors give thanks to all the workshop participants and the CRRA chapter of AFIS for supporting this research.

REFERENCES

- Baker, L., Clemente, P., Cohen, B., Permenter, L., Purves, B., & Salmon, P., 1996. Model Driven System Design Working Group: foundational concepts for model driven system design. *INCOSE International Symposium*, 6(1), p.1179–1185.
- Blessing, L. T., & Chakrabarti, A., 2009. DRM, a design research methodology. *Springer Science & Business Media*.
- Crisp, H. E. 2007. Systems engineering vision 2020. Seattle, Washington
- Douglass, B. P. 2015. Agile systems engineering. *Morgan Kaufmann*.
- Frank, M., Broodney, H., Orion, U., & Kordova, S. K. 2016. From Common Strategies and Approaches to Virtual Integration. In *INCOSE International Symposium*, 26, pp. 1988-1999).
- Friedenthal, S., Moore, A., & Steiner, R. (2014). *A Practical Guide to SysML: The Systems Modeling Language*. Morgan Kaufmann.
- INCOSE, 2015. Systems engineering handbook: A guide for system life cycle processes and activities. Ed. Walden, D. D., Roedler, G. J., Forsberg, K., Hamelin, R. D., & Shortell, T. M. – Wiley
- Laing, C., 2017. Towards the integration of verification in Model-Based Systems Engineering. *Master thesis dissertation Grenoble-INP*.
- Maropoulos, P. G., & Ceglarek, D. 2010. Design verification and validation in product lifecycle. *CIRP Annals - Manufacturing Technology*, 59(2), p. 740–759.
- Murphy, L., & Collopy, P. 2012. A work-centered perspective on research needs for systems engineering with models. *Procedia Computer Science*, 8, p. 315–320.
- NDIA, 2011. Final Report of the Model Based Engineering Subcommittee of the National Defense Industrial Association (NDIA).
- Schamai, W., 2013. Model-Based Verification of Dynamic System Behavior against Requirements: Method, Language, and Tool. *Doctoral dissertation, Linköping University*
- Smith, W. B., 2014. Re-Use Libraries Leveraging Model-Based Systems Engineering to greatly increase engineering productivity. *INCOSE International Symposium*, 24(1), p.298–312.

Id	Verification type	Model Type	Verified Element	C1		C2		C3		C4		C5		C6		C7	
				SOTA	AFIS	SOTA	AFIS	SOTA	AFIS	SOTA	AFIS	SOTA	AFIS	SOTA	AFIS	SOTA	AFIS
V1.1	Semantic review	Semantic Diagrams	Completeness of problem space	+	+					+	+/-	+	+			+	+
V2.1	Semantic review	Behavior diagram	Stakeholder and needs identification completeness	+	+					+	+/-	+	+			+	+
V3.1	Semantic review	Requirement diagram	Completeness of system requirements	+	+					+	+	+	+			+	+
V3.2	Testing	Executable form of semantic diagram(s)	System requirement specification		+							+				+	+
V3.3	Formal - Model Checking	Executable form of semantic diagram(s)	System requirement specification	+						+						+	+
V4.1	Semantic review	Semantic architecture	System architecture quality	+						+						+	-
V4.2	Semantic review	Requirement diagram	Requirement traceability	+	+					+	+					+	+
V4.3	Testing	Executable form of semantic diagram(s)	System requirements specification	+						+						+	+
V4.4	Formal - Model checking	Executable form of semantic diagram(s)	System requirement specification	+						+						+	
V4.5	Formal - Ontology	Semantic architecture	System architecture correctness	+	+					+	+					+	+
V5.1	Semantic review	Requirement diagram	Requirement traceability	+						+	+					+	+
V5.2	Semantic review	Parametric diagram	Requirement specification	+						+						+	
V6.1	Test	Multiphysics system model	System requirement specification	+/-						+/-						+	+
V6.2	Test	Multiphysics system model	System requirements	+/-						+/-						+	
V6.3	Review (3D model)	Discipline simulation model	System requirement specification	+/-						+/-						+	
V6.4	Test	Discipline simulation model	System requirements	+/-						+/-						+	
+	Positive impact on criteria	Negative impact on criteria	Negative impact on criteria	Barriers exist to obtain a positive impact	Unknown impact on criteria	Identified in state of the art	Identified by AFIS organization										

Table 3 Influence of techniques on success criteria

Reconfigurable transfer line balancing problem: A new MIP approach and approximation hybrid algorithm.

Y. LAHRICHI, L. DEROUSSI, N. GRANGEON, S. NORRE

LIMOS / CNRS UMR 6158
Campus Universitaire des Cézeaux
1 rue de la Chebarde
63178 AUBIERE CEDEX - FRANCE

youssef.lahrichi@isima.fr, laurent.deroussi@uca.fr, Nathalie.grangeon@uca.fr, sylvie.norre@uca.fr

ABSTRACT: We consider the problem of balancing reconfigurable transfer lines. The problem is quite recent and motivated by the growing need of reconfigurability in the new *INDUSTRY 4.0* context. The problem consists into allocating different tasks necessary to machine a single part to different workstations placed into a serial line. The workstations can contain multiple machines operating in parallel and the tasks allocated to a workstation should be sequenced since the machines considered are mono-spindle head CNC machines and setup times between operations are needed to perform tool changes. Besides precedence constraints between operations are considered. In this article we suggest a new MIP approach based on formulation presented by Andrés et al.(2008) for a resembling problem, we suggest effective pre-processing procedures and a lower bound, and we introduce a novel hybrid approach that approximates the optimal solution when the setup times are bounded by the completion times. Experimentation are performed on benchmark instances and the methods are compared with those suggested in the literature.

KEYWORDS: Transfer line balancing, Reconfigurability, Industry 4.0, Mixed integer programming, Hybrid algorithms, approximation algorithms

1 INTRODUCTION

New consuming trends, global competition and growing variety in demand in the actual economical context raises an important issue in transfer line design. In fact, shortening life cycle times imposes the consideration of reconfigurability in transfer line design. The modern transfer line should be easily and cost-effectively reconfigurable to address two different issues: the variability in production size and the variability in the product specifications. While the first one imposes a variation in cycle time, the second one is linked to the set of tasks involved.

To address this issue, Koren et al.(1999) suggested in late 1990s the novel concept of **Reconfigurable Manufacturing System (RMS)**. A RMS could be seen as a serial line of workstations (corresponding to the stages in Figure.1). Each workstation is equipped by multiple machines operating in parallel. Part units are moved from a workstation to another thanks to a *conveyor*. The part is delivered then to the first available machine in a workstation from the *gantry*.

The RMS highly addresses the issue of production size variability. Indeed the ability to add or remove

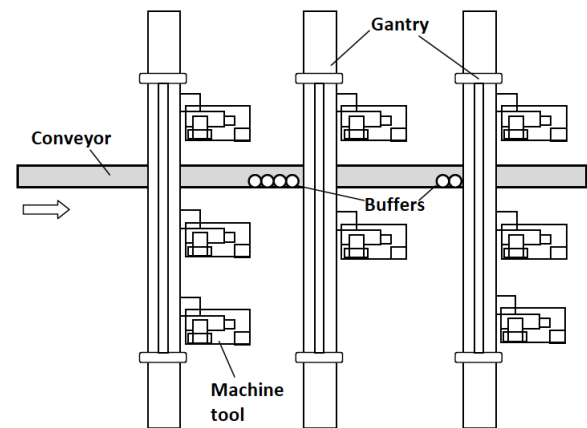


Figure 1 – Reconfigurable Manufacturing System: [Koren 2010]

a machine in a workstation allows monitoring the cycle time with high granularity which is referred to as **scalability** [Wang 2017]. Besides RMS offers a good trade off between productivity and flexibility while *Dedicated Manufacturing System* are highly productive but very poorly flexible and *Flexible Manufacturing System* highly flexible but very expensive and

almost never profitable.

Despite being scalable and profitable, RMS could not address the issue of variability in product specifications only if equipped with **mono-spindle head machines**. Indeed, those machines can perform a huge set of operations, each machine being equipped with a **tool magazine**. To perform an operation, a machine needs a specific tool. Thus, setup times between operations must be considered in addition to operation times in order to perform tool changing.

Once equipped with mono-spindle head machines, RMS addresses both the production size and product specifications variability issues: whenever one or both of these elements comes to change, the manufacturer can easily adjust the production by performing a *reconfiguration* of the system which can be seen as the process of balancing the transfer line taking into consideration setup times and multiple parallel machines at each workstation. Once this operation performed, machines are then added or removed from workstations if necessary, and machines are remotely configured to perform the new sequence of operations; RMS allowing to perform those two steps rapidly and cost-effectively. There is no need of physical machine reconfiguration, since all machines are equipped with the same tool magazine and can thereby perform the same set of operations.

The problem of balancing RMS appears then to be of strategic importance for the manufacturer. In section 2 we define this problem and introduce basic notations, in section 3 the related work is described then a lower bound and an example are introduced in section 4. A new MIP approach is presented in section 5 and a novel approximation hybrid algorithm is then introduced in section 6. An experimental study was also performed, it is showing highly promising results. It is covered in the section 7 of the paper.

2 PROBLEM DEFINITION

The instance of the optimization problem could be described by the following data:

- The set of operations, the corresponding times, setup times and precedence relations.
- A maximum number of workstations to be used placed serially.
- A maximum number of machines per workstation.
- A cycle time.
- A maximum number of operations to be allocated to a workstation.

The optimization problem consists then in finding an allocation of the operations to the workstations and determining a number of machines per workstation while **minimizing** the number of machines used and respecting the following **constraints**:

- The sum of the times of the operations and the induced setup times of the sequence allocated to a workstation divided by the number of machines in that workstation must not exceed the cycle time.
- Precedence constraints must be respected: when an operation i precedes an operation j , the workstation to which the operation i is allocated must be less (placed before in the line) or as the workstation to which the operation j is allocated.
- The number of workstations must not exceed the maximum number of workstations.
- The number of operations allocated to a workstation must not exceed the maximum number of operations per workstation.
- The number of machines in a workstation must not exceed the maximum number of machines per workstation.

For the rest of the paper, we use the notations presented in Table 1.

N	Set of operations, indexed on $\{1, 2, \dots, n\}$.
S	Set of workstations, indexed on $\{1, 2, \dots, s_{max}\}$ s_{max} denoting the maximum number of workstations.
P	Set of couples $(i, j) \in N \times N$ such that i precedes j (also denoted: $i \ll j$).
M	Maximum number of operations to be allocated to a workstation.
M'	Maximum number of machines to be in a workstation.
C	Cycle time.
t_i	Completion time of operation i .
$t_{i,j}$	Set-up time to be considered when operation i is performed just before operation j in some workstation

Table 1 – Table of notations.

3 RELATED WORK

This problem could be seen as an assembly line balancing problem. Those problems have been well studied in the literature; however those considering parallel machines or sequence dependent setup times have rarely been considered. The originality of the problem comes from the consideration of both elements.

We can only list two papers dealing with this issue with an exact approach: (Essafi et al., 2010) and (Borisovsky et al., 2014).

Both approaches fail to solve the problem for medium to large scale instances. Essafi et al.(2010) suggest a MIP approach while Borisovsky et al.(2014) uses a set partitioning model coupled with a constraint generation algorithm.

The MIP approach of Essafi et al. (2010) lies on modelling the overall sequence constituted of the concatenation of the sequences of all the workstations :it uses the variables $x_{i,q}$ (i for the operation and q for the position in the sequence). It can solve instances with 15 operations while the other approach (Borisovsky et al., 2014) can solve instances with up to 50 operations.

4 EXAMPLE AND LOWER BOUND

We first introduce a lower bound for the problem, we then draw an example and its optimal solution.

Taking into consideration the fact that for every workstation j , its workload time (W_j) and its number of machines M_j should satisfy:

$$W_j \leq C.M_j$$

Then, the total workload $W = \sum_{j \in S} W_j$ and the total number of machines $m = \sum_{j \in S} M_j$ must satisfy:

$$W = \sum_{j \in S} W_j \leq \sum_{j \in S} C.M_j = C. \sum_{j \in S} M_j = C.m$$

i.e:

$$W \leq C.m \tag{1}$$

Let us now assume that $n > s_{max}$. Then the workload W must satisfy:

$$W \geq \sum_{i \in N} t_i + \lambda_{1+n-s_{max}} \tag{2}$$

where $\lambda_{1+n-s_{max}}$ denotes the sum of the $1+n-s_{max}$ smallest setup times.

Indeed, (2) is true because the workload is composed of the operations times ($\sum_{i \in N} t_i$) and the induced setup times. And since $n > s_{max}$, there must be at least $n - s_{max}$ operations that are "not alone" at the workstation they are affected to. Any solution must then consider at least $n - s_{max} + 1$ setup times.

Then, we have from (1) and (2):

$$\sum_{i \in N} t_i + \lambda_{1+n-s_{max}} \leq C.m$$

From this equation we deduce a lower bound for the number of machines used if $n > s_{max}$:

$$z_{lb} = \left\lceil \frac{\sum_{i \in N} t_i + \lambda_{1+n-s_{max}}}{C} \right\rceil \tag{3}$$

If $n \leq s_{max}$, the classical lower bound for SALBP is still available :

$$z_{lb} = \left\lceil \frac{\sum_{i \in N} t_i}{C} \right\rceil$$

Let us now give the example of a small instance, compute the lower bound and give an optimal solution.

The instance is described by the following data:

- The part requires the execution of 7 operations numbered from 1 to 7 ($n = 7$).
- At most 3 stations can be used. ($s_{max} = 3$).
- Precedence constraints are given by:

$$P = \{(1, 3), (2, 3), (3, 4), (4, 5), (5, 6), (5, 7)\}$$

represented in Figure.2.

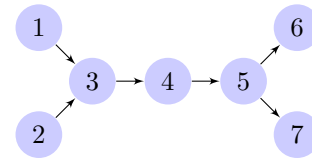


Figure 2 – Precedence graph.

- $M = 3$, Maximum number of operations that could be allocated to a workstation.
- $M' = 3$, Maximum number of machines that could be hosted by a workstation.
- Completion times are represented in Table 2:

i	1	2	3	4	5	6	7
t_i	1.5	1	3.5	1.5	2.5	3	1

Table 2 – Operations times.

- Setup times are represented in Table 3.
- $C = 2.5$, cycle time.

$t_{i,j}$	$j = 1$	2	3	4	5	6	7
$i = 1$	X	0.5	10	10	10	10	10
2	10	X	0.5	10	10	10	10
3	0.5	10	X	10	10	10	10
4	10	10	10	X	0.5	10	10
5	10	10	10	0.5	X	10	10
6	10	10	10	10	10	X	0.5
7	10	10	10	10	10	0.5	X

Table 3 – Setup times.

Computing the lower bound (equation 3) gives: $z_{lb} = 7$.

We describe a realizable solution by allocating to the first workstation the sequence 1, 2, 3 and 3 machines, to the second workstation the sequence 4, 5 and 2 machines and to the third workstation the sequence 6, 7 and 2 machines.

The solution is realizable since the workload divided by the number of machines for each station does not exceed the cycle time:

For station 1: $(t_1 + t_2 + t_3 + t_{1,2} + t_{2,3} + t_{3,1})/3 = 2.5$

For station 2: $(t_4 + t_5 + t_{4,5} + t_{5,4})/2 = 2.5$

For station 3: $(t_6 + t_7 + t_{6,7} + t_{7,6})/2 = 2.5$

Since the solution is realizable and its cost equals the lower bound, it is optimal.

5 A MIP FORMULATION AND PRE-PROCESSING PROCEDURES

In this section, we describe a new MIP approach based on a formulation for the sequence-dependent setup times assembly line balancing problem introduced by Andrés et al.(2008) that has not been experimented and does not consider parallel machines in workstations. Besides we suggest novel pre-processing procedures that computes:

- $e(i)$: The earliest station to which operation i could be affected.
- $l(i)$: The latest station to which operation i could be affected.
- M_j : The maximum number of operations that could be affected to workstation j .

We first give the pre-processing algorithms, we then describe the MIP approach for our problem.

While x denotes the earliest station to ensure that the workload of i and all its predecessors is done, the condition $x < y$ captures the fact that x stations are

Algorithm 1 $e(i)$ ($i \in N$: opération)

$P_i := \{i\} \cup \{j \in N; (j, i) \in P\}$ (i and its predecessors).

$$x := \left\lceil \frac{\sum_{i \in P_i} t_i}{C.M'} \right\rceil$$

$$y := \left\lceil \frac{\sum_{i \in P_i} t_i + \lambda_{1+|P_i|-x}}{C.M'} \right\rceil$$

while $x < y$ **do**

$$x := x + 1$$

$$y := \left\lceil \frac{\sum_{i \in P_i} t_i + \lambda_{1+|P_i|-x}}{C.M'} \right\rceil$$

end while

Return x

not enough to ensure that the induced setup times are also processed. The second algorithm computing the latest workstation is quite similar to the first one but considering the successors instead of predecessors and counting from s_{max} instead of 1.

Algorithm 2 $l(i)$ ($i \in N$: opération)

$S_i := \{i\} \cup \{j \in N; (i, j) \in S\}$ (i and its successors).

$$x := \left\lceil \frac{\sum_{i \in S_i} t_i}{C.M'} \right\rceil$$

$$y := \left\lceil \frac{\sum_{i \in S_i} t_i + \lambda_{1+|S_i|-x}}{C.M'} \right\rceil$$

while $x < y$ **do**

$$x := x + 1$$

$$y := \left\lceil \frac{\sum_{i \in S_i} t_i + \lambda_{1+|S_i|-x}}{C.M_0} \right\rceil$$

end while

Return $s_{max} - x + 1$

Once having computed $e(i)$ and $l(i)$, we could review the maximum number of operations that could be affected to any workstation j as: $M_j = \text{Min}(M, |\{i, e(i) \leq j \leq l(i), i \in N\}|)$.

Another pre-processing procedure is run to remove redundant precedence constraints : $(i, j) \in P$ is removed if there exists a non trivial path between i and j in the precedence graph.

We could now describe the MIP approach. This approach is based in modelling the workstations sequences. It uses the following binary **variables**:

-

$$x_{i,j,s} = \begin{cases} 1 & \text{If operation } i \text{ is affected to workstation } j \text{ in the } s^{th} \\ & \text{position of its sequence.} \\ 0 & \text{If not.} \end{cases}$$

•

$$y_j = \begin{cases} 1 & \text{If at least one operation is} \\ & \text{affected to workstation } j \\ 0 & \text{If not.} \end{cases}$$

•

$$z_{i,j,k} = \begin{cases} 1 & \text{If the operation } i \text{ is processed} \\ & \text{just before operation } j \text{ at} \\ & \text{workstation } k. \\ 0 & \text{If not.} \end{cases}$$

•

$$w_{i,j} = \begin{cases} 1 & \text{If the operation } i \text{ is affected to} \\ & \text{the last position of workstation } j. \\ 0 & \text{If not.} \end{cases}$$

•

$$v_{j,k} = \begin{cases} 1 & \text{If } k \text{ machines are affected to} \\ & \text{workstation } j. \\ 0 & \text{If not.} \end{cases}$$

We consider the **objective** of minimizing the number of machines used: $Min \sum_{j \in S} \sum_{k=1}^{M'} k.v_{j,k}$ under the **constraints**:

$$\sum_{j=e(i)}^{l(i)} \sum_{s=1}^{M_j} x_{i,j,s} = 1, \forall i \in N \quad (4)$$

This set of constraints ensures that every operation is affected to exactly one workstation at a unique position of its sequence.

$$\sum_{i \in N, e(i) \leq j \leq l(i)} x_{i,j,s} \leq 1, \forall j \in S, s = 1, \dots, M_j \quad (5)$$

This set of constraints ensures that at most one operation is affected to each position of the sequences of the workstations.

$$\sum_{i \in N, e(i) \leq j \leq l(i)} x_{i,j,s+1} \leq \sum_{i \in N, e(i) \leq j \leq l(i)} x_{i,j,s} \quad (6)$$

$\forall j \in S, s = 1, \dots, M_j - 1$

This set of constraints ensures that no position $s + 1$ in any workstation is taken by any operation unless the position s is also taken by some operation.

$$\sum_{k=1}^{M'} v_{j,k} = y_j, \forall j \in S \quad (7)$$

This set of constraints ensures that only one number of machines is chosen for every used workstation.

$$y_{j+1} \leq y_j, \forall j = 1, \dots, s_{max} - 1 \quad (8)$$

This set of constraints ensures that no workstation is used unless its precedent workstation is also used.

$$\sum_{j=e(i)}^{l(i)} \sum_{s=1}^{M_j} (M.(j-1) + s)x_{i,j,s} \leq \sum_{j=e(i')}^{l(i')} \sum_{s=1}^{M_j} (M.(j-1) + s)x_{i',j,s}, \forall (i, i') \in P \quad (9)$$

This set of constraints ensure that precedence constraints are satisfied.

$$\sum_{i \in N, e(i) \leq j \leq l(i)} \sum_{s=1}^{M_j} t_i.x_{i,j,s} + \sum_{i, i' \in N^2; i \neq i', e(i) \leq j \leq l(i), e(i') \leq j \leq l(i')} t_{i,i'}.z_{i,i',j} \leq C. \sum_{k=1}^{M'} k.v_{j,k}, \forall j \in S \quad (10)$$

This set of constraints ensure that the cycle time is not exceeded in any workstation.

$$x_{i,k,s} + x_{i',k,s+1} \leq 1 + z_{i,i',k}, \forall i, i' \in N^2, i \neq i', k \in \{e(i), \dots, l(i)\} \cap \{e(i'), \dots, l(i')\}, s = 1, \dots, M_k - 1 \quad (11)$$

This set of constraints ensures that if operation i' is followed by operation i at station k then $z_{i,i',k}$ is put to 1.

$$x_{i,j,s} - \sum_{i' \in N; i' \neq i, e(i') \leq j \leq l(i')} x_{i',j,s+1} \leq w_{i,j}, \forall i \in N, \forall j \in \{e(i), \dots, l(i)\}, s = 1, \dots, M_j - 1 \quad (12)$$

$$x_{i,j,M_j} \leq w_{i,j}, \forall i \in N, \forall j \in \{e(i), \dots, l(i)\} \quad (13)$$

The constraints (13) and (14) ensure that $w_{i,j}$ is put to one whenever operation i is positioned in the last occupied position of workstation j .

$$w_{i,j} + x_{i',j,1} \leq 1 + z_{i,i',j}, \forall i \in N, i' \in N, i \neq i', j \in \{e(i), \dots, l(i)\} \cap \{e(i'), \dots, l(i')\} \quad (14)$$

This constraint ensures that if operation i is positioned in the last occupied position of workstation j and operation i' positioned in the first position of workstation j then $z_{i,i',j} = 1$ and consequently the setup time $t_{i,i'}$ is considered in (10).

The MIP is experimented in Section 7.

6 AN HYBRID APPROACH

Let's now describe a novel hybrid approach for our problem. Besides, we show that the algorithm is a 2-approximation if we assume that:

$$t_{i,j} \leq t_k, \forall i, j, k \in N \quad (15)$$

which seems to be a very reasonable and realistic assumption while considering industrial instances. (see section 7)

The algorithm consists of two steps. The first consists of solving a MIP while the second one is an exact dynamic programming algorithm for solving the ATSP:

- *Step 1:* Perform a parallel line balancing without taking setup times into consideration using a MIP approach.
- *Step 2:* For each workstation, perform a sequencing of the operations using a dynamic programming algorithm then if the number of machines in the workstations is insufficient to fulfill the cycle time, add the necessary ones. If the maximum number of machines is violated, then we return to step 1 by adding a constraint that forbids the assignment of the set of operations to the workstation. This step is described in more details in section 6.2.

The first subsection is devoted to the first step and the second subsection is devoted to the second step. We refer to this method as "BFSL" (*Balance First, Sequence Last*)

Let us now remark that the solution outputted by the algorithm is feasible and its overall cost (c) is given by the cost of the solution outputted by the step1 (c_1) plus the number of machines added in step 2 (m). i.e

$$c = c_1 + m$$

besides we have $c_1 \leq c^*$ where c^* denotes the optimal solution of the RMS balancing problem (because c_1 does not take setup times into consideration).

And thanks to (15) we have $m \leq c_1$ (because the number of setup times for each workstation is less or equal to the number of operations and then the workload involved by the setup times in each workstation is less or equal to the workload involved by the operations times). Those two inequations ($c_1 \leq c^*$, $m \leq c_1$) finally give:

$$c \leq 2.c^*$$

which shows the approximation ratio.

6.1 Step 1: Balancing without setup times

In this step we are concerned with balancing the RMS but without taking into consideration setup-times. This is done by the following MIP that is derived from the precedent one by removing unnecessary variables and constraints. We use the following **variables**:

- $$x_{i,j} = \begin{cases} 1 & \text{If operation } i \text{ is affected to workstation } j. \\ 0 & \text{If not.} \end{cases}$$
- $$y_j = \begin{cases} 1 & \text{If at least one operation is affected to workstation } j \\ 0 & \text{If not.} \end{cases}$$
- $$v_{j,k} = \begin{cases} 1 & \text{If } k \text{ machines are affected to workstation } j. \\ 0 & \text{If not.} \end{cases}$$

We consider the **objective** of minimizing the number of machines used: $Min \sum_{j \in S} \sum_{k=1}^{M'} k.v_{j,k}$

under the **constraints**:

$$\sum_{j=e(i)}^{l(i)} x_{i,j} = 1, \forall i \in N \quad (16)$$

This set of constraints ensures that every operation is affected to exactly one workstation.

$$\sum_{k=1}^{M'} v_{j,k} = y_j, \forall j \in S \quad (17)$$

This set of constraints ensures that only one number of machines is chosen for every used workstation.

$$\sum_{i \in N, e(i) \leq j \leq l(i)} x_{i,j} \leq M_j, \forall j \in S \quad (18)$$

This set of constraints ensures that the maximum number of operations to be allocated to a workstation is respected.

$$y_{j+1} \leq y_j, \forall j = 1, \dots, s_{max} - 1 \quad (19)$$

This set of constraints ensures that no workstation is used unless its precedent workstation is also used.

$$\sum_{j=e(i)}^{l(i)} j.x_{i,j} \leq \sum_{j=e(i')}^{l(i')} j.x_{i',j}, \forall (i, i') \in P \quad (20)$$

This set of constraints ensures that precedence constraints are satisfied.

$$\sum_{i \in N, e(i) \leq j \leq l(i)} t_i.x_{i,j} \leq C. \sum_{k=1}^{M'} k.v_{j,k}, \forall j \in S \quad (21)$$

This set of constraints ensures that the cycle time is not exceeded in any workstation.

The above MIP introduces far less variables and constraints than the first MIP (Section 5). Experimentation show that it can solve instances with very large set of operations.

6.2 Step 2: Sequencing operations in every workstation and adding necessary machines

From step 1, we are given the affectation of operations to the workstations with a number of machines at each workstation. We are now concerned with sequencing the operations in every workstation and since the setup times were not considered in step 1 we may be obliged to add some machines at some workstations to fill the cycle time constraint.

The sequencing problem is an ATSP where operations represent cities and set-up times distances between cities. This operation is performed with an exact dynamic programming algorithm introduced in Held and Karp (1962). We compute then the workload for every workstation and determine easily the number of machines to be added. If more machines than the the maximum authorized is required then we return to step 1 and a constraint forbidding the allocation of this set of operations to the workstation is added to the linear model.

7 EXPERIMENTAL RESULTS

We describe in this section the experimentation being held on a 16Go RAM computer with JAVA 8 and CPLEX (v12.7.0). Three sets of instances are considered, provided by Borisovsky et al.(2014) and described to be of industrial importance. However those instances consider additional constraints: accessibility, inclusion and exclusion. Those constraints are taken into consideration by adding the necessary linear constraints. Experimentation are summarized in Table 4. z^* and z_{lb} denote respectively the optimal solution and the lower bound. "MIP sol." denote the solution outputted by CPLEX after "MIP time" seconds. The duality gap is denoted by "MIP Gap". If the optimal solution is not found after 10,000 seconds, CPLEX is stopped and the actual best known solution is taken with its corresponding duality gap. d denotes the density of the precedence graph while d_s denotes the Scholl density:

$$d_s = \frac{2 \cdot 100 \cdot \sum_{i \in N} |P_i|}{n \cdot (n-1)} \text{ where } P_i \text{ denotes the predecessors of } i.$$

"BFSL" and "BFSL time" denote respectively the solution and the time of the hybrid approach.

All instances have 14 operations. First set of instances (p14-10) have a Scholl density in [5,15]. 6 instances out of 9 have been solved to optimum. With the MIP from Essafi et al.(2010) only 4 out of 9 instances have been solved to optimum with a time limit of 10,000 seconds. Second set of instances (p14-25) have a Scholl density in [15,25]. All instances have been solved to optimum. With the MIP from Essafi et al.(2010) only 5 out of 10 instances have been solved to optimum with a time limit of 10,000 seconds. Third set of instances (p14-40) have a Scholl density in [25,40]. All instances have been solved to optimum. With the MIP from Essafi et al.(2010) only 8 out of 10 instances have been solved to optimum with a time limit of 10,000 seconds.

The MIP was enable to solve instances with 20 operations within 10,000 seconds. All the instances were solved by the set partitioning model (Borisovsky et al., 2014), besides the set partitioning model was able to solve instances with 50 operations.

8 CONCLUSION AND PERSPECTIVES

We have presented in this paper a new MIP, and a novel hybrid approximation algorithm and held experimentation on both Benchmarks. The results are quite promising. We can make some remarks from the experimentation:

- The MIP is more efficient with instances having bigger Sholl density.
- The lower bound that we present is on average at 15% of the optimal solution.
- The hybrid approach gives medium results and is very fast.
- The MIP that we present is more efficient than the one presented in Essafi et al.(2010) and less efficient than the algorithm presented in Borisovsky et al.(2014).

However, we could take many directions as a continuation of this research:

- Improvement of the BFSL algorithm with posterior local search improvement algorithms for example.
 - The use of polyhedral approaches lying on the MIP.
 - Research could be done to show a better approximation ratio if there exists some $\lambda \in [0, 1]$ s.t:
- $$t_{i,j} \leq \lambda \cdot t_k, \forall i, j, k \in N$$
- Studying the problem in an uncertain context is a must to fill with INDUSTRY 4.0 requirements.

Instance	d	d_s	z^*	z_{lb}	MIP sol.	MIP time	MIP Gap	BFSL Sol.	BFSL time
p14-10-1	7.69	8.76	11	9	11	4512	0	14	0.78
p14-10-2	7.69	9.89	10	8	10	9558	0	11	0.69
p14-10-4	7.69	9.89	-	9	10	10000	17.04	12	1.93
p14-10-5	7.69	9.89	-	8	11	10000	26.06	13	0.80
p14-10-6	7.69	9.89	-	8	11	10000	25.29	13	0.80
p14-10-7	7.69	9.89	9	8	9	1482	0	10	1.52
p14-10-8	8.79	12.08	11	9	11	9491.69	0	13	0.43
p14-10-9	8.79	9.89	10	9	10	1021	0	12	0.17
p14-10-10	7.69	8.79	9	8	9	13428.47	0	13	4.48
p14-25-1	15.38	16.48	10	8	10	7336	0	13	1.91
p14-25-2	14.28	21.97	10	8	10	406	0	12	1.18
p14-25-3	10.98	21.97	10	8	10	585	0	12	2.07
p14-25-4	10.98	18.68	10	9	10	1543	0	11	0.52
p14-25-5	14.28	19.78	11	8	11	511	0	13	3.86
p14-25-6	15.38	25.27	10	8	10	498	0	10	0.96
p14-25-7	14.28	18.68	10	9	10	2772	0	12	1.82
p14-25-8	13.18	20.87	9	9	9	1822	0	13	0.22
p14-25-9	13.18	23.07	10	8	10	636	0	14	1.18
p14-25-10	12.08	17.58	10	8	10	2658	0	11	1.38
p14-40-1	19.78	36.26	11	8	11	229	0	13	1.31
p14-40-2	21.98	29.67	9	7	9	1778.32	0	12	0.93
p14-40-3	17.58	24.17	10	8	10	4006	0	12	0.41
p14-40-4	18.68	29.67	10	8	10	322	0	11	0.72
p14-40-5	14.28	26.37	10	8	10	838	0	12	1.78
p14-40-6	17.58	38.46	10	8	10	104	0	11	1.22
p14-40-7	17.58	26.37	9	7	9	1368	0	11	0.39
p14-40-8	19.78	26.37	9	8	9	491	0	11	0.40
p14-40-9	16.48	27.47	10	8	10	333.97	0	12	1.32
p14-40-10	16.48	29.67	9	8	9	409	0	10	0.41

Table 4 – Experimentation with benchmark instances.

ACKNOWLEDGMENTS

The authors acknowledge the support received from the *Agence Nationale de la Recherche* of the French government through the program "Investissements d'Avenir"(16-IDEX-0001 CAP 20-25).

REFERENCES

- Andrés, C., M. Cristóbal, and R. Pastor, 2008. Balancing and scheduling tasks in assembly lines with sequence-dependent setup times. *European Journal of Operational Research*, Vol.187, No. 3, pp. 1212-1223.
- Borisovsky, P. A., Delorme, X., and A. Dolgui, 2014. Balancing reconfigurable machining lines by means of set partitioning model. *International Journal For Production Research*, Vol. 52, pp. 4026-4036.
- Essafi, M., X. Delorme X., A. Dolgui A., and O. Guschinskaya, 2010. A MIP approach for balancing transfer line with complex industrial constraints. *CIRP Journal of Manufacturing Science and Technology*, Vol. 58, No.2, pp. 176-182.
- Held, M., and R. M. karp. 1962. A Dynamic Programming Approach to Sequencing Problems. *Journal of the Society for Industrial and Applied Mathematics*, Vol.10, pp. 196-210.
- Koren, Y., U. Heisel, F. Jovane, T. Moriwaki, G. Pritschow, G. Ulsoy, and H. Van Brussel, 1999. Economic benefits of reconfigurable manufacturing systems. *Annals of CIRP*, Vol. 42, No. 2, pp. 527-540.
- Koren, Y., 2010. The global manufacturing revolution - product-process-business integration and reconfigurable systems. John Wiley & Sons.
- Wang, W., Y. Koren, and X. Gu, 2017. Value creation through design for scalability of reconfigurable manufacturing systems. *International Journal For Production Research*, Vol. 55, No. 5, pp. 1227-1242.

REVISITING DIGITAL MOCK-UP FOR SME INVOLVED IN SYSTEMS ENGINEERING DEPLOYMENT

Vincent Chapurlat

IMT Mines Alès - LGI2P
site Louis Leprince-Ringuet, 7 avenue Jules Renard, 30100 Alès, France
vincent.chapurlat@mines-ales.fr

Maxence Lafon

CEA, Nuclear Energy Division
DMRC, SA2I
F-30207 Bagnols sur Cèze, France
maxence.lafon@cea.fr

Blazo Nastov

Axellience
4 Avenue des Saules, 59160 Lille
blazho.nastov@axellience.com

ABSTRACT: *This article revisits the concept of Digital Mock-Up to successfully deploy Model Based System Engineering (MBSE) in Small and Medium-sized Enterprises (SMEs) and to reassure the transition and the digital continuity of SMEs during and after deployment. For this purpose, we propose a formalized adaptation of Digital Mock-Up considering MBSE principles and SMEs needs, practices and activities. We believe that this new adaptation can ease, accelerate and improve the deployment of MBSE in SMEs remaining interoperable and coherent with their information systems.*

KEYWORDS: *Model Based System Engineering, Modelling, Verification, Validation, Evaluation, Digital Mock-Up, Digital Twin*

1 INTRODUCTION

Nowadays, the deployment of Systems Engineering (SE) in Small and Medium-sized Enterprises (SMEs) is becoming a new trend (PISOC 2017), for various legitimate reasons of continuous improvement. SE is a multi-disciplinary and collaborative method for the engineering of complex systems (e.g. the design of mechatronics products and related services) (BKCase 2017). In a slightly less pronounced way, it also promotes project management aspects (PMBok 2013). However, SE relies on concepts that seem difficult to handle in the core business of SMEs. To bridge this gap, SE promotes modelling principles known as Model-Based Systems Engineering (MBSE) (INCOSE 2008). MBSE has several undeniable advantages in various sectors (Kopach-Konrad 2007), but still requires a certain level of mastery. For instance, the (ISO 15288) business and management processes are often considered too complex for SMEs, and the (ISO 29110) is a relevant adaptation at their scale.

A Digital Mock-Up refers to a 3D geometrical model of a system (e.g. a technical product or a building) providing opportunities for virtualized manipulation and simulation. The goal is to reduce development time and to promote cheaper virtual prototyping (Chabal et al., 2016). For instance, during development, stakeholders can use Digital Mock-Ups to study, evaluate and validate alternative solutions, to verify if a system meets the expected requirements and properties such as system maintainability or operational performance, and then to understand and validate systems.

This notion of Digital Mock-Up for complex products and associated services is currently becoming one of the essential factors for efficient and successful deployment and use of MBSE in the SME community.

This article aims to revisit the notion of Digital Mock-Up, on the one hand, to enhance it with the necessary rigor imposed by the SE, and on the other hand, to adapt it to the needs and practices of SMEs. The expected long-term result is a tool-equipped variant of the MBSE approach for SMEs that is interoperable and in line with their tools and practices. This is particularly important to harmoniously ensure the transition and digital continuity in a SME environment as proposed by the Industry of the Future program (EFFRA 2017).

2 NEEDS, PRACTICES AND EXPECTATIONS

Numerous experiences have shown that SE deployment can help SMEs improve performances and collaboration capabilities (Honour 2013). SMEs can then reduce costs, improve control over business and financial risks and improve customers' relations. They can also collaborate with other partners to respond to new business opportunities that a single SME cannot address, while maintaining uncompromised their core business expertise and intellectual properties.

Considering previous works on SE deployment in SME (Laporte et al. 2012), this article focuses on the following needs that emerge during deployment:

- **Need 1 – modelling:** provide means for modelling (multi-viewpoint, multi-paradigm and collaborative modelling languages e.g. (Dirix et al. 2013), modelling patterns, anti-patterns and reference models, modelling tools, etc.) that aid and guide engineers during modelling activities.
- **Need 2 – analysis:** provide means to verify, validate and evaluate designed models (simulators, model-checkers, evaluation processes, etc.).
- **Need 3 – agile engineering:** move forward step by step using these modelling and analysis means. The goal

is to minimize risks and to reassure financial stakes by (1) promoting dynamic interaction with customers to manage their values and requirements, and by (2) evaluating modelled solution to limit mistakes, errors or omissions. These are the principles of the early verification and validation activities (early V&V) as part of an agile engineering approach (Pesole 2010). The early V&V consists in:

- checking as early as possible and whenever necessary all system models developed during design. The aim is to answer the following question: "Have I modelled correctly the expected system considering the requirements of all stakeholders, good business practices and modelling rules? "
- validating progressively the design by successively considering alternative solutions. This process is a careful revision of design and includes all stakeholders: the business actors involved in the system life-cycle, the customers and the end-users. Any decision at this level must be made with confidence. So, the aim is to answer the following question: "Have I modelled the right system for each of my stakeholders or, at least, have I obtained a certain level of approval and acceptance from them? "

- **Need 4 – data management and exchange:** provide means, mechanisms and processes for the management and exchange of data and knowledge (e.g., models). These must be automated and assisted. They consume and generate data and knowledge. This implies ad-hoc solutions for interoperability, both conceptual, methodological (intra and inter-business) and technical (between and with the existing tools available in the digital information system of the company), considering all levels of the SME: organization, processes, tools, data (Chen et al., 2007). In the long-term, the point is in ensuring and promoting the digital transition of data, knowledge, and know-how from system design to system end of life. This is mandatory for the integration of Digital Mock-Ups into Product Life-cycle Management (PLM) solutions without compromising the information systems and technical choices of the SME. Finally, data management and exchange include the SME collaboration that preserve when necessary its knowledge and its intellectual property.

Due to lack of space, the focus here is on system design and production phases. System operating and dismantling phases are out of scope. So, we dissociate the concepts of Digital Mock-Up (relevant during system design and production phases) from Digital Twin (relevant during system operating and dismantling phases) (Gartner 2017).

3 DIGITAL MOCK-UP (DMU)

3.1 Digital Mock-Up vs. Digital Twin

Digital Mock-Ups are used by architects and design experts to understand the products and associated services that compose the expected system S (i.e., a system under design and realization), to improve S, test it, evaluate its performance, justify choices or generate docu-

mentation. During deployment, Digital Mock-Ups are used by integration experts to verify and validate (V&V) the integration of S before making the transfer to the customer site where S is prepared and put into operation. At this moment the SME can be also involved in the expected services ensuring the maintenance of S in operational conditions, while it is preferable to consider the concept of Digital Twin, denoted as DT(S) in the next. The purpose is then to describe the processes and operating procedures, to monitor the evolution of S in use for various purposes such as preventive maintenance, and to evolve S until its decommissioning. The Digital Twin inherits of course the structure and some of the properties of the Digital Mock-up presented in the next sections. It will nevertheless be differentiated by its content and the needs it responds to. It will indeed be used during the operation of the system, for example for control purposes, configuration change management, optimization or monitoring.

So, we define simply a Digital Mock-Up of a system S, denoted as DMU(S) in the following, as a *global model of S*. DMU(S) results from the federation, composition and alignment of all data, information and knowledge concerning the system S, including all the business models presented below, which are developed, manipulated and used during S design. Indeed, designing S consists in defining iteratively the problem for which S has to be designed then the expected solution for S. The problem is modelled by defining the frontiers of S, organizing, checking and managing the set of stakeholders' and system requirements. The solution is then modelled by defining step by step S components, devices, human actors and subsystems, and how they must interact to fulfil S mission in an operational environment in order to meet the whole set of requirements i.e. requirements baseline. The system S is called in the next System Of Interest (SOI) defined as "the system whose life cycle is under consideration" (ISO 15288). The next defines step by step and formalize a Digital Mock-Up DMU(SOI) meeting the needs stated above.

3.2 DMU(SOI): modelling operational environment of SOI

Figure 1 schematizes this environment.

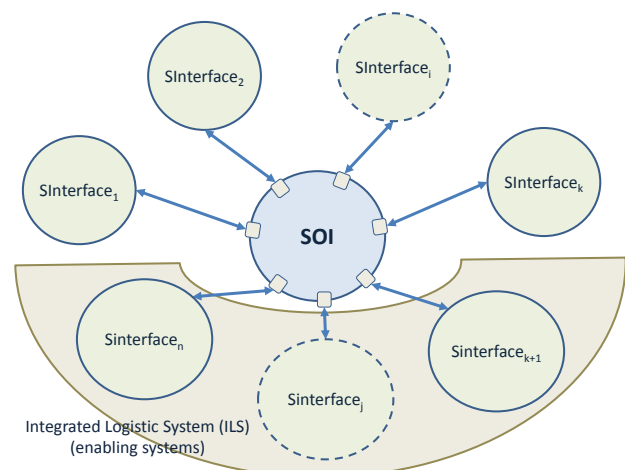


Figure 1: (external vision) SOI positioned in its environment

The SOI maintains and exchanges services with other systems called then *interfaced systems* (denoted S_{Interfaced}). Some of these systems, then grouped in the Integrated Logistics Support system (ILS), are called *enabling systems* (ISO15288). They make available for the SOI the services it needs to be and remain maintained in operational conditions all along its life cycle, i.e. allowing SOI to fulfil its mission with a proven level of performance, while complying with the specified requirements and under the conditions laid down in the design.

The DMU(SOI) must therefore, firstly, take into account this external vision which precisely positions the boundaries of the SOI, i.e. the operational context in which it operates. The environment E of a SOI is thus a system itself, composed of the SOI and of all the interfaced systems including the systems that compose the ILS.

It is further characterized by a set of properties (Chapurlat 2013a) (Micouin 2014) that constrains the SOI evolution and the different services exchanged through interactions between all these systems whatever may be their role all along SOI life cycle. For example, a handling robot operates in a workshop, at the interface with human operators, workflows, or machines on which it picks or drops parts. Its environment is further characterized by a brightness level, a degree of humidity, or levels and types of dust or radiation that must be taken into account because they can impact the services rendered or expected from the robot vis-à-vis the human operators for instance.

Secondly, it is necessary for DMU(SOI) to link modelling elements related to the management of the development project the SME must define and carry out to design and then realize the SOI. For this, we recommend using the standard (OMG 2008) and following the guidelines of the Project Management Body of Knowledge (PMBok 2013).

3.3 DMU(SOI): Modelling Views

In a MBSE approach, and more broadly by following systemic principles, modelling the SOI request to move to an internal vision. This means a multi-viewpoints approach as proposed by (Penalva 1997) (Feliot 2007). We retain hereafter the following viewpoints:

- *System Viewpoint* (what seems to be the SOI or what is the SOI currently? What are the relevant and available data, information and knowledge about the SOI?). This viewpoint defines the set of data, information and knowledge usable to characterize the SOI. These remain not covered by other views but must be shared and exchanged by all the others.

- *Requirements, Values and Properties viewpoint* (what should the SOI answer to?) defines stakeholders and system requirements, values and expectations applied to the SOI. First it allows understanding stakeholders' positions and aims; legacy rules and constraints, roles. Second it permits guiding design process. The goal is to structure, provide and manage (e.g. updating, assuming traceability of changes...) a Requirements Baseline.

- *Functional viewpoint* (what is the mission and what should the SOI do, what services should it render?) defines the system's functional architecture, specifying system's functions and their sub-functions. A function defines the requested and provided services by the system as a transformation of input flows into output flows performed by the system to achieve its mission. The viewpoint shows how functions are dynamically arranged, their execution sequenced and how conditions for control or data flows are taken into consideration to satisfy the Requirements Baseline. By the principle of iterative design, such functional architecture may evolve considering hereafter described architectures. This view point defines also different solutions of system's logical architecture, i.e., variations of arrangements of functions and their sub-functions highlighted in functional architecture and their interfaces (internal and external). In other words, a logical architecture shows how the system's functions can be logically associated for instance by grouping their input and output flows to optimize their future allocation to physical components, or by considering requested modularity.

- *Physical viewpoint* (what does the SOI consist of to fulfil this mission?) allows first representing various solutions of logical architecture implementation, i.e., arrangement of physical elements (components, subsystems and physical interfaces) which provides a possible solution for the SOI. It is then intended to satisfy one of the proposed logical architectures and respecting by definition the whole Requirements Baseline. Second, this viewpoint allows to define the organic architecture that is similar and thus often confused with the physical architecture. The organic architecture highlights technical and configured components representing the final solution of the SOI put into its operational context.

- *Behavioural viewpoint* (how does the SOI evolve and change in order to fulfil its mission?) emphasizes various aspects of the expected behaviour of the SOI and particularly its operational modes (Chapurlat et al. 2013c), the various operational scenarios that illustrate this behaviour when considering the SOI in a particular mode, and the relations of dependence to be taken into consideration between all the data, information and knowledge allowing actors to precise how the SOI changes can impact its behaviour and properties.

Each of these viewpoints is therefore populated with SOI models specifically dedicated to the modelling objectives pointed out in the view. Finally, each involved actor must be able to establish these models by working at different levels of detail according to his own modelling and analysis objectives e.g. the evaluation of performance, availability or safety.

3.4 DMU(SOI): Domain Specific Modelling Languages

Building DMU(SOI) needs integrating the different models that make up each of these viewpoints. It requires, on the one hand, knowing how to federate, compose or align these models. Each model is established in accordance with a Domain Specific Modeling Language

(DSML (Nastov 2016a)) as shown schematically in Figure 2.

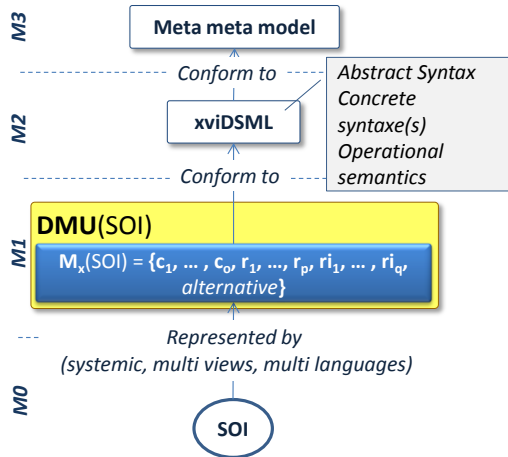


Figure 2: A model is conform to its metamodel e.g. here considered to a xviDSML (Nastov 2016a)

A DSML allows models to be established in a viewpoint covering the objectives of an actor and highlighting various levels of detail of the SOI. These DSMLs are eventually based on functional paradigms such as eFFBD (CORE 2017), organic such as PBD or object such as SysML (SysML 2010) with a broader coverage since SysML itself is a DSML composed of DSMLs participating in several views or MODELICA (MODELICA 2017).

On the other hand, building DMU(SOI) requests to take attention to various models V&V strategies that can mainly be classified into: *Model Appraisal, Guided Modelling, Simulation and Formal Proof* (Chapurlat 2013b). DMU(SOI) must then authorizes architectural and requirements modelling patterns usage (Pfister 2012)(Lemazurier 2017), properties proof and simulation techniques to support users and help them to revisit these models for improvement or to assess the properties of the SOI as it is modelled when operating in its operational context. The process is iterative and repeated until the models reach a certain level of quality and validity. DMU(SOI) becomes then a relying and accurate source of data, information and knowledge concerning the SOI that allow SMEs experts to make decisions with confidence.

For these two purposes (modelling and analysing), we propose:

- To base our approach on works developed by (Nastov 2016) (Combemale 2016) (Combemale et al. 2012). Particularly, (Nastov et al. 2016b) (Nastov et al. 2016c) propose tool-equipped approaches for the design of executable, verifiable and interoperable DSMLs (xviDSMLs). Besides modelling of the chosen viewpoints, they allow direct verification, support validation then evaluation of the models, conform to xviDSML, that compose DMU(SOI).

- To establish models (being not conform to the same principle of xviDSML) federation, composition and alignment rules by using existing standards such as Functional Mock-Up Interface (FMI 2017) or (DMAO

2013), HLA principles (HLA 2010) or simulation bus, such as used for instance in (Bouffaron 2016), that promote mechanisms, rules and tools to encapsulate a model in order to make it interoperable with other models from different nature, and authorizing a federated simulation.

- To use transformation rules promoted and supported by tools such as (Kahani et al. 2015), allowing the transformation of models conform to classical DSMLs into models conform to other and tooled DSMLs for simulation or properties checking such as xviDSMLs. This allows authorizing then proving and simulating them without huge efforts and transparent for the SME users.

3.5 DMU(S): synthesis

As a first synthesis, a DMU(S), whatever may be the system S (SOI, or another system for instance SInterface_i, the environment E or ILS) is defined as:

$$\text{DMU}(S) ::= \mathbf{M}_S(S) \cup \mathbf{M}_E(S) \cup \mathbf{M}_{PM}(S)$$

Where

$\mathbf{M}_S(S) ::=$ System S model by highlighting multi-views approach as proposed and formalized in the next.

$\mathbf{M}_E(S) ::=$ Environment model, including the representation of each interfaced and enabling system that composes the Integrated Logistic System (ILS). ILS is itself composed of enabling and interfaced systems requested in order to maintain the SOI in operational conditions

$\mathbf{M}_E(S) ::= \cup_{i=1..n} [\mathbf{M}_S(\text{SInterface}_i) \cup \text{interactions}(\mathbf{M}_S(\text{SInterface}_i), \mathbf{M}_S(S)]$ where SInterface_i are systems from the Environment of S with which S interacts

$\mathbf{M}_{PM}(S) ::=$ project management model.

$\mathbf{M}_{PM}(S) ::= \mathbf{PM}(S) \cup \mathbf{Plan}(S)$

Where

$\mathbf{Plan}(S) ::=$ Set of IVTV plans (Integration, Verification, Transition, and Validation plans) and Justification Reports

$\mathbf{PM}(S) ::=$ Set of models requested for project management purpose

$\mathbf{PM}(S) ::= \mathbf{MP}(S) \cup \mathbf{MR}(S) \cup \mathbf{M}_{BPM}(S) \cup \mathbf{WBS}(S)$

Where

$\mathbf{M}_P(S) ::=$ Project data model (textual)

$\mathbf{M}_R(S) ::=$ Project resources model (graphical)

$\mathbf{M}_{BPM}(S) ::=$ Business processes models (graphical)

$\mathbf{WBS}(S) ::=$ Set of alternative Work Breakdown Structure models (structured list of activities, tasks and operations to be performed considering various alternatives of solutions represented in $\mathbf{M}_S(S)$)

Figure 3 schematizes $\mathbf{M}_E(\text{SOI})$ and $\mathbf{M}_{PM}(\text{SOI})$ that compose DMU(SOI). Each model noted $\mathbf{M}_S(\text{SInterface}_i)$ is the model of the i^{th} interfaced system that interacts with the SOI at a moment of its life-cycle. This model is therefore constructed respecting exactly the same principles and way used for building $\mathbf{M}_S(\text{SOI})$ model as described below.

3.1 DMU(S): detail of system S model

Figure 4 summarizes the model of a system S that is the third and last element that composes DMU(S).

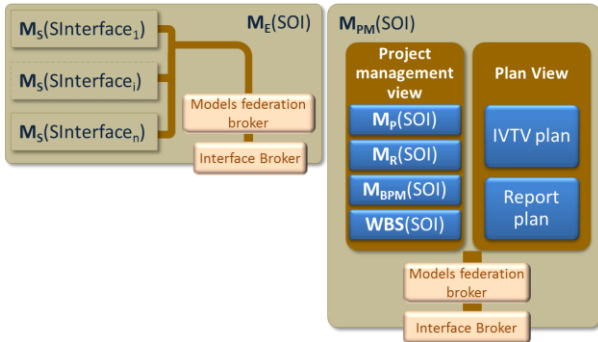


Figure 3: SOI Environment and Project Management modeling (OMG 2008) (PMBok 2013)(Chapurlat et al. 2014)

The *Interface Broker* allows encapsulating and defining formally what are the inputs / outputs data as proposed in Components Based Architecture environments such as DEDAL (Mokni et al. 2015). The *Models Federation Broker* implements models federation, composition or alignment mechanisms and rules when requested.

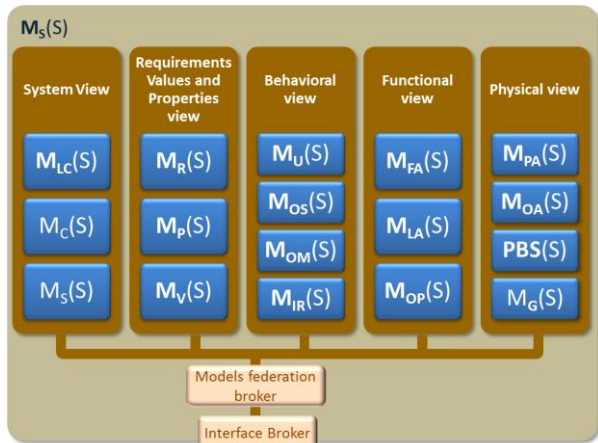


Figure 4 : (internal vision) Model of a system S $M_S(S)$

All models conform to a xviDSML are of course executable (i.e. could be simulated) independently or together in conjunction with other models conform to other xviDSMLs being interoperable by nature. In this case, Models Federation Broker organizes and synchronizes the exchanges of Inputs / Outputs flows between models as it is formalized in the meta model of the DSML. From a more formalized point of view, we note:

$M_S(S) ::= S$ model composed by various models from the five modelling views

$$M_S(S) ::= \text{SYSTEM}(S) \cup \text{RVP}(S) \cup \text{BV}(S) \cup \text{FV}(S) \cup \text{PV}(S)$$

Where:

SYSTEM(S) ::= Set of data, information and knowledge usable to characterize other dimensions of S not covered in views and shareable with all the other views

$$\text{SYSTEM}(S) ::= M_{LC}(S) \cup M_C(S) \cup M_S(S)$$

Where:

$M_{LC}(S)$::= Life Cycle model making appear life cycle phases and milestones (textual / graphical)

$M_C(S)$::= Context model making appear Interfaced Systems and assuming the link with other model of these systems (graphical)

$M_S(S)$::= Set of data models making appear attributes, mission description, etc. and data, information and business knowledge coming from external sources (e.g. Internet or distant data bases) that characterize the system S (data representation models / ontological resources / textual)

$\text{RVP}(S)$::= Set of models requested in **Requirements, Values and Properties view**

$$\text{RVP}(S) ::= M_R(S) \cup M_P(S) \cup M_V(S)$$

Where:

$M_R(S)$::= Stakeholders Requirements and alternatives of System Requirements baseline model - textual / graphical)

$M_P(S)$::= (optional) Properties baseline model (modelling / system properties - textual / graphical)

$M_V(S)$::= (optional) Stakeholders and System Values baseline model (textual / graphical)

$\text{BV}(S)$::= Set of models requested in **Behavioural view**

$$\text{BV}(S) ::= M_U(S) \cup M_{OS}(S) \cup M_{IR}(S) \cup M_{OM}(S)$$

Where:

$M_U(S)$::= (optional) Set of stakeholders' usages models (textual / graphical)

$M_{OS}(S)$::= Set of Operational Scenarios models (textual / graphical)

$M_{AI}(S)$::= (optional) Set of Influence Relations models (textual / graphical)

$M_{OM}(S)$::= Set of Operational Modes (diagram) models

$\text{FV}(S)$::= Set of models requested in **Functional view**

$$\text{FV}(S) ::= M_{FA}(S) \cup M_{LA}(S) \cup M_{OP}(S)$$

Where

$M_{FA}(S)$::= Set of alternative Functional Architectures models (graphical)

$M_{LA}(S)$::= Set of alternative Logical Architectures models (graphical)

$M_{OP}(S)$::= (optional) set of alternative operational processes models (graphical)

$\text{PV}(S)$::= Set of models requested in **Physical view**

$$\text{PV}(S) ::= M_{PA}(S) \cup M_{OA}(S) \cup M_G(S) \cup \text{PBS}(S)$$

Where:

$M_{PA}(S)$::= Set of alternative Physical Architectures models (graphical)

$M_{OA}(S)$::= Set of alternative Organic Architectures models (graphical)

$M_G(S)$::= Set of alternative Geometric models corresponding to each alternative Organic architecture (graphical)

$\text{PBS}(S)$::= Set of alternative Physical Breakdown Structure models (textual / graphical)

DMU(SOI) is then schematized as shown in Figure 5.

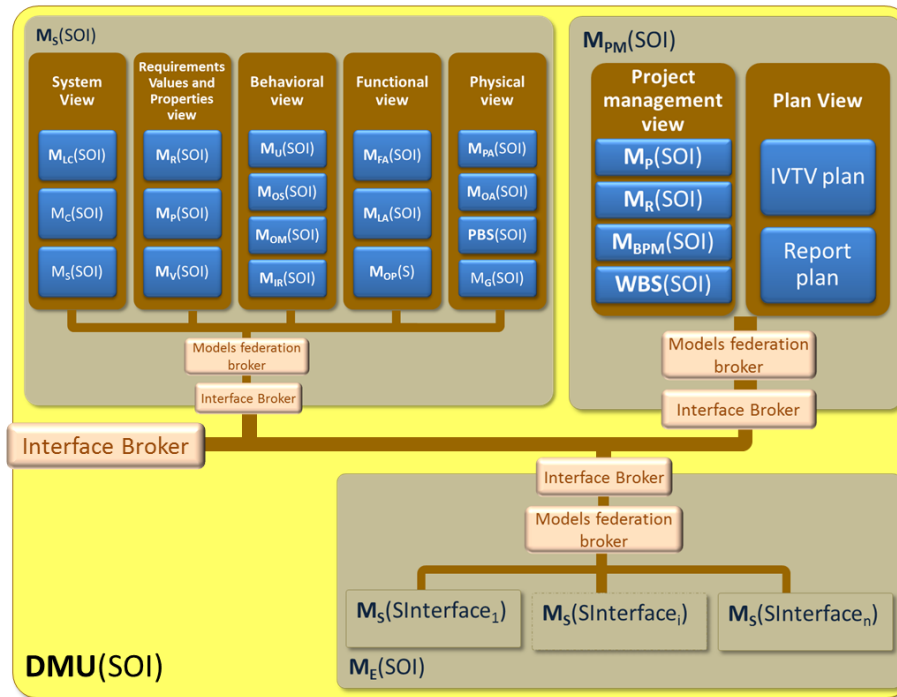


Figure 5: Digital Mock-Up synthesis

4 DIGITAL MOCK-UP: EXAMPLE IN NUCLEAR DECOMMISSIONING DOMAIN

The field of nuclear decommissioning brings together a large number of actors, including SMEs, to design, carry out and operate many products, techniques, processes, tools, etc. useful for the dismantling of a nuclear installation and the management of its radioactive waste. Nowadays, a digital transformation of the nuclear industry is ongoing (SFEN 2017) and the notion of DMU is burgeoning within the extended decommissioning enterprise (all project stakeholders). New technologies are already making it possible to model then perform various simulations at the early stages of a project (operations, radiological monitoring, project costs and indicators, waste streams, training, etc.), e.g. 3D modelling (CAD), virtual reality, or augmented reality, federated thanks to BIM tools (Chabal et al., 2016). Furthermore, the large number of stakeholders, in particular SMEs, involved in each project justifies the interest in the deployment of new interoperable and collaborative technologies. The systemic approach and the SE aim to facilitate and guide the future design of an interoperable and collaborative DMU, in particular through the formalization of a Decommissioning and Dismantling System, which is allowed, among other things, by MBSE (Lafon et al. 2018). This design work, in progress, actually handles many models compliant with many DSML, such as those illustrated in Figure 6.

5 CONCLUSION

This article revisits and lays down the foundations of a Digital Mock-Up enriched with numerous concepts allowing and facilitating the modelling of complex systems in their environment, the use of proven checking techniques and federated simulation techniques and standards. A DMU is designed to finally facilitate the deployment of a MBSE resolutely adapted to the size of a SME or a network of collaborative SMEs. The work undertaken illustrates the implementation and the use of the DMU in the multidisciplinary nuclear decommissioning field, which is strongly constrained by a set of requirements and practices.. Perspectives are numerous. It is of course necessary at least to study and formalize the transition from a DMU to a Digital Twin (DT) that can then be used during the operation phase and future dismantling of a system. This implies working on the consideration of the DMU and the DT in tools for Product Life cycle Management (PLM). Finally, the use of high-performance computing resources in the cloud has become an important development area for SMEs today. The current offer have been enriched as well as the interest of these means are shown in the Fortissimo (Fortissimo 2014), CloudSME (CloudSME 2014) and CloudFlow (Cloudflow 2014) projects. These two-coupled concepts of Digital Mock-Up and Digital Twin are thus an additional offer made available for SMEs to face the digital shift.

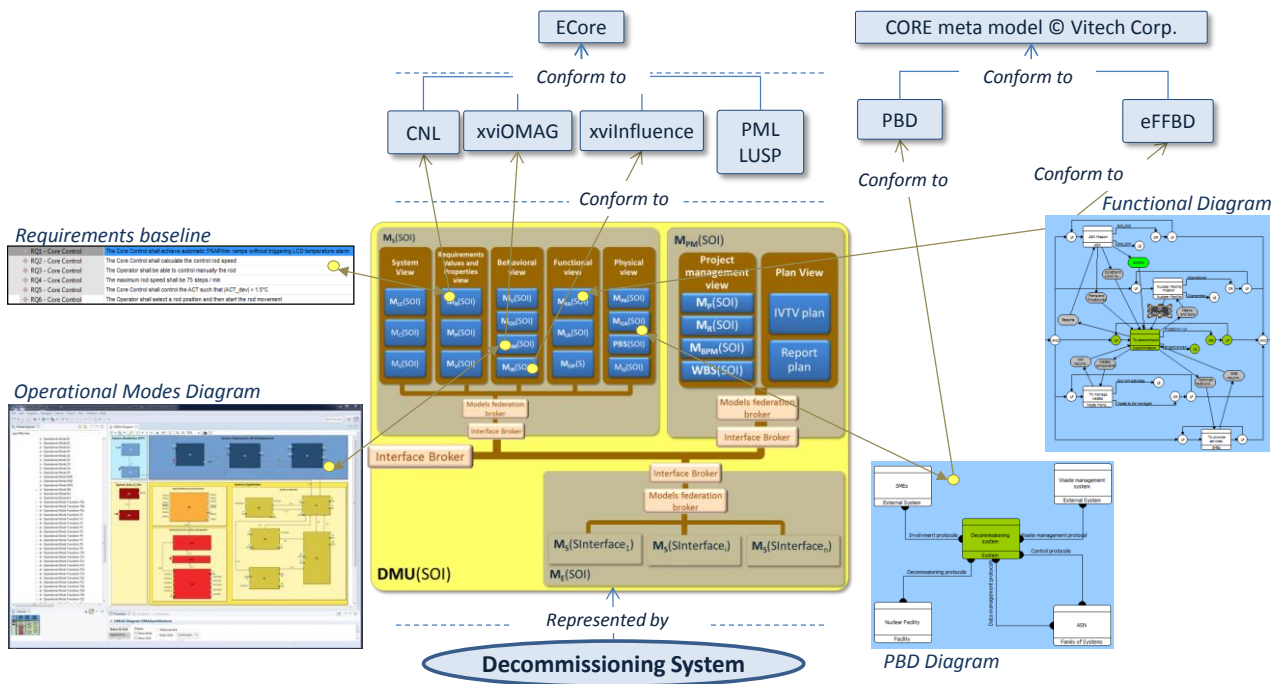


Figure 6: Examples of models and their associated DSML to design the DMU of the Decommissioning System

6 REFERENCES

- (BKCase 2017) BKCASE Editorial Board. 2017. The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 1.8. R.D. Adcock (EIC). Hoboken, NJ: The Trustees of the Stevens Institute of Technology. (www.sebokwiki.org - last visited 2017-12-26).
- (Bouffaron 2016) F.Bouffaron, Co-spécification système exécutable basée sur des modèles - Application à la conduite interactive d'un procédé industriel critique, PhD thesis, Université de Lorraine, 2016, HAL reference=<tel-01263624> (in French)
- (Chabal et al. 2016) Chabal C., Soulabaille Y., Virtual Reality: a way to prepare and optimize operations in decommissioning projects, International Symposium on PREparation for DECommissioning, February 2016, Lyon, France
- (Chapurlat 2013a) Chapurlat, V., (2013) "UPSL-SE: A Model Verification Framework for Systems Engineering." Computers in Industry 64(5): 581-97.
- (Chapurlat 2013b) Chapurlat V., Property concept and modelling language for Model-Based Systems Engineering (MBSE) context, Internal Research Report (access on demand)
- (Chapurlat et al. 2013c) Chapurlat V., Daclin N., Proposition of a guide for investigating, modeling and analyzing system operating modes: OMAG, International Conference on Complex System Design and Management CSDM 2013 December 2013, Paris
- (Chapurlat et al. 2014) V.Chapurlat, E.Bonjour, From Model Based Systems Engineering to Model Based System Realization: role and relevance of IVTV Plan, International Conference Advances in production Manufacturing Systems APMS 2014, September 21th to 24th, Ajaccio, France
- (FMI 2017) Functional Mock-up Interface (FMI), <http://fmi-standard.org/>, (Accessible le 25/10/2017)
- (Chen et al. 2007) Chen, D., Dassisti, M., Elveaeter, B.: Enterprise interoperability framework and knowledge corpus – final report. Interop deliverable DI.3, may (2007)
- (CloudFlow 2014) CloudFlow project (last accessed: October 14, 2016): <http://www.eu-cloudflow.eu/>
- (CloudSME 2014) CloudSME project (last accessed: October 14, 2016): <http://cloudsme.eu/content/project>
- (Combemale 2016) Combemale, B. The GEMOC Initiative: on the globalization of modeling languages. Retrieved from <http://gemoc.org/> (2016)
- (Combemale et al. 2012) Combemale, B., Crégut, X., & Pantel, M. (2012). A Design Pattern for Executable DSML. The 19th Asia-Pacific Software Engineering Conference (APSEC), 282-287.
- (CORE 2017) System Definition Guide, see http://www.vitechcorp.com/support/documentation/genesy_s/500/systemdefinitionguide.pdf (Last accessed December 2017)
- (Crawley 2007) Ed Crawley, Introduction to System Architecture: Architecture to Value, Rev 2.0, January 5, 2007
- (Dirix et al. 2013) Dirix, M., Muller, A., & Aranega, V. (2013). GenMyModel: an online UML case tool. In ECOOP.
- (DMAO 2013) DMAO, IEEE Recommended Practice for Distributed Simulation Engineering and Execution Process Multi-Architecture Overlay (DMAO) in IEEE Std 1730.1-2013, vol., no., pp.1-91, Nov. 1 2013 doi: 10.1109/IEEESTD.2013.6654219
- (EFFRA 2017) EFFRA, Factories of the Future, Multi-annual roadmap for the contractual PPP under Horizon 2020, European Community
- (Félot 2007) C.Félot, Toward a formal theory of systems, Colloque d'Automne du LIX 2007 - CAL07 Complex Systems: Modelling, Verification and Optimization, Paris, Carré des Sciences, 3rd and 4th October 2007, <http://www.lix.polytechnique.fr/~liberti/cal07/presentations/> (last visited 11th July 2012)

- (Fortissimo 2014) Fortissimo project (last accessed: October 14, 2016): <http://www.fortissimo-project.eu/>
- (Gartner 2017) Gartner Top 10 Strategic Technology Trends for 2018, see <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/> (last accessed: December 2017)
- (HLA 2010) IEEE 1516-2010 – “IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) -- Framework and Rules”, <https://standards.ieee.org/findstds/standard/1516-2010.html>, IEEE Standard Association, NY, 2010, (last accessed october 2017)
- (Honour 2013) Honour, Eric C., Systems engineering return on investment, PhD, D&S, School of Electrical and Information Engineering, Australia, 2013.
- (INCOSE 2008) INCOSE, Survey of Model-Based Systems Engineering (MBSE) Methodologies, INCOSE-TD-2007-003-01, Version/Revision: B, 10 June 2008
- (ISO 15288) ISO/IEC 15288:2008(E) / IEEE Standards 15288.2008 – Systems engineering – System life cycle processes (2nd edition), February 2008
- (ISO 29110) ISO/IEC 29110:2015, Software engineering - Lifecycle profiles for Very Small Entities (VSEs), 2015
- (Kahani 2015) Kahani, N. & R. Cordy, J., 2015. Comparison and Evaluation of Model Transformation Tools, Ontario, Canada. Available at: <http://research.cs.queensu.ca/TechReports/Reports/2015-627.pdf> (last accessed: December 2017)
- (Kopach-Konrad 2007) Renata Kopach-Konrad, Mark Lawley, Mike Criswell, Imran Hasan, Santanu Chakraborty, Joseph Pekny, Bradley N. Doebbeling, Applying Systems Engineering Principles in Improving Health Care Delivery, Journal of General Internal Medicine, Volume 22, Supplement 3, pp. 431-437, 2007
- (Lafon et al. 2018) M.Lafon, V.Chapurlat, J-F. Milot, C.Moitrier, The D&D of Nuclear Facilities: a Systemic Approach for Project Organization and Monitoring – 18099, WM2018 Conference, March 18-22, 2018, Phoenix, Arizona, USA
- (Laporte 2012) Laporte, C.Y., Palza Vargas, E., The Development of International Standards to facilitate Process Improvements for Very Small Enterprises. Dans: Software Process Improvement and Management: Approaches and Tools for Practical Development, IGI Global Publisher. USA. 2012 , pp 34-61.
- (Lemazurier et al. 2017) L. Lemazurier, V. Chapurlat, A. Grossetête, An MBSE Approach to Pass from Requirements to Functional Architecture, International World Congress IFAC, IFAC 2017, Toulouse, July 2017
- (Micouin 2014) Micouin, P., Model-Based Systems Engineering: fundametnales and methods, “Control, System and Industrial Engineering” series, ISTE/Wiley, 2014
- (MODELICA 2016) MODELICA - A Unified Object-Oriented Language for Physical Systems Modeling Language Specification Version 3.2, March 24th, 2010 <http://www.modelica.org/> (last visited 2014-07-16)
- (Mokni et al. 2015) A.Mokni, M.Huchard, C.Urtado, S.Vauttier, Y. Zhang. An evolution management model for multi-level component-based software architectures . SEKE: Software Engineering and Knowledge Engineering, Jul 2015, Pittsburgh, United States. 27th International Conference on Software Engineering and Knowledge Engineering, 2015
- (Nastov 2016a) Nastov, B. (2016). Contribution à une méthode outillée pour la conception de langages de modélisation métier interopérables, analysables et prouvables pour l’Ingénierie Système Basée sur des Modèles, University of Montpellier (Phd in English)
- (Nastov 2016b) Nastov, B., Chapurlat, V., Dony, C., (2016) Towards Semantical DSMLs for Complex or Cyber-Physical Systems, Proceedings of the 11th International Conference on Evaluation of Novel Software Approaches to Software Engineering (ENASE 2016). Rome, Italy: Scitepress publishing.
- (Nastov 2016c) Nastov, B., Chapurlat, V., Dony, C., (2016) “Towards V&V Suitable Domain Specific Modeling Languages for MBSE: A Toolled Approach.” In The 26th Annual INCOSE International Symposium (IS 2016). Edinburgh, Scotland: Wiley publishing.
- (OMG 2008) OMG, Software & Systems Process Engineering Meta-Model Specification, V2.0, 2008
- (Penalva 1997) Penalva Jean-Michel, La modélisation par les systèmes en situations complexes. Thèse de Doctorat, Université de Paris Sud, France, 1997 (in French)
- (Pesole 2010) Pesola, J-P., Building Framework for Early Product Verification and Validation”. Espoo 2010. VTT Publications 736 <http://www.vtt.fi/inf/pdf/publications/2010/P736.pdf> (Accessed October 2017), (2010)
- (Pfister et al. 2012) F.Pfister, V.Chapurlat, M.Huchard, C.Nebut, J.-L.Wippler, A proposed meta-model for formalizing Systems Engineering knowledge, based on functional architectural patterns, International Journal Systems Engineering, INCOSE Eds., 2012
- (PISOC 2017) The PISOC initiative in France, Workshop EMEA, INCOSE, September 2017
- (PMBoK 2013) Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Fifth Edition, 2013. ISBN13: 9781935589679
- (SFEN 2017) SFEN Convention, Le Nucléaire Accélère sa Transformation Numérique, March 2017, Paris, see <https://www.sfen-convention2017.org/> (Last accessed December 2017) (in French)
- (SysML 2010) System Modelling Language SysML (<http://www.sysml.org/>), 2010

Robustesse, résilience : une brève synthèse des définitions au travers d'une analyse structurée de la littérature

A. CLEMENT¹, F. MARMIER^{1,2}, D. KAMISSOKO¹, D. GOURC¹, L. WIOLAND³,
V. GOVAERE³, J. CEGARRA⁴

¹Université de Toulouse - IMT Mines Albi
Campus Jarlard - Route de Teillet
81013 Albi Cedex 09 - France
{prenom.nom}@mines-albi.fr

²Strasbourg University
CNRS BETA
61 avenue de la Forêt Noire
67085 Strasbourg Cedex, France

³INRS
Working Life Department
54500 Vandoeuvre, France
{prenom.nom}@inrs.fr

⁴Laboratoire Sciences de la Cognition, Technologie
Ergonomie (SCoTE EA 7420), Université de Toulouse
INU Champollion, Albi, France
{prenom.nom}@univ-jfc.fr

RÉSUMÉ : *La réduction des délais des cycles conception-commercialisation, l'essor des technologies informatiques et l'arrivée des organisations relevant de l'industrie 4.0 font vivre aux entreprises de plus en plus de perturbations qu'elles doivent être capable d'identifier, d'anticiper et de gérer. Les notions de robustesse, de résilience sont plus que jamais pregnantes dans les travaux de recherche. Dans ce contexte, il convient de correctement définir ces concepts et les métriques qui permettent de les évaluer pour le pilotage des organisations et systèmes industriels. Cet article présente les résultats d'une revue de la littérature sur les définitions des termes robustesse et résilience dans divers domaines telles que présentées dans des articles publiés entre 1975 et 2017. Les articles sont classés par domaines et disciplines. L'analyse des définitions montre une similitude dans les définitions produites dans les différents secteurs.*

MOTS-CLÉS : *comparaison, définition, résilience, risque, robustesse*

1 INTRODUCTION

Les entreprises de production de biens tout autant que les entreprises de production de services sont exposées à des perturbations imprévisibles, par exemple des changements de la demande client, de la disponibilité des matières premières, de la disponibilité des ressources. Pour faire face à ces perturbations, les décideurs/ managers doivent mettre en place des procédures et des systèmes de pilotage aptes à anticiper, gérer ces perturbations dans le respect des exigences de performance et de rentabilité. La mesure du respect des exigences et de l'efficacité des mesures prises nécessite de mettre en place et suivre des indicateurs spécifiques. De nombreux travaux de recherche traitent de ce sujet et proposent divers indicateurs comme ceux permettant de mesurer la robustesse, la résilience, la stabilité du système. (Pavard, Dugdale, Saoud, Darcy & Salembier 2009) indiquent que ces notions sont devenues centrales dans de nombreux domaines scientifiques allant de l'informatique à la biologie en passant par l'écologie ou la finance. D'après

les auteurs, il n'existe pas de définition unique de cette notion et ceci d'autant plus qu'elles interfèrent les unes les autres.

Face à ce constat, l'objectif de cet article, au moyen d'une analyse des travaux issus de divers secteurs d'activités, est de faire le point en particulier sur les concepts de robustesse et résilience, les définitions consensuelles et les indicateurs proposés pour les évaluer par une analyse des travaux issus de divers secteurs d'activités. En vue de préparer des définitions et des indicateurs de mesure dans le secteur de la logistique et du Transport Routier de Marchandises.

L'article est structuré en 5 sections. Après une introduction (section 1), la section 2 présente la méthode de recherche qui a été utilisée pour mener l'analyse de la littérature. L'analyse globale des articles et des revues identifiées selon divers critères est présentée dans la section 3. La section 4 présente les résultats détaillés de l'analyse bibliographique des définitions et discute les définitions par domaine en tentant d'identifier les points communs et divergences. Une conclusion est proposée en fin d'article.

2 METHODE DE RECHERCHE

L'analyse et la comparaison des définitions associées aux concepts de robustesse et de résilience se sont appuyées sur une revue de la littérature. La méthode de recherche utilisée pour effectuer cette analyse est présentée sous la forme d'un processus (voir Figure 1), et comporte trois phases principales :

- collecte des articles,
- sélection des articles,
- analyse et classification des définitions.

2.1 Collecte des articles

Pour identifier les articles proposant des définitions des concepts de robustesse et résilience, une recherche structurée a été menée. La recherche d'articles a été menée en utilisant une combinaison des termes "robustesse" ET "résilience" dans leur titre. L'utilisation de ces critères de recherche avait pour objectif d'identifier des articles qui pouvaient potentiellement traiter des deux concepts dans leur développement et proposer les définitions respectives. Dans un premier temps, il nous est apparu important de focaliser la recherche sur des articles qui de manière explicite faisaient référence à ces deux termes directement dans le titre. Ainsi, cette première analyse a pu être menée à partir d'une base réduite d'articles pertinents¹. Une seconde étape devra être menée, comme nous le verrons dans la section conclusion, en étendant la recherche à d'autres articles afin de confirmer les premiers résultats.

Cet article ne traite que de la première étape avec une sélection d'articles intégrant ces deux termes dans le titre.

La recherche a été effectuée, le 10 janvier 2018, sur la totalité de la base de données Web of Science (WOS) de Thomson Reuters. Il a été identifié 51 articles, tous les articles sélectionnés sont en langue anglaise.

2.2 Sélection des articles

La phase de sélection a pour objectif de filtrer les articles, sur la base d'une lecture de l'abstract et/ou du texte complet, afin de retenir ceux qui de manière explicite proposent des définitions des deux termes conjointement. Ainsi, seuls les articles pertinents pour

¹A titre d'illustration, une recherche d'articles utilisant ces deux termes dans les champs Titre et/ou résumé et/ou mots clés fait ressortir plus de mille trois cent articles (1327 articles répondent à ce critère). Il est important de noter que bon nombre des articles ainsi identifiés ne fournissent pas de définitions explicites et/ou ne permettent pas de mener une analyse comparative pertinente. Une telle approche nécessiterait alors un travail long et fastidieux d'analyse et de sélection des articles.

notre analyse sont conservés. Les autres articles, non pertinents, sont exclus de la base, et ne sont pas considérés dans l'étape suivante. La lecture des articles a permis d'identifier d'autres références pertinentes.

Sur les 51 articles identifiés initialement, le processus de sélection, mettant en évidence les articles pertinents au regard de notre objectif et de nos critères de recherche, conduit à retenir 45 articles. Le contenu de ces articles est analysé dans la phase suivante (voir Figure 1).

2.3 Analyse et classification

L'objet de la phase d'analyse est d'identifier les définitions et de les comparer au regard de divers critères, comme par exemple : le domaine de recherche, le type d'application... Au final, nous souhaitons identifier s'il existe des convergences, points communs ou différences dans les définitions de ces termes au cours du temps ou selon les domaines.

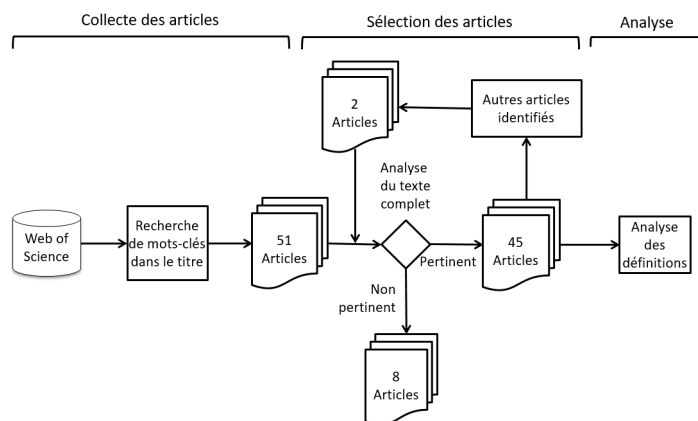


Figure 1 – Processus de recherche

3 RÉPARTITION DES PUBLICATIONS

Dans un premier temps, la recherche a eu pour objectif de savoir à partir de quelle période la comparaison, ou la différenciation entre la résilience et la robustesse a commencé à se faire. L'utilisation des deux termes est présente dès 1987 (voir Figure 2) mais ce n'est qu'à partir de 2012 que ces deux concepts sont régulièrement associés et discutés en même temps dans un même papier. Environ 80% des articles de la recherche ont été publiés sur les 6 dernières années.

Le type de publication (voir Table 2) montre leur diversité sur ce sujet. Trois revues ont publié au moins 2 articles sur cette utilisation des deux concepts. Il s'agit de : Ecology and Society (4 articles), Journal of the knowledge economy (2 articles), Journals of Gerontology series a-biological sciences and medical

Table 1 – Répartition des articles par domaines

Domaine	Nombre d'articles	Référence
Ingénierie	11	(Balzari & Balzari 2017) (Schlichtmann 2016) (Ibanez et al. 2016) (Okoh & Haugen 2015) (Zhu & Basar 2015) (Alenazi & Sterbenz 2015) (Bongard & Lipson 2014) (Amantini et al. 2012) (Wybo 2008) (Brailean et al. 1999) (Fryer et al. 1987)
Informatique	11	(Camara et al. 2017) (Mercer et al. 2016) (Amann & James 2015) (de Souza & Zhou 2015) (Ponnambalam et al. 2014) (Martin & Ludek 2013) (Sha & Panchal 2013) (Daouda et al. 2013) (Greco et al. 2012) (Read 2005) (Klau & Weiskircher 2005)
Environnement	9	(Dragicevic 2016) (Nair & Howlett 2016) (Mumby et al. 2014) (Witten 2014) (Krupa et al. 2014) (Anderies et al. 2013) (Domptail et al. 2013) (Anderies & Hegmon 2011) (Fleischman et al. 2010)
Biologie	6	(Wong et al. 2017) (Phan et al. 2016) (Prostova et al. 2015) (Aschbacher et al. 2014) (Butler & Silver 2009) (Levin & Lubchenco 2008)
Business, économie	5	(Carayannis et al. 2017a) (Carayannis et al. 2017b) (Tieman 2017) (Edgeman 2015) (Brandon-Jones et al. 2014)
Agriculture	3	(Perez-Hedo et al. 2017) (Urruty et al. 2016) (Sauvant & Perez 2010)
Gériatrie	3	(Arbeev et al. 2016) (Ukrantseva et al. 2016) (Whitson et al. 2016)
Zoologie	1	(Rollo & Shibata 1991)
Administration publique	1	(Capano & Woo 2017)
Sciences sociales	1	(Scholz et al. 2012)

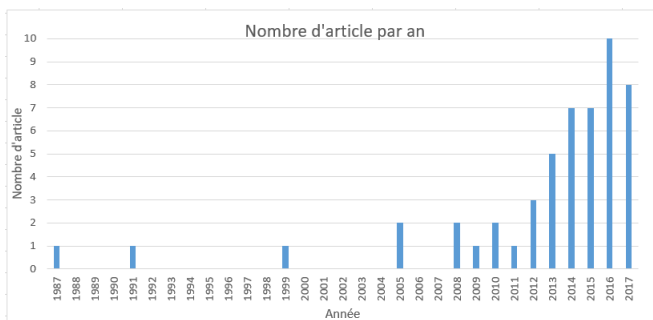


Figure 2 – Distribution des articles par année

sciences (2 articles). Cette analyse montre finalement une grande diversité dans l'expression de cette confrontation de concepts car elle est présente dans de nombreux journaux.

Cette diversité de supports de publication est également observée dans les domaines de recherche (voir Table 1). Dix domaines sont aujourd'hui représentés dans l'ensemble des articles traités au cours de la recherche. L'ingénierie, l'informatique et l'environnement sont largement dominants mais cette différenciation robuste-résilience se fait également dans d'autres domaines comme la biologie, l'économie, l'agriculture et la gériatrie. On retrouve également des articles dans des domaines moins at-

Table 2 – Répartition par type

Type	Nombre
Article	31
Article de conférence	8
Editorial	3
Revue de la littérature	3
Lettre	2
Résumé	2
Correction	1

tendus comme la zoologie, l'administration publique ou encore les sciences sociales.

4 ANALYSE DES DEFINITIONS

Tout d'abord, commençons par étudier et développer les sources étymologiques de ces deux termes. Dans un deuxième temps, nous analysons les définitions utilisées aujourd'hui dans les divers domaines afin d'identifier si elles sont conformes au sens originel et/ou si des éléments supplémentaires sont venus compléter ces définitions.

Le terme robuste est dérivé du terme « robuste », issu du latin *robustus*, qui exprime le caractère « solide comme le chêne » et de *robur* (« chêne »). Il exprime la capacité à résister (force, vigueur) et la

capacité « de supporter l'effort, la fatigue, le mal »².

D'après Rigaud (2011), sur la base des travaux de Tisseron³, le terme résilience s'est étymologiquement construit à partir des racines latines « re » et « salire » qui signifient respectivement « en arrière » et « sauter ». Le mot résilience est utilisé dès le XVII^e siècle en Angleterre pour désigner la capacité de réaction après un choc. Il s'agirait d'exprimer la capacité de rebondir, de résister à un choc ou une déformation.

4.1 Définitions

La table 3 recense les définitions des termes robustesse et résilience en fonction des domaines auxquels les articles se réfèrent (voir Table 1). Certains articles ne donnent pas de définitions explicites mais utilisent un exemple pour expliquer les différentes visions de la robustesse et de la résilience.

En plus des définitions présentes dans la table 3, deux revues de la littérature ((Urruty et al. 2016), (Brandon-Jones et al. 2014)) donnent de nombreuses définitions complémentaires que nous exploitons pour la discussion.

La première revue de la littérature, celle de (Urruty et al. 2016) se base uniquement sur le domaine de l'agriculture et de la biologie avec des définitions de la robustesse et de la résilience qu'ils généralisent ainsi :

- pour la robustesse: *"Robustness is ability to maintain desired levels of agricultural outputs despite the occurrence of perturbations"*
- pour la résilience: *Resilience is ability to absorb change and to anticipate future perturbations through adaptive capacity.*

Cette généralisation leur permet de donner une représentation graphique des différents concepts de robustesse et de résilience (voir Figure 3). Cette figure nous permet d'observer la différence entre la robustesse et la résilience. La notion de robustesse se base sur le principe selon lequel la perturbation ne modifie pas l'état initial dans le cas où le système est très robuste. La notion de résilience repose sur le principe d'adaptation selon lequel il va d'abord y avoir une phase d'absorption de la perturbation puis un retour à l'état initial.

On peut parler ainsi de système robuste ou non, voire de système plus ou moins robuste. Il en est de même pour la résilience. Par exemple, la temps que met le système à absorber la perturbation et revenir à

l'état initial peut être un indicateur du niveau de résilience. Concernant la robustesse, le niveau de perturbation (son intensité) que le système peut accepter sans modifier l'état initial peut exprimer le niveau de robustesse du système.

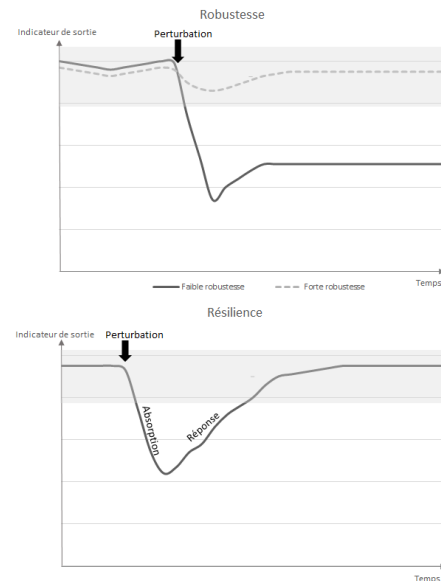


Figure 3 – Illustration des concepts inspiré de (Urruty et al. 2016)

La seconde revue de la littérature de (Brandon-Jones et al. 2014) s'intéresse aux définitions de robustesse et de résilience dans le domaine de la supply chain. Ces deux définitions sont :

- pour la robustesse: *Supply Chain Robustness: Robustness refers to the ability of a supply chain to withstand disruption and continue operating.*
- pour la résilience: *Supply Chain Resilience: As discussed in the literature review, resilience references the ability of a supply chain to bounce back from a disruption.*

4.2 Discussion

Dans l'ensemble, les définitions de la robustesse et de la résilience ne sont pas forcément identiques dans tous les domaines mais on remarque qu'elles ont quand même des fortes similitudes.

Pour la robustesse, les auteurs parlent d'un état initial et du fait que la robustesse se qualifie par rapport à la perturbation qui ne doit pas le modifier ou modifier ses fonctionnalités principales. Pour la résilience, les auteurs établissent un consensus sur la capacité d'adaptation du système lors d'une perturbation jusqu'au retour à l'état initial du système. Cette adaptation est souvent décrite en deux phases la réaction à la perturbation puis la réponse pour un retour à l'état d'origine.

²Dictionnaire le Littré

³Tisseron, S. (2009). La Résilience. Presses Universitaires de France. isbn : 978-2-13-057795-9, 128 pages.

Table 3 – Définitions dans les différents domaines

	Référence	Robustesse	Résilience
INGENIERIE	(Ibanez et al. 2016)	<i>A system is robust with respect to an event if it is resilient for that event under all defined states</i>	<i>Resilience is the ability to minimize and recover from the consequences of an adverse event, [...] for a given state of the system</i>
	(Zhu & Basar 2015)	<i>System's ability to withstand a known range of uncertain parameters or disturbances. it's a pre-event concept</i>	<i>The resilience aspect of a system, which refers to the system's ability to recover online after adversarial events occur.</i>
	(Wybo 2008)	<i>Robustness corresponds to the ability of the organization to survive and stay under control by the emergence of new organizational patterns.</i>	<i>Resilience corresponds to the ability of the organization (at any level) to keep achieving its tasks by adapting its functioning to hazardous situations, uncertainty, time pressure and threats</i>
INFORMATIQUE	(Greco et al. 2012)	<i>Robustness is the capability of a system of maintaining given performance levels in the presence of unfavorable variations of operating conditions, is investigated and applied to real water distribution systems</i>	<i>Resilience indexes have been introduced in order to evaluate the usefulness of pressure head surplus in normal operating conditions to allow the network to overcome critical operating conditions</i>
	(Read 2005)	<i>Robustness is the maximum transient force that a system can receive before system collapse.</i>	<i>A distinction needs to be made between robustness and resilience of a system through the way the system responds to transient impulses. Human systems are complex due to the ability of humans not only to react and respond to their environment, but to construct abstract/symbolically based "environments" – cultures – that are as real in terms of their affect on behavior as the material environment in which individuals are embedded.</i>
ENVIRONNEMENT	(Dragicevic 2016)	<i>A system is considered to be robust when, in presence of external perturbations, it continues to function without fundamental changes to the original configuration.</i>	<i>It is considered to be resilient when it adapts to external perturbations while continuing to function, be it at the expense of changes in configuration.</i>
	(Mumby et al. 2014)	<i>Robustness measures the ability of a system to maintain itself within a narrow range of function and is ideally suited to problems that require careful setting of upper and lower bounds for system properties, such as optimal fisheries yield.</i>	<i>Resilience is the ability of a system to remain able to exhibit recovery trajectories even though disturbance may push the reef closer to a set of thresholds that mark the point where coral mortality outweighs recovery.</i>
	(Anderies et al. 2013)	<i>Robustness explicitly links the dynamics of systems to performance measures</i>	<i>The capacity to sustain a shock and continue to function and, more generally, cope with change</i>
	(Domptail et al. 2013)	<i>Robustness designates a change in the structure of the SES to maintain its function</i>	<i>Resilience is the capacity of a system to sustainably manage short disturbances (shocks) and long-term pressures (stresses) is related to its properties,</i>

	Référence	Robustesse	Résilience
BIOLOGIE	(Prostova et al. 2015)	<i>Robustness is the the invariance of phenotypes in the face of perturbation.</i>	<i>This three-step recovery of wild-type phenotype after its impairment – temporary acquisition of functional conformation, procurement of correcting mutations and negative selection</i>
	(Aschbacher et al. 2014)	<i>Robustness is defined as a property that allows a system to maintain its functions despite internal and external perturbations</i>	<i>Resilience has been variously defined as thriving in the face of adversity, or alternatively, the capacity to maintain homeostasis/allostasis under adversity</i>
	(Butler & Silver 2009)	<i>Robustness is the ability of a system to sustain function in the face of variations or perturbations in the operating environment.</i>	<i>Resilience is the ability to maintain an acceptable level of function in the face of various insults that challenge full, normal operation.</i>
BUSINESS	(Edgeman 2015)	<i>Robustness is enterprise resistance or immunity gained through strategies, policies, partnerships, and practices that maintain or advance competitive position when shocks or challenges arise.</i>	<i>Resilience is enterprise ability to self-renew through innovation, changing and reinventing itself by adapting its responses to political, social, economic and other competitive shocks or challenges.</i>
GER	(Ukrainitseva et al. 2016)	<i>Robustness is the ability to resist deviation from the original state</i>	<i>Resilience is the ability to recover after such deviation</i>

La différence entre les définitions est la même pour les deux concepts. Dans certains domaines les auteurs ne considèrent pas un seul état d'origine mais plutôt un intervalle de tolérance. Un système est donc considéré robuste par rapport à une perturbation lorsque la perturbation ne le fait pas sortir de cet intervalle de tolérance et pour la résilience lorsque le système retourne dans cet intervalle de tolérance et y reste. Cela explique que dans certains cas, la robustesse et la résilience sont considérées comme identiques.

Pour sa part, (Woods 2006) précise que la notion de résilience introduit un critère sur l'efficacité du retour à une normalité.

5 CONCLUSION

L'objectif de cette revue de littérature était de distinguer les notions de robustesse et de résilience dans les différents domaines pour, à l'avenir, identifier une définition de ces deux concepts dans le domaine de la logistique. Pour améliorer cette revue de littérature, il est envisagé d'étendre les critères de filtrage (robustesse OU résilience), d'affiner la recherche par mots clés ou par le résumé ainsi que d'étendre ces recherches à d'autres bases de données (Wiley, Taylor and Francis, Elsevier...) de manière à confirmer nos premiers résultats. Des courants comme le Resilience Engineering ou la vaste littérature sur la robustesse en Recherche Opérationnelle ou en Informatique pourront alors permettre de raffiner les concepts. De plus, des concepts proches comme celui d'homéostasie du risque pourront être mieux distingués (Wilde 1982).

Il sera également intéressant d'analyser les travaux menés dans des domaines dont les développements prennent un essor important ces dernières années, comme par exemple ceux portant sur la gestion des crises humanitaires (Vargas Florez, Lauras, Okongwu & Dupont 2015). Ce raffinement des définitions permettra également de proposer une définition de ces deux concepts plus spécifiquement dans le domaine de la logistique.

REMERCIEMENTS

Les travaux présentés dans cet article ont été financés par l'ANR dans le cadre du projet de recherche collaborative SMART PLANNING (ANR-16-CE26-0017).

References

- Alenazi, M. J. F. & Sterbenz, J. P. G. (2015). Evaluation and Comparison of Several Graph Robustness Metrics to Improve Network Resilience, *2015 7th International Workshop on Reliable Networks Design and Modeling (rndm) Proceedings* pp. 7–13.
- Amann, P. & James, J. I. (2015). Designing robustness and resilience in digital investigation laboratories, *Digital Investigation* **12**: S111–S120.
- Amantini, A., Choras, M., D'Antonio, S., Egozcue, E., Germanus, D. & Hutter, R. (2012). The human role in tools for improving robustness and resilience of critical infrastructures, *Cognition Technology & Work* **14**(2): 143–155.
- Anderies, J. M., Folke, C., Walker, B. & Ostrom, E. (2013). Aligning Key Concepts for Global

- Change Policy: Robustness, Resilience, and Sustainability, *Ecology and Society* **18**(2): 8.
- Anderies, J. M. & Hegmon, M. (2011). Robustness and Resilience across Scales: Migration and Resource Degradation in the Prehistoric US Southwest, *Ecology and Society* **16**(2): 22.
- Arbeev, K., Ukraintseva, S., Bagley, O., Duan, M., Arbeeveva, L., Zhbannikov, I., Cohen, A. & Yashin, A. I. (2016). Physiological dysregulation as promising measure of robustness and resilience in aging studies, *Gerontologist* **56**: 180–181.
- Aschbacher, K., Rodriguez-Fernandez, M., van Wietmarschen, H., Tomiyama, A. J., Jain, S., Epel, E., Doyle, F. J. & van der Greef, J. (2014). The hypothalamic-pituitary-adrenal-leptin axis and metabolic health: a systems approach to resilience, robustness and control, *Interface Focus* **4**(5): UNSP 20140020.
- Balzari, U. & Balzari, A. (2017). Robustness and resilience of a passive control solution assembling buffer and cladding panels, *Smart Structures and Systems* **20**(5): 637–640.
- Bongard, J. & Lipson, H. (2014). Evolved Machines Shed Light on Robustness and Resilience, *Proceedings of the Ieee* **102**(5): 899–914.
- Brailean, J. C., Sikora, T. & Miki, T. (1999). Special issue on error resilience and robustness for video coding technology, *Signal Processing-Image Communication* **14**(6-8): 443–445.
- Brandon-Jones, E., Squire, B., Autry, C. W. & Petersen, K. J. (2014). A contingent resource-based perspective of supply chain resilience and robustness, *Journal of Supply Chain Management* **50**(3): 55–73.
- Butler, M. P. & Silver, R. (2009). Basis of Robustness and Resilience in the Suprachiasmatic Nucleus: Individual Neurons Form Nodes in Circuits that Cycle Daily, *Journal of Biological Rhythms* **24**(5): 340–352.
- Camara, J., de Lemos, R., Laranjeiro, N., Ventura, R. & Vieira, M. (2017). Robustness-Driven Resilience Evaluation of Self-Adaptive Software Systems, *Ieee Transactions on Dependable and Secure Computing* **14**(1): 50–64.
- Capano, G. & Woo, J. J. (2017). Resilience and robustness in policy design: a critical appraisal, *Policy Sciences* **50**(3): 399–426.
- Carayannis, E. G., Grigoroudis, E. & Stamati, D. (2017a). Re-visiting BMI as an Enabler of Strategic Intent and Organizational Resilience, Robustness, and Remunerativeness, *Journal of the Knowledge Economy* **8**(2): 407–436.
- Carayannis, E. G., Grigoroudis, E. & Stamati, D. (2017b). Re-visiting BMI as an Enabler of Strategic Intent and Organizational Resilience, Robustness, and Remunerativeness (vol 8, pg 407, 2017), *Journal of the Knowledge Economy* **8**(2): 437–437.
- Daouda, K., Francois, P. & Pascale, Z. (2013). Technological networks robustness and resilience assessment, *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (ieee-Iesm 2013)* pp. 447–453.
- de Souza, R. & Zhou, R. (2015). *Improve Robustness and Resilience of Networked Supply Chains*.
- Domptail, S., Easdale, M. H. & Yuerlita (2013). Managing Socio-Ecological Systems to Achieve Sustainability: A Study of Resilience and Robustness, *Environmental Policy and Governance* **23**(1): 30–45.
- Dragicevic, A. Z. (2016). From robustness to resilience: A network Price identity approach, *Ecological Complexity* **28**: 47–53.
- Edgeman, R. (2015). Strategic resistance for sustaining enterprise relevance A paradigm for sustainable enterprise excellence, resilience and robustness, *International Journal of Productivity and Performance Management* **64**(3): 318–333.
- Fleischman, F. D., Boenning, K., Garcia-Lopez, G. A., Mincey, S., Schmitt-Harsh, M., Daedlow, K., Lopez, M. C., Basurto, X., Fischer, B. & Ostrom, E. (2010). Disturbance, Response, and Persistence in Self-Organized Forested Communities: Analysis of Robustness and Resilience in Five Communities in Southern Indiana, *Ecology and Society* **15**(4): 9.
- Fryer, P., Paterson, W. & Slater, N. (1987). Robustness of fouling heat-exchanger networks and its relation to resilience, *Chemical Engineering Research & Design* **65**(3): 267–271.
- Greco, R., Di Nardo, A. & Santonastaso, G. (2012). Resilience and entropy as indices of robustness of water distribution networks, *Journal of Hydroinformatics* **14**(3): 761–771.
- Ibanez, E., Lavrenz, S., Gkritza, K., Mejia-Giraldo, D. A., Krishnan, V., McCalley, J. D. & Somani, A. K. (2016). Resilience and robustness in long-term planning of the national energy and transportation system, *International Journal of Critical Infrastructures* **12**(1-2): 82–103.
- Klau, G. W. & Weiskircher, R. (2005). Robustness and resilience, in U. Brandes & T. Erlebach (eds), *Network Analysis: methodological Foundations*, Vol. 3418, pp. 417–437.
- Krupa, M. B., Chapin, F. S. & Lovcraft, A. L. (2014). Robustness or resilience ? Managing the intersection of ecology and engineering in an urban Alaskan fishery, *Ecology and Society* **19**(2): 17.
- Levin, S. A. & Lubchenco, J. (2008). Resilience, robustness, and marine ecosystem-based management, *Bioscience* **58**(1): 27–32.
- Martin, H. & Ludek, L. (2013). *The status and importance of robustness in the process of critical infrastructure resilience evaluation*.

- Mercer, E., Rungta, N. & Gillan, D. J. (2016). Special Issue on Systematic Approaches to Human-Machine Interface: Improving Resilience, Robustness, and Stability, *Ieee Transactions on Human-Machine Systems* **46**(2): 169–173.
- Mumby, P. J., Chollett, I., Bozec, Y.-M. & Wolff, N. H. (2014). Ecological resilience, robustness and vulnerability: how do these concepts benefit ecosystem management?, *Current Opinion in Environmental Sustainability* **7**: 22–27.
- Nair, S. & Howlett, M. (2016). From robustness to resilience: avoiding policy traps in the long term, *Sustainability Science* **11**(6): 909–917.
- Okoh, P. & Haugen, S. (2015). Improving the robustness and resilience properties of maintenance, *Process Safety and Environmental Protection* **94**: 212–226.
- Pavard, B., Dugdale, J., Saoud, N. B.-B., Darcy, S. & Salembier, P. (2009). *Conception de systèmes socio techniques robustes*, Octarès Editions.
- Perez-Hedo, M., Suay, R., Alonso, M., Ruocco, M., Giorgini, M., Poncet, C. & Urbaneja, A. (2017). Resilience and robustness of IPM in protected horticulture in the face of potential invasive pests, *Crop Protection* **97**: 119–127.
- Phan, T., Beare, R., Ma, H. & Srikanth, V. (2016). Network Robustness and Resilience among Different Acute Stroke Team Models in the Delivery of Acute Reperfusion Therapy: Less Is More from Spoke to Hub and Telestroke, *Cerebrovascular Diseases* **42**: 157–157.
- Ponnambalam, L., Long, D. H., Sarawgi, D., Fu, X. & Goh, R. S. M. (2014). Multi-Agent Models to Study the Robustness and Resilience of Complex Supply Chain Networks, *Proceedings of 2014 Ieee International Conference on Intelligent Autonomous Agents, Networks and Systems (inagentsys)* pp. 7–12.
- Prostova, M. A., Gmyl, A. P., Bakhmutov, D. V., Shishova, A. A., Khitrina, E. V., Kolesnikova, M. S., Serebryakova, M. V., Isaeva, O. V. & Agol, V. I. (2015). Mutational robustness and resilience of a replicative cis-element of RNA virus: Promiscuity, limitations, relevance, *Rna Biology* **12**(12): 1338–1354.
- Read, D. (2005). Some observations on resilience and robustness in human systems, *Cybernetics and Systems* **36**(8): 773–802.
- Rollo, C. & Shibata, D. (1991). Resilience, robustness, and plasticity in a terrestrial slug, with particular reference to food quality, *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **69**(4): 978–987.
- Sauvant, D. & Perez, J.-M. (2010). Robustness, hardness, flexibility, plasticity, resilience ... new quality criteria for animals and farming systems Foreword., *Productions Animales* **23**(1): 3–3.
- Schlichtmann, U. (2016). The next frontier in ic design: determining (and optimizing) robustness and resilience of integrated circuits and systems, *2016 China Semiconductor Technology International Conference (cstic)* .
- Scholz, R. W., Blumer, Y. B. & Brand, F. S. (2012). Risk, vulnerability, robustness, and resilience from a decision-theoretic perspective, *Journal of Risk Research* **15**(3): 313–330.
- Sha, Z. & Panchal, J. H. (2013). Towards the design of complex evolving networks with high robustness and resilience, in C. J. J. Paredis, C. Bishop & D. Bodner (eds), *2013 Conference on Systems Engineering Research*, Vol. 16, pp. 522–531.
- Tieman, M. (2017). Halal risk management: combining robustness and resilience, *Journal of Islamic Marketing* **8**(3): 461–475.
- Ukrainitseva, S., Yashin, A. I. & Arbeev, K. G. (2016). Resilience Versus Robustness in Aging, *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **71**(11): 1533–1534.
- Urruty, N., Tailliez-Lefebvre, D. & Huyghe, C. (2016). Stability, robustness, vulnerability and resilience of agricultural systems. A review, *Agronomy for Sustainable Development* **36**(1): 15.
- Vargas Florez, J., Lauras, M., Okongwu, U. & Dupont, L. (2015). A decision support system for robust humanitarian facility location, *Engineering Applications of Artificial Intelligence* **46**: 326–335.
- Whitson, H. E., Duan-Porter, W., Schmader, K., Morey, M., Cohen, H. J. & Colon-Emeric, C. (2016). Response to Ukrainitseva et al. Letter: Resilience Versus Robustness in Aging, *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **71**(11): 1535–1536.
- Wilde, G. J. S. (1982). The Theory of Risk Homeostasis: Implications for Safety and Health, *Risk Analysis* **2**(4): 209–225.
- Witten, T. M. (2014). It's Not All Darkness: Robustness, Resilience, and Successful Transgender Aging, *Lgbt Health* **1**(1): 24–33.
- Wong, A. S. Y., Pogodaev, A. A., Vialshin, I. N., Helwig, B. & Huck, W. T. S. (2017). Molecular Engineering of Robustness and Resilience in Enzymatic Reaction Networks, *Journal of the American Chemical Society* **139**(24): 8146–8151.
- Woods, D. D. (2006). Essential characteristics of resilience, *Resilience engineering: Concepts and precepts* pp. 21–34.
- Wybo, J.-L. (2008). The role of simulation exercises in the assessment of robustness and resilience of private or public organizations, in H. J. Pasman & I. A. Kirillov (eds), *Resilience of Cities to Terrorist and Other Threats: Learning from 9/11 and Further Research Issues*, pp. 491–507.
- Zhu, Q. & Basar, T. (2015). Game-theoretic methods for robustness, security, and resilience of cyber-physical control systems games-in-games principle for optimal cross-layer resilient control systems, *Ieee Control Systems Magazine* **35**(1): 46–65.

RREFLEX, AN INNOVATIVE TOOL FOR THE SYNTHESIS OF FLEXIBLE HEAT EXCHANGER NETWORK

L. PAYET, G. HETREUX, R. HETREUX, P. FLOQUET

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, 31432 Toulouse, France

lucille.payet@ensiacet.fr, gilles.hetreux@ensiacet.fr, raphaele.thery@ensiacet.fr, pascal.floquet@ensiacet.fr

ABSTRACT: *The massive introduction of connected technologies will revolutionize the way of designing and controlling industrial systems. Associated with the concept of Process Integration (PI), they should significantly improve the energy performance of production facilities through the introduction of Heat Exchanger Networks (HEN). In this context, the purpose of the RREFlex methodology is to synthesize configurable and flexible HEN. For production systems, flexibility comes in many aspects. This communication focuses on the adaptability property, i.e. the ability of a network to perform its function suitably for a set of N operating points. The MILP models implemented for network synthesis are described, as well as the additional constraints needed for adaptability purpose. Different alternatives are presented and applied to an example.*

KEYWORDS: *configurable heat exchanger networks, adaptability, MILP formulation, process integration*

1 INTRODUCTION

The concept of *Factory of the Future* (or Industry 4.0), technologically speaking, originated in the booming of connected systems in the industry, which is changing drastically our ways of designing and operating production facilities. Many definition of this concept can be found in the literature as it covers several aspects. Among those, the definition commonly acknowledged by both industrial stakeholders and academics is from the firm Fives, an industrial engineering group deeply involved in the development of this concept. According to Fives Group (“Innovation,” 2017), the *Factory of the Future* will have to be more environmentally friendly, using production unit consuming fewer resources (raw material, energy) and generating less pollution (wastes, CO₂), smarter and with more sophisticated production equipment. More flexible, using reconfigurable production facilities, factories will be able to offer supplies closer to markets demand.

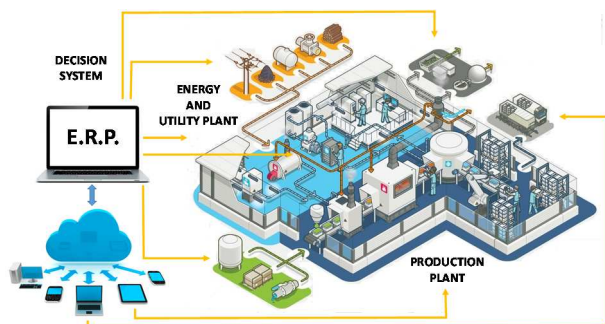


Figure 1: Integrated management of production and energy

According to this definition, one of the cornerstones of Industry 4.0 is to provide a concrete answer to the energy

issue which is a major challenge in the coming decades for manufacturers. By exploiting an information system built on a communication network and constant and instantaneous exchanges, this management can become more efficient by optimizing the coordination between the needs and availabilities of each part of the site (figure 1). Thus, in addition to generating new productivity gains, a rigorous energy management will enable the industrial sector to reduce its environmental footprint.

One of the most effective strategies to improve the energy efficiency of industrial sites is to maximize heat recovery via *Energy Integration (EI)*, as still pointed out recently by (Pereira *et al.*, 2017). Indeed, *EI* is an essential concept which consists in coupling the hot and cold streams internal to the process rather than consuming external utilities. The *EI* methodology is based on the *Pinch Analysis* (Linnhoff and Hindmarsh, 1983) and leads to the *Heat Exchanger Network (HEN)* as a final solution to achieve a significant reduction of energy consumption. On the other hand, it is common sense to note that the more integrated the system, the less flexible it is. Indeed, by nature, a *HEN* is designed for a specific and stable operating point. In practice, however, plants often operate at different conditions. Typically, a chemical plant will either process different feedstocks or operate at various capacity levels. Due to interactions between the process and the heat exchangers network, variations of operating conditions at the chemical plant will imply different values of flow rates and temperatures for the streams that participate in the *HEN*. As a consequence, the problem of synthesizing flexible network configurations which can handle a finite number of N different scenarios of operation becomes an important design problem. From an operating point of view, the control of such reconfigurable energy systems is very complex and the use of numerical technologies seems inevitable. Those connected devices will enable the decision-makers to follow the needs in real-time and to

reconfigure automatically the topology of the HEN in order to optimize the energy consumption of the production system. Designing such systems is precisely what the RREFlex project aims at.

The work presented in this communication concerns one of the steps of the methodology carried out in this software, the design of HEN with adaptability capabilities. The remainder of this paper is organized as follows. The main principles of Energy Integration are briefly recalled and the outlines of the RREFlex methodology are described in section 2. Since the one-scenario MILP model is extended to the multi-scenarios problem, a review of this model is presented first in section 3. The extension of the model for multi-scenarios considerations is discussed in section 4. Finally in section 5, the methodology is applied on a case study and results are discussed.

2 ENERGY INTEGRATION

2.1 Principle

The concept of energy integration is based primarily on the principle of "energy recycling". When it is implemented at the scale of an industrial site, the plant is considered as a complex network in which coexist hot process streams that have to be cooled and cold process streams that have to be heated. A relevant coupling of compatible hot and cold streams can reduce drastically the need for utilities, thereby improving the overall energy efficiency of the facility.

2.2 Equations of heat exchangers

The fundamental components of the recovery system are the counter-current heat exchangers (

Figure 2).

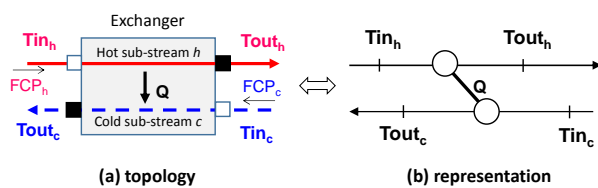


Figure 2: heat exchanger e

In a counter-current heat exchanger e , the heat load Q_e exchanged between a hot sub-stream h and a cold sub-stream c is calculated according to equation (1).

$$Q_e = FCP_h \cdot (Tin_h - Tout_h) = FCP_c \cdot (Tout_c - Tin_c) \quad (1)$$

FCP_h and FCP_c are the heat capacity flow rate (flowrate \times specific heat) of respectively hot stream h and cold stream c . Tin_h and $Tout_h$ are respectively the inlet and outlet temperatures of hot stream h while Tin_c and $Tout_c$ are the inlet and outlet temperatures of cold stream c . A characteristic feature of an exchanger e is its heat exchange area A_e , given by the equation (2):

$$A_e = \frac{Q_e}{U_e \Delta T_{lm_e}} \quad (2)$$

where U_e is the overall heat transfer coefficient of heat exchanger e between hot and cold streams and ΔT_{lm_e} is the logarithmic average temperature difference between hot and cold streams evaluated by equation (3).

$$\Delta T_{lm} = \frac{(Tin_h - Tout_c) - (Tout_h - Tin_c)}{\ln\left(\frac{Tin_h - Tout_c}{Tout_h - Tin_c}\right)} \quad (3)$$

Another characteristic values are the heat exchange approach temperature defined as $\Delta T_1 = Tin_h - Tout_c$ and $\Delta T_2 = Tout_c - Tin_c$. The Heat Exchange Minimum Approach Temperature (HEMAT) of a HEN is linked to the technology of heat exchangers.

2.3 Design of heat exchanger networks

Based on the choice of the HEMAT, the Pinch Analysis, detailed in (Kemp, 2007), provides the minimum energy requirement (MER) which correspond to the minimum consumption. Then, the final decisive step is the design of the Heat Exchanger Network that achieves this energy target. There is a substantial literature on HEN design methods, as reflected in (Furman and Sahinidis, 2002), which lists and analyses as many as 461 articles from the late 40's to 2000. This study highlights different methodologies that were carried out. The approaches concern algorithmic methods, methods based on heuristics or mathematical programming techniques. Nevertheless, in recent years, the latter predominates and the need for more robust formulations and more efficient global optimization algorithms is all the more crucial. Recently, (Silva Gercio P. et al., 2017) highlighted the need of a simultaneous method to provide HEN for multiproduct plants taking into account fluctuations on the operating conditions. The MILP formulation was selected here to ensure the optimality of the solutions.

2.4 Philosophy of the RREFlex methodology

The HEN that achieves the MER is generally not the ultimate target and many criteria can influence the final topology of a "viable" HEN. The challenge with designing viable HEN is that it is often not possible to identify and formulate all constraints out of hand. As highlighted by (Smith et al., 2010), retrofitting a flexible HEN to an operating plant remains a complex and critical problem. (Sreepathi and Rangaiah, 2014) concluded that it would be better to provide several "good" solutions and let experienced engineers select the most suitable for its own case. Based on this philosophy, the computational framework developed as part of the RREFlex project proposes an original approach, which consists in providing the site engineer with several HEN solutions and letting him post-evaluate the performance of each one using key performance indicators (KPIs) based on economical, topologi-

cal, practical and operational considerations. The methodology consists in an iterative process relying on 4 major modules:

- an enhanced data collection based on the analysis of process historical data,
- the energy diagnosis based on the *Pinch Analysis*,
- the proposal and evaluation of several configurations of HEN achieved by an optimization approach using *Mixed Integer Linear Programming (MILP)* formulation in order to guarantee the efficient calculation of several configurations in case of industrial scale problem. *CPLEX* kernel is used through *ILOG IDE* software (*IBM*).
- the a posteriori evaluation of the performances of these configurations: performance indicators are relative to on-site considerations such as topology of the HEN or its flexibility in addition to more traditional indicators like energy consumption and global costs. Then, the analysis of these solutions by the engineer leads to the formulation of new constraints. Note that the methodology for enhanced data acquisition and post-evaluation of the robustness of an integrated system can be found in (Payet et al., 2017).

The method therefore assigns the end-user a pivotal deciding role, and new improved HEN solutions are designed iteratively. With such a scheme, one HEN solution provides the basis for the next one, and the design process stops when *KPIs* can no longer be improved and the HEN configurations fulfill the industrial on-site constraints correctly.

3 SYNTHESIS OF ONE-SCENARIO HEN

3.1 Problem statement

The HEN synthesis problem addressed in this section can be stated as follows. Given is the set S of process stream to be integrated, divided in a subset H of hot process streams which have to be cooled and a subset C of cold process streams which have to be heated ($S = H \cup C$). Specified with fixed values are also each stream's heat capacity flow rates FCP_s , the inlet temperature $T_{in,s}$ and outlet temperature $T_{out,s}$ targeted. Under this assumption, this case corresponds to the synthesis of HENs for one scenario of operation. The objective is to determine the HEN which satisfies the two following criteria:

- minimum utility cost,
- minimum number of heat exchangers units.

The two criteria are based on the observation that a HEN that required minimum operating cost (criterion 1) and fewest number of heat exchanger units (criterion 2) is very close to the economic optimal solution ((Linnhoff and Hindmarsh, 1983). The solution has to define the network by providing the following:

- hot and cold utilities required,
- network configurations i.e. streams matches and the number of units,

- heat loads and operating temperatures of each exchangers,

The following basic assumptions will be made for modelling the HEN synthesis problem:

- the enthalpy of the process streams is a linear function of temperature,
- the minimum temperature approach DT_{min} is specified by the design engineer.

3.2 Modelling of the HEN synthesis problem

Each heat exchange is characterized by the hot and cold sub-streams passing through the considered exchanger. Each sub-stream is bounded by a *temperature interval* defining the inlet and outlet temperatures of the corresponding hot and cold sub-streams. Figure 2 shows an exchanger (figure 2.a) and its representation (figure 2.b). A process stream can be heated up or cooled down by one or several heat exchangers in series (figure 3).

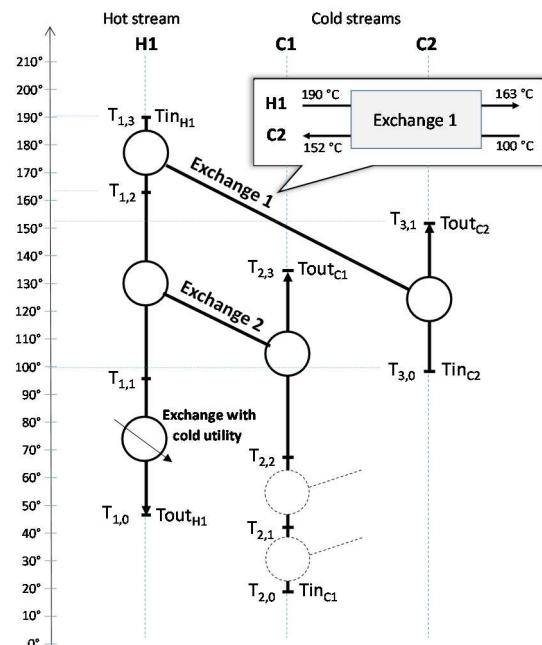


Figure 3: notion of sub-stream and temperature interval

The number of exchange of each stream is a priori unknown, as well as the temperature defining the bounds of the sub-streams. Determining these variables is the goal of the optimization model. For instance, figure 3 illustrates a solution for the hot stream **H1**: it exhibits two exchanges respectively with cold stream **C1** (exchange 2) and **C2** (exchange 1) and one exchange with a cold utility. Then, it is split into three sub-streams bounded respectively by $[47, 98]$, $[98, 163]$ and $[163, 190]$. The operating conditions of the exchanger implementing the exchange 1 (for instance) are indicated on figure 3.

The modelling of the HEN synthesis problem is based on a *bipartite flow graph* and a *routing graph*. The bipartite flow graph couples the heat load of each sub-stream and the matches between a hot stream and a cold stream.

Thanks to such representation, it can be seen as a *transshipment problem* of energy from hot streams (on the left) toward cold streams (on the right). The routing graph reflects the respect of constraints of potential (temperature) resulting from the 1st and 2nd principles of thermodynamic. The use of such a graphical and formal support allows a systematic and unambiguous construction of the *MILP* model.

3.3 Model formulation

3.3.1 Nomenclature

In order to formulate the proposed *MILP* model, the following definitions are necessary:

(a) Sets:

- S : set for index of streams such that $S = H \cup C$
- H : set for index of hot streams, $NH = \text{card}(H)$
- C : set for index of cold streams, $NC = \text{card}(C)$
- I_s : set for index of intervals with stream s
- I_h : set for index of intervals of hot stream h
- I_c : set for index of intervals of cold stream c

(b) Indexes:

- s : index for streams, $s \in S$,
- h : index for hot streams, $h \in H$,
- c : index for cold streams, $c \in C$
- i : index for intervals of stream s , $i \in I_s$
- k : index for intervals of hot stream h , $k \in I_h$
- l : index for intervals of cold stream c , $l \in I_c$

(c) Parameters:

- Tin_s : inlet temperature of stream s (K),
- $Tout_s$: outlet temperature of stream s (K),
- FCP_s : heat capacity flow rate of stream s ($kW.K^{-1}$),
- $DTmin$: temperature approach of the considered process (K),
- N_s : maximum number of exchange desired for stream s , $N_s = \text{card}(I_s)$,
- CUH : cost per unit for hot utility (€/kW),
- CUC : cost per unit for cold utility (€/kW),
- CFE : fixed cost per exchanger (€),
- MQ : “big M ” value for heat load,
- MT : “big M ” value for temperature.

Binary variables are introduced to designate the existence of each potential heat exchanger in the network and to model fixed cost charges for the exchangers. Continuous variables are assigned to temperatures and heat loads.

(d) Decision variables :

- $Qex_{h,k,c,l}$: heat load exchanged between interval k of hot stream h and interval l of cold stream c (kW),
- $Q_{s,i}$: heat load available in interval i of stream s (kW),

- $QU_{s,i}$: heat load exchanged between interval i of stream s and a utility (kW),
- $T_{s,i}$: upper bound temperature of interval i of stream s and a utility (K),
- $y_{h,k,c,l}$: binary variable stating whether the interval k of hot stream h exchanged with interval l of cold stream c ,
- $z_{s,i}$: binary variable stating whether the interval i of stream s exchanged with a utility,

3.3.2 MILP model

With these definitions, the formulation can now be presented. The general model involves overall heat balances for each stream, stream energy balances at each interval, assignment of known stage temperatures, calculation of hot and cold utility loads and logical constraints. In this first formulation, no assumption is made on the type of hot and cold utility.

Assignment of inlet and outlet temperature of streams

$$T_{c,0} = Tin_c \quad \forall c \in C \quad (4)$$

$$T_{c,N_c} = Tout_c \quad \forall c \in C \quad (5)$$

$$T_{h,0} = Tout_h \quad \forall h \in H \quad (6)$$

$$T_{h,N_h} = Tin_h \quad \forall h \in H \quad (7)$$

In this actual model, fixed inlet and outlet temperatures are assumed and assigned respectively to the temperature variable $T_{h,k}$ bounding each stream. For a hot stream h , the inlet temperature Tin_h corresponds to the variable $T_{h,k}$ located at index $k = N_h$, while for a cold streams c , the inlet temperature Tin_c is assigned to the temperature variable $T_{c,l}$ located at $l=0$.

Feasibility of temperatures

$$T_{s,i} \geq T_{s,i-1} \quad \forall s \in S, \forall i \in I_s \quad (8)$$

$$T_{c,l} \leq T_{h,k} - DTmin + (1 - y_{h,k,c,l}) \cdot MT \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c \quad (9)$$

$$T_{c,l-1} \leq T_{h,k-1} - DTmin + (1 - y_{h,k,c,l}) \cdot MT \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c \quad (10)$$

Constraints (8) are needed to specify a monotonic increase of temperature at each successive interval i of a stream s . Note that $T_{s,i} = T_{s,i-1}$ when $Q_{s,i} = 0$. To ensure feasible driving forces for exchangers selected during the optimization procedure, constraints (6) and (7) verify that a minimum temperature difference $DTmin$ exists between the inlet and outlet temperature of the exchanger. As these constraints have to be activated only if a match exists, the binary variables $y_{h,k,c,l}$ are used. When a match $((h,k),(c,l))$ occurs, $y_{h,k,c,l} = 1$ and the constraint becomes active so that the approach temperature is properly considered. However, when the match does not occur, $y_{h,k,c,l} = 0$ and the contribution of the “big M ” on the right-hand side deems the equation inactive. The corresponding big “ M ” value is

set to the highest temperature of the streams of the process, multiplied by two.

Multiple hot and cold utilities are not considered in this paper. Nevertheless, note that similar constraints are used for utility exchangers when the outlet temperature of the utility stream, UT_{out} , is not strictly higher (for hot utilities) or lower (for cold utilities) than the outlet temperature of the process streams.

Heat balance for each sub-stream

$$Q_{h,k} = \sum_{c \in C} \sum_{l \in I_c} Q_{ex_{h,k,c,l}} + QU_{h,k} \quad \forall h \in H, \forall k \in I_h \quad (11)$$

$$Q_{c,l} = \sum_{h \in H} \sum_{k \in I_h} Q_{ex_{h,k,c,l}} + QU_{c,l} \quad \forall c \in C, \forall l \in I_c \quad (12)$$

$$Q_{s,i} = FCP_s * (T_{s,i} - T_{s,i-1}) \quad \forall s \in S, \forall i \in I_s \quad (13)$$

Energy balances are defined at the upper bound of each interval i of a stream s . So, constraints (13) determine the temperature $T_{s,i}$ taking into consideration the fact that for two adjacent intervals of a stream, the temperature to the upper bound of the previous interval corresponds to the temperature to the lower bound of the next interval. Constraints (11) (resp. (12)) are needed to ensure sufficient cooling (resp. heating) for each interval k (resp. l) of a hot (resp. cold) stream h (resp. c). These constraints specify that the heat transfer requirement of each interval of a hot (resp. cold) stream must equal the sum of the heat it exchanges with a cold (resp. hot) stream or with a cold (resp. hot) utility stream.

Overall heat balance for each stream

$$FCP_s(T_{s,N_s} - T_{s,0}) = \sum_{i \in I_s} Q_{s,i} \quad \forall s \in S \quad (14)$$

Defined for each stream s , constraint (14) is an overall heat balance to ensure that the total heat load of a stream must equal the sum of the heat exchanged by each interval i of this stream s .

Allocation constraints of each exchange

Allocation constraints (15) and (16) are needed to determine the existence of a match between process streams and also any match invoking utility streams.

$$z_{h,k} + \sum_{c \in C} \sum_{l \in I_c} y_{h,k,c,l} \leq 1 \quad \forall h \in H, \forall k \in I_h \quad (15)$$

$$z_{c,i} + \sum_{h \in H} \sum_{k \in I_h} y_{h,k,c,l} \leq 1 \quad \forall c \in C, \forall l \in I_c \quad (16)$$

An integer value of one for a binary variable designates that the match is present in the network. Then, these constraints define that at most one exchange is possible for each sub-stream.

It should be noted that several other logical constraints are introduced into the model in order to improve the convergence rate of the model but are not described in this paper.

Existence of exchanges

$$QU_{s,i} \leq MQ \cdot z_{s,i} \quad \forall s \in S, \forall i \in I_s \quad (17)$$

$$Q_{ex_{h,k,c,l}} \leq MQ \cdot y_{h,k,c,l} \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c \quad (18)$$

Constraints (17) and (18) force heat load variables to zero when no match exists between process streams or with a utility. The upper bound (big "M" value) is set to the biggest heat content of the streams of the process, multiplied by two.

Objective function

Equation (19) is the objective function and has to minimize the total number of heat exchangers units (equation (20) and (21)) and the global utility cost.

$$\min z = NbTotEx + \left(CUC \cdot \sum_{h \in H} \sum_{k \in I_h} QU_{h,k} + CUH \cdot \sum_{c \in C} \sum_{l \in I_c} QU_{c,l} \right) \quad (19)$$

where:

$$NbIntEx = \sum_{h \in H} \sum_{k \in I_h} \sum_{c \in C} \sum_{l \in I_c} y_{h,k,c,l} \quad (20)$$

$$NbTotEx = NbIntEx + \sum_{s \in S} \sum_{i \in I_s} z_{h,k} \quad (21)$$

3.4 Stream splitting

When plants exhibit a significant imbalance in the number of streams of each kind (hot/cold), it may be appropriate to split sub-streams into several parallel subdivisions to increase the opportunities of match. Parameter $Ndiv_s$ define the maximum number of subdivision that the N_s sub-streams of a stream s can be subject to. To take into account the stream splitting, constraints (15) and (16) are replaced by the following constraints (22) and (23).

$$Ndiv_h \cdot z_{h,k} + \sum_{c \in C} \sum_{l \in I_c} y_{h,k,c,l} \leq Ndiv_h \quad \forall h \in H, \forall k \in I_h \quad (22)$$

$$Ndiv_c \cdot z_{c,i} + \sum_{h \in H} \sum_{k \in I_h} y_{h,k,c,l} \leq Ndiv_c \quad \forall c \in C, \forall l \in I_c \quad (23)$$

These constraints define that at most $Ndiv_s$ exchanges with other process streams are allowed for each sub-stream or only one for a match with a utility. An example of stream splitting is shown on figure 4 where a hot sub-stream is split into two subdivisions around the node $HS_{h,k}$.

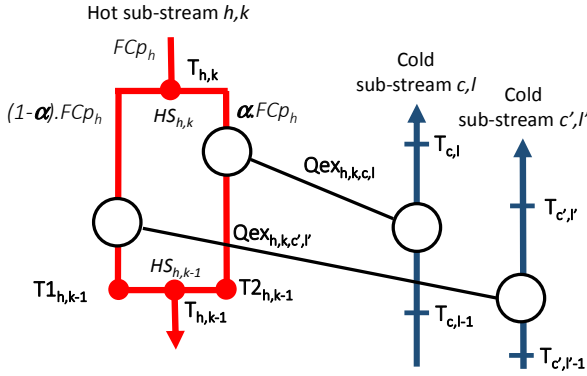


Figure 4: split of a sub-stream into two subdivisions

Here, each sub-division of sub-stream h,k exchanges a heat load given by equations (24) and necessitates equation (25) to calculate the outlet temperature $T_{h,k-1}$ of the sub-stream (node $HS_{h,k-1}$)

$$\begin{aligned} Qex_{h,k,c,i} &= \alpha \cdot FCP_h \cdot (T_{h,k} - T_{2h,k-1}) \\ Qex_{h,k,c',i'} &= (1 - \alpha) \cdot FCP_h \cdot (T_{h,k} - T_{1h,k-1}) \end{aligned} \quad (24)$$

$$T_{h,k-1} = \alpha \cdot T_{2h,k-1} + (1 - \alpha) \cdot T_{1h,k-1} \quad (25)$$

The introduction of variables α makes non-linear the equations (24) and (25). For this reason, the assumption of isothermal mixing at node $HS_{h,k-1}$ is imposed. This simplification induces that $T_{1h,k-1} = T_{2h,k-1} = T_{h,k-1}$ and variables α are no longer required in the model. As a result, it eliminates the requirement for nonlinear heat balances (24) around each exchanger as well as nonlinear heat mixing equations (25). Instead, only the heat balances (11), (12) and (13) around each sub-stream are needed.

For network structures where splits are present, it should be noted that the assumption for isothermal mixing may lead to an overestimation of the area cost. In some cases, this assumption might exclude network structures which are only feasible with non-isothermal mixing. However, for larger problems, this restriction is less frequent since greater flexibility in matching and selection of configuration can usually ensure an equally good network without this particular split structure.

4 SYNTHESIS OF MULTI-SCENARIOS HEN

4.1 Problem statement

Specifically, the multi-scenarios synthesis problem which is to be addressed in this communication can be stated as follows. As previously, given is a set H of hot streams which have to be cooled, and a set of cold streams C which have to be heated. Flow rates, inlet and outlet temperatures are specified for these streams at N scenarios of operation. Auxiliary heating and cooling are available from a set of utilities. The problem is then to synthesize a HEN configuration which remains feasible for the finite set of N scenarios of operation and which minimizes:

- the number of heat exchanger units,
- the utility costs for each scenarios of operation.

It should be noted that the first criterion cited above is independent of the scenarios while the second one is dependent of the duration of each scenario considered in the problem. Therefore, duration need to be specified in order to evaluate correctly the utility costs. Moreover, it is assumed that the dynamic effect of changing flow rates, inlet and outlet temperatures from one scenario to another is neglected.

4.2 Procedure for the multi-scenarios problem

The following assumptions are made on the heat exchangers units:

- each unit is assigned to the same pair of hot sub-streams h,k and cold sub-streams c,l at each scenario of operation. This assumption is stated for practical convenience, because otherwise extensive piping may be required to use a given unit for different pairs of sub-streams.
- each unit can handle variable heat loads. This implies the availability of by-passes in each heat exchanger unit to adjust the desired heat load.

To formulate mathematically the multi-scenario problem, the following additional indices and parameters are needed:

- P : set of scenario of operation,
- p : index for scenario of operation, $p \in P$,

Under these assumptions, the synthesis of flexible HENs for multi-scenarios operation is realized in 3 main stages:

1. first, *MILP* model proposed in section 3 is extended for multi-scenarios operation. Let call *MS* this *MILP* model. It determines the required sub-streams matches and the heat that must be exchanged at each match in each scenario of operation. Let E be the set of the exchangers e assigned to each match.
2. then, for each match, the corresponding area $A_{e,p}$ is calculated for each scenario p . The effective area A_e of the exchanger e corresponds to the largest one, i.e.:
$$A_e = \max_{p \in P} (A_{e,p}) \quad \forall e \in E \quad (26)$$
3. Finally, for each scenario p and for each exchanger e , the coefficients $\alpha_{h,p}$ and $\alpha_{c,p}$ are calculated by solving a *NLP* model named *MFR* $_{e,p}$,

In order to evaluate the quality of the solution, the consumption in hot utilities HUC_{global} and in cold utilities CUC_{global} of the network found in step 1 is compared respectively with the maximum consumption $HUMAX_{global}$ and $CUMAX_{global}$ and the minimum consumption $HUMER_{global}$ and $CUMER_{global}$ where:

$$HUMAX_{global} = \sum_{p \in P} \sum_{c \in C} FCP_{c,p} \cdot (T_{out_{c,p}} - T_{in_{c,p}}) \quad (27)$$

$$CUMAX_{global} = \sum_{p \in P} \sum_{h \in H} FCP_{h,p} \cdot (T_{in_{h,p}} - T_{out_{h,p}}) \quad (28)$$

$$HUMER_{global} = \sum_{p \in P} HUMER_p \quad (29)$$

$$CUMER_{global} = \sum_{p \in P} CUMER_p \quad (30)$$

$CUMER_p$ and $HUMER_p$ are the MER respectively in cold and hot utilities obtained for the scenario p if the matches could be independent between each scenario.

4.3 MILP formulation of model MS

As indicated in the previous subsection, the $MILP$ model proposed in section 3 is extended for determining the required streams matches and the heat that must be exchanged at each match in each scenario of operation. Thus, parameters FCP_s , Tin_s and $Tout_s$ become $FCP_{s,p}$, $Tin_{s,p}$ and $Tout_{s,p}$ where index p defines the scenario. Similarly, variables $Q_{s,i}$, $QU_{s,i}$, $Qex_{h,k,c,l}$ and $T_{s,i}$ become respectively $Q_{s,i,p}$, $QU_{s,i,p}$, $Qex_{h,k,c,l,p}$ and $T_{s,i,p}$. On the contrary, binary variables $y_{h,k,c,l}$ and $z_{s,i}$ remain independent of the scenario to keep identical the network topology. The model is then defines as follow:

$$T_{c,0,p} = Tin_{c,p} \quad \forall c \in C, \forall p \in P \quad (31)$$

$$T_{c,N_c,p} = Tout_{c,p} \quad \forall c \in C, \forall p \in P \quad (32)$$

$$T_{h,0,p} = Tout_{h,p} \quad \forall h \in H, \forall p \in P \quad (33)$$

$$T_{h,N_h,p} = Tin_{h,p} \quad \forall h \in H, \forall p \in P \quad (34)$$

$$T_{s,i,p} \geq T_{s,i-1,p} \quad \forall s \in S, \forall i \in I_s, \forall p \in P \quad (35)$$

$$T_{c,l,p} \leq T_{h,k,p} - DTmin + (1 - y_{h,k,c,l}) \cdot MT \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c, \forall p \in P \quad (36)$$

$$T_{c,l-1,p} \leq T_{h,k-1,p} - DTmin + (1 - y_{h,k,c,l}) \cdot MT \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c, \forall p \in P \quad (37)$$

$$Q_{h,k,p} = \sum_{c \in C} \sum_{l \in I_c} Qex_{h,k,c,l,p} + QU_{h,k,p} \quad \forall h \in H, \forall k \in I_h, \forall p \in P \quad (38)$$

$$Q_{c,l,p} = \sum_{h \in H} \sum_{k \in I_h} Qex_{h,k,c,l,p} + QU_{c,l,p} \quad \forall c \in C, \forall l \in I_c, \forall p \in P \quad (39)$$

$$Q_{s,i,p} = FCP_{s,p} * (T_{s,i,p} - T_{s,i-1,p}) \quad \forall s \in S, \forall i \in I_s, \forall p \in P \quad (40)$$

$$FCP_{s,p} (T_{s,N_s,p} - T_{s,0,p}) = \sum_{i \in I_s} Q_{s,i,p} \quad \forall s \in S, \forall p \in P \quad (41)$$

$$z_{h,k} + \sum_{c \in C} \sum_{l \in I_c} y_{h,k,c,l} \leq 1 \quad \forall h \in H, \forall k \in I_h \quad (42)$$

$$z_{c,i} + \sum_{h \in H} \sum_{k \in I_k} y_{h,k,c,i} \leq 1 \quad \forall c \in C, \forall i \in I_c \quad (43)$$

$$QU_{s,i,p} \leq MQ \cdot z_{s,i} \quad \forall s \in S, \forall i \in I_s, \forall p \in P \quad (44)$$

$$Qex_{h,k,c,l,p} \leq MQ \cdot y_{h,k,c,l} \quad \forall h \in H, \forall k \in I_h, \forall c \in C, \forall l \in I_c, \forall p \in P \quad (45)$$

Equation (46) is the objective function and has to minimize the total number of heat exchangers units and the global utility cost, where D_p is the duration of scenario p .

$$\min z = NbTotEx + CUC \cdot \sum_{h \in H} \sum_{k \in I_h} \sum_{p \in P} D_p \cdot QU_{h,k,p} + CUH \cdot \sum_{c \in C} \sum_{l \in I_c} \sum_{p \in P} D_p \cdot QU_{c,l,p} \quad (46)$$

4.4 NLP formulation of models $MFR_{e,p}$

In solution of the $MILP$ model MS , each variable $y_{h,k,c,l}$ such that $y_{h,k,c,l} = 1$ indicates the existence of an exchanger e at location $(h_e, k_e, c_e, l_e) = (h, k, c, l)$ in each scenario p . This area A_e is evaluated with equation (2), (3) and (26). Now, considering as parameters $Qex_{h,k,c,l,p}$, $T_{h,k,p}$, $T_{c,l,p}$, $T_{h,k-1,p}$ and $T_{c,l-1,p}$ (renamed respectively $Qex_{e,p}$, $Thin_{e,p}$, $Tcout_{e,p}$, $Thout'_{e,p}$ and $Tcin_{e,p}$), it is possible to determine the values of variables $\alpha_{h_e,p}$ and $\alpha_{c_e,p}$ that govern the operation of the valves coupled with bypasses (figure 5).

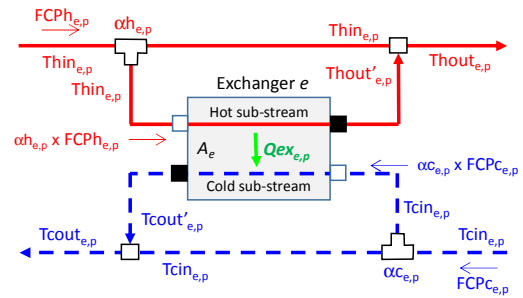


Figure 5: heat exchanger e with by-passes

For each exchanger e and each scenario p , the four variables $\alpha_{h_e,p}$, $\alpha_{c_e,p}$, $Tcout'_{e,p}$ and $Thout'_{e,p}$ are the solutions of the following NLP model $MFR_{e,p}$:

$$\alpha_{h_e,p} \cdot FCP_{h_e,p} \cdot (Thin_{e,p} - Thout'_{e,p}) = Qex_{e,p} \quad (47)$$

$$\alpha_{c_e,p} \cdot FCP_{c_e,p} \cdot (Tcout'_{e,p} - Tcin_{e,p}) = Qex_{e,p} \quad (48)$$

Constraints (47) and (48) determine the heat flow rate passing through the exchanger e .

$$\alpha_{h_e,p} \cdot Thout'_{e,p} + (1 - \alpha_{h_e,p}) \cdot Thin_{e,p} = Thout_{e,p} \quad (49)$$

$$\alpha_{c_e,p} \cdot Tcout'_{e,p} + (1 - \alpha_{c_e,p}) \cdot Tcin_{e,p} = Tcout_{e,p} \quad (50)$$

Constraints (49) and (50) calculate the intermediate temperatures $Tcout'_{e,p}$ and $Thout'_{e,p}$ in order to reach the outlet temperature targeted $Tcout_{e,p}$ and $Thout_{e,p}$.

$$U \cdot A_e \cdot \frac{(Thout'_{e,p} - Tcin_{e,p}) - (Thin_{e,p} - Tcout'_{e,p})}{\ln \left(\frac{Thout'_{e,p} - Tcin_{e,p}}{Thin_{e,p} - Tcout'_{e,p}} \right)} = Qex_{e,p} \quad (51)$$

Constraints (51) take into account the area A_e of the exchanger e .

$$z = \max \alpha_{h_e,p} + \alpha_{c_e,p} \quad (52)$$

The objective (52) seeks to obtain solutions preferentially using only one of the two by-passes.

5 APPLICATION

This example problem consists of two hot streams and two cold streams. There are three scenarios of operation of the same duration in which the flow rates, the inlet and outlet temperatures take different values. The data for these scenarios are shown in Table 1 and a minimum temperature approach $\Delta T_{min} = 10^\circ\text{C}$ is assumed. The MER results are reported in Table 2. The objective is to derive a HEN which is feasible for the three scenarios of operation, and requires minimum utility cost with the fewest number of heat exchangers.

Table 1: Data for example

	Stream	kind	FCp (kW/°C)	Tin (°C)	Tout (°C)
Scenario 1	H1	hot	2	180	60
	H2	hot	3	150	30
	C1	cold	2	20	135
	C2	cold	3	80	140
Scenario 2	H1	hot	2.5	130	60
	H2	hot	1	180	20
	C1	cold	4	60	160
	C2	cold	1.5	75	100
Scenario 3	H1	hot	1.5	170	95
	H2	hot	2.5	120	25
	C1	cold	2.5	40	145
	C2	cold	5	80	140

Table 2: Energy diagnosis of example: MER values

Scenario	Hot Utility (kW)	Cold Utility (kW)	Tpinch (°C)
01	0	190	-
02	177.5	75	65
03	275	62.5	85

In the solution, eight units are predicted for the flexible network configuration (obtained in less than 4s with *ILOG CPLEX 12.6.2*). The MER is reached for each scenario. Figure 6 shows the HEN compatible with all the scenarios established by the MILP model MS. The values of heat load and temperatures are for Scenario 2 in particular on figure 6.

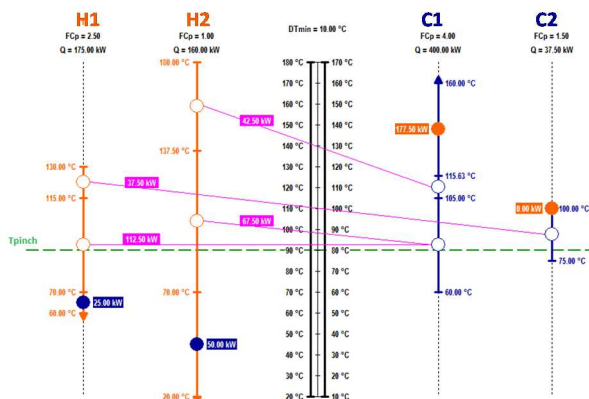


Figure 6: Scenario 2 HEN

The solution presented has reached the MER consumption but includes 8 heat exchangers as the energy efficiency was prioritized in the model. The objective function can be weighted in order to prioritize solutions that minimize the number of heat exchangers and thus investment costs.

6 CONCLUSION

The proposed procedure provides a systematic approach for synthesizing flexible heat exchanger networks for multi-scenario operation. This procedure involves the solution of a first MILP model which determines the feasible matches by assuming that the location of heat exchangers are the same for all scenarios. Area of each exchanger is then calculated and a NLP model determines the control variables of by-passes.

ACKNOWLEDGMENTS

This work is part of the *RREFlex* project whose general goal is the development of software dedicated to the synthesis of flexible Heat Exchanger Networks. The LGC greatly acknowledge TOTAL and ADEME (French Environment and Energy Management Agency) for the financial support of this project.

REFERENCES

- Furman, K.C., Sahinidis, N.V., 2002. A Critical Review and Annotated Bibliography for Heat Exchanger Network Synthesis in the 20th Century. *Ind. Eng. Chem. Res.* 41, 2335–2370. <https://doi.org/10.1021/ie010389e>
- Innovation: Building the plants of the future [WWW Document], 2017. URL <http://www.fivesgroup.com/fr/aucoeur-de-fives/innovation/innovation-construire-lusine-du-futur.html> (accessed 1.15.18).
- Kemp, I.C., 2007. Pinch analysis and process integration: a user guide on process integration for the efficient use of energy, 2nd ed. ed. Butterworth-Heinemann, Amsterdam ; Boston.
- Linnhoff, B., Hindmarsh, E., 1983. The pinch design method for heat exchanger networks. *Chem. Eng. Sci.* 38, 745–763. [https://doi.org/10.1016/0009-2509\(83\)80185-7](https://doi.org/10.1016/0009-2509(83)80185-7)
- Payet, L., Thery Hétreux, R., Hétreux, G., Bourgeois, F., Floquet, P., 2017. Flexibility Assessment of Heat Exchanger Networks: From a Thorough Data Extraction to Robustness Evaluation. *Chem. Eng. Res. Des.* <https://doi.org/10.1016/j.chemd.2017.11.036>
- Pereira, P.M., Fernandes, M.C., Matos, H.A., Nunes, C.P., 2017. FI2EPI: A heat management tool for process integration. *Appl. Therm. Eng.* 114, 523–536. <https://doi.org/10.1016/j.applthermaleng.2016.11.131>
- Silva Gercio P., Miranda Camila B., Carvalho Esdras P., Ravagnani Mauro A. S. S., 2017. A simultaneous approach for the synthesis of multiperiod heat exchanger network using particle swarm optimization. *Can. J. Chem. Eng.* 0. <https://doi.org/10.1002/cjce.23065>
- Smith, R., Jobson, M., Chen, L., 2010. Recent development in the retrofit of heat exchanger networks. *Appl. Therm. Eng., Selected Papers from the 12th Conference on Process Integration, Modelling and Optimisation for Energy Saving and Pollution Reduction* 30, 2281–2289. <https://doi.org/10.1016/j.applthermaleng.2010.06.006>
- Sreepathi, B.K., Rangaiah, G.P., 2014. Review of Heat Exchanger Network Retrofitting Methodologies and Their Applications. *Ind. Eng. Chem. Res.* 53, 11205–11220. <https://doi.org/10.1021/ie403075c>

Safety Analysis of Pilot-System Interaction

Vatsal PANT, Jean-Charles CHAUDEMAR

ISAE-SUPAERO
Toulouse - France

vatsal.pant@isae.fr, jean-charles.chaudemar@isae.fr

Hamid DEMMOU

LAAS-CNRS, Université de Toulouse, CNRS, UPS
Toulouse - France

hamid.demmou@laas.fr

ABSTRACT: *This paper presents on-going work in the scope of development of a new methodology of safety analysis taking into account pilot's behavior. The initial part of the research deals with understanding pilot error and to model his behavior against required tasks. The paper presents the methodology and the tool developed for this purpose.*

KEYWORDS: *STAMP, STPA, MBSE, MBSA, Process Model, Flight Control, Safety*

1 INTRODUCTION

Flight travel has become safer and safer over the years thanks to the various state-of-the-art technologies used in design and development of the aircraft (Clinton V. Oster Jr. 2013). But, the human error as cause of accident still remains at the same ratio. It was shown in a research by Wiegmann that the cause of human error due to bad decision making amounts to around 25 percent (Wiegmann & Shappell 2011). Researchers are trying to study what more (or in some case - less) information should be given to the pilot in order to improve his decision making and reduce any chances of error. Our research has the same objective, i.e, to provide the pilot with information about what might go wrong based on the current situation he is in. For example, based on the descend trend at a particular instant, an automated analysis is performed and then the pilot is informed of the possible hazardous situation he might enter into. In this process, the time is of the utmost importance in order to perform a possible corrective maneuver. One of our objectives in this study is to develop a safety analysis method that takes into account pilot's task behavior. To do this a hazard analysis technique developed in MIT, USA called STAMP-STPA has been used (Leveson 2011).

The paper presents the above mentioned hazard analysis process in the second section. Third section explains the methodology used to meet that objective. Fourth sections details about modeling the pilot task behavior, moving on to the fifth section where the use of ADS-B data in our project has been described. The following section introduces the tool developed for carrying out the aforementioned analysis. Some results and conclusions are presented at the end.

2 STAMP and STPA

This section defines the STAMP process model, the hazard analysis process based on STAMP called STPA, and a brief introduction to use STPA.

System Theoretic Accident Modeling Processes (STAMP) is a methodology used to model a system and/or a process. It is based on the concepts of Systems Engineering and Control Theory. The modeling is done on the base of functional structure rather than the physical component structure. The entire process is classified in different level of complexities starting from the most complex at the top and then going down to the least complex at the bottom. This complexity corresponds to the functions performed at each level. Each level has a controller and a process as shown in Figure 1. Each controller gives some control actions as input to the process and receives feedback from the process. (Thomas 2017)

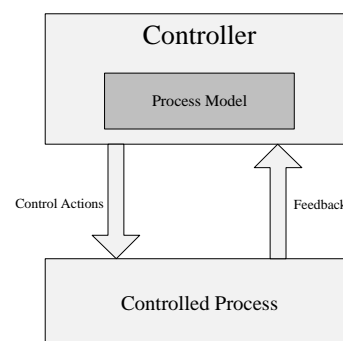


Figure 1 – Control based STAMP model

According to this modeling, an associated definition of failure is proposed, thus failure is defined as loss

of control since the focus is on early functional architecture instead of physical one. The basic ideology is that there have been incidents in the past where neither components failed nor humans made error yet accidents occurred (Nancy Leveson & Marais 2003). Another useful concept in safety is the hazardous condition depicting the result of a failure in the STAMP framework.

Moreover, STAMP offers a collection of processes, including a System Theoretic Process Analysis (STPA), which is of interest for our study. STPA is a top-down method that makes use of a control structure. It is performed in following two steps (Leveson & Thomas 2013) :

- Define Accidents and Hazards (associated with these accidents) and obtain loss of control that can lead to a hazardous state
- Identify how these unsafe control actions could occur in the control structure.

Unsafe Control Actions (UCA) are the commands given by the controller to the process that can result in a hazardous situation. *UCA* is defined as:

$$\textit{Context} + \textit{ControlAction} + \textit{Hazard}$$

Here, *Context* is defined as the scenario or condition or constraint that helps us in assigning safe or unsafe nature to a control action (Thomas 2013). For example, one can imagine that the pilot lowers the landing gear, this is a control action but it becomes an unsafe control action if it is lowered at a high speed and at a high altitude, as it increases the drag on the aircraft and it becomes even more critical if there is a critical situation like fuel starvation. This might result in stall. Hence it is very important to carefully consider the context.

According to STPA there are four types of possible unsafe control actions:

- Providing a control action will cause a hazard. In this case, the control action may be tagged as "inappropriate".
- Not providing a control action will cause a hazard. This is the absence of a required control action.
- Providing it too early/too late will cause a hazard. The focus is on the time at which a control action is provided.
- Providing it for too long or stopping it too soon will cause a hazard (this is more applicable when the input is continuous). The focus is on the duration of the control action.

Given these STAMP and STPA concepts, our study relates to an original approach enabling to extend STPA analysis for on-line safety analysis.

3 PROPOSED METHODOLOGY

This section defines the approach used in order to integrate the concept of unsafe control actions with safety analysis.

As aforementioned, the *UCA* concept becomes a central component of our methodology, which requires to formalize it. For that, a formal tabular specification (Russo, Miller, Nuseibeh & Kramer 2000) is used to define the unsafe control action uca_{ij} , such as for any i, j :

$$uca_{ij} = \langle C_i, UC_j, u_{ij} \rangle$$

where:

- C represents the *Context*, i.e. the current state variables characterizing the system behavior with its given environment; then, C_i is the i th element of this list of context.
- UC stands for the type of the control action as defined in STAMP; we assume there are 6 types as described below.
- u corresponds to the textual description of the hazard of the concerned unsafe control action; i, j is for the position of the cell in the table.

An illustration of STPA table structure is shown in table 1, where $n > m$.

	UC ₁	UC ₂	UC ₃	UC ₄
C ₁	u_{11}	u_{12}	u_{13}	u_{14}
.
C _n	u_{n1}	u_{n2}	u_{n3}	u_{n4}
.
C _m	u_{m1}	u_{m2}	u_{m3}	u_{m4}

Table 1 – *Unsafe Control Actions* table

On the one hand, this UCA table structure implicitly reveals a time characteristic for a control action (noted ca , later on). Let be t_0 the starting time whose discrete value belongs to $ok, +, -, \infty$, ok is the appropriate time, $+$ too late time, $-$ too early, and ∞ for absence of ca . Similarly, td the duration whose value belongs to $ok, +, -, \infty$, ok is the appropriate duration, $+$ too long, and $-$ too short.

On the other hand, the types of UCA 3 and 4 are split hereafter, since their description states two distinct features. Therefore, the type rearrangement is given as follows:

- UC_1 : for providing ca which is inappropriate, noted $uca(t0 = ok, td = ok)$.
- UC_2 : for absence of ca , noted $ca(t0 = \infty, td = ok)$.
- UC_3 : for providing ca too early, noted $ca(t0 = -, td = ok)$.
- UC_4 : for providing ca too late, noted $ca(t0 = +, td = ok)$.
- UC_5 : for stopping ca too soon, noted $ca(t0 = ok, td = -)$.
- UC_6 : for a long duration of ca , noted $ca(t0 = ok, td = +)$.

Contrary to the UCA types, the *Context* relies on actual data extracting from aircraft actual behavior. To be relevant, these data are given by actual records of several flights, in approach phase for our study case. The process using this data base is described hereafter.

Figure 2 explains the proecess of our study.

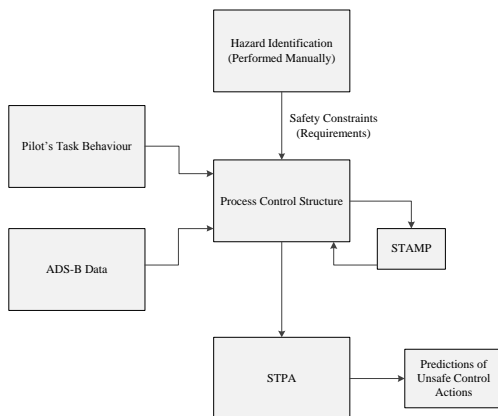


Figure 2 – Methodology

Our proposed methodology deals with studying a very critical phase of a flight, e.g. approach phase. This methodology consists of several steps depicted as follows:

Step1 : To build an initial STPA table in order to obtain control actions.

Step2 : To collect data in order to understand how pilot actually perform successful approaches. To do this, ADS-B was used.

Step3 : To develop and use a tool that takes current position of the aircraft and improves STPA analysis in order to list unsafe control actions that a pilot might give.

As far as the prediction of UCA is concerned, STPA table and the ADS-B data base are necessary for the tool to be developed, however the performance of the prediction would actually be improved if we are able to add a model of pilot tasks, which reduces the number of possible irrelevant unsafe control actions. For multiple failure, we assume for instance that an UCA_4 can not be immediately followed by an UCA_3 , thus, the time analysis is taken into account (Bruggemann & Campbell 2017).

The next section shows an ideal pilot task model that we would like to generate automatically from the STPA table, ADS-B and official procedures.

4 MODELLING PILOT TASK BEHAVIOUR

The first part of the research deals with modeling the behavior of pilot in response to tasks that he/she has to perform. To be able to do that the available manuscripts that aid pilot in training and subsequently performing certain tasks, e.g., FCTM, FCOM, and SOPs are used. (To be noted that these are not the official documents from aircraft manufacturers but the ones available online. ref: www.airbusdriver.net) Since our study is based on one of the most critical phase of flight, i.e., approach and landing, we will, for the moment, look only at those actions that will be performed during this phase. Figure 3 gives an idea of the task performed during approach phase.

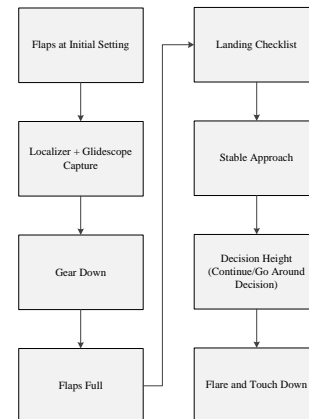


Figure 3 – Approach Procedure

These tasks are to be performed by the pilots at certain altitude, speed and at required distance from the Touch Down Zone. It can be seen that these tasks correspond to some actions that the pilots have to perform. Each time they perform these tasks they need to make certain decisions based on current state of the aircraft, environmental conditions and other external factors (like Air Traffic Controllers instruc-

tions). We tried to study the hazardous conditions that might exist and based on it we constructed a flowchart (Figure 4). This flowchart enables us to improve STPA analysis by providing important control actions. It also gives some idea on some decisions that the pilots make based on their mental model (experience, fatigue, stress level, etc.). These are different from the actions he performs based on the prescribed procedures. We call these two type of decisions as *procedure based decisions* (represented in Figure 4 as white diamonds) and *human factors based decisions* (represented as black diamonds)

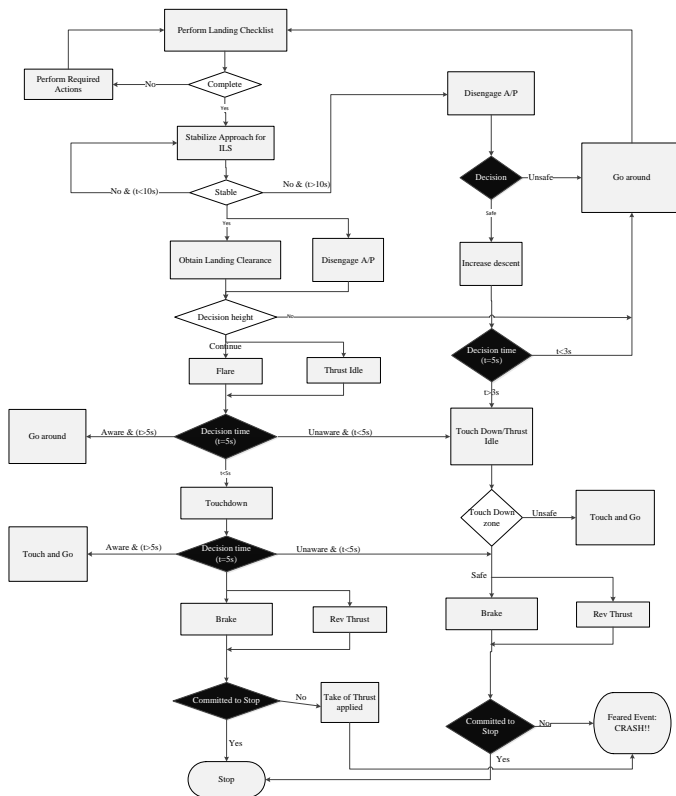


Figure 4 – Pilot Task Flowchart

Based on this model we obtain control actions that we use next. The next section gets an insight into the utilization of ADS-B for actual pilot behavior.

5 ADS-B DATA

ADS-B stands for Automated Dependent Surveillance - Broadcast, it is a technology that helps an aircraft obtain its position based on satellite navigation and broadcasts it. This allows ground based receivers to track an aircraft. This is very useful for ATC and other aircraft to know exact position of other aircraft in order to maintain safe separation. The ADS-B data packet contain information about the track, ground speed, call-sign (it is a type of aircraft identifier) and position of the aircraft. ADS-B data help on-line flight tracking platforms to track air traffic in

real time.

We have used the data from various approach patterns of Airbus A320 aircraft in Toulouse, Blagnac airport (ICAO: LFBO). We extract data that enables us to understand how different pilots react in certain situations while performing the required task. This was done by studying how the speed and altitude changes with time (Figure 5).

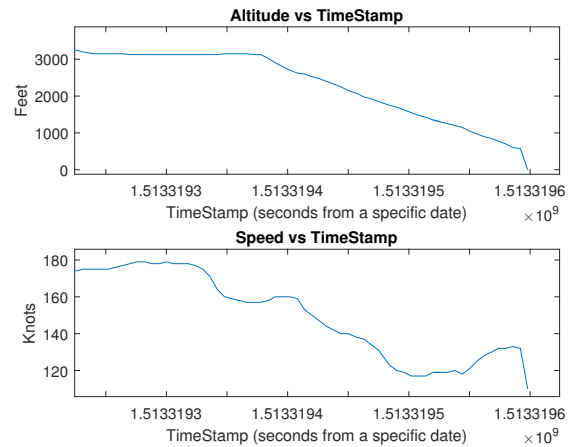


Figure 5 – Altitude and Speed vs Time

Studying these graphs over a wide range of data will provide us an understanding of the normal trends of altitude and speed change for successful landings that might help us define our safety margin.

Another interesting plot that was extracted was Altitude vs Speed. This was particularly interesting as this eliminated time variable. Since not all pilot perform similar actions at same time, it is interesting to find a way to find a common performance graph.

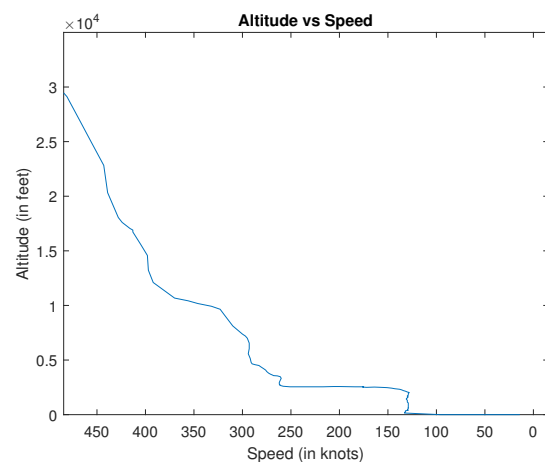


Figure 6 – Altitude vs Speed

Figure 6 shows an example Altitude vs Speed graph. It is normal to have an approach curve like this, it should not be confused with unstable flight as the

curve doesn't represent the actual position of the aircraft. It shows the trends in altitude and speed. Some sudden spikes and drops in the curves arise mainly due to missing data or some changes that occur due to operational reasons like extension of landing gears, etc.

6 INTRODUCTION TO THE TOOL

The authors decided to develop a tool that can automatically perform the analysis. MATLAB based GUIDE (Graphical User Interface Development Environment) is used to build this tool. The tool takes ADS-B data as the input and predicts unsafe control actions. It also plots an altitude vs speed diagram of the approach in order to compare it to safe approach patterns. Figure 7 shows the layout of the tool. It has a mode selection switch that decides whether data will be taken automatically or must be entered manually. There is an input box that allows data to be entered manually. The output box presents results from the analysis. And the nominal performance curve plots the fitted curve based on the data points that are input.

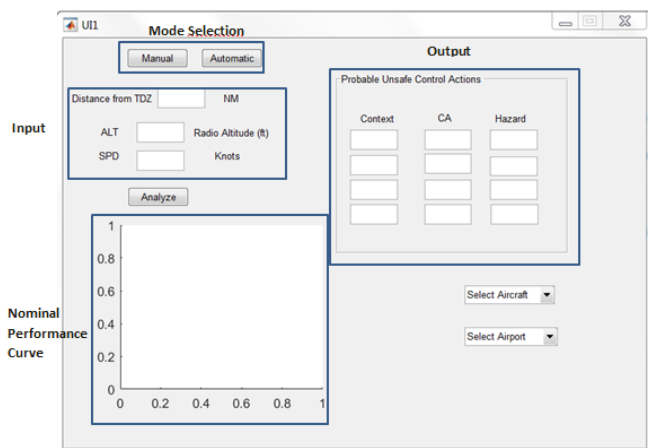


Figure 7 – Screenshot of the Tool

The tool carries out following tasks.

- Taking position, speed and heading values from ADS-B receiver and importing it in a .csv file (done in "Automatic" mode). This can also be input manually by selecting "Manual" from the Mode Selection Switch.
- Plot a nominal approach Altitude vs. Speed curve for a particular aircraft type and a particular airport. It is done in the "Nominal Performance Curve".
- Based on the values of Altitude and Speed assume which step of procedure the pilot will perform next. Compare the current action to actions in

future and list down possible unsafe control actions in the output box.

In order to performance the analysis following steps are performed:

- Define the nominal approach curve based on the values obtained from ADS-B data.
- Select a point for which we want to study the scenario, or perform our analysis. This is crucial as the aircraft is constantly moving instead of performing the analysis every second it might be interesting to do it over a time zone for which the conditions remain same, and change it over to the next zone as soon as it passes to the next scenario.
- We use data from the aircraft to obtain a safety margin. This is obtained by plotting all the safe landings over a period of time and then based on extreme points making a safe envelope.
- Calculate the distance of our selected point from the safety margin.
- Based on this distance define the context for unsafe control action (e.g., Low/high speed (if the speed difference is too much) and same for altitude). It is done by obtaining the factor (altitude or speed) that is farthest from the safe margin and use that as "critical" because it will be difficult to achieve.

- Based on the context and current position populate the list of unsafe control action on the tool.

7 RESULTS

Once all the Unsafe Control Actions were defined, they were entered in the tool for analysis of an aircraft in real time. The tool provides us with results in the output box in $uca_{ij} = \langle C_i, UC_j, u_{ij} \rangle$ format. For example in Figure 8, based on current altitude and speed performing the next control action will result in stall due to increase drag. Here the context is low speed and the corresponding hazard is Stall.

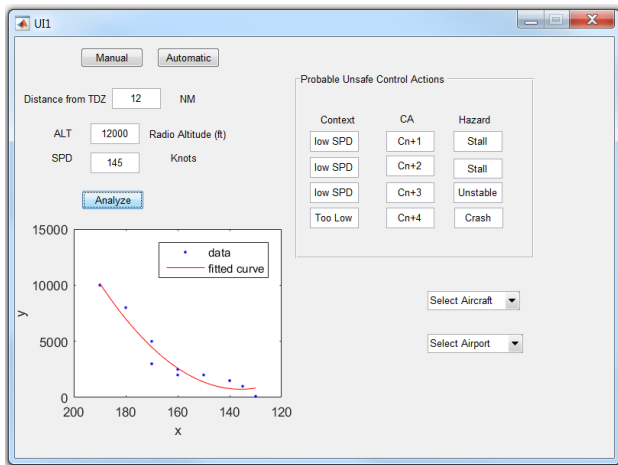


Figure 8 – An output on the Tool

It can be seen that the safety margin is obtained as an envelope and is presented in dotted lines in Figure 9.

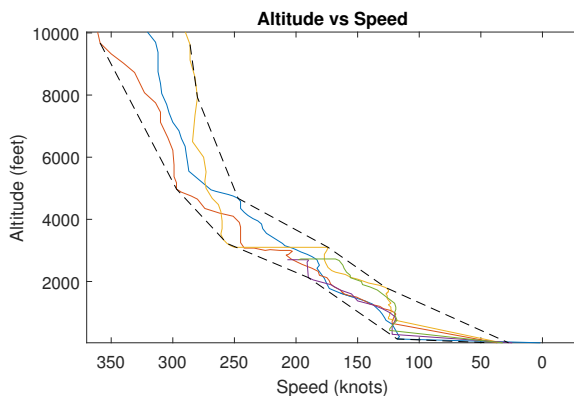


Figure 9 – Safety Margin

8 CONCLUSION AND FUTURE SCOPE

This tool was necessary to better visualize the study for ourselves and to be able to distribute as a stand-alone application to Engineers and pilots to collect data and have their feedback. As mentioned the tool is in the preliminary stage although it can perform some analysis at this level too. A mid-term progress would be to perform some real time tests on simulator to evaluate the performance of the tool and to improve it. At present a safety margin has been obtained and the tool can base its analysis on this safety margin. However, as part of current work, machine learning is being introduced in order to have a safety envelope that is more accurate. Machine Learning/Neural Networks can also be implemented to automatically select a context based on the position of the aircraft, pilot's previous actions, and the environment in real time (wind, etc..). This tool will also be proposed to a discussion with pilot's and have their feedback. This is also very important in order to validate the tool.

ACKNOWLEDGMENTS

The ongoing research is being funded by Dassault Aviation. The project is under the ISAE-SUPAERO and Dassault Aviation chaire.

REFERENCES

References

- Bruggemann, T. S. & Campbell, D. (2017). Analysing the reliability of multi UAV operations, *17th Australian International Aerospace Congress: AIAC 2017*, Barton, A.C.T., pp. 406–417.
- Clinton V. Oster Jr., John S. Strong, C. K. Z. (2013). Analyzing aviation safety: Problems, challenges, opportunities, *Research in Transportation Economics* (43): 148–164.
- Leveson, N. G. (2011). *Engineering a Safer World: Systems Thinking Applied to Safety*, The MIT Press.
- Leveson, N. & Thomas, J. (2013). Stpa primer, <http://sunnyday.mit.edu/STPA-Primer-v0.pdf>.
- Nancy Leveson, Mirna Daouk, N. D. & Marais, K. (2003). Applying stamp in accident analysis, *Technical report*, 2nd Workshop on the Investigation and Reporting of Accidents.
- Russo, A., Miller, R., Nuseibeh, B. & Kramer, J. (2000). An Abductive Approach for Handling Inconsistencies in SCR Specifications, *(ICSE2000) International Workshop on Intelligent Software Engineering*, Limerick, Ireland.
- Thomas, J. (2013). *Extending And Automating A Systems-Theoretic Hazard Analysis For Requirements Generation And Analysis*, PhD thesis, Massachusetts Institute of Technology.
- Thomas, J. (2017). Stamp/stpa beginner introduction, *STAMP Workshop 2017*.
- Wiegmann, D. & Shappell, S. (2011). A human error analysis of commercial aviation accidents using the human factors analysis and classification system (hfacs), *Technical report*, University of Illinois at Urbana-Champaign, Institute of Aviation and FAA Civil Aeromedical Institute.

SCHEDULING IN ADDITIVE MANUFACTURING

Anne-Lise ANTOMARCHI, Séverine DURIEUX

Université Clermont Auvergne, CNRS, SIGMA
Clermont, Institut Pascal, F-63000 Clermont–Ferrand,
France
anne-lise.antomarchi@sigma-clermont.fr
severine.durieux@sigma-clermont.fr

Emmanuel DUC

Université Clermont Auvergne, CNRS, SIGMA
Clermont, Institut Pascal, F-63000 Clermont–Ferrand,
France
emmanuel.duc@sigma-clermont.fr

ABSTRACT: *The specificities of additive manufacturing change the methods usually used in production. At the operational level, additive manufacturing changes the scheduling methods. The goal of this article is to propose a method to distribute the different parts on the plates, to allocate production orders to a machine and to sequence them in each machine. To achieve this aim a double problem must be solve: a bin packing problem and a scheduling problem. The proposed approach is to solve the both problems together using the cross entropy. The method developed is based on a construction heuristic, the cross entropy method and the Analytic Hierarchy Process method for the multicriteria evaluation.*

KEYWORDS: *Scheduling, Bin Packing, Cross Entropy, Additive Manufacturing, AHP*

1 INTRODUCTION

Additive manufacturing (AM) is a field on the rise. This new technology can represent an industrial revolution. So far, AM was used for prototyping. Now, AM is used for spare parts and industrials wonder on the interest of AM for complex and expensive parts in a more standardized production.

This article deals with the optimization of the scheduling of AM process. Particularly, the goal is to propose a method for the distribution of different parts on machine and for the allocation and the sequencing of production orders on machines.

A state of the art is presented on AM and associated specificities, on scheduling in AM, on bin packing problem and on scheduling problem. The treated problem is positioned on these different fields. After having formalized the bin packing and the scheduling problems, the method is presented and developed step by step: the construction heuristic, the cross entropy and the multicriteria evaluation realized using the AHP method and a simulation model.

2 ADDITIVE MANUFACTURING

AM is a process of joining materials which allows manufacturing parts from data of a 3D model, layer upon layer on a plate. Usually, AM is considered in opposition with the subtractive process [1] [2].

Gao [1] classifies the various processes according to their technologies, the materials and their power sources. He puts forward their strengths and their downsides.

In this research work, we focus on the Selective Laser Melting (SLM) technology using metallic powders.

2.1 AM process

Croft [3] identifies the Value Chain of production (Figure 1).

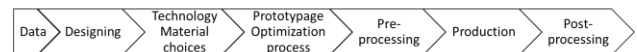


Figure 1. Value chain of AM production

The value chain of AM is simplified [4]. However, each part has to be designed or redesigned by taking into account requirements of the process of AM. Besides, a part made by the process SLM requires necessarily post-processing (surface treatment, milling, etc.)

Further to different interviews with industrials, we modelled the complete process of manufacturing on the figures 2 thanks to the IDEF3 formalism.

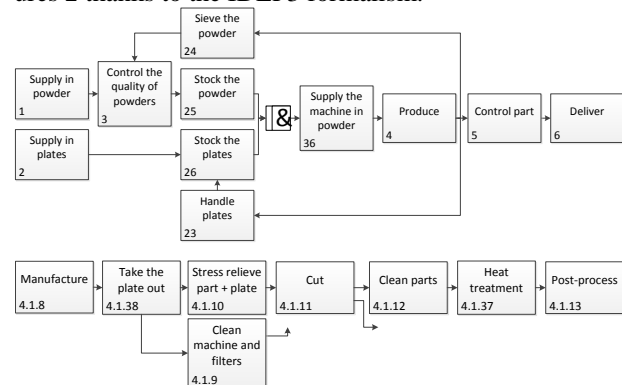


Figure 2. AM Process modelling

2.2 Cost study

The attraction for the AM is due to the decrease of numerous costs ([5], [6], [7]) while increase the customization of the products, by simplifying the process of production and by making more complex parts.

Nevertheless, the costs of SLM machines and metallic powders remain very high. Thus it is advisable to have a decision-making tool to make a go/nogo on AM in comparison to a subtractive manufacturing [4]. Lindemann [5] proposes a method to access to the costs throughout the product life cycle.

Mellor [8] presents the principal costs of AM (Figure 3). An efficient scheduling allows working on factors in grey.

Operation times	Machine costs	Labour costs	Materials costs
Number per platform	Machine and ancillary equipment	Machine operator cost per hour	Material per part
Platform build time	Depreciation cost per year	Set-up time to control machine	Support material per part
Hours per year in operation	Maintenance per year	Post-processing time per build	Material cost per kg

Figure 3. The main cost factors used in AM costing [8]

3 AM PROBLEMATICS DEALING WITH OPERATIONAL DECISION LEVEL

3.1 Specificities

Many researches try to optimize the print feed rate and to adjust new properties [9]. However, the AM times are still long. When the post processes are added, the AM becomes the bottleneck of the global workshop. Furthermore, these times are incompatible with the short deadlines waited by the customers.

Added to the manufacturing time, the powder change and the cleaning of the machine are very long: from several hours to several days depending on the machines. Moreover, the human intervention is useless during the manufacturing time but crucial between two productions. However, for economic reasons an operator cannot be attached at each machine. Thus, the operator is a shared resource which can generate blocking times on the AM machines.

Finally, the composition of the plate (which part and the quantity) has a direct impact on the lasing time but also on all the rest of the post-process.

The way to start production of the plate of parts on the various AM machines will thus have a significant impact on the efficiency of the global system and on the respect to delivery dates.

3.2 Scheduling problematics

From the specificities of AM and discussions with industrialists, allocation and scheduling steps are identified (Figure 4).

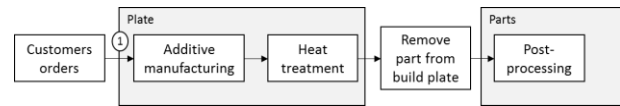


Figure 4. Allocation and scheduling steps in AM

The planner receives several customer orders (CO). Every CO can contain several part references, each in different quantities.

Then, he has to define (Figure 5):

- Allocation and quantity of parts to plates;
- Allocation of plates to machines;
- Order to launch the different plates.

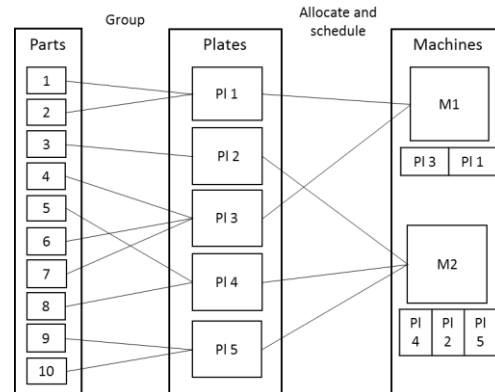


Figure 5. Grouping, allocation and scheduling

The objective is to respect the customer due date, to decrease production costs and to maximize the utilization rates of AM machines.

Others constraints concern the plates surface, the powder as well as the dimensional constraints of the AM machines.

There are few articles about the scheduling in AM and the allocation of parts on plates. We found only the works of Li and al [10]. The particularity of their work is to allocate parts on machine and then to group parts on jobs without taking into account the due dates. However, we think to solve the problem in its entirety and not in a sequential way will allow obtaining more successful solutions.

Thus the objective of these works is to solve both problems of allocation and sequencing in the same approach.

4 BIN PACKING

The grouping of parts on plates can correspond to a bin packing problem: the packing of a set of items into a minimum number of bins.

4.1 Typology of cutting and packing problems

The typology of cutting and packing problems is proposed in [11]. It is composed by four parameters: the problem dimension, the allocation type, the nature of containers and the nature of components.

Wascher and al [12] modified the typology to integrate the recent cutting and packing problems. The 2D bin packing problem (2BP) has two specificities: the orientation (R for the non-oriented case and O for the oriented case) and the guillotine constraint (G for the presence of the guillotine constraint, F otherwise).

Here, the bin packing problem is a 2-dimensional problem with all objects and a selection of bins. The bins have different sizes and the objects have different shapes. Parts are non-oriented and the guillotine constraint is not used. Thus, this article deals with a 2BP|R|F problem.

4.2 Methods

Various metaheuristics are used to solve the bin packing problems. Jakobs [13] uses a genetic algorithm for the packing of polygons. Faroe [14] develops a guided local search for 3D bin packing problem. Dowsland [15] proposes a simulated annealing techniques for packing problems. Hong [16] uses a hybrid heuristic algorithm for the 2D variable-sized bin packing problem and Cui and al [17] implements a sequential heuristic for the 2D bin packing problem.

5 SCHEDULING

In manufacturing process, a scheduling is the organization of the different activities on the time and the allocation of machines and human resources to these activities. This allocation depends on the duration of the activity, the predecessor activity and the resource availability.

5.1 Definitions

A set of tasks or activities is characterized by a duration, an execution mode and a set of machines which can execute it.

A set of dedicated resources or not. Their quantity can be limited or not.

The goal of a scheduling is to determine start times, duration or finish times of the tasks and identity and sequencing of allocated resources.

A scheduling must respect some time constraints: availability date, deadlines, simple or general precedence and pre-emptive mode or not.

In non-pre-emptive mode, a scheduling attributes, at each task, an execution interval and a machine such as:

- Time constraints are satisfied
- At each moment, a machine execute at the most one task
- At each moment, a task is executed by at the most one machine

In pre-emptive mode, a scheduling attributes at each task: several execution intervals and a machine for each interval such as the three constraints for the non-pre-emptive mode are satisfied.

A criteria joins a cost at each scheduling to determine the scheduling with the minimal cost. A criteria is regular if it is not decreased about the end of task date.

5.2 Scheduling problems

Several kinds of scheduling problems exist [18] (Figure 6).

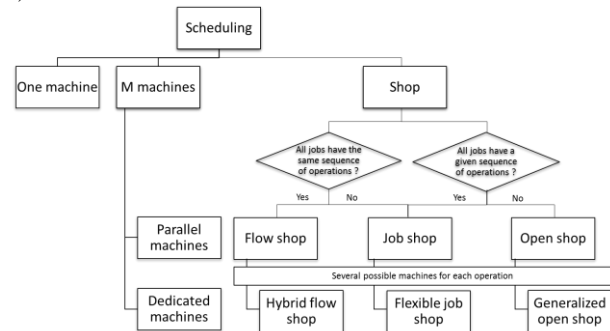


Figure 6. Scheduling problems

In a one machine problem, a permutation of jobs must be determined. Some rules are used to solve these problems: Smith's rule, Jackson's rule and Moore-Hodgson's algorithm.

In a m machines problem, besides the permutation of jobs, the jobs must be assign to a machine (identical or not).

Shop scheduling problems can be separated in principal categories: flow shop, job shop and open shop [18].

5.3 Methods

For one machine problems, the optimization methods used are: greedy heuristics, neighborhood based metaheuristics and genetic algorithms. The genetic algorithms are developed with crossover: in one or two points and mutations: permutation of 2 jobs, move of one job or a sequence of jobs and moves of local search [19]. For parallel machines problems, some optimization methods used for one machine problems can be used too. Moves of neighborhood can be added: assign a job to another machine, move 2 jobs on two machines [20]. For shop scheduling problems, optimization methods used are: tabu search, simulated annealing, evolutionary algorithms and others like dynamic programming or constraints programming.

5.4 Scheduling in AM

The scheduling realized in this article, is a parallel machine scheduling with eligibility constraint.

Among the classical constraints in scheduling, this problem must respect:

- The eligibility constraint: a plate is not compatible with all machines,
- No precedence: there is only one operation,
- No-pre-emptive mode: it is impossible to interrupt an operation on a machine,
- All the machines are available at $t=0$,
- Machines capacity: one plate simultaneously,
- Setup constraints: between two plates, there is a cleaning time or a time to change the powder,
- Auxiliary resources: operators.

6 MATHEMATICAL FORMALISATION

The aim is to obtain a distribution of different parts on plates and a sequenced list of production orders (PO) assigned to each machine. The study is realized in a given workshop where layout and management rules are determined.

The operators' planning is not taking into account.

6.1 Notations

i	Index of the order
j	Index of the part reference
l	Index of the PO
k	Index of the plate reference
m	Index of the AM machines
nb_order	Number of orders
nb_ref	Number of references
nb_pror	Number of PO
nb_pla	Number of plates
nb_ma	Number of machines

6.2 Input data

To schedule activities, four types of input data are necessary:

- Customer orders list with due date, and the list of part references and the quantity for each;
- The list and characteristics of AM machines;
- The list of plates with their characteristics;
- For each part reference, the process planning, the geometric specifications and the type of powder.

6.2.1 Customer orders

d_i	Due date of the order i
$nb_row_ord_i$	Number of rows in the order i

A row c of the order i is characterized by the 2-tuple $\{ref_{c,i}^{ord}; qty_{c,i}^{ord}\}$ with $ref_{c,i}^{ord}$ the part reference of the row c of the order i and $qty_{c,i}^{ord}$ the quantity associated.

6.2.2 Machines

$h_m^{machine}$	Maximal height accepted by the machine m
OC_m	Operation cost of the machine m
$speed_m^{vol}$	Production speed laser (cm^3/h)

6.2.3 Plates

w_k^{pla}	Width of plate k
l_k^{pla}	Length of plate k
$plate_{k,m} = 1$	if the plate k is compatible with machine m

6.2.4 Parts

p_j	Powder reference of the part j
qp_j	Powder quantity for the part j
w_j^{part}	Width of part j

l_j^{part}	Length of part j
h_j^{part}	Height of part j
$surf_j$	Surface of part j
v_j	Volume of the part j $v_j = s_j * h_j^{part}$
h_j^{layer}	The thickness of the layer of the part j
ht_time_j	Heat treatment time of the part j

6.3 Bin packing problem

The bin packing formalization is inspired by the formalization proposed by Li and Zhang [21].

6.3.1 Notations

n	Index of part
nb_total_parts	Total number of parts
ref_n^{part}	Part reference of the part n
(x_n, y_n)	Coordinates of the bottom left corner of part n

6.3.2 Decision variables

$X_{n,l}$	$= 1$ if the part n is assigned to PO l $= 0$ otherwise
Y_l	$= 1$ if the PO l is formed $= 0$ otherwise
$PL_{n,z}$	$= 1$ if the part n is placed to the left of the part z $= 0$ otherwise
$PB_{n,z}$	$= 1$ if the part n is placed below to the part z $= 0$ otherwise
$O_{n,k}$	$= 1$ if the part n is placed in such way that its width is parallel to the width of the plate k used for the PO l $= 0$ otherwise

6.3.3 Constraints

A part cannot be assigned to a PO that is not formed.

$$\forall n \in \{1, \dots, nb_total_parts\},$$

$$\forall l \in \{1, \dots, nb_pror\}, X_{n,l} \leq Y_l \quad (1)$$

Each part is assigned to one and only one PO.

$$\forall n \in \{1, \dots, nb_total_parts\}, \sum_l X_{n,l} = 1 \quad (2)$$

A part cannot be placed outside of the plate.

$$\forall n \in \{1, \dots, nb_total_parts\},$$

$$x_n + w_{ref_n^{part}}^{part} O_{n,k} + l_{ref_n^{part}}^{part} (1 - O_{n,k}) \leq w_k^{pla} \quad (3)$$

$$y_n + l_{ref_n^{part}}^{part} O_{n,k} + w_{ref_n^{part}}^{part} (1 - O_{n,k}) \leq l_k^{pla}$$

The surface of all parts of a PO must not exceed the surface of the assigned plate.

$$\forall l \in \{1, 2, \dots, nb_order\},$$

$$\forall \alpha \in \{1, 2, \dots, nb_row_pror\}$$

$$\sum_{\alpha, l} s_{ref_pror, \alpha, l} \leq l_l^{pla} * w_l^{pla} \quad (4)$$

Two parts cannot overlap in positions if they are placed on the same plate.

$$\forall z, n \in \{1, \dots, nb_total_parts\}, z \neq n, \\ x_z + w_{ref_z^{part}}^{part} O_{z,k} + l_{ref_z^{part}}^{part} (1 - O_{z,k}) \leq x_n \quad (5) \\ y_z + l_{ref_z^{part}}^{part} O_{z,k} + w_{ref_z^{part}}^{part} (1 - O_{z,k}) \leq y_n$$

$$\forall z, n \in \{1, \dots, nb_total_parts\}, z < n, \\ \forall l \in \{1, \dots, nb_pror\}, \\ PL_{z,n} + PB_{z,n} + PL_{n,z} + PB_{n,z} \geq X_{z,l} + X_{n,l} - 1 \quad (6)$$

All the parts of the same PO must use the same powder.

$$\forall l \in \{1, 2, \dots, nb_pror\}, \forall \alpha \in \{1, 2, \dots, nb_row_ord_l - 1\}, \\ p_{ref_{\alpha,l}^{ord}} = p_{ref_{\alpha+1,l}^{ord}} \quad (7)$$

All the orders must be completed.

$$\forall i \in \{1, 2, \dots, nb_order\}, \forall c \in \{1, 2, \dots, nb_row_ord_i\} \\ qty_{c,i}^{ord} = \sum_{\alpha,l} qty_{\alpha,l}^{pror} \cdot Z_{\alpha,l,i,c} \quad (8) \\ Z_{\alpha,l,i,c} = \begin{cases} 1 & \text{si } ref_{\alpha,l}^{pror} = ref_{c,i}^{ord} \\ 0 & \text{otherwise} \end{cases}$$

The decision variables are binary variables.

$$\forall n \in \{1, \dots, nb_total_parts\}, \forall l \in \{1, \dots, nb_OF\}, \\ X_{n,l} \in \{0, 1\} \\ \forall l \in \{1, \dots, nb_pror\}, Y_l \in \{0, 1\} \quad (9) \\ \forall z, n \in \{1, \dots, nb_total_parts\}, PB_{z,n}, PL_{z,n} \in \{0, 1\} \\ \forall n \in \{1, \dots, nb_total_parts\}, O_{n,k} \in \{0, 1\}$$

Coordinates are positive variables.

$$\forall n \in \{1, \dots, nb_total_parts\}, x_n, y_n \geq 0 \quad (10)$$

6.4 Scheduling problem

The scheduling is crucial for the cleaning time at each powder change. The allocation of plates must respect the production volume of AM machines.

6.4.1 Decision variables

The decision variables are the sequenced allocation of the PO to each machine.

ma_l Number of the machine which the PO l is assigned

pla_l Plate reference of the PO l

$nb_row_pror_l$ Rows number of the PO l

A row a of the PO l is characterized by the 3-tuples $\{ref_{a,l}^{pror}; qty_{a,l}^{pror}; order_{a,l}\}$ with $ref_{a,l}^{pror}$ the part reference of the row a of the PO l , $qty_{a,l}^{pror}$ the quantity associated. This quantity of this part reference is produced for the $order_{a,l}$.

From these decision variables, following values are calculated:

v_l Global volume of the PO l
 p_l Powder of the PO l
 $lasing_time_{l,m}$ Lasing time of PO l on machine m
 ht_time_l Heat treatment time of the PO l

$cutting_time_l$ Cutting time of the PO l
 $cooling_time_l$ Cooling time of the PO l
 $cleaning_time_l$ Cleaning time of the PO l
 $recovery_powder_time_l$ Powder recovery time of the PO l

The volume of the PO l is calculated from the equation (11).

$$v_l = \sum_n X_{n,l} * v_{ref_n^{part}} \quad (11)$$

The lasing time of the PO l on the machine m is calculated from the equation (12).

$$lasing_time_{l,m} = \frac{v_l}{speed_m^{vol}} \quad (12)$$

$aff_{l,m} = 1$ if the PO l is allocated to the machine m .

6.4.2 Scheduling Constraints

The scheduling must respect some scheduling and AM constraints.

The height of parts of an OF must not surpass the maximal height accepted by the machine.

$$\forall l \in \{1, 2, \dots, nb_pror\}, \forall \alpha \in \{1, 2, \dots, nb_row_ord_l\} \\ h_{ref_{\alpha,l}^{part}}^{part} \leq h_{ma_l}^{machine} \quad (13)$$

The chosen plate for a PO must be compatible with the chosen machine.

$$\forall l \in \{1, 2, \dots, nb_pror\}, plate_{pla_l, ma_l} = 1 \quad (14)$$

6.5 Objective function

To be efficient, a scheduling must respect some conditions as the due dates given by the customers, the maximization of the utilization rate of the machines and the minimization of the cost. These conditions ensure the economical optimization of the production.

The chosen performance indicators are: the respect of the due date given by customers, the utilization rate of resources and overall production cost of the PO.

6.5.1 Respect of the due dates

The respect of the due date of an order i is calculated from the exit dates of the PO l (ds_l) which contains the parts of the order i . The date when the order i is ended (df_i) is the exit date of the PO l which contains the last part of the order i . The goal is to minimize the sum of all the delays.

$$\forall i \in \{1, \dots, nb_order\}, delay = \sum_l delay_i \quad (15)$$

$$\forall i \in \{1, \dots, nb_order\},$$

$$delay_i = \begin{cases} df_i - d_i & \text{if } df_i - d_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

6.5.2 Utilization rate

The powder changes take a long time and strongly impact the utilization rate. The goal is to maximize the average of all the utilization rates $rate_m$.

$$\forall m \in \{1, \dots, nb_ma\}, utilization_rate = \frac{\sum_m rate_m}{nb_ma} \quad (17)$$

6.5.3 Production cost

The overall production cost of the PO l C_{prod}^{tot} is the sum of the production cost of each PO C_{prod_l} and the setup cost of each machine m C_{setup_m} . It has to be minimized. Moreover, the set-up time needed for machine m (ST_m) and the cost of human work per time unit HC are used.

$$C_{prod}^{tot} = \sum_l C_{prod_l} + \sum_m C_{setup_m} \quad (18)$$

$$C_{prod_l} = aff_{l,m} * lasing_time_{l,m} * OC_m \quad (19)$$

$$C_{setup_m} = ST_m * HC \quad (20)$$

7 PROPOSED APPROACH

The objective of is to optimize conjointly:

- The allocation of the parts to the plates
- The allocation of the plates to the AM machines
- The sequencing of the plates on the corresponding AM machine.

The optimized solution has to respect the set of the constraints expressed above.

The optimization has to be multicriteria.

7.1 Cross Entropy

For the combinatorial problems, various methods can be used. Among them, the simulated annealing, the tabu search, the genetic algorithm, ant colonies and the cross entropy.

The cross entropy (CE) is a recent method created in 1997 by Rubinstein [22]. This optimization method was created to estimate probabilities of rare events in complex stochastic networks. This continue method can be adapted to combinatory problems changing a deterministic model in a stochastic model. The CE is an iterative procedure where each iteration is cutting in two phases: the generation of a random sample of solutions under a probabilistic mechanism and the updating of the probabilistic parameters conserving only the best solutions.

The steps of the CE method are [23]: the initialization of the probability distribution, the draw of a value sample under this distribution, the evaluation of the drawn values, the selection of the bests and the deformation of the probability distribution according to the best values. These steps are repeated until the stop criterion is reached.

The stop criterion can be a maximal number of iterations or a number of iterations where the quantile is the same.

This method demonstrated its efficiency in research works concerning the industrial problems [23]. Furthermore, the CE allows learning about these sets from the best solutions. Therefore we choose this optimization method to resolve our problem.

7.2 Principle of the proposed approach

7.2.1 Solution coding

The solutions are coded by a vector where each part matches a plate on a machine (Figure 7).

$P_{l,m}$ corresponds to the affectation of the part n to the plate k on the machine m .

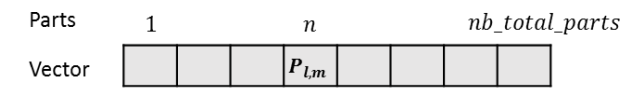


Figure 7. Coding

7.2.2 Method explication

The Figure 8 illustrates the algorithm of the proposed approach.

To begin the algorithm, the probability matrix is initialized. A set of x possible solutions is created by a construction heuristic. These solutions are evaluated by a multicriteria evaluation using the equations 13,15 and 16. The y best solutions are kept among the x solutions evaluated and the y solutions kept in the previous iteration. If the stop criterion is not achieved, the probability matrix is updated and a new population is created.

A probability matrix must be defined. Each case corresponds to the number of times where a part is in a plate allocated to a machine divided by the number of best solutions kept y .

The parameters are the number of solutions created x , the number of best solutions kept y and the number of iterations corresponding to the stop criterion.

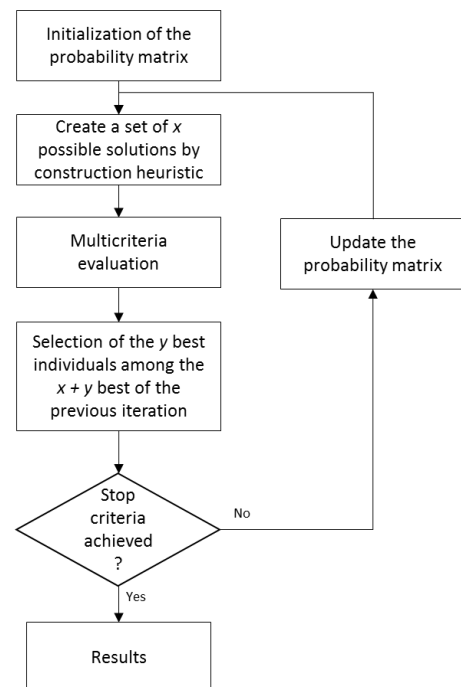


Figure 8. Cross entropy method

7.2.3 Construction heuristic

An allocation of each part to a plate is done according to the probability matrix (Figure 9).

Cui and al [17] implement a sequential heuristic for the 2D bin packing problem. They prove the efficiency of their method. The heuristic is based on the notions of possible positions and envelope. The possible positions correspond to the points where the lower left corner of the parts can be placed. The envelope is the area available to place new parts. The best position is one that minimizes the difference between the envelope surface before and after the placement of the part. If, when a part is placed, one of its limits defines with the limits of the other parts or with the limits of the plate an area too small for a part to be placed, the surface of this area is deduced from the surface of the envelope. When a part is allocated to a plate, the bin packing constraints are checked. If one of them is not respected, the part is allocated to another plate.

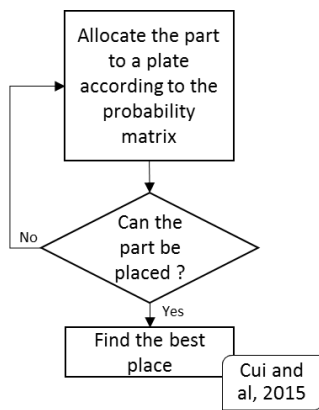


Figure 9. Construction heuristic

7.3 Multicriteria evaluation

7.3.1 Simulation model

The performance evaluation resting on the complete workshop and the production flows being complex, we cannot use analytical methods. Thus, it will be based on a model of flow simulation developed on Arena© software.

The model represents a multi-part, multi-powder, multi-plate and multi-machine shop. It considers the process from the stock of plates and powders to the post-processing and the ended-part.

Several assumptions are made. The workshop is considered without failure. The model is deterministic. Operators are present on the whole opening time of the workshop. Excepting the AM machines, all the operations are managed in FIFO.

The input data are times of each operation, number of resources, the process planning of each part. The input variables are the composition of each PO, and the sequenced list of PO for each machine. The input variables are the result of the cross entropy algorithm, they change for each tested solution.

The output data used to evaluate solutions are:

- The output date of each PO l : ds_l
- The utilization rate of each machine m : $rate_m$.
- The set-up time needed for machine m : ST_m

The global evaluation is realized thanks to the equations (15,17,18).

The simulation model reads the input data and variables in an Excel spreadsheet and, at the end of the simulation, the statistics are collected in another Excel spreadsheet (Figure 10). The tool is coded with VBA code.

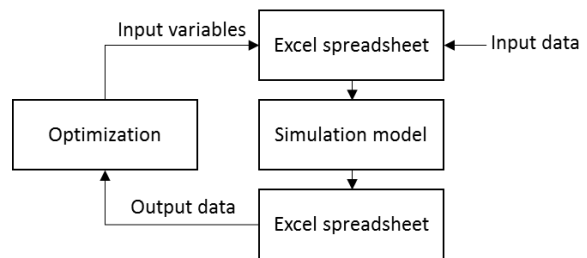


Figure 10. Use of simulation model

7.3.2 Multicriteria evaluation method

For the multicriteria decision making, various methods can be used. Among them, ELECTRE, PROMETHEE [24] and AHP [25].

Analytic Hierarchy Process (AHP) allows modelling the human decision mechanism. AHP is the only method which allows the measurement of the coherence of the decision maker preferences [25]. This method takes into account the dependence and the interdependence of the considered evaluation criteria and the hierarchy between them [26]. Moreover, the AHP has demonstrated robustness across a range of application domains [27].

Thus, the AHP method is chosen.

Dubromelle and al [27] model the problem on a hierarchical structure with 4 levels: the global objective level, the criteria level, the indicators level and the alternatives level. The AHP decision structure for our problem is presented by the figure 11.

The AHP is implemented through two phases: configuration and evaluation [26]. The configuration phase consists in the pair-wise comparison to assess the relative importance of two elements. The analysis of the matrixes obtained allows verifying the coherence. If the result is coherent, the matrixes are converted into priority vectors. The exploitation phase allows to obtain the vector giving the relative importance of the alternatives compared to the global objective.

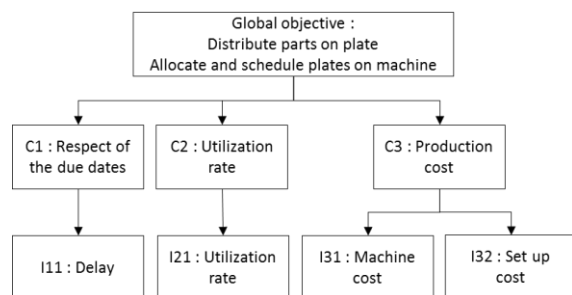


Figure 11. AHP Decision structure

8 CONCLUSION

The specificities of AM impact the scheduling methods. In fact, the composition of the plate and the way to start production of the plates on the various AM machines have a significant impact on the efficiency of the global shop. The objective is to optimize the distribution of different parts on plates, the allocation of PO to a machine and the sequencing of them in each machine. Few articles deals with the problem of the sequencing but none deals with the problem in its entirety. Therefore we proposed a new approach which allows solving both problems together using the cross entropy method. As further work, the method will be test on industrial cases. To improve this method, the planning of operators can be taken into account, the indicator formula can evolve, especially for the production cost and other optimization methods can be tested.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the SIGMA Clermont Foundation.

REFERENCES

- [1] W. Gao *et al.*, “The status, challenges, and future of additive manufacturing in engineering,” *Comput. Des.*, vol. 69, pp. 65–89, 2015.
- [2] ASTM International, *F2792-12a - Standard Terminology for Additive Manufacturing Technologies*. 2013.
- [3] A. Croft, “Advances in production technology,” *Concr.*, vol. 30, no. 3, pp. 39–48, 1996.
- [4] G. White and D. Lynskey, “Economic Analysis of Additive Manufacturing for Final Products: an Industrial Approach,” *Pittsburgh, Univ. Pittsburgh, Swanson Sch. Eng.*, pp. 1–11, 2013.
- [5] C. Lindemann, U. Jahnke, M. Moi, and R. Koch, “Analyzing product lifecycle costs for a better understanding of cost drivers in additive manufacturing,” *Int. Solid Free. Fabr. Symp.*, vol. 23, pp. 177–188, 2012.
- [6] V. Bhasin and M. Bodla, “Impact of 3D Printing on Global Supply Chains by 2020,” p. 82, 2014.
- [7] M. Baumers, P. Dickens, C. Tuck, and R. Hague, “The cost of additive manufacturing: Machine productivity, economies of scale and technology-push,” *Technol. Forecast. Soc. Change*, vol. 102, pp. 193–201, 2016.
- [8] S. Mellor and E. March, “An Implementation Framework for Additive Manufacturing,” *Dr. Philos. Submitt. to Univ. Exet.*, no. March, pp. 1–262, 2014.
- [9] D. Cohen, M. Sargeant, and K. Somers, “3D printing takes shape,” *McKinsey Q.*, pp. 1–6, 2014.
- [10] Q. Li, I. Kucukkoc, and D. Z. Zhang, “Production planning in additive manufacturing and 3D printing,” *Comput. Oper. Res.*, vol. 83, pp. 157–172, 2017.
- [11] U. Dyckhoff and Finke, “Cutting and Packing in Production and Distribution: Typology and Bibliography,” *Springer-Verlag, New York*, 1992.
- [12] G. Wäscher, H. Haußner, and H. Schumann, “An improved typology of cutting and packing problems,” *Eur. J. Oper. Res.*, vol. 183, pp. 1109–1130, 2007.
- [13] S. Jakobs, “On genetic algorithms for the packing of polygons,” vol. 2217, no. 94, 1996.
- [14] O. Faroe, D. Pisinger, and M. Zachariasen, “Guided Local Search for the Three-Dimensional Bin-Packing Problem,” no. April 2014, 2003.
- [15] K. A. Dowsland, “Theory and Methodology Some experiments with simulated annealing techniques for packing problems,” *Eur. J. Oper. Res.*, vol. 68, pp. 389–399, 1993.
- [16] S. Hong, D. Zhang, H. Chuin, X. Zeng, and Y. Si, “Discrete Optimization A hybrid heuristic algorithm for the 2D variable-sized bin packing problem,” *Eur. J. Oper. Res.*, vol. 238, no. 1, pp. 95–103, 2014.
- [17] Y. Cui, Y. Cui, and T. Tang, “Sequential heuristic for the two-dimensional bin-packing problem,” *Eur. J. Oper. Res.*, vol. 240, no. 1, pp. 43–53, 2015.
- [18] H. Abedinnia, C. H. Glock, and M. D. Schneider, “Machine scheduling in production: A content analysis,” *Appl. Math. Model.*, vol. 50, no. March, pp. 279–299, 2017.
- [19] A. Sioud, “Approches hybrides pour la résolution d’un problème d’ordonnement industriel,” 2011.
- [20] A. Bitar, “Ordonnement sur machines parallèles appliqué à la fabrication de semi-conducteurs: ateliers de photolithographie,” 2016.
- [21] X. Li and K. Zhang, “Single batch processing machine scheduling with two-dimensional bin packing constraints,” *Int. J. Prod. Econ.*, vol. 196, no. November 2017, pp. e10–e10, 2018.
- [22] R. Y. Rubinstein, “Optimization of computer simulation models with rare events *,” *Eur. J. Oper. Res.*, no. 96, 1997.
- [23] P. De Boer, D. Kroese, S. Mannor, and R. Rubinstein, “A Tutorial on the Cross-Entropy Method,” pp. 19–67, 2005.
- [24] B. Roy and D. Bouyssou, *Aide multicritère à la décision : méthodes et cas*. 1993.
- [25] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw Hil. New York, 1980.
- [26] P. Pujo and F. Ounnar, “Pull System Control For Job Shop Via A Holonic , Isoarchic & Multicriteria Approach,” *IFAC Proc. Vol.*, vol. 41, no. 2, pp. 15799–15804, 2008.
- [27] Y. Dubromelle, T. Louati, F. Ounnar, and P. Pujo, “AHP / ANP a Decision Making Service in PROSIS Model,” *IFAC Proc. Vol.*, 2009.

Scoring Risk of Investment Projects using Nominal Classification

Ayeley Tchangani

LGP / Université de Toulouse
47 Avenue d'Azeirex 65016 Tarbes
ayeley.tchangani@iut-tarbes.fr

ABSTRACT: *Projecting, to launch a new product by a manufacturing company, to take an engagement to finance an innovation project by a bank or a financial institution, to fund a research project by a private company or a public institution, to engage in the construction of an infrastructure such as a high way, a power plant, by a government, etc. are decisions that convey some risks because of uncertainty inherent to such projects. Taking into account these risks when making decisions is therefore important for any corporate; for this purpose, corporates generally classify each project into predefined risk classes or categories that constitute an ordinal scoring, for their appropriate treatment. On another side, projects will be characterized by many attributes (cost, duration, quality, etc.) which evaluation will account for their classification into the appropriate class. The process of classifying objects characterized by many attributes into predefined classes is known as nominal classification. In this communication, a framework referred to as fuzzy nominal classification will be developed taking into account inherent uncertainty, be it related to attributes evaluation, risk classes characterization, or decision makers' attitude, to be used as a decision support for projects risk assessment*

KEYWORDS: *Project Risk Assessment, Nominal Classification, Bipolar Analysis, Fuzzy Uncertainty, Choquet Integral.*

1 Introduction

As expressed in the preface of (Newton, 2015), "every project involves risks and every project needs to have a management strategy for dealing with threats and opportunities represented by each risk". One of important stage in risk management processes consist in classifying or categorizing projects in order to derive strategic actions to apply to each project. Today, an ever-increasing number of professionals and managers in industry, government, and academia are devoting a larger portion of their time and resources to the task of improving their approach to, and understanding of, risk-based decision making. Integration of risk factor in decision making or risk informed decision making is receiving a great attention by researchers and decision makers in many domains such as engineering (designing technical systems that meet some requirements in terms of safety), finance (set up norms to monitor finance activities in order to avoid companies collapse), environment (develop sustainable agriculture and natural resources extraction), science and medical research (monitoring scientists activity by the society to avoid creating new threats). But to address correctly important risks, decision makers must consider discriminating, categorizing or scoring them; this process consists basically in assigning them to predefined classes, categories, or scores that is nominal classification problem; a sub-field of multi-criteria or multi-attributes (attributes of the object to classify) and multi-objectives (multi-

features classes, categories, or scores to choose) decision making problems; multi-criteria/multi-objectives decision making is now a well established research area with a broad literature; however one can cite (Steuer, 1986) and (Roy and Bouyssou, 1993) as ones of classical references. The purpose of this communication is to present a risk analysis framework in terms of risks identification, assessment, categorization or scoring, and possibly risk management (mitigation and/or prevention actions) that can be used as a support in risk informed decision making in different domains in general and projects (mainly international investment projects) management in particular. Indeed, globalization of economic activities gives opportunities to investors to consider to diversify their investment portfolio in investing in different countries for instance but also the risk of ill appreciation of the targeted country. The remainder of this communication is organized as follows: in the second section, the process of obtaining necessary materials to be used to evaluate risk, going from risky attributes elicitation to their evaluation (that is assigning them a measure) will be presented; third section is devoted to deriving scoring algorithm that consists basically in using previous obtained measures of risky attributes to build a one to one relationship between predefined classes and the considered project; in the fourth section, possible risk management derivation based mainly on identified risky attributes clustering will be briefly mentioned; section five considers an illustrative application to show usefulness of the established approach

and the sixth section concludes the communication.

2 Risky Attributes elicitation and evaluation

2.1 Elicitation

In terms of risk assessment, attributes that characterize projects may come from different horizons; so before a project start, one must identify all potential threats using different methods and tools among which the past experience, the opinion of members of project team, etc.. Identification process is a purely analytic activity that can be based on proven methods such as the following ones.

- **Documentation reviews.** This consists in structured reviews of all project documentation including plans, assumptions, preview projects files, contracts, etc. to possibly identify risky attributes of the project.
- **Assumptions analysis.** In practice, each project is based on a set of hypotheses, scenarios, or assumptions; the purpose of assumptions analysis is to explore the validity of assumptions in order to detect possibly inaccuracy, inconsistency or incompleteness that may constitute sources of risk.
- **Fishbone Diagrams.** Fishbone diagrams are risk diagramming techniques that include cause and effect diagrams useful for identifying risky attributes of a project.
- **SWOT (Strength, Weakness, Opportunities, Threats) analysis.** SWOT analysis consists in looking at the project from the perspective of its internal strengths and weakness as well as external opportunities, and threats; this is a very useful approach for risky attributes identification of a project.

The main purpose of these approaches is to have an answer to the following triple questions (Haimes et al., 2002).

- What can go wrong ? Identification of all events or sequence of events (or scenarios) that may have an (negative) effect on the projects that we referred to here as risky attributes of the projects.
- What is the likelihood that it would go wrong ? The probability of occurrence of that events or attributes.
- And, what are the consequences ? Characterization of the impact on the project if undesired events do really occur.

Answers to these questions help risk analysts identify, measure, quantify, and evaluate risks and their consequences and impacts. Once the identification process is finished, one must consider evaluating different identified risky attributes. Technically speaking risk is jointly associated with the likelihood (probability) of something (a risky attribute or event) happening and the impact (severity, gravity of that thing) which arises if it does actually happen. Thus and formerly, the risk $R_P(a_i)$ related to an attribute a_i of project P is defined by two measures: the probability of occurrence $Pr(a_i)$ of that attribute and its severity $\sigma(P/a_i)$ (impact of event a_i on the project P). The severity measures the negative impact on the project that are generally expressed by the amount of some losses (economic loss, lives loss, etc.) or by the probability of no satisfaction of objectives. In the literature risk is defined by the measure known as the criticality that is the product of the probability of occurrence and the severity; thus the measure $x_i(P)$ of risk related to attribute a_i on project P is given by equation (1).

$$x_i(P) = R_P(a_i) = Pr(a_i)\sigma(P/a_i) \quad (1)$$

Finally each project P will be completely characterized by its risk vector $x(P)$ as given by equation (2) if m primary risky attributes have been identified.

$$x(P) = [x_1(P), x_2(P), \dots, x_m(P)] \quad (2)$$

The following subsection go further in the process of evaluating the severity $\sigma(P/a_i)$ of the attribute a_i on the considered project P .

2.2 Evaluation of severity of risky attributes

The process of evaluating the severity $\sigma(P/a_i)$ of an attribute a_i on the project P may be not straightforward in general; indeed, elicitation process ends up most of time by a primary list of risky attribute which evaluation will characterize the concerned project as defined by equation (2). But, in practice one may need to split each primary attribute into many other sub attributes hierarchically until an operational level (a level where one can assign a score to attributes' severity) is reached as shown by Fig. 1. As shown on Fig. 1, at each level there may exist interactions between corresponding attributes so that for a sound evaluation these interactions must be taken into account. One possible tool in the literature that permit to evaluate interactions between entities is the so called "Decision-Making Trial and Evaluation Laboratory" or DEMATEL technique for short (Fontela and Gabus, 1976), (Gabus and Fontela, 1972), (Gabus and Fontela, 1973) which is briefly described in the following steps. As shown on Fig. 1, at each level there may exist interactions between corresponding attributes so that for

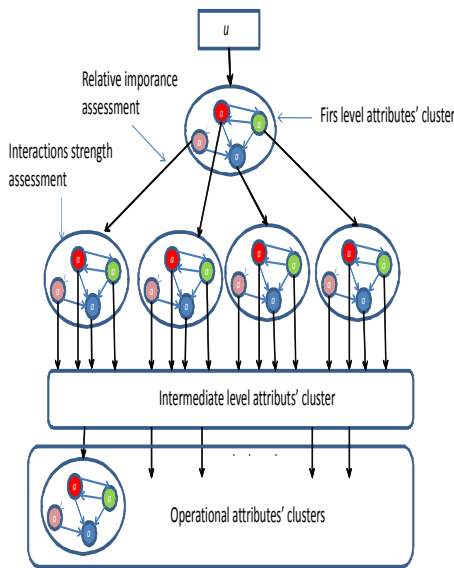


Figure 1 – Hierarchy in severity evaluation of risky attributes

a sound evaluation these interactions must be taken into account. One possible tool in the literature that permit to evaluate interactions between entities is the so called "Decision-Making Trial and Evaluation Laboratory" or DEMATEL technique for short (Fontela and Gabus, 1976), (Gabus and Fontela, 1972), (Gabus and Fontela, 1973) which is briefly described in the following steps.

- Step 1: Generating the direct-relation matrix by measuring the relationship between elements, attributes, criteria or entities in four levels: 0 (no influence), 1 (very low influence), 2 (low influence), 3 (very high influence) and 4 (very high influence). Next, experts make sets of the pairwise comparisons in terms of influence and direction between the direct-relation matrix, which is an $m \times m$ matrix A , in which a_{ij} is denoted as the degree to which the attribute a_i affects (influences) attribute a_j .
- Step 2: Normalizing the direct-relation matrix. On the basis of the direct-relation matrix A , the normalized direct relation matrix X can be obtained as given by equation (3)

$$X = \frac{A}{\max \left\{ \max_i \left\{ \sum_{j=1}^n a_{ij} \right\}, \max_j \left\{ \sum_{i=1}^n a_{ij} \right\} \right\}} \quad (3)$$

- Step 3: Attaining the total-relation matrix. In

fact influence will circulate among elements until a steady state is reached; so one needs to compute the total relation matrix. Once the normalized direct-relation matrix X is obtained, the total relation matrix T is calculated as shown by equation (4)

$$T = X(I - X)^{-1} \quad (4)$$

where I is the identity matrix.

From the total interaction matrix, one can determine some index (namely the dispatching index d_i and receiving index r_i for attribute i) that can be used to estimate the power or the influence degree of each attribute; these index are given by equation (5).

$$d_i = \sum_{j=1}^n T[i, j], r_i = \sum_{j=1}^n T[j, i] \quad (5)$$

The importance of attributes at each level can be analyzed using $d_i + r_i$ and $d_i - r_i$ as they convey following properties.

- $d_i + r_i$ represents relationships importance or connectivity degree of attribute a_i within the network either as a dispatcher or a receiver;
- $d_i - r_i$ represent the influence degree and characterizes the attribute a_i as a dispatcher or a receiver: indeed, if $d_i - r_i > 0$ then the corresponding attribute influences more other attributes than it is influenced and constitute therefore a dispatcher of influence; it will be consider as a receiver of influence otherwise.

Connectivity degree $d_i + r_i$ and influence degree $d_i - r_i$ can be used to categorize attributes for management purpose as described by Fig. 3 in the section dealing with risk management possibilities. In terms of evaluation, once interactions parameters are determined at each level, one will propagate the values going from operational attributes for which one can evaluate until first level attributes that constitute the final evaluation vector x of the considered project P , see Fig. 1 is reached. Decision regarding the considered project will depend on the risk category, class, or score it belongs to as defined by the organization. Thus when risky attributes are elicited and evaluated, one must consider classifying the project into appropriate class. This issue is not an obvious one because of

- Definition and characterization of risk classes; this issue is a matter of top management decision mainly in what concerns the number of classes or categories.

- Defining relationship between classes and attributes according to their evaluation score; in general, relying on experience of managers and/or using experts knowledge one can arrive at a suitable characterization.
- Aggregating the relationship strength between classes and attributes may be a challenging task as attributes may be evaluated using heterogeneous units and scales; so the analyst needs to transform these evaluations into a unique transparent evaluation scale that allow sound aggregation procedure.

The concerns raised here will be addressed in the next section that defines the process of fuzzy nominal classification as a framework and tool to classifying appropriately a project characterized by its risky attributes vector $x(P)$ into the appropriate risk class or category.

3 Bipolar fuzzy nominal classification

Formally nominal classification problem considered in this communication is defined by the following materials.

- A project P to be classified is characterized by a set of m risky attributes or criteria and the value of attribute l such as risk evaluation or score obtained in the previous section, is given by x_l and belong to an ordered (continuous or discrete) set X_l ($x_l \in X_l$); in many cases and for continuous attributes, X_l will be a subset of real numbers set R ; notice that we are therefore not considering categorical (yes/no) evaluated attributes.
- The former defined project must be assigned to one of the n risk classes or categories of the set $C = \{c_1, c_2, \dots, c_n\}$ using the attributes vector x .

In practice it is rather rare that relationships between attributes and classes be determined exactly. Most of the time they will be described only up to some degree of uncertainty with linguistic characterization such as if budget overtaking is high, low, mean, approximatively, near to, etc. then one should privilege a particular class compared to other classes for the inclusion of the corresponding project; that is they will be described fuzzily leading to what we name fuzzy nominal classification. The purpose of fuzzy nominal classification methods or algorithms is then to establish a procedure that select the most appropriate class where to include the considered project; one may notice that this is an absolute decision making process as objects to classify are not compared with each other. The approach adopted in this communication to derive such classification procedures rely on the notion

of bipolarity, see (Tchangani, 2010), (Tchangani and Pérés, 2010), (Tchangani et al., 2012); indeed, bipolar reasoning is pervasive in decision analysis and constitutes a sort of divide to better apprehend paradigm; this bipolarity results often in uncertainty in decision making processes. For instance a medical doctor who receives a patient with a certain level of body temperature must decide if the patient is suffering of a fever or not; in general doctor may know with certainty the category of the patient if the temperature range in a certain interval and will be less certain for other range of temperature. Many other decision problems such as classification of projects according to risks they convey can be considered in the same way. The stepping stones in bipolar analysis approach, for nominal classification, are the classifiability and rejectability measures $\mu_C^j(P)$ and $\mu_R^j(P)$ given a class c_j and a project P ; so their derivation is an important step towards a sound classification algorithm. These measures must be established considering the performance of the considered project with regard to the considered class. As mentioned above, relationships between attributes characterizing projects and risk classes are most of time fuzzily defined. In this communication, we consider the following fact: if one were to classifying or avoid classifying a project into the class c_j using only its attributes l measure x_l , there will be some values of its range X_l from which the inclusion will be immediate and that from which rejection also will be immediate; and some intermediate values from which decision of inclusion or not is not obvious. To characterize this issue; given a class c_j and an attribute l , the range X_l of the later will be partitioned into disjunctive classifiability range $X_l^{C,j}$ and the rejectability range $X_l^{R,j}$; let us denote by $m_l^{C,j}(x_l(\cdot))$ and $m_l^{R,j}(x_l(\cdot))$ classifying and rejecting membership functions with regards to the couple class c_j and attribute l value $x_l(\cdot)$; these functions verify following conditions, equations (6)-(9)

$$m_l^{C,j}(x_l(\cdot)) = 1; m_l^{R,j}(x_l(\cdot)) = 0, \forall x_l(\cdot) \in X_l^{C,j} \quad (6)$$

$$m_l^{C,j}(x_l(\cdot)) = 0; m_l^{R,j}(x_l(\cdot)) = 1, \forall x_l(\cdot) \in X_l^{R,j} \quad (7)$$

$$0 < m_l^{C,j}(x_l(\cdot)) < 1, \forall x_l(\cdot) \in X_l - X_l^{C,j} \cup X_l^{R,j} \quad (8)$$

$$0 < m_l^{R,j}(x_l(\cdot)) < 1, \forall x_l(\cdot) \in X_l - X_l^{C,j} \cup X_l^{R,j} \quad (9)$$

that are schematically illustrated on Fig.2

Finally, in terms of risk assessment materials, a project P will be characterized by two $m \times n$ (where m is the number of attributes and n the number of classes) membership matrices $M^C(P)$ and $M^R(P)$ defined by equation (10)

$$M_P^C(l, j) = m_l^{C,j}(x_l(P)); M_P^R(l, j) = m_l^{R,j}(x_l(P)) \quad (10)$$

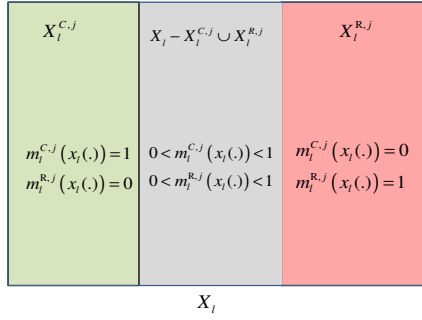


Figure 2 – Example of influential network relation map (INRM)

where $x_l(P)$ is the value of the attribute l for project P . The classifiability and rejectability measures $\mu_C^j(P)$ and $\mu_R^j(P)$ are then obtained as equations (11) and (12)

$$\mu_C^j(P) = \frac{G(M_P^C(:,j))}{\sum_{k=1}^n \{G(M_P^C(:,k))\}} \quad (11)$$

$$\mu_R^j(P) = \frac{G(M_P^R(:,j))}{\sum_{k=1}^n \{G(M_P^R(:,k))\}} \quad (12)$$

where G is an aggregation operator. As there are many features characterizing each class, their membership functions must be aggregated to derive a global classifiability and rejectability measures. A sound and useful aggregation operator is therefore needed. Given the synergy obtained by considering separately, classifiability and rejectability zone, it is obvious of considering a synergistic aggregation operator. Furthermore, features characterizing classes do not necessarily have the same importance in the classification process and experts may be able to weight them in terms of normalized vector ω . In this case, Choquet integral associated with a weighted cardinal fuzzy measure (wcfm) (Tchangani, 2013), is a suitable aggregation operator that permit to overcome difficulties due to the dimension of the vector to aggregate encountered in the literature (Grabisch, 1996) as a straightforward formula does exist. Given a numerically valued n dimension vector θ with a relative importance vector ω , Choquet integral of x associated to a weighted cardinal fuzzy measure with relative importance vector ω is give by equation (13)

$$C_\omega^{wcfm}(x) = \sum_{k=1}^n \{\varphi(A_k) (\theta_{\sigma(k)} - \theta_{\sigma(k-1)})\} \quad (13)$$

where $\varphi(A_k)$ is a weighted cardinal fuzzy measure of

A_k given by equation (14)

$$\varphi(A_k) = \left(\frac{n - (k - 1)}{n} \right) \left(\sum_{j \in A_k} \omega_j \right) \quad (14)$$

and A_k is defined as equation (15)

$$A_k = \{\sigma(k), \sigma(k + 1), \dots, \sigma(n)\} \quad (15)$$

where σ is a permutation over x as given by equation (16)

$$x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}; x_{\sigma(0)} = 0 \quad (16)$$

There is no difficulty to verify this straightforward formula; in this case the operator $G(\cdot)$ of equations (11) and (12) is given by $G(\cdot) = C_{\omega_j}^{wcfm}(\cdot)$ where ω_j is the relative importance weights vectors of features characterizing the class c_j . From classifiability and rejectability measures $\mu_C^j(P)$ and $\mu_R^j(P)$ one can define the classifying set $C(P)$ of the project P as given by equation (17)

$$C_q(P) = \{c_j : \mu_C^j(P) \geq q(\mu_R^j(P))\} \quad (17)$$

where q is a non decreasing function representing the caution or boldness attitude of decision maker; it can be reduced to just a proportional index. One should notice that many categories may be qualified for inclusion of a given project so that the ultimate class in which to include the project can be selected by optimizing index such as $\mu_C^j(P) - q(\mu_R^j(P))$ or other.

For the purpose of risks management (resources allocation for risks sharing, avoiding, preventing, or mitigating, etc.), the first step is to assess risks namely to identify and to evaluate risks; these processes have been considered in the previous sections, in the following section, we will briefly sketch how risk management actions can be undertaken.

4 Risks management

Risk management is decision making under uncertainty; it consists in defining control variables or management actions that are things or procedures that can be realized in order to achieve objectives; it seeks therefore answering questions such as the following ones (Haimes et al., 2002): what can be done ? what are available options ? what are associated options, etc.. Risk management actions will act primarily on risky attributes identified in the previous section. For this purpose DEMATEL analysis presented in the previous section can be very useful to optimize and

organize how to effectively control risky attributes. Indeed, from the influence degree $d - r$ and connectivity degree $d + r$ obtained from DEMATEL analysis and by setting a threshold (C^{th}) (that corresponds in general to the mean value) on the connectivity degree, one can obtain four categories of attributes as depicted on the Fig.3

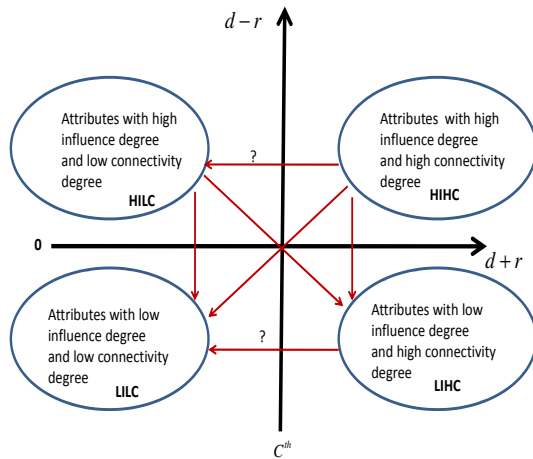


Figure 3 – Categorization of attributes according to their influence and connectivity

Appropriate treatment of attributes as management action may rely on this categorization as described in the following items.

- *Attributes with high influence degree and high connectivity degree (HIHC attributes)*. These attributes have a great influence degree and inverter-connection degree. These causal attributes have high prominence, so their occurrence invokes a wide range of influences and interactions, such that they may trigger or exaggerate the other attributes and induce feedback. With such a high risk level, top priority management interventions are required. Response strategies should be proactive, such as avoidance strategies to eliminate the risk factors or mitigation efforts to lessen their potential impacts to an acceptable level.
- *Attributes with high influence degree and low connectivity degree (HILC attributes)*. Their emergence produces effects on only certain attributes, so improving these attributes might have limited impacts on overall adjustments or optimization of the project. The best response strategy might be acceptance, with contingent plans, or mitigation to minimize their impacts.
- *Attributes with low influence degree and high connectivity degree (LIHC attributes)*. These at-

tributes have high prominence; they are among the most important attributes and demand greater attention. Their risk levels are associated with their dependence on others and how well their cause attributes (HIHC and/or HILC attributes) can be managed. In addition, these attributes might generate feedback effects, due to their high levels of interaction with other attributes, which should affect management actions. Accordingly, these attributes are avoided or mitigated through careful monitoring of the source attributes by which they are influenced.

- *Attributes with low influence degree and low connectivity degree (LILC attributes)*. These attributes are characterized by their minor influences and interactions with other factors. They are independent and isolated, so they may be judged as acceptable in ordinary situations.

In the following section we will consider a problem of risk scoring in the domain of international trade (selecting suppliers from foreign countries, outsourcing activities in foreign countries, investing in infrastructures development in foreign countries, etc.). To undertake such projects, investors must measure the risk on their investment performance given the situation of the targeted country.

5 Illustrative application

The illustrative application considers the problem of classifying countries in terms of risks some actors such as investors may face by investing in these countries. In international trade context, risk classification of countries relies mainly on the Worldwide Governance Indicators that determine the level of risks related to Governance in sourcing countries. There are 6 attributes, dimensions, or indicators of governance identified by the World Bank for this purpose that are briefly described in the following points.

1. Voice and Accountability (**VA**): perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.
2. Political Stability and Absence of Violence/Terrorism (**PV**): perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.
3. Government Effectiveness (**GE**): capturing perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality

of policy formulation and implementation, and the credibility of the government’s commitment to such policies.

4. Regulatory Quality (**RQ**): perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
5. Rule of Law (**RL**): capturing perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.
6. Control of Corruption (**CC**): perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

So a project consisting in considering investing in a country to classify will be characterized by a 6 dimension vector $x(P)$ given by equation (18)

$$x = [VA, PV, GE, RQ, RL, CC]. \tag{18}$$

Data concerning 209 countries were evaluated and registered in 2014 by Foreign Trade Association(FTA) (a Brussels based on organization working for European enterprises) and their statistics are shown on Table 1 (FTA, 2014); how these data are obtained, the units and scales of the evaluation, are not the matters of this communication, we just use them for illustration purpose.

Table 1 – Statistics of registered data of the considered application

	VA	PV	GE	RQ	RL	CC
min	0	0	0	0	0	0
max	100	100	100	100	100	100
mean (m)	49.7	49.6	50.3	50.3	49.8	50.0
std (σ)	29.1	29.1	29.1	29.1	29.1	29.3

Two classes were considered by (FTA, 2014): risk countries and low-risk countries; here we consider three categories of risks namely low risk (L), medium risk (M), and high risk (H) with the membership function of each of 6 attributes or dimensions given by Fig. 4. We also consider that all dimension are equally important so that the weighting vector ω to be used in Choquet integral (reduced in this case to the mean value) of x associated to a weighted cardinal fuzzy measure of equation (13) is given by

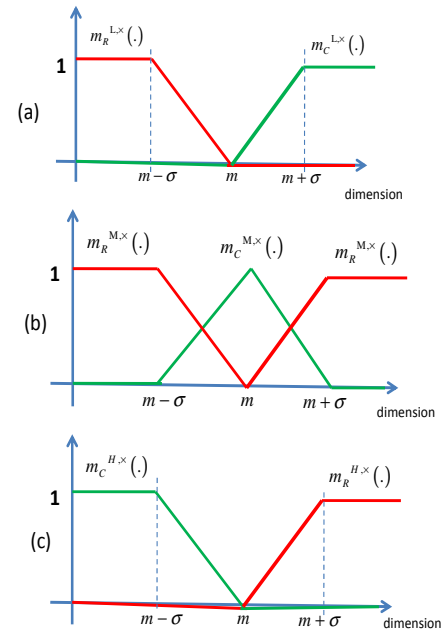


Figure 4 – Classifiability and rejectability membership functions of each dimension with regard to each class of the considered application

$\omega = 1/6 * ones(6, 1)$. Let us consider three countries classified (based on the mean value of x) by FTA as risk countries and characterized by following vectors, equations (19)-(21)

$$x_1 = [12.2, 45.8, 58.3, 48.3, 46.9, 69.7]; \tag{19}$$

$$x_2 = [23.0, 38.7, 7.6, 17.5, 23.5, 16.6]; \tag{20}$$

$$x_3 = [65.7, 48.1, 64.9, 65.9, 58.7, 59.2]. \tag{21}$$

Results obtained by applying the approach derived in this communication are given by Table 2. One should notice that country characterized by vector x_3 can be classified as medium or low risk country because its classifiability degree for these categories is greater than its rejectability degree; but if we consider maximizing the index $\mu_C^j(u) - \mu_R^j(u)$, the ultimate class of this country will be low risk class. The approach adopted in this communication is more richer than relying just on the mean value of attributes to classify as done in (FTA, 2014) because in terms of management one may find more precisely attributes to address for risk reduction purpose by doing sensitivity analysis for instance and/or optimal allocation of resources by concerned countries to improve their score.

Table 2 – Examples of results obtained by the proposed approach

	Low (L)		Medium (M)		High (H)		$C_1(P)$
	μ_R	μ_C	μ_R	μ_C	μ_R	μ_C	
x_1	0.22	0.08	0.64	0.80	0.14	0.12	M
x_2	0.50	0.00	0.50	0.05	0.00	0.95	H
x_3	0.00	0.34	0.51	0.66	0.49	0.00	M, L

6 Conclusion

The issue of scoring risk convey by a project in general and investment project in particular has been addressed in this communication. Indeed, for risks management (mainly in terms of resources allocation for risk sharing, risk mitigation, risk prevention, etc.), managers need to score or assign projects to pre-defined classes or categories for their appropriate treatment. To this end, a framework relying on bipolar fuzzy nominal classification, where risky attributes characterizing projects are fuzzily defined, has been developed. The established procedure goes from the attributes vector of a project that is converted to bipolar fuzzy membership functions (classifiability and rejectability memberships) based on fuzzy characterization of risk classes with regard to risky attributes; then these membership measures are aggregated using a Choquet integral associated to a weighted cardinal fuzzy measure (wcfm) to take into account synergy between attributes, to finally obtain classifiability and rejectability measures. The main contribution of the developed approach concerns the fact that raw data (that is the original attributes vector) are not needed to be normalized nor to be evaluated on the same scale as the transformation into membership functions is transparent. Furthermore, with this approach many classes may be qualified for a given project rendering this approach very flexible as decision makers' attitude such as hesitation can be revealed. The application of the developed approach framework to a real world problem in terms of classification of countries by the potential risks investors in these countries may face show interesting possibilities of this framework. The short term task regarding the approach proposed in this communication is to validate it on a real world application.

Fontela E. and Gabus A., 1976. The DEMATEL observer, *DEMATEL 1976 Report*, Switzerland Geneva: Battelle Geneva Research Center.

Gabus A. and Fontela E., 1972. World problems and invitation to further thought within the framework of DEMATEL, Switzerland Geneva: Battelle Geneva Research Centre.

Gabus A. and Fontela E., 1973. Perceptions of

the World problematic: Communication procedure, communicating with those bearing collective responsibility (DEMATEL Report No. 1)? Switzerland Geneva: Battelle Geneva Research Centre.

FTA, 2014. Countries' Risk Classification - *Public Document V.1*

Grabisch M., 1996. The application of fuzzy integrals in multicriteria decision making, *European Journal of Operational Research*, 89 (3), p. 445-456.

Haimes Y.Y., Kaplan S., and Lambert J. H., 2002. Risk Filtering, Ranking, and Management Framework Using Hierarchical Holographic Modeling *Risk Analysis*, Vol. 22, No. 2, P; 383-397.

Newton P., 2015. Managing Project Risk. Project Skills. *Free Management E-Books*.

Roy B. and Bouyssou D., 1993. Aide Multicritere a la Decision: Methodes et Cas, Edition Economica.

Steuer R. E., 1986. Muticriteria Optimization: Theory, Computation, and Application, Wiley, New York.

Tchangani A. P., 2013. Bipolar Aggregation Method for Fuzzy Nominal Classification using Weighted Cardinal Fuzzy Measure (WCFM) *Journal of Uncertain Systems*, 7: 2., p. 138-151.

Tchangani A. P., 2010. Considering Bipolarity of Attributes with regards to Objectives in Decisions Evaluation, *Inzinerine Ekonomika - Engineering Economics*, Vol. 21, No. 5., p. 475-484.

Tchangani A. P., Bouzarour Y., and Pérès F., 2012. Evaluation Model in Decision Analysis: Bipolar Approach, *INFORMATICA: an International Journal*, 23 (3), p.461-485.

Tchangani A. P. and Pérès F., 2010. BOCR Framework for Decision Analysis, *Proceedings of 12th IFAC Symposium on Large Scale Complex Systems Theory and Applications*, Volume 43, Issue 8 Edited by: Borne Pierre and Filip Florin Gheorghie. p. 507-513 IFAC PapersOnLine.

Semigreedy Heuristics for the Flexible Jobshop Scheduling Problem with Multiple Process Plans and Sequence Dependent Setup Times

Nicolau Santos^{1,3}
nsantos@inesctec.pt

Gonalo Figueira^{1,2}
goncalo.figueira@fe.up.pt

Joo Pedro Pedroso^{1,3}
jpp@fc.up.pt

¹INESC TEC

²FEUP

³FCUP

R. Dr. Roberto Frias
4200-465 Porto - Portugal

R. Dr. Roberto Frias
4200-465 Porto - Portugal

R. Campo Alegre 1021/1055
4169-007 Porto - Portugal

ABSTRACT: *In this work we present a study on semigreedy heuristics for the Flexible Jobshop Scheduling Problem with multiple process plans and sequence dependent setup times. We evaluate different combinations of (selection rule, greedy parameter, objective function) in a set of instances generated from data from a real production system. We also compare the performance of these heuristics in a benchmark from the literature and conclude that, although simple, the algorithms perform better than more complex methods and still have margin for further improvement.*

KEYWORDS: *Combinatorial Optimization, Flexible Jobshop, Heuristics, Multiple Process Plans.*

1 INTRODUCTION

In the Jobshop Scheduling Problem (JSP) we have a set of n jobs to perform on m machines. Each job has a sequence of technological operations (route or process plan) and each operation is associated to a given machine. At a given time a machine can only process a single operation and job preemption is not allowed. To solve the JSP we need to find a feasible sequence of operations in each machine in order to optimize a given objective function. The JSP is naturally extended with the introduction of parallel machines, resulting in the Flexible Jobshop Scheduling Problem (FJSP) for which it is possible to produce each operation in one or more machines. In our work we deal with a more general model where the route of each job is not unique: more precisely, each job may have more than one process plan, and the number of operations in the different routes is not necessarily the same. This situation arises in real production systems due to characteristics of the machines or directly from the skills of workers. The presented work is derived from the study of a metal sheet printing shop where the flexibility derives from the number of colors that printing machines are able to apply. As an example, suppose we have a job whose production stages consist in the sequential application of the following products: preparation varnish, color A, color B and finishing varnish. Now, suppose the available machines are the following, with the given characteristics: M1 is able to apply preparation varnish; M2 has the capacity to print 2 colors; M3 and M4 both can print a single color; and M5 is able to apply fin-

ishing varnish. In Figure 1 we present 2 alternative process plans for the given example. Nodes 0 and 6 represent dummy source and sink nodes, while nodes 1 and 5 represent the preparation and finishing varnish operations. Nodes 2, 3 and 4 represent color operations and there are two alternative sequences of production, namely $1 \rightarrow 2 \rightarrow 5$ and $1 \rightarrow 3 \rightarrow 4 \rightarrow 5$. Note that the 2 colors in the initial production stages are condensed in a single operation (2) for the first route, while in the second production plan they are split in 2 operations (3 and 4). All these created operations are independent, each having its own processing time. A given operation may have more than one job successor, but only one of them will be chosen.

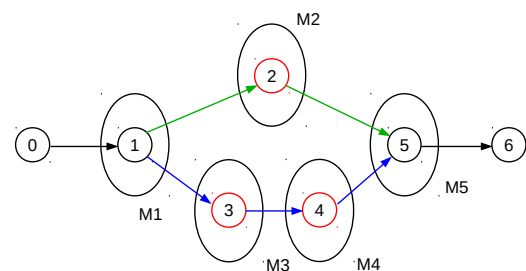


Figure 1 – Example of a job with two alternative process plans with different number of operations.

In the presented model we also consider the existence of a sequence dependent setup time $S_{i,k}$ between any two consecutive operations i and k in a given machine and denominate the resulting problem as Flexible

Jobshop Scheduling Problem with Multiple Process Plans and Sequence Dependent Setup Times (FJSP-MPP-SDS). To solve the FJSP-MPP-SDS we need to select a route for every job and schedule the respective operations in order to optimize a given objective function. Let i be an operation and j a given job. We denote p_i as the processing time of operation i and d_j due date of job j . Following closely the nomenclature of (Taillard 1994), let J_i be the job to which i belongs and M_i the machine which processes i . We define the operation PJ_i as the job predecessor of operation i , i.e., the operation of job J_i immediately before operation i in the considered route and the operation PM_i as the machine predecessor of i , i.e., the operation scheduled on machine M_i immediately before operation i . With these definitions we now can calculate C_i , the completion time of operation i as:

$$C_i = \max(C_{PJ_i}, C_{PM_i} + S_{PM_i,i}) + p_i \quad (1)$$

Equation 1 states that no operation can be started before its job and machine predecessors have both been completed, including the setup time from the latter. To obtain C_i , we add p_i to the maximum of the previous values. We define $PJ_i = 0$ for the first operation of job j and $PM_i = 0$ for the first operation in machine M_i . In the second case $S_{PM_i,i}$ should be replaced by zero or a given constant, depending on the characteristics of the considered production system. In the remainder of this work we assume the machines are initially prepared to process the first operation and consider this value to be 0.

Let F_j be the last operation in the selected route of job j . We define the *Makespan*, denoted as C_{\max} , as $\max(F_j : j = 1 \dots n)$ and the *total flowtime*, denoted as C_{sum} , as $\sum_{i=1}^n F_j$. Finally we define the *tardiness* of job j as $T_j = \max(F_j - d_j, 0)$ and the *total tardiness*, defined as TT , as $\sum_{i=1}^n T_j$.

Although references to the FJSP-MPP can be traced to as early as (Iwata, Murotsu, Oba, Uemura & Okushima 1978) there are very few works on this subject. In (Kim & Egbelu 1998) the authors present a mathematical model for the FJSP-MPP and discuss preprocessing approaches to reduce the dimension of the problem before using a solver. In (Kim & Egbelu 1999) the same authors extend their results with a heuristic that combines branch and bound with a dispatching rule; the presented algorithms are tested in instances up to 10 jobs and 10 machines. More recently (Golmakani & Namazi 2012) present an Artificial Immune Algorithm and (Golmakani & Birjandi 2013) propose a two-phase algorithm. To our knowledge there are no publications on the FJSP-MPP-SDS, and all the articles on the FJSP-MPP deal with C_{\max} minimization.

In this work we present a simple heuristic method for the FJSP-MPP-SDS that is easily adaptable for different objective functions, and scales well to large

problems with hundreds of jobs. In Section 2 we describe the proposed method. Section 3 presents a computational study of the proposed method considering several objective functions. Finally, in Section 4 we present some conclusions and future work.

2 CONSTRUCTION HEURISTICS

Construction heuristics are among the simplest methods to find solutions to a given optimization problem, however it is not straightforward how to proceed in the case of the FJSP-MPP-SDS. One option would be to choose a route for each job in a first step and schedule the resulting operations afterwards. However we will proceed in an integrated fashion, implicitly selecting a process plan by scheduling one of its operations, i.e., assigning a start and finish time to the chosen operation in its machine. At each step, the construction heuristic iteratively selects an operation o from a candidate list X and builds a schedule Y . The process is repeated until there are no more candidates available. Pseudo code for the proposed heuristic is presented in Algorithm 1. We now present

Data: an instance for the FJSP-SDS-MPP

Result: a feasible schedule

$X \leftarrow \emptyset$

$Y \leftarrow \emptyset$

initialize(X)

while $X \neq \emptyset$ **do**

score(X)

$o \leftarrow \textit{choose}$ (X)

schedule(Y, o)

update(X, o)

end

Algorithm 1: Construction Heuristic for the FJSP-SDS-MPP

a description of the above mentioned procedures:

- *initialize*(X) - let the candidates list X initially contain every possible first operation of each job;
- *score*(X) - for each element in X we assign a score based on a evaluation function; alternative evaluation functions are presented in Section 2.1;
- *choose*(X) - at each step of execution, we need to choose an operation to schedule. Several alternatives are discussed in Section 2.2;
- *schedule*(o) - assigns a start and end time to the selected operation o and stores it in Y . Calculations for bookkeeping are made according to Equation 1;
- *update*(X, o) - in update we first remove from X every operation with the same job as the previously scheduled operation o . Afterwards we add to X every possible successor of o .

Note that, depending on the possible routes, each job may have more than one operation in X at a given time. The final schedule presents a Topological Sort of the chosen operations.

2.1 Scoring Rules

One of the major steps in a scheduling process is to decide how to prioritize one operation over another. A straightforward way to make the decision is to assign a score to every operation, according to a given rule, and make an informed choice afterwards. Although a great variety of rules is found in the literature we will focus on the following three:

- Earliest Available Operation (EAO): score is given by the earliest possible starting time for each operation;
- Earliest Finishing Operation (EFO): to the earliest starting time of each operation we sum its production time, obtaining its score;
- Modified Operation Due Date (MOD): score determined by the maximum between the finishing time of an operation and the due date of the corresponding job. This rule prioritizes operations which have completion time closer to the job's due date.

The EAO rule aims at the minimization of idle time between operations, so it is a criterion associated with C_{\max} . The EFO rule prioritizes operations with smallest production times, so it is a rule related to C_{sum} . Finally, the MOD rule takes in to account the due date of the scheduled jobs, and is naturally associated with TT .

We refrain from testing selection rules such as shortest processing time, earliest due date or shortest setup time, since in the context of the presented framework they yield very poor results. This comes from the fact that, unlike the previous rules, these do not take in to account the state of the system. In Section 3.2 we present a comparison between our method and dispatching rules implemented in the literature.

2.2 Selection Rules

The two classical methods to select an element from a candidates list are: the greedy approach, where we select the element with the best score, and the random choice. However these options might be too rigid or loose in practice. To overcome these limitations (Hart & Shogan 1987) and (Feo & Resende 1989) developed a semigreedy procedure that aims to find a more balanced candidate through the use of a Restricted Candidate List (RCL), where the choice is limited to the most promising candidates. In this

procedure we first need to get the minimum a and maximum b of the scores $\{s_o : o \in X\}$ of the candidate operations. With these values and a predefined $\alpha \in [0, 1]$ we calculate the value cut that is used to limit the selection, where:

$$a = \min(\{s_o : o \in X\})$$

$$b = \max(\{s_o : o \in X\})$$

$$cut = a + \alpha(b - a)$$

After these calculations we are able to define the RCL according to α .

$$RCL = \{o \in X : s_o \leq cut\}$$

We then choose randomly an operation from this set. With $\alpha = 0$ we have a greedy approach while with $\alpha = 1$ we have a random choice. Note that even with $\alpha = 0$ the search may be diversified, since a candidate is chosen randomly from ties with minimum score. In order to further understand the influence of this diversification we also implement a purely greedy static variant where, in case of ties, we always select the operation with lowest index from the job of lowest index.

3 RESULTS

In this section we present computational results to evaluate the performance of the proposed heuristic. Section 3.1 presents results obtained for instances generated with distributions identical to those found on a real production system. Section 3.2 presents results for the well known benchmark of (Brandimarte 1993), developed for the standard FJSP. All algorithms were implemented in C++ and compiled with Clang 800.0.42.1. The machine used for the tests has an Intel Xeon W3520 processor at 2.67 GHz and 16GB of RAM running OS Darwin 15.6.0. Only a single thread was used for each run.

3.1 Tests with Generated Instances

To compare the performance of the proposed methods we generated 60 instances with production times, setup times, due dates and routes sampled from the distributions of a real production system. The dimensions of the generated data have $n \in \{10, 20, 40, 80, 160, 320\}$ and $m = 8$. We tested Algorithm 1 with the EAO, EFO and MOD scoring rules, the static variant (s), and semigreedy variants with $\alpha \in \{0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0\}$. Considered objectives are C_{\max} , C_{sum} and TT . For each combination we performed 100000 runs, which were enough to analyze convergence of the algorithms. The CPU times ranged from an average of 1.6 seconds in the smaller instances with 10 jobs, to 1116 seconds

in the larger instances with 320 jobs. For each (instance, α , rule) combination we calculated the error E defined as

$$E = (V - B)/B, \quad (2)$$

where V is the objective value of the considered run and B is the best value obtained for the corresponding instance in all tests.

In Tables 1, 2 and 3 we present the average error for the different combinations of scoring rules, alpha and objective function on each group of instances. As expected, EAO performs better for the C_{max} objective; however MOD exhibits the best results for both C_{sum} and TT objectives. This is explained by the fact that the generated instances have a very high number of late jobs and the overall objective values for each instance for the C_{sum} and TT objectives are also similar. An unexpected result is the variation of the errors with α . For the smaller instances, with 10 jobs, the best values are obtained with $\alpha = 0.1$ or $\alpha = 0.2$. However, as the instance size grows, the best results are obtained with $\alpha = 0$. This indicates that when the number of jobs is small the heuristic needs a higher degree of randomness to find the best sequence of operations in each machine. We also observe that for $\alpha = 0$ and $\alpha = 0.05$ the semigreedy method consistently performs better than the static variant. Even though the improvement is more marked for the C_{max} , it is always beneficial to add some variety to the search. In the above analysis it is not clear what is the number of runs needed by the semigreedy implementations to overcome the static variant. Besides, a good solution may be obtained in the proposed number of runs by mere chance. To answer these questions we developed an experiment to analyze the evolution of the expected value of the minimum with bootstrapping (Efron 1992). The 100000 observation of each experiment were used as a sample for analyzing the distribution of the solution found by each of the score rules. We analyzed the behavior of: the EAO in the C_{max} minimization problem; the EFO in the C_{sum} minimization problem; and the MOD in the TT minimization problem. For each experiment, we have chosen 1000 random subsamples, with replacement, of 10, 100, 1000 and 10000 objective values and stored the best value of each sample. In Figure 2 we display the evolution of the error, determined with Equation 2, of the expected value of the minimum for the C_{max} objective and EAO rule. We evaluate the evolution of the results with the size of the sample, for different values of α , against the static variant. We observe that for samples of size 10 and $\alpha \in \{0, 0.05, 0.1\}$ the semigreedy variants already outperform the static variant. We also observe that the results of sampling are consistent with the results of Table 1, as the results with $\alpha = 0$ are better than for $\alpha = 0.05$, and results for $\alpha = 0.05$ are better than for $\alpha = 0.1$, independently of the size of the sample considered. Also

consistent with Tables 2 and 3 are Figures 3 and 4, where we present the evolution of the EFO rule for the C_{sum} objective and the evolution of the MOD rule for the TT objective, in our bootstrapping experiment. Again, a better α in the initial results is also better in the bootstrap. However, for different values of α , there are now visible differences in the speed of convergence: for $\alpha = 0$ again we only need samples of size 10 to perform better than in the static variant; however, for $\alpha = 0.05$ a sample of size 10000 is now needed to be, on average, better than the static variant.

3.2 Brandimarte Instances

Although the presented algorithms were developed for the FJSP-MPP-SDS, they can be directly applied to the regular FJSP. In Table 4 we compare the results with the test instances proposed in (Brandimarte 1993). Column *best* presents the best known solutions retrieved from (Quintiq 2014); values in bold are optimal. Column *ts* refers to the Tabu Search procedure of (Gambardella & Mastrolilli 1996). Columns *spt*, *lwkr* and *mwkr* are the simple dispatching rules "shortest processing time", "least work remaining", and "most work remaining", as implemented in (Brandimarte 1993). *tab1* and *tab2* are two distinct Tabu Search algorithms, also from (Brandimarte 1993). Finally, *sg0.05* is our proposed heuristic with the EAO rule and $\alpha = 0.05$. For each instance Mk01 to Mk10, we present the best objective found by each algorithm; row *av* presents the average error for each implementation. As we can observe, the well-known dispatching rules *spt*, *lwkr* and *mwkr* perform poorly and are easily overcome by our simple semigreedy heuristic. Interestingly, our simple method is also able to outperform *tab1* and *tab2* even though these elaborated algorithms consider critical paths in their search procedure. Among the methods presented, *ts* presents the best results, with an average error of only 0.85.

4 CONCLUSIONS

In this work we presented a study on semigreedy heuristics for the Flexible Jobshop Scheduling Problem with Multiple Process Plans and Sequence Dependent Setup Times. Through an series of tests in generated data and instances from the literature, we analyzed the behavior of the method for different combinations of (selection rule, greedy parameter, objective function). We also showed that with these simple methods we can outperform well-known dispatching rules and more complex methods. Nevertheless, we also note that there may be room for improvement, as results are not on par with state of the art implementations of more complex metaheuristics, even though these were developed for a more specif

problem (Flexible Jobshop Scheduling Problem with C_{\max} minimization as objective). A straightforward alternative for improving the results is to add a Local Search (LS) procedure and improve the algorithm into a Greedy Randomized Adaptive Search Procedure (Feo & Resende 1995). However, LS is typically tied to the objective function under evaluation, and we may lose the generality of our implementation. Other alternative under study is to develop better scoring rules with the aid of Machine Learning techniques. Finally it is not clear how the different methods would behave under a more realistic scenario of jobs arriving in different periods of time. For this purpose we are developing a simulation framework to analyze the dynamic variant of the problem.

ACKNOWLEDGMENTS

This work is financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project «POCI-01-0145-FEDER-006961», and by National Funds through the Portuguese funding agency, FCT - Fundaao para a Ciencia e a Tecnologia as part of project «UID/EEA/50014/2013».

References

Brandimarte, P. (1993). Routing and scheduling in a flexible job shop by tabu search, *Annals of Operations research* **41**(3): 157–183.

Efron, B. (1992). Bootstrap methods: another look at the jackknife, *Breakthroughs in statistics*, Springer, pp. 569–593.

Feo, T. A. & Resende, M. G. (1989). A probabilistic heuristic for a computationally difficult set covering problem, *Operations research letters* **8**(2): 67–71.

Feo, T. A. & Resende, M. G. (1995). Greedy randomized adaptive search procedures, *Journal of global optimization* **6**(2): 109–133.

Gambardella, L. & Mastrolilli, M. (1996). Effective neighborhood functions for the flexible job shop problem, *Journal of scheduling* **3**(3): 3–20.

Golmakani, H. & Birjandi, A. (2013). A two-phase algorithm for multiple-route job shop scheduling problem subject to makespan., *International Journal of Advanced Manufacturing Technology* **67**.

Golmakani, H. R. & Namazi, A. (2012). An artificial immune algorithm for multiple-route job shop scheduling problem, *The International Journal of Advanced Manufacturing Technology* **63**(1): 77–86.

Hart, J. P. & Shogan, A. W. (1987). Semi-greedy heuristics: An empirical study, *Operations Research Letters* **6**(3): 107–114.

Iwata, K., Murotsu, Y., Oba, F., Uemura, T. & Okushima, K. (1978). Optimization of selection of machine tools, loading sequence of parts and machining conditions in job-shop type machining systems, *Annals of the CIRP* **27**(1): 447–451.

Kim, K.-H. & Egbelu, P. (1999). Scheduling in a production environment with multiple process plans per job, *International Journal of Production Research* **37**(12): 2725–2753.

Kim, K.-H. & Egbelu, P. J. (1998). A mathematical model for job shop scheduling with multiple process plan consideration per job, *Production Planning & Control* **9**(3): 250–259.

Quintiq (2014). Flexible job shop scheduling problem. [Online; accessed 12-January-2018].
URL: <http://www.quintiq.com/optimization/flexible-job-shop-scheduling-problem-results.html>

Taillard, E. D. (1994). Parallel taboo search techniques for the job shop scheduling problem, *ORSA journal on Computing* **6**(2): 108–117.

rule	n	s					α				
		0.0	0.05	0.1	0.2	0.3	0.4	0.5	0.75	1.0	
EAO	10	0.21	0.04	0.03	0.01	0.00	0.00	0.01	0.01	0.02	0.10
	20	0.30	0.03	0.01	0.01	0.02	0.06	0.07	0.08	0.12	0.31
	40	0.25	0.01	0.01	0.02	0.06	0.10	0.13	0.16	0.22	0.48
	80	0.18	0.00	0.01	0.04	0.11	0.16	0.19	0.21	0.27	0.57
	160	0.12	0.00	0.03	0.07	0.15	0.22	0.26	0.28	0.34	0.76
	320	0.18	0.00	0.03	0.09	0.19	0.28	0.32	0.34	0.40	0.94
	av	0.21	0.01	0.02	0.04	0.09	0.14	0.16	0.18	0.23	0.53
EFO	10	0.38	0.37	0.19	0.14	0.07	0.05	0.04	0.04	0.04	0.10
	20	0.44	0.37	0.21	0.18	0.17	0.17	0.17	0.18	0.19	0.31
	40	0.44	0.33	0.23	0.25	0.26	0.28	0.29	0.31	0.35	0.48
	80	0.28	0.21	0.20	0.24	0.29	0.32	0.34	0.34	0.40	0.57
	160	0.27	0.19	0.23	0.28	0.34	0.37	0.39	0.41	0.49	0.76
	320	0.20	0.14	0.21	0.28	0.37	0.40	0.43	0.45	0.54	0.94
	av	0.34	0.27	0.21	0.23	0.25	0.26	0.28	0.29	0.33	0.53
MOD	10	0.41	0.40	0.19	0.12	0.07	0.05	0.03	0.03	0.03	0.10
	20	0.43	0.35	0.21	0.18	0.17	0.18	0.18	0.18	0.18	0.31
	40	0.43	0.32	0.23	0.24	0.25	0.28	0.30	0.30	0.34	0.48
	80	0.27	0.21	0.21	0.24	0.29	0.31	0.34	0.34	0.40	0.57
	160	0.29	0.19	0.23	0.29	0.35	0.37	0.39	0.41	0.50	0.76
	320	0.21	0.14	0.21	0.28	0.37	0.40	0.44	0.45	0.54	0.94
	av	0.34	0.27	0.21	0.23	0.25	0.27	0.28	0.29	0.33	0.53

Table 1 – Average errors for *Makespan* minimization under different instance size, scoring rule and selection rule combinations.

rule	n	s					α				
		0.0	0.05	0.1	0.2	0.3	0.4	0.5	0.75	1.0	
EAO	10	0.35	0.09	0.05	0.03	0.03	0.04	0.05	0.06	0.09	0.21
	20	0.56	0.15	0.14	0.13	0.16	0.21	0.24	0.27	0.35	0.60
	40	0.49	0.15	0.15	0.18	0.23	0.30	0.36	0.40	0.52	0.93
	80	0.60	0.26	0.27	0.32	0.44	0.54	0.60	0.64	0.76	1.25
	160	0.52	0.30	0.34	0.40	0.55	0.68	0.75	0.78	0.90	1.57
	320	0.61	0.30	0.37	0.46	0.64	0.82	0.89	0.93	1.02	1.86
	av	0.52	0.21	0.22	0.25	0.34	0.43	0.48	0.51	0.61	1.07
EFO	10	0.12	0.09	0.03	0.01	0.01	0.02	0.03	0.05	0.08	0.21
	20	0.10	0.08	0.00	0.02	0.08	0.12	0.18	0.24	0.38	0.60
	40	0.07	0.03	0.01	0.05	0.15	0.23	0.29	0.37	0.58	0.93
	80	0.04	0.01	0.06	0.17	0.32	0.42	0.49	0.58	0.82	1.25
	160	0.05	0.01	0.09	0.23	0.41	0.51	0.59	0.67	0.94	1.57
	320	0.04	0.00	0.11	0.28	0.50	0.60	0.68	0.75	1.00	1.86
	av	0.07	0.04	0.05	0.13	0.24	0.32	0.38	0.44	0.64	1.07
MOD	10	0.12	0.09	0.03	0.01	0.01	0.02	0.03	0.04	0.08	0.21
	20	0.10	0.07	0.00	0.02	0.08	0.13	0.18	0.25	0.38	0.60
	40	0.07	0.02	0.01	0.05	0.15	0.23	0.30	0.36	0.58	0.93
	80	0.05	0.01	0.06	0.16	0.32	0.42	0.49	0.56	0.81	1.25
	160	0.05	0.00	0.09	0.24	0.41	0.51	0.59	0.67	0.93	1.57
	320	0.05	0.01	0.11	0.29	0.50	0.60	0.68	0.76	1.01	1.86
	av	0.07	0.03	0.05	0.13	0.24	0.32	0.38	0.44	0.63	1.07

Table 2 – Average errors for *Total Flowtime* minimization under different instance size, scoring rule and selection rule combinations.

rule	n	s					α				
		0.0	0.05	0.1	0.2	0.3	0.4	0.5	0.75	1.0	
EAO	10	0.36	0.09	0.05	0.03	0.03	0.05	0.05	0.06	0.10	0.21
	20	0.57	0.15	0.14	0.14	0.16	0.21	0.25	0.28	0.36	0.61
	40	0.49	0.16	0.15	0.18	0.23	0.30	0.36	0.41	0.53	0.94
	80	0.61	0.26	0.27	0.32	0.44	0.54	0.61	0.64	0.76	1.26
	160	0.52	0.30	0.34	0.41	0.55	0.68	0.75	0.79	0.90	1.57
	320	0.61	0.31	0.37	0.46	0.64	0.82	0.89	0.93	1.02	1.86
	av	0.53	0.21	0.22	0.26	0.34	0.43	0.48	0.52	0.61	1.08
EFO	10	0.12	0.10	0.03	0.01	0.01	0.02	0.03	0.05	0.08	0.21
	20	0.11	0.08	0.00	0.02	0.08	0.13	0.18	0.25	0.39	0.61
	40	0.07	0.03	0.01	0.05	0.15	0.24	0.30	0.37	0.59	0.94
	80	0.04	0.01	0.06	0.17	0.32	0.43	0.49	0.58	0.83	1.26
	160	0.05	0.01	0.09	0.23	0.41	0.51	0.59	0.67	0.95	1.57
	320	0.04	0.00	0.11	0.28	0.50	0.61	0.69	0.76	1.01	1.86
	av	0.07	0.04	0.05	0.13	0.25	0.32	0.38	0.45	0.64	1.08
MOD	10	0.13	0.09	0.04	0.01	0.01	0.02	0.03	0.04	0.08	0.21
	20	0.10	0.07	0.00	0.02	0.08	0.13	0.19	0.25	0.39	0.61
	40	0.08	0.03	0.01	0.05	0.15	0.23	0.30	0.37	0.59	0.94
	80	0.05	0.01	0.06	0.16	0.32	0.42	0.49	0.57	0.82	1.26
	160	0.05	0.00	0.09	0.24	0.41	0.51	0.60	0.67	0.93	1.57
	320	0.05	0.01	0.11	0.29	0.50	0.61	0.69	0.76	1.01	1.86
	av	0.07	0.03	0.05	0.13	0.24	0.32	0.38	0.44	0.64	1.08

Table 3 – Average errors for *Total Tardiness* minimization under different instance size, scoring rule and selection rule combinations.

instance	best	ts	spt	lwkr	mwkr	tab1	tab2	sg _{0.05}
Mk01	40	40	65	76	54	49	42	41
Mk02	26	26	44	48	41	41	32	35
Mk03	204	204	397	506	296	263	211	204
Mk04	60	60	109	121	89	89	81	67
Mk05	172	173	231	291	188	188	186	180
Mk06	57	58	128	143	139	128	86	81
Mk07	139	144	188	238	262	188	157	177
Mk08	523	523	670	1042	558	523	523	523
Mk09	307	307	573	723	536	444	369	337
Mk10	193	198	536	627	524	363	296	265
av		0.85	79.46	117.52	68.06	45.93	21.20	17.00

Table 4 – Best results obtained in the Brandimarte dataset for different algorithms in the literature and our Semigreedy implementation.

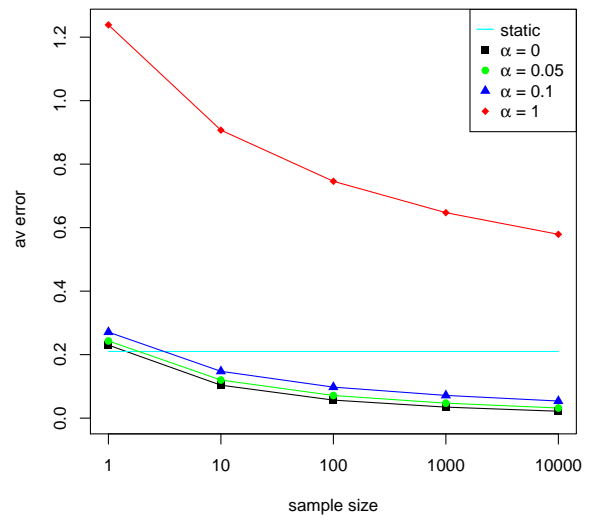


Figure 2 – Evolution of the expected value of the error with the increase of sample size for the *Makespan* objective and the *EAO* rule.

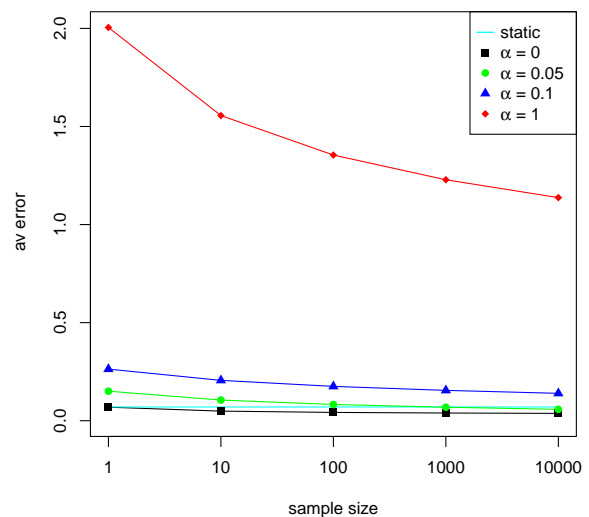


Figure 3 – Evolution of the expected value of the error with the increase of sample size for the *Total Flowtime* objective and the *EFO* rule.

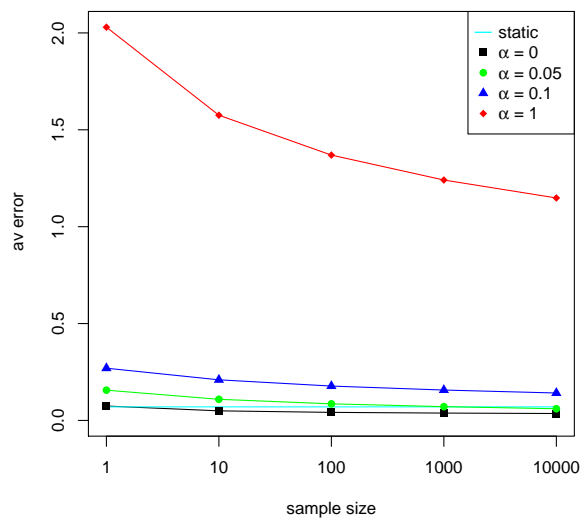


Figure 4 – Evolution of the expected value of the error with the increase of sample size for the *Total Tardiness* objective and the *MOD* rule.

SIMULATION OF COMMUNICATION BASED TRAIN CONTROL FOR ROBUSTNESS ASSESSMENT OF A MIXED TRAFFIC MANAGEMENT SYSTEM

S. CORNET, P. BOUVAREL

SNCF Réseau
34 rue du Commandant Mouchotte
75014 Paris - France
selim.cornet2@reseau.sncf.fr, paul.bouvarel@reseau.sncf.fr

C. BUISSON

IFSTTAR - LICIT
Rue Maurice Audin
69518 Vaulx en Velin - France
christine.buisson@entpe.fr

F. RAMOND

SNCF Innovation & Recherche
40 avenue des Terroirs de France
75012 Paris - France
francois.ramond@sncf.fr

J. RODRIGUEZ

IFSTTAR - ESTAS
20 rue Elisée Reclus
59666 Villeneuve d'Ascq - France
joaquin.rodriguez@ifsttar.fr

ABSTRACT: *In order to achieve higher capacity on existing infrastructures with dense traffic, railway companies are deploying Communication Based Train Control (CBTC) systems. These systems would considerably facilitate the implementation of new ways for managing traffic, such as headway-based control. However, for commercial reasons, traffic management systems using the timetable adherence principle have to be maintained on low traffic areas; both traffic management systems should hence have to coexist on some lines. The aim of this paper is to present a simulation method for assessing the robustness of operations on such a line, located both in low and dense traffic area. More precisely, the capability of a headway-based control system to resist to disturbances occurring out of its control perimeter is evaluated. The simulation tool reproduces train traffic evolution, disturbances and dispatching decisions, allowing to estimate the quality of service for various disturbance scenarios. Numerical experiments are carried out on one of the most saturated lines in the Paris suburban area. The obtained results prove the practical feasibility of such a mixed traffic management up to a critical level of disturbances.*

KEYWORDS: *Communication-Based Train Control, headway-based control, railway operations, real-time traffic management*

1 INTRODUCTION

1.1 Railway operations in dense urban areas

Most railway companies operating services in densely populated areas have been experiencing an increase of demand for passenger transportation during last years. As a result, railway networks reach saturation, making it difficult to answer the demand. In addition, as the number of trains grows, operations become more sensitive to disturbances, leading to a poor quality of service for passengers. The so-called three-aspects signaling system, used in many countries for ensuring the operations safety, plays a particular role in disturbances propagation in the network. This system allows to maintain a sufficient headway

between two consecutive trains. It works the following way. The track is divided into blocks, each of them protected by a signal. A green signal indicates to the train driver that the next two blocks are free, and that he can proceed at the same speed. A yellow signal means that the next block is free but the following one is not, and the driver must therefore prepare to stop before the limit of the current block. Finally, a red signal imposes the driver to stop as the next block is occupied.

Figure 1 shows how this system may lead a delay to propagate as knock-on delays along a line. If a train dwells in a station for a longer time than the nominal one (due e.g. to a jammed door), the following trains are likely to catch up and meet a yellow signal. This

will force them to slow down and result in even higher delays for these trains.

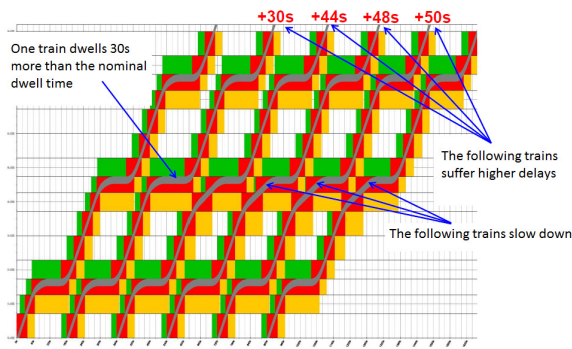


Figure 1 – Propagation of delays by a fixed signaling system

When designing timetables, railway companies usually insert buffers between consecutive trains to prevent the propagation of delays. But this comes at a price, as it has the effect of reducing the capacity of an already saturated network while not guaranteeing against delay propagation. Operating companies seek therefore to improve the system capacity and reliability without resorting to heavy and costly infrastructure modifications.

1.2 Communication-Based Train Control

Communication-Based Train Control (CBTC) systems are intelligent signaling and vehicle control systems that help to reduce the effect of the phenomena described in the previous section (IEEE Standard 1474.1 (2004)). These kind of systems have been widely implemented on metro lines (such as line 14 in Paris or line L in New York). Vehicles equipped with CBTC are precisely located on the network thanks to telecommunication systems, independently of the track detection devices. CBTC also integrates an Automated Train Operation (ATO) system, that controls speed and acceleration of the train during its driving phases, and an Automated Train Supervision (ATS) system that performs automatic traffic management decisions. As a result, trains are able to automatically adapt their speed depending on the position on the preceding train. Train protection is hence ensured by a moving block signaling system. That allows much shorter headways than the fixed three-aspects signaling system, and reduces the phenomenon of delay propagation described in the previous subsection, resulting in higher capacity.

1.3 Headway-based traffic control

Most railway companies operate trains according to a timetable, even in dense and highly circulated areas. When a disturbance occurs that results in delays, train drivers and traffic managers are required to be-

have in such a way that the trains arrival and departures occur as close as possible to their planned times. However, this is not always the most efficient way to operate commuter trains. Indeed, surveys show that when the train frequency gets sufficiently high, passengers do not pay attention to schedule and present themselves at the station with the purpose of boarding the first train available (Luethi et al. (2007)). In addition, this operation mode does not guarantee a good quality of service to passengers. This is particularly true at points where a good synchronization is required, such as junctions.

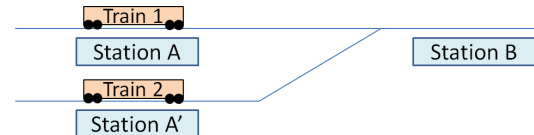


Figure 2 – A junction

	Train 1	Train 2
Station A	8:00	
Station A'		8:05
Station B	8:05	8:10

Table 1 – Drawbacks of schedule-based traffic control

Consider two trains running on the network described in Figure 2 according to the schedule given by Table 1. Suppose that train 1 suffers a 5-minute delay and that train 2 is on schedule. Then a conflict for infrastructure utilization will occur at the junction, and one of the trains will have to stop on the track, causing inconvenience to passengers and increasing delays. It might be better to allow train 2 to depart ahead of schedule if possible, or to hold it at Station A' for a time long enough to prevent the conflict.

In addition, Van Breusegem et al. (1991) proved that uneven time intervals between train can damage the stability of operations if no dispatching decision is made. Indeed, a train arriving at a station a long time after the previous train will have to dwell a longer time, as more passengers will wish to board. The headway between trains will consequently grow, leading to an even higher dwell time at the next station and a higher delay, while the following train is likely to catch up with the delayed one. This phenomenon is well-known of bus passengers and operators, where it is called “bus bunching” (Newell et al. (1964)). Finally, we can note that reducing the headways variance leads to a smaller expected waiting time for passengers (Ding et al. (2001)).

For those reasons, many bus and light rail operators have given up schedule-based control to prefer headway-based traffic management. In this framework, vehicles do not have a pre-planned schedule, but are monitored in real-time in order to maintain even headways between them. Up to our knowledge,

this concept has been little applied to railway services so far ; the implementation of such a framework would be considerably facilitated by the CBTC systems. However, it does not seem feasible to equip a whole suburban line, due to the high cost of CBTC systems. Nor is it practicable to operate a line with branches with headway-based control in totality. Indeed, the capacity limitations of the common section would result in lower frequencies on the branches, where a schedule is needed for commercial reasons. As a result, the option of operating a line with headway-based control coupled with a CBTC system on the dense common section, and schedule-based control with manual driving on the other parts of the line, is considered.

The aim of this paper is to assess the robustness of such a traffic management scheme. For that purpose, we present a simulation method for reproducing railway traffic under headway-based traffic management performed by a CBTC system. We specifically study the robustness of such an operations system, i.e. its capability to resist to small disturbances occurring outside the control area. The remainder of this paper is organized as follows. In section 2, we briefly review some related work. Section 3 is dedicated to the description of the modeling framework and the simulation method we developed. Numerical experiments are presented in section 4, where we apply the method to a saturated line of Paris suburban network. Section 5 concludes the paper.

2 RELATED WORK

The real time railway traffic management problem has been widely studied and many approaches have been proposed for solving it ; a review of these methods is presented in Cacchiani et al. (2014). Among them, some consist in recovery algorithms after minor disturbances. In that case, only rescheduling decisions (such as re-timing arrival and departures events, re-ordering trains at junctions) need to be taken. Other consider the case of recovering a feasible schedule after a major disruption, for which the rolling stock and crew have to be rescheduled as well, but this falls out of the scope of our study. Most of the designed algorithms rely on a mathematical programming formulation of the problem with a no-wait job shop scheduling problem as studied by Mascis et al. (2002). D’Ariano et al. (2007) propose a branch-and-bound algorithm for computing a conflict-free schedule in real time after a perturbation. Pellegrini et al. (2015) consider the case where decisions can be taken in a fixed control area and describe a method for reacting to disturbances occurring outside this area. However, most of the works address the case where train protection is ensured by a fixed block signaling system and trains are operated according to a sched-

ule. D’Ariano et al. (2008) investigate the concept of flexible timetables to improve real-time traffic management, making thus a first step towards headway-based control in railway systems.

Headway-based traffic management, using in particular holding strategies, have been studied for urban transportation systems such as buses (see for example Sun et al. (2008), Daganzo (2009), Berrebi et al. (2017)). The aim is to provide methods for keeping headways as even as possible so that approximately the same amount of passengers board each bus at a given stop. However, the constraints are not the same as for railway systems ; the specific problem of managing junctions is not considered. This issue is addressed by Schanzenbächer et al. (2017), who propose a model based on max-plus algebra to prove that under some hypothesis, the traffic evolves toward an equilibrium, and to compute its characteristics.

The development of CBTC systems has opened a new area of research, concerning management and simulation of such systems. Ning et al. (2015) propose a control method for minimizing headway variance and energy consumption by selecting adapted speed profiles. Pochet et al. (2016) designed a genetic algorithm to reschedule trains when those are sharing infrastructure with other vehicles that are not equipped with a CBTC system. A tool for simulating traffic on a corridor is presented by Chen et al. (2010).

Finally, note that several railway traffic simulation tools already exist. They are used for various purposes, such as ensuring that a timetable is conflict-free, passenger information or robustness assessment of a timetable. The commercial softwares OpenTrack and RailSys rely on a microscopic description of the infrastructure and signaling system, and are notably used for scheduling. However, these softwares require a high computation time and are more adapted to lines operated with according to a schedule. Li et al. (2005) propose a simple cellular automaton model for simulating railway traffic with a moving block signaling system.

3 MODELING RAILWAY OPERATIONS AND HEADWAY BASED CONTROL

3.1 Description of trains and infrastructure

In this preliminary work, we restricted ourselves to the simulation of the area where trains are operated according to a headway-based control scheme. We model the infrastructure at a mesoscopic level, by a network (N, A) . The set of arcs represents homogeneous segments of tracks, linking nodes that are the singular points (stopping points in stations, junctions, points where the speed limit changes). To each arc $a \in A$, we associate a length ℓ_a , a speed limit s_a

and a capacity c_a (meaning the highest number of trains that can run on this segment simultaneously). In addition, following Pellegrini et al. (2015), we allow some arcs to be reserved by a train before this train enters it, guaranteeing that no other train will enter the reserved infrastructure during the reservation. This is used for determining train orders at junctions.

The set of trains to be operated is denoted T . To each train $\theta \in T$ we associate its physical characteristics that are its length ℓ^θ , its acceleration coefficient α^θ and its braking coefficient γ^θ . The service of train θ is described by the following parameters: an itinerary I^θ that is the ordered sequence of arcs taken by the train, a servicing D^θ that is the sequence of the scheduled dwell times at each node of the network (these times being null for the nodes that do not represent a station), and the time at which the train is scheduled to enter the network $start^\theta$. The train evolution is described by: the state of the train (running, dwelling, not operating), the arc a^θ on which the head of the train is located, its position x^θ on the arc, its speed s^θ , its preceding train $prec^\theta$ and its current dwell time $dwell^\theta$.

3.2 Eulerian scheme

During the simulation, the trains evolution is computed using an Eulerian scheme, with a time horizon T_f and a time step dt . The core algorithm can be synthetically described in the following way:

```
t=0 ;
while t < T_f do
  forall trains  $\theta \in T$  do
    Compute safety distance ahead of train  $\theta$  ;
    Update position and speed for train  $\theta$  ;
  end
  Enforce dispatching decisions ;
  t  $\leftarrow$  t + dt ;
end
```

Algorithm 1: Description of the core algorithm

We subsequently present each of the three functions performed by this algorithm. First, as train protection is ensured by a moving block signaling system, it is necessary to compute the distance between a train and the next point where it has to stop. We can then deduce the maximal speed which allows the train to stop before that point. The computation of this safety distance is performed by Algorithm 2. We assumed here the arcs of I^θ to be ordered according to the itinerary of train θ .

The trains' position is updated along the simulation using an Eulerian scheme. At each time iteration t , a new speed instruction is computed for each running

```
if arc  $a^\theta$  contains another train ahead of  $\theta$  then
   $\theta' \leftarrow$  closest train ahead of  $\theta$  ;
   $dist \leftarrow x^{\theta'} - x^\theta - \ell^{\theta'}$  ;
else
   $dist \leftarrow \ell_{a^\theta} - x^\theta$  ;
  forall arc  $a \in I^\theta$  not yet traveled by  $\theta$  do
    if  $a$  contains a train  $\theta'$  then
       $dist \leftarrow dist + \ell_a - x^{\theta'} - \ell^{\theta'}$  ;
      break ;
    else
      if  $a$  is reserved by a train  $\theta' \neq \theta$  or  $a$  is a
      station where  $\theta$  is scheduled to stop then
        break ;
      else
         $dist \leftarrow dist + \ell_a$  ;
      end
    end
  end
end
return dist ;
```

Algorithm 2: Computation of safety distance of train θ

train: that is the speed the train will reach at time $t + dt$ (assuming a constant acceleration between t and $t + dt$). The position is then updated accordingly. Algorithm 3 describes the performed operations. Note that if a train cruises at speed s , according to the law of uniform deceleration, the minimum distance for it to stop is $s^2/2\gamma$. Therefore, if a train has safety distance λ , the maximum speed it can reach while still being able to stop before that point is $\sqrt{2\gamma\lambda}$. However, as speed cannot be adapted instantly, we compute a speed instruction for the next time step; the safety distance should be therefore reduced to $\lambda - sdt$, and the maximum speed ensuring safety $\sqrt{2\gamma(\lambda - sdt)}$. We distinguish between two cases:

- Either the train is able to speed up while still satisfying the safety constraints. Then it accelerates to the highest possible speed according to its acceleration coefficient, maximum speed of the track segment and safety conditions. Note that the maximum speed instruction satisfying the latter is the only solution of equation $s = \sqrt{2\gamma(\lambda - sdt)}$, that is $\sqrt{2\gamma\lambda + \gamma^2(dt)^2} - \gamma dt$.
- Or the train has to slow down, and brakes with deceleration coefficient γ .

3.3 Disturbances

Let us recall that our goal is to provide a simulation method for assessing the robustness of a traffic management system where a transition between schedule-based and headway-based control is operated. Our

method seeks therefore to reproduce the disturbances that are susceptible to occur in such a system. We identified two categories of disturbances:

- Disturbances occurring outside the headway-based control area. We do not delve into the details of such disturbances, and model them by an arrival of trains at the entrance of this area with uneven time intervals. This is done by adding to the nominal arrival times the realizations of independent uniform random variables, with a magnitude to be chosen as a parameter.
- Prolonged dwell times at stations located in the dense area (due e.g. to passengers blocking doors or driver waiting for passengers to board).

We introduced these disturbances as realizations of independent random variables following uniform distributions. This choice is motivated by the lack of available data for estimating the probabilistic distribution of such phenomena ; this is indeed a prospective work and no railway system is currently operated in such a paradigm. We therefore used ad hoc hypothesis coming from operators' experience. However, some refinements are possible: Yuan (2006) established that, under some hypothesis, arrival times of trains follow a log-normal distribution.

3.4 Heuristics for railway traffic management

Last, the simulation method must be able to reproduce traffic management decisions. For the sake of simplicity and performance in terms of computation time, we implemented sub-optimal rule-based heuristics rather than more elaborated methods for solving the real-time traffic management problem. The possible decisions are the following:

- All trains may depart stations as soon as possible (*No Holding policy*), or some trains might be held longer in stations in order to re-adjust the headways between trains in case of disturbances (*Holding policy*). This policy might be particularly efficient in stations located ahead of junctions, thus preventing trains to compete for the use of infrastructure and to be forced to wait for the junction to be free.
- When passenger data is unknown and traffic is homogeneous (all trains stop in the same stations), it seems reasonable to manage junctions according to a *First In First Out (FIFO) policy*. However, when direct and omnibus trains share the same infrastructure, our method proposes a *Threshold policy*: when a direct and an omnibus train are likely to compete for passing the junction, the direct train is given priority if

```

 $\bar{s}^\theta \leftarrow s^\theta ;$ 
 $\lambda \leftarrow \text{SafetyDistance}(\theta) ;$ 
if Train  $\theta$  is running then
  if  $\bar{s}^\theta < \sqrt{2\gamma(\lambda - \bar{s}^\theta dt)}$  then
     $s^\theta \leftarrow \min(\bar{s}^\theta + \alpha^\theta dt, s_{a^\theta}, \sqrt{2\gamma\lambda + \gamma^2(dt)^2} - \gamma dt)$ 
  else
     $s^\theta \leftarrow \max(0, \bar{s}^\theta - \gamma dt) ;$ 
  end
   $x^\theta \leftarrow x^\theta + (\bar{s}^\theta + s^\theta)dt/2 ;$ 
  if  $x^\theta > \ell_{a^\theta}$  then
     $x^\theta \leftarrow x^\theta - \ell_{a^\theta} ;$ 
     $a^\theta \leftarrow a^\theta + 1 ;$ 
  end
  if  $s^\theta == 0$  and  $a^\theta$  is a serviced station then
    Set train  $\theta$ 's state to "Dwelling" ;
     $dwell^\theta \leftarrow dt$ 
  end
else
  if Train  $\theta$  is dwelling then
    if  $dwell^\theta \geq dwellTime(a^\theta)$  then
      Set  $\theta$ 's state to "Running"
    else
       $dwell^\theta \leftarrow dwell^\theta + dt$ 
    end
  else
    if  $t == start^\theta$  then
      if  $\lambda > 0$  then
        Set  $\theta$ 's state to "Running"
      else
         $start^\theta \leftarrow start^\theta + dt$ 
      end
    end
  end
end

```

Algorithm 3: Eulerian scheme for updating train data

the resulting delay for the omnibus train remains below a fixed threshold.

4 NUMERICAL EXPERIMENTS

4.1 Case study: line D of Paris suburban network

We applied our method to assess the feasibility of such a mixed traffic control on one of the most saturated lines in Paris suburban area, namely line D of the RER (*Réseau Express Régional*, the local express network). A map of the dense part of this line is given by Figure 3. We assumed this part to be equipped with CBTC and operated with headway-based control.

Line D crosses Paris for linking Orry-la-Ville (on the north) to Corbeil-Essonnes and Melun (on the south). This line presents several peculiarities. First, it shares

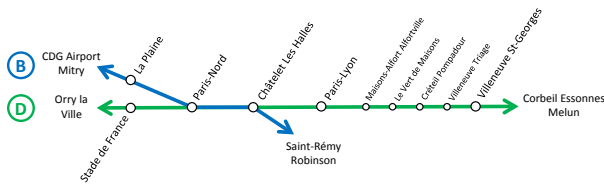


Figure 3 – Line D of Paris RER

its infrastructure with line B of the RER network, another saturated line, between the stations Paris-Nord and Châtelet-les-Halles. For that reason, this common segment is one of Europe's busiest track sections, with a traffic reaching 32 trains per hour during the peak period. Second, in order to increase the offer on the southern part of line D despite the saturation of the common section, some trains coming from the south end their journey at Paris-Lyon station and operate from there a new service towards the south. Two tracks are dedicated for these turnaround services, while two other tracks are used for those running through the station. Figures 4 and 5 present respectively a map of these track sections.

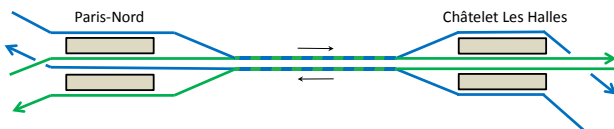


Figure 4 – Common section between Paris-Nord and Châtelet

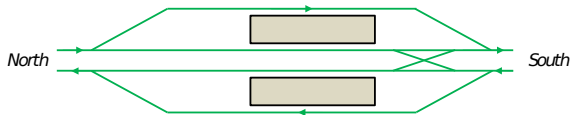


Figure 5 – Paris-Lyon station

The transport offer is organized as follows in peak hours:

- 20 trains per hour on line B in each direction ;
- 12 trains per hour on line D from the north to the south and from the south to the north ;
- 8 trains per hour on line D from the south to Paris-Lyon and from Paris-Lyon to the south.

The nominal time interval between trains in the common section is 90 seconds. Trains of line B are scheduled every three minutes, while trains of line D follow a 6 min/6 min/3 min pattern in the common section. South of Paris-Lyon, the nominal time interval is three minutes. Figure 6 represents on a time-space diagram the nominal trajectories of trains from North to South ; the blue, green and red lines are associated respectively to the trains of line B, the trains of line

D from the north to the south and the trains of line D from Paris-Lyon to the south. Trains of line D and line B are respectively 246 meters and 208 meters long, and we assumed both acceleration and braking coefficients equal to 0.7 m/s^2 .

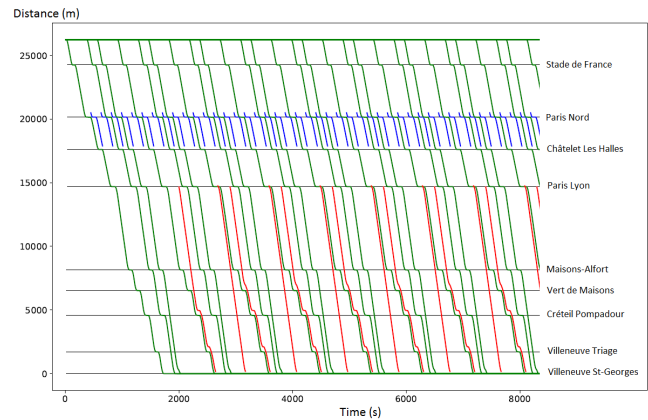


Figure 6 – Nominal trajectories from north to south

4.2 Results

Our aim is to assess the capability of the system to resist to disturbances that may occur upstream of the headway-based control area. Figure 7 illustrates what may happen if trains enter this area without complying by the nominal interval (because of disturbances that were not recovered by traffic management). We can observe on that figure some phenomena of congestion ahead of the junction between lines B and D at the station of Paris-Nord: indeed, some trains are forced to stop ahead of the station. This results in an increase of travel time for passengers. We can also note that some direct trains (red lines) get stuck behind omnibus trains south of Paris-Lyon and are therefore slowed down as well. For example, unlike Figure 6, there is no spare time window for allowing the first red train to insert on the south branch of the line. Consequently, this train has to follow the fifth green train that is a slower one, and gets delayed.

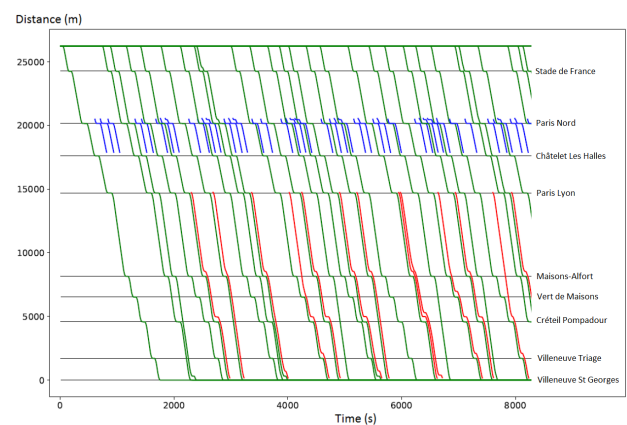


Figure 7 – Disturbed trajectories of trains

In order to quantify how the quality of service de-

creases with the magnitude of upstream disturbances, we define the delay of a train in a given disturbance scenario as the difference between its total running time in this scenario and its running time in the nominal case (with no disturbances).

We ran several simulations with various magnitudes of uniform disturbances, and computed the average delay over all trains that completed their mission, in each scenario. In all scenarios, we assumed the dwell times in stations to follow independent random variables with uniform values in $[40, 70]$ at stations inside Paris and $[20, 50]$ at stations outside Paris (the nominal dwell times being respectively 50 and 30 seconds). The simulation horizon is 3 hours (the duration of the peak period), and the time step is 5 seconds (chosen as a compromise between computation time and accuracy of the results).

For each level of disturbance and each policy, we performed 100 replications. The Holding policy is applied in the stations located upstream of junctions only, and according to the following rule (we reuse the notations of Figure 2). We denote by Δ_{AB} (respectively $\Delta_{A'B}$) the nominal travel time between station A (respectively station A') and station B , and I the nominal time interval between both trains. If train 1 leaves station A at time dep_A , train 2 will not be allowed to leave station A' before the time $dep_{A'} = dep_A + (\Delta_{AB} - \Delta_{A'B}) + I$, thus ensuring that train 2 will not arrive at station B with a shorter interval than the nominal one. Results are plotted on Figure 8.

The junctions of Châtelet les Halles and Paris-Nord are managed according to a FIFO policy while the junction of Paris-Lyon follows a Threshold policy (priority given to the fastest trains up to a 30s consequent delay for the other train). Indeed, both direct and omnibus trains use the infrastructure south of Paris-Lyon while all trains are omnibus on the northern part of the line.

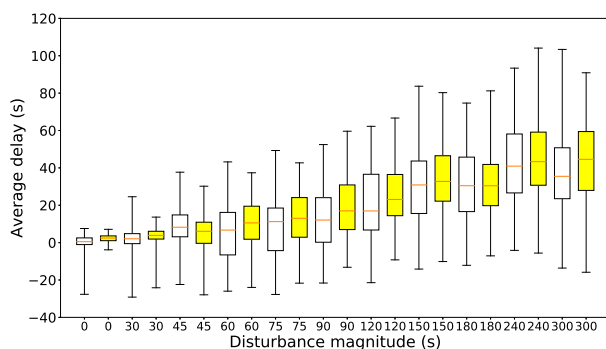


Figure 8 – Delays due to disturbances - No Holding Policy (white) and Holding Policy (yellow)

The distributions of delays using respectively the Holding and the No Holding policies are plotted on

figure 8. We note that for both policies, delays remain limited both in average and extreme values for small levels of disturbances (up to approximately 45 seconds), but grow and become more unpredictable for higher levels. However, delays grow at a non-exponential rate: this is mostly due to the moving block signaling system that is known to alleviate the risk of delay propagation from one train to the next, compared to a fixed signaling system.

In addition, up to a certain point the Holding Policy allows to reduce extreme values of delays, but in most cases this comes at the price of a higher travel time in average. For greater levels of disturbances the Holding Policy seems to become unable to cope even with the extreme values. This work consequently does not appear conclusive on the relevancy of implementing such a policy and further studies seem to be needed.

This study also underscores the effects of stochasticity on the quality of operations in the dense area. A specific attention should therefore be paid to the schedule adherence in the upstream low traffic areas, where the system is less sensitive to disturbances, as small delays in these parts of the network can lead to instability in the common section. In case of delays occurring in the branches, traffic management decisions such as canceling a whole train or part of its service could be made to prevent propagation to the dense area, but this possibility was not investigated in this work.

5 CONCLUSION

Our approach demonstrates the practical applicability of such a mixed traffic management for a railway line crossing a dense area, the system being able to resist to disturbances of small magnitude. It also highlights the benefits and drawbacks of a Holding policy to adjust headways.

Yet, some questions remain to be answered. First, the obtained results emphasize the need of an efficient traffic management ; improved control strategies should therefore be tested. In particular, the possibility of adjusting headways by controlling the trains speed between stations in addition to their dwell times could be explored. Second, random disturbances upstream of the dense area were assumed to be independent and uniformly distributed, which in practice is usually not the case: our results might benefit to be confirmed by a refined model. Last, no studies have been made so far on how to build schedules and manage traffic in real time in the upstream and downstream areas to facilitate the transition between schedule-based and headway-based control. All those topics could be matters for further research.

REFERENCES

- Berrebi S., Hans E., Chiabaut N., Laval J., Leclercq L., Watkins K., 2017. Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C*, to appear.
- Cacchiani V., Huisman D., Kidd M., Kroon L., Toth P., Veelenturf L., Wagenaar J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. *Transportation Research Part B*, 63, p. 15 - 37.
- Chen R., Guo J., 2010. Development of the New CBTC System Simulation and Performance Analysis. *WIT Transactions on the Built Environment*, 114, p. 497 - 507.
- D'Ariano A., Pacciarelli D., Pranzo M., 2007. A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183 (2), p. 643 - 657.
- D'Ariano A., Pacciarelli D., Pranzo M., 2008. Assessment of flexible timetables in real-time traffic management of a railway bottleneck. *Transportation Research Part C*, 16 (2), p. 232 - 245.
- Daganzo C., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B*, 43, p. 913 - 921.
- Ding Y., Chien S., 2001. Improving Transit Service Quality and Headway Regularity with Real-Time Control. *Transportation Research Record: Journal of the Transportation Research Board*, 1760, p. 161 - 170.
- Li K., Gao Z., Ning B., 2005. Cellular automaton model for railway traffic. *Journal of Computational Physics*, 209, p. 179 - 192.
- Luethi M., Weidmann U., Nash A., 2007. Passenger arrival rates at public transport stations. *86th Transportation Research Board Annual Meeting 2007. Institute for Transport Planning and Systems, ETH Zurich*.
- Mascis A., Pacciarelli D., 2002. Job-shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research*, 143 (3), p. 498 - 517.
- Newell G.F., Potts R.B., 1964. Maintaining a bus schedule. *Proceedings of the 2nd Australian Road Research Board*, vol. 2, p. 388 - 393.
- Ning B., Xun J., Gao S., Zhang L., 2014. An Integrated Control Model for Headway Regulation and Energy Saving in Urban Rail Transit. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), p. 1469 - 1478.
- Pellegrini P., Marlière G., Pesenti R., Rodriguez J., 2015. RECIFE-MILP: An effective MILP-based heuristic for the real-time railway traffic management problem. *IEEE Transactions on Intelligent Transportation Systems*, 16 (5), p. 2609 - 2619.
- Pochet J., Baro S. and Sandou G., 2016. Supervision and rescheduling of a mixed CBTC traffic on a suburban railway line. *IEEE International Conference on Intelligent Rail Transportation (ICIRT 2016)*, Birmingham, United Kingdom, p. 32 - 38.
- Schanzenbächer F., Farhi N., Christoforou Z., Leurent F., Gabriel G., 2017. A discrete event traffic model explaining the traffic phases of the train dynamics in a metro line system with a junction. *Proceedings of 2017 IEEE 56th Conference on Decision and Control (CDC)*, Melbourne, Australia, p. 6283 - 6288.
- Sun A., Hickman M., 2008. The Holding Problem at Multiple Holding Stations. *Computer-aided Systems in Public Transport. Lecture Notes in Economics and Mathematical Systems*, 600, Springer, Berlin, Heidelberg.
- Van Breusegem V., Campion G., Bastin G., 1991. Traffic modeling and state feedback control for metro lines. *IEEE Transactions on automatic control*, 36(7), p. 770 - 784.
- Yuan J., 2006. Stochastic modelling of train delays and delay propagation in stations. *PhD thesis, Delft University of Technology*.
- IEEE Standard for Communications-Based Train Control (CBTC) Performance and Functional Requirements, IEEE Standard 1474.1, 2004.

SIMULATION OF VALUE FOR ENVIRONMENTALLY ADAPTABLE PACKAGING SOLUTIONS ARCHITECTURE

CARLO LEARDI

Tetra Pak Packaging Solutions
carlo.leardi@tetrapak.com

ABSTRACT: *Environmental awareness increase is characterizing the last decades in the liquid food packaging environment. All the stakeholders from raw materials producers, to consumers and recyclers address new and more pointy needs regarding environmental impact. The technical translation for this need is a significant reduction in CO₂e footprint. A paradigmatic shift from fossil to renewable, e.g. sugar cane, based raw materials, is the strategic direction identified by the packaging solutions developers. It is anyhow necessary to identifying a holistic, global and sustainable compromise among technical, sociological and geographical challenges. An adaptable and upgradable architecture for packaging solutions involving the opportunity to progressively increase, with minimal or null impact on the value flow, the percentage of renewable materials constitute the solution. The ENPV, Expected Net Present Value represents the KPI tracked during the overall packaging solution development and the usage phase. A combination of Bayesian inference and Montecarlo simulation leveraging on the adaptable architecture and its operative scenario items allows to estimate the ENPV and its uncertainty at each decision point. The support to the program's decisions is so furnished to the System Engineering organization. An industrial implementation of the simulation approach is exemplified and discussed.*

KEYWORDS: *Systems architectures, Adaptability, Value, ENPV, simulation*

1 INTRODUCTION

In liquid food packaging industry, environmental impact awareness is rapidly increasing as an effect of diaries, consumer's organizations and consumers compulsive needs.

The systems engineers translated this generic need into a requirement targeted to a significant CO₂e carbon dioxide equivalent footprint reduction target.

The strategy to fulfil this system requirement is a paradigmatic shift from fossil to renewable raw materials, alias natural resources that grow back, such as wood and sugar cane.

The final target, a fully renewable package, is the result of a technical compromise among all the technical and functional requirements of the packaging solution along the overall system life-cycle.

The main challenges regard the regional availability of renewable raw materials, like as the sugar cane waste called "bagasse" in Brazil, a globally sustainable environmental effort, the technical and industrial enablers implementation, the packaging solutions cost targets satisfaction and a return on investment affordable by the diaries as an effect of consumer's tangible appreciation.

The measures are the ENPV, the Expected Net Present Value and a shift of market shares from fossil-based to renewable raw materials based packaging solutions. A

mid-term and world-wide vision has so to be taken into consideration.

The more suitable technical solution is an adaptable package architecture which allows gradual replacements of renewable raw materials based components, like as the caps or the protective and decorative plastic layers that constitute the semi-finished packaging material. This shift is implemented by sustainable architecture upgrades as soon as intermediate market shares targets are reached.

This article addresses a generic Montecarlo simulation framework that allows the monetary value estimation in mid-term and within a wide geographical and configuration framework aiming at supporting the decision about the optimal degree of adaptability required to the packaging solution.

One of the addressed aspects regards the necessity to make each upgrade as transparent as possible to the diaries, e.g. mitigating the impact on industrial costs and on consolidated process best-practices. The final consumer shall so find an increasing offer of packages satisfying its environmentally friendly need on the shelf, at a price aligned with the market state-of-the-art without severe changes challenges required to the producing and the retailing structure.

This category of simulations, began many years ago, is continuously evolving by including more technical, economic and market shares elements. The accuracy of the predictions has been confirmed by the first short terms results and shall be furtherly validated during next steps.

2 ENVIRONMENTALLY FOCUSED NEEDS

The trend of more and more complex Systems of Systems is addressed by Incose in its 2025 vision: *“The vision for systems engineering in 2025 is shaped by the global environment, human and societal need, policy and business challenges, as well as the technologies that underlie systems. The evolving work environment, following global trends, both constraints and enables the way systems engineering is practiced.”*

is also focusing on systems which can significantly reduce the already overwhelming environmental impact with a global view on the overall socio-economical scenario.

In the liquid food packaging industry new, appealing, affordable and performing packages are foreseen. One of the key of success is finding a compromise among an improved environmental profile, global sustainability and enhanced functionalities.

It is possible to quantify by a unique measure the environmental impact as the carbon footprint, CO₂e, the equivalent CO₂ resumed value.

Other measures regard the amount of materials used for the packages, the recyclability of the used packages into high value products and the energy involved by each phase of the overall life-cycle.

This article will focus on the materials renewability, without forgetting the other environmental aspects by means of specific functional requirements fulfilment.

The current focus of packaging materials developers regards the shift from fossil, oil in this case, to renewable, e.g. sugar-cane, raw materials.

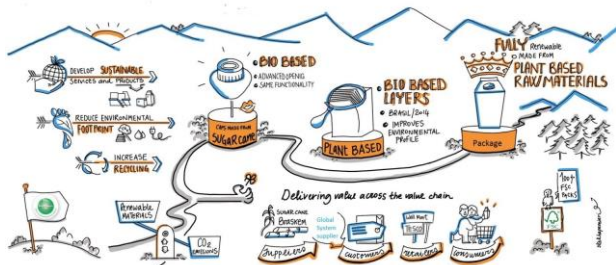


Figure 1: Environmental aspects in Liquid Food Packaging industry, a naïve representation

Such a view includes the overall stakeholders and life-cycle needs as depicted in figure n.1. The raw materials producers, the converting plants which produces the packaging material ready for the diaries to be filled and sealed. A peculiar role is played by retailers which makes available the packages on the shelf to the consumers and often state the toughest quality requirements. All these stakeholders manage and enable the “push” effect to the market.

The consumers and their association realize the pull effect that complement and make effective the success of the new environmental friendly packaging solutions allowing a positive ENPV return for the entire value chain.

3 PACKAGING SOLUTIONS ARCHITECTURAL INSIGHTS

A packaging solution realized by only renewable and recyclable raw materials, see picture 2, is the final target intended to satisfy of all the stakeholders.



Figure 2: Package renewable elements

An optimal equilibrium among the Co₂e footprint reduction, a holistic environmental sustainability and the expected service provided by the packaging solution has anyhow to be envisioned.

CO₂e, or carbon dioxide equivalent, is a standard unit for measuring carbon footprints. It expresses the impact of each different greenhouse gas in terms of the amount of carbon dioxide that would create the same amount of warming. That way, a carbon footprint consisting of lots of different greenhouse gases can be resumed and estimated as a single number.

This strategic option faces although severe technical, geographical and sociological challenges like as:

- Renewable raw materials availability
- Local and Global sustainability for the overall renewable materials life-time, from growth to recycling.
- Legacy, actual and future performance requirements.
- National and international legislation
- Sustainable industrial cost impact.

Replacing fossil-derived plastic with natural, renewable raw materials like as sugar cane residual waste named in brazil “bagasse” depicted in figure 3:



Figure 3: Sugar cane transformation residual pulp, the “bagasse”

is one of the technical opportunities to sensibly increase the packaging solutions renewability percentage. Bagasse is the fiber residue after the sugarcane has been crushed to extract the juice in the sugar mill. The fiber can be burnt to produce heat and energy or used as an organic fertilizer for further growth. It is however possi-

ble to extract ethanol and converting it, by dehydration, into ethylene, the precursor for polyethylene.

It is abundantly available in Brazil but also in India and China. Its growth does not threaten in any way the local environmental profile. The extensions, climate and the industrial best practices in that continent allow a substantial abundant availability of this fully renewable and recyclable non-fossil raw material.

The bagasse can so be transformed into polyethylene with properties satisfying the packaging industry requirements.

It allows, at the polyethylene production site, a 75% CO₂e reduction in comparison with the fossil based polyethylene.

The transformation of these raw materials is localized in Brazil and the transportation into other continents involves a partial disruption of the improved environmental profile.

In general, the availability of alternative raw materials, like as soy or grass, is not sufficient to satisfy the overall environmental need. The vegetable productions should be well integrated into the local productions. They should not replace or eliminate other traditional fundamental productions and the legacy agricultural processes shall be respected so to avoiding un-expected negative impacts.

The recycling processes of the packages must be implemented flexible enough to efficiently sustaining growing percentages of new materials mixed with fossil ones. In some cases, separation processes could be required to avoid loss of value related to mixing recyclable materials with different intrinsic value concepts.

Liquid food industrial examples are the renewable protective and decorative layers, the caps, reported in figure 2; their collection and further recycling processes.

Liquid food solutions are always challenged by increasingly pointed technical requirements while the actual requirements are considered “as given” and fundamental to maintain the brand’s fidelity. The replacement of fossil with recyclable materials involves the satisfaction of all the overall specification of technical and not technical requirements.

Legacy systems constitute, due to the relevant investment and the refurbishment opportunities, a relevant part of the installed equipment and transformation processes baseline. The new packaging solutions must be handled without impact by the overall technical baseline, including the legacy systems.

The liquid food packaging world is driven by international regulatory organizations like as FDA or EU regulations. Their aim is to protect all of us as consumer. The new architectural choices for the packaging solutions cannot overcome anyone of the reference standards and test results.

National legislation plays also a role in defining the industrial background constraints for each installation during the overall system life-cycle.

Finally, the industrial cost of this environmental shift paradigm must be sustainable to all the stakeholders. The raw materials producers shall need a progressive increase

of production, the transforming plants shall adequately their procurement, transformation and transfer processes without losses of efficiency. The dairies and the retailers have the same key performance indexes of the transforming plants but at the same time they must maximise the Return on Investment of the innovation effort.

Finally, every person involved, starting from the packaging solutions architectural system engineers ending the packaging machine operator should experience a positive impact.

A global environmental improvement shall leave, by necessity, a positive impact for all the involved stakeholders, including every one of us intended as consumers and inhabitants of this unique but finite earth.

A totally adaptable architecture, e.g. every one of the packages elements can be replaced by a renewable one, will lead to an architectural solution not sustainable from procurement, development effort and industrial cost and finally to an un-satisfactory Return on Investment.

A too rigid or geographically focused architecture implementation shall lead to a limited fulfilment of the global need with threatening challenges to furtherly increase up-to the desired levels the market shares shifts.

In practice, the proposed simulation approach allows to determine the which one of the fossil raw materials based elements can be replaced by renewable ones, in which sequence and at which intermediate steps.

The main question is so “which elements, where and by which rate to be replaced by renewable raw materials?” “Which ones are the KPIs and the TPMs to be tracked during this transformation process?”

4 ROLE OF SIMULATIONS FOR AN ENVIRONMENTAL SHIFT PARADIGM

The systems engineering development processes, e.g. the VEE process proposed in figure 4, involve the definition of several decision points covering the system life-cycle and, in particular way, the development one.

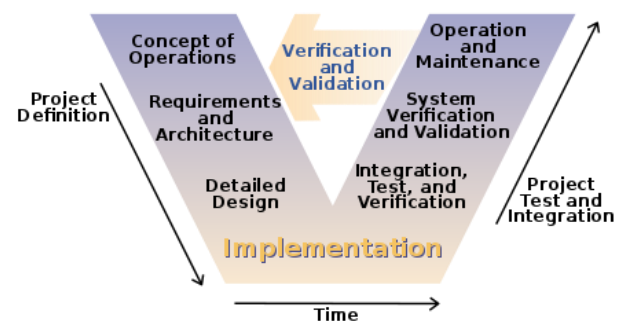


Figure 4: Generic VEE model, including operation and maintenance

It is essential to define, just after the idea definition, the validation strategy and its applicative plan.

In other words, which ones shall be the Key and the Technical Performance Indexes, KPIs, and TPMs which shall be validated when the packaging solution shall be available on the market. Each one of these aspects are

tracked and estimated at each decision point answering to the following questions:

“Is the packaging solution technically good enough?”
“Shall the packaging solution be still profitable during its usage phase?”

“Are the advancements on line with the program schedule?”

The proposed simulation considers the following aspects: Key Performance Indexes addressing value: ENPV, technological evolution assessments, market composition and evolution, architecture adaptability and upgradability.

The simulations, in this industrial case a combination of Bayesian inference and montecarlo runs, play a fundamental role until the system is verified and ready to be validated on the market.

A Model Based System Engineering paradigm is applied. At each one of the development steps the simulation furnishes one of the relevant supports for the project advancements decision.

5 AN INDUSTRIAL APPLICATION CASE OF SIMULATIONS.

This article proposes the design and the implementation of one combination of montecarlo simulation runs combined with Bayesian inference.

At each one of the development steps the likelihood about each one of the technical and market components are inserted and the prior models are updated into posterior ones, modified into their location and dispersion parameters or rejected due to fundamental behavior changes. A montecarlo simulation is then run so to re-estimate the ENPV value of the packaging solution.

The simulation estimates cover a mid-long period of the usage phase including a wide as possible scope.

The model is based on the package architecture proposed in figure 5:



Figure 5: Packaging material structure and renewable materials

For each one of the package components the renewability opportunities of materials are applied. Different percentages of fossil plastic are estimated.

Six sources of information and uncertainty, reported into figure 6 are concurrently simulated:

- Raw materials availability: actual, future and cost trends
- Technical Performance Measures: safety, quality and functionality areas
- Market shares: fossil and potential replacement opportunities available in the market.
- Industrial costs and final prices strategies, including advertising and brand’s environmental attitude.
- Customers behavior: shifting from fossil to renewable packages according to loyalty and environmental awareness.
- Regulatory, standardization and other external factors influences.

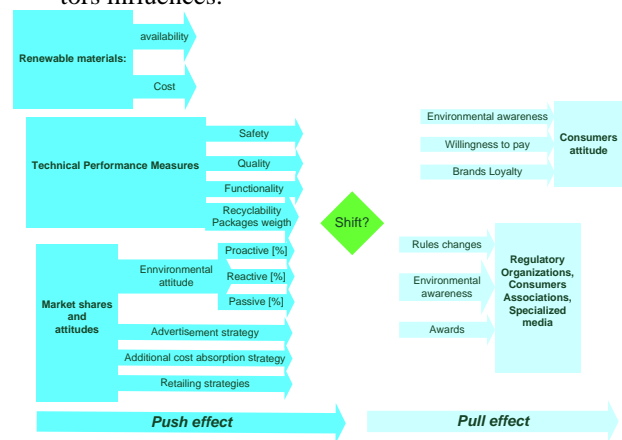


Figure 6: Push and pull effects and their contributions

The push effect, the one that enables the greatest possible opportunities to be caught, is represented and simulated by the following sources of information.

The Raw materials availability is one key enabler for the return of value. Un-sufficient production shift shall not deliver enough value limiting the success to local trials or to geographical regions with a positive attitude versus environmental sustainability without fundamentally impacting the overall market trend.

A relevant lack of availability could block the increasing shift trends and hamper the overall strategy for limited or mid periods.

Exceeding capabilities shall however return un-sustained industrial and retailing costs. E.g. high value un-sold lots of material and consequent serious threats regarding the next shift opportunities.

Only a carefully balanced introduction of renewable packages shall allow the initially increasing and then the maintain return of value at the expected values.

The simulations involve the use of two or three parameters exponentially based prior distributions. These parameters cover the expected quantities of raw materials available by the location parameter, the hopes to cover higher requests by the shape parameter and the minimum availability by the cut parameters. Lambda, Weibull, low or high extreme value focused distributions characterize

the development phases, while triangular or uniform distributions are more suitable to describe the initial assumptions.

At each technical step during development the packaging solution Technical Performance Measures are simulated. The functional requirements tested are the one that characterize every package: safety, preservation of the nutritional and organoleptic properties during the overall shell-life, appearance and opening functionalities.

The used distribution is the normal and, for the exponential behavior, the log-normal one. The likelihood estimations are the modeled results of the Verification process evidences. The expected result is that no deviations have to be experienced by the producer, the retailer, the consumers and the recycling stakeholders as an effect of the shift from fossil to renewable raw materials introduction. All the technical parameters have to be equally satisfied. Eventual minor deviations should be modeled and inserted into the overall simulation. Eventual improvements are not considered as a sell argument because the only advertised argument will be the environmentally improved profile.

The actual market shares are the base for the simulation. The initial diffusion of the renewable packaging solutions is based initially and on short terms on replacement of fossil raw materials based packages with renewable ones. Only in a mid-long term, the overall shares are expected to be increased by replacing other typology of packages.

The brands producers are characterized according their innovation level as: “proactive”, the ones that introduce and gain the maximum advantage from the introduction of the innovation. The “reactive” ones introduce the innovation as pressure effect of more innovative or environmental prone brands. They are more gradual in the introduction and typically playing a defensive strategy. Finally, the “passive” brands, sometimes owning large market shares, need long terms to adapt to strongly mutated situations. They will not adopt the environmental innovation if not forced by severe risks of significant market shares losses. In any case their answer will be more on the price level and on the tradition consolidation.

Each one of the brands is associated with different percentages of renewable packages replaced and an advertising strategy according to the respective classes.

The price policies complete the framework in the hands of the packages producers for their brands. The producer can maintain the same price level of the fossil package or re-charge partially or totally the industrial cost increase.

New competitors entrance and brands absorption are simulated by means of Poisson distributions that modify once a year the market shares composition.

The raw materials industrial costs, including initial and sustain investment, procurement, transformation processes and best practices impact are carefully simulated. Raw materials costs variations are simulated by the triangular distribution. Making them available to each industrial process are better described by stable increases of industrial costs, e.g. by exponential distributions characterized by shape parameters around or below one. This

means that an initial uncertainty about industrial cost increase is compensated by a stabilization effect as soon as the packaging solutions gain a solid market share and enter by right into the consolidated industrial best practices.

The pull effect is furnished by the consumer behavior. Each consumer is characterized, according to the availability and the final prices policies defined by the previous steps and the personal environmental attitude, by a probability to buy a renewable package instead than a fossil based one. The brand fidelity plays a relevant role as well as the advertisement launch and sustainment strategies. Market research are performed and periodically updated to estimate the change probabilities nodes as well as to allocating consumers to different categories.

Some of the customers are available to spend more than usual for the same product with a declared and certified improved environmental impact. A larger consumers population accept the change favorably only at the same price level and a limited part is indifferent and selects randomly when the choice is proposed on the shelf.

Consumers choices can also be impacted by consumers organizations and social networks. Fake news can also influence locally the launch campaign.

The role of regulatory organizations and of sector media is modeled. A sudden restriction on the usage of one raw material can severely impact the overall scenario. An advertised award for an environmentally friendly package can move the percentages of customers prone to change at similar of increased price.

The recycling companies can otherwise influence the market trends by impacting directly or indirectly one of the previous parameters. E.g. the increased industrial cost of separation of the renewable packages is charged to the retailers and the final price is increased otherwise the previous industrial costs are maintained at the same level.

These impacts are simulated by conditional binomial probabilities by means of Bayesian inference.

E.g. one customer loyal to his favorite brand and tendentially prone to shift, is posed in front of the choice among his usual product packaged in a fully renewable package and the fossil one in the traditional appearance. The advertisement plays its role and the innovative package is well visible in front of his eyes, while the fossil one, although still available, is on a lower shelf level.

The probabilities of a shift are very high and the pull effect is strong bringing to a further reduction of the fossil availability in that retailing structure.

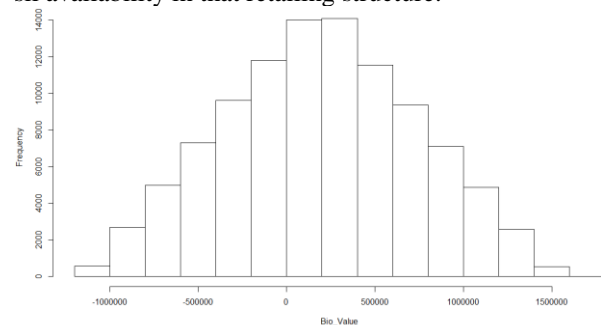


Figure 7: Histogram of one simulated scenario ENPV

It should be considered that many of the simulated sources of information are not independent. The more relevant correlation factors, like as the cost and environmental impact increase related to the energy increase due to distance from the raw materials and the distribution locations, are simulated according to the software capabilities and the complexity of the simulations.

6 SIMULATIONS CHARACTERISTICS

As stated by the previous chapter a great number of variables are settled to their prior state, updated according eventual changes to posterior evidences, correlated and the returned ENPV is furnished for different architectures.

Special attention has been given to tailoring the simulation for the application area.

One Montecarlo simulation framework would have driven to not representative assumptions. Time updates play a relevant role. Quality parameters are continuously addressed by normal and log-normal distributions.

The number and the importance of the players on the market are updated on yearly basis.

The basic simulation node is the decision one, implemented by Bayesian conjugates, where the push effect proposed by the producers and retailers to the market encounters the pull one as an effect of each single consumer's choice.

The probability: P(Choose Bio vs. Fossil package/environmental awareness, availability to pay, brand loyalty)

At least 40 mio montecarlo simulations are run for each architecture and the results are graphically summarized for decision making.

The ENPV estimations are expressed according to the distributions variability and the statistical parameters uncertainty (figure 7 and 8).

All the simulations are run on r codes by using the MC2D package, allowing parameters variability and uncertainty management and advanced sensitivity analysis.

```
> tmp$Bio_Value
      mean      sd 99.5%    50% 99.9%  nsv Na's
median 194784.3 542101.2 1386478 188552.2 1465187 1001 0
mean   191577.6 542950.6 1388749 190220.9 1467424 1001 0
2.5%   155864.8 523246.9 1331539 152402.9 1403373 1001 0
97.5%  227641.8 563938.8 1450051 230836.9 1555705 1001 0
attr(,"type")
[1] "vu"
```

Figure 8: Example of ENPV results including distributions variability and parameters uncertainty.

A simplified tool version is then proposed on excel base using @Risk@ tool. The compared results are shown in figure n. 9.

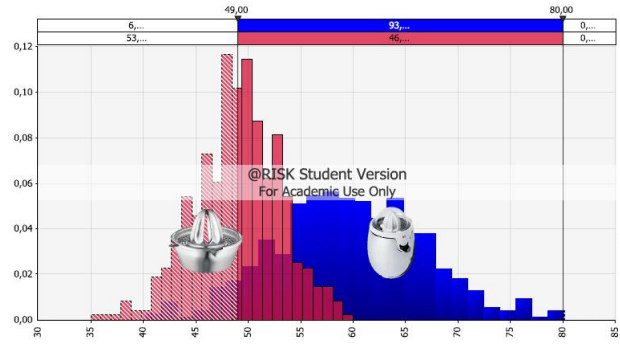


Figure 9: Exemplified version of @risk simulated scenarios comparison

7 RESULTS AND FURTHER EVOLUTION

The proposed simulations allowed supporting the development of the first packaging solutions and the short-term returns are aligned with the simulation results.

It is so possible to estimating, by a combination of Bayesian inference and MonteCarlo simulations, the ENPV monetary expression of value related to environmental impact reduction of an adaptable and upgradable liquid food packaging solutions architecture as deployed in a midterm and wide scope exploitation case.

The simulation framework requires at the same a simplification effort to be manageable and further returns from the market to identify and validate the relevant parameters.

ACKNOWLEDGMENTS

This simulation work was originated as one of the deliverables of the European Funded project AMISA and from the cooperation with the Technical University of Munich for the development of the VASA methodology.

REFERENCES

- Bartolomei, J.E. et al., 2006. Screening for Real Options "In" an Engineering System: A Step Towards Flexible System Development--PART I: The Use of Design Matrices to Create an End-to-End Representation of a Complex Socio-Technical System. In *INCOSE Conference on System Engineering Research, Los Angeles, CA*.
- Browning, T.R., 2003. On customer value and improvement in product development processes. *Systems Engineering*, 6(1), pp.49–61. Available at: <http://doi.wiley.com/10.1002/sys.10034>
- Browning, T.R. & Honour, E.C., 2008. Measuring the life-cycle value of enduring systems. *Systems Engineering*, 11(3), pp.187–202. Available at: <http://dx.doi.org/10.1002/sys.20094>.
- Cardin, M.-A. et al., 2007. Extracting value from uncertainty: A methodology for engineering systems design. In *17th Symposium of the International Council on Systems Engineering, San Diego, CA, United States*. Citeseer.
- Engel, A. & Browning, T.R., 2008. Designing systems for adaptability by means of architecture options. *Sys-*

- tems Engineering*, 11(2), pp.125–146. Available at: <http://dx.doi.org/10.1002/sys.20090>.
- Fitzgerald, M.E., 2010. *MANAGING UNCERTAINTY IN SYSTEMS WITH A VALUATION APPROACH FOR STRATEGIC CHANGEABILITY*.
- Fletcher, D., Brennan, R.W. & Gu, P., 2009. A Method for Quantifying Adaptability in Engineering Design. *Concurrent Engineering*, 17(4), pp.279–289.
- Fricke, E. & Schulz, A.P., 2005. Design for changeability (DfC): Principles to enable changes in systems throughout their entire lifecycle. *Systems Engineering*, 8(4), pp.342–359.
- Gu, P., Xue, D. & Nee, A., 2009. Adaptable design: concepts, methods, and applications. ... *of the Institution of Mechanical Engineers ...*, 223(11), pp.1367–1387. Available at: <http://pib.sagepub.com/content/223/11/1367.short> [Accessed November 10, 2013].
- Hashemian, Mehdi, 2005. *Design for adaptability*. Saskatchewan, Canada. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.134.2831> [Accessed October 11, 2013].
- Hashemian, M., 2005. *Design for adaptability*. Saskatchewan, Canada.
- Kissel, M. & Schrieverhoff, P., 2012. Design for Adaptability – Identifying Potential for Improvement on an Architecture Basis.
- McManus, H. & Hastings, D., 2006. A framework for understanding uncertainty and its mitigation and exploitation in complex systems. *IEEE Engineering Management Review*, pp.1–19. Available at: <http://mit.sustech.edu/NR/rdonlyres/Engineering-Systems-Division/ESD-85JFall-2005/695980D5-FDEE-4C4D-A505-14B8E9C95841/0/uncertainty.pdf> [Accessed November 7, 2013].
- De Neufville, R. et al., 2009. Identifying real options to improve the design of engineering systems H. B. Nembhard & M. Aktan, eds. *Real Options in Engineering Design Operations and Management*, 02138(July), pp.1–37.
- De Neufville, R., 2003. Real options: dealing with uncertainty in systems planning and design. *Integrated Assessment*, 4(1), pp.26–34.
- De Neufville, R. & Scholtes, S., 2011. *Flexibility in Engineering Design (Engineering Systems)*, The MIT Press. Available at: <http://www.amazon.com/Flexibility-Engineering-Design-Systems/dp/0262016230> [Accessed November 24, 2012].
- Orloff, M.A., 2006. *Grundlagen der klassischen TRIZ: Ein praktisches Lehrbuch des erfinderischen Denkens für Ingenieure (VDI-Buch) (German Edition)*, Springer. Available at: <http://www.amazon.com/Grundlagen-klassischen-TRIZ-praktisches-erfinderischen/dp/3540340580> [Accessed February 28, 2013].
- Ross, A.M., Rhodes, D.H. & Hastings, D.E., 2008. Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value. *Systems Engineering*, 11(3), pp.246–262.
- Schulz, A.P. et al., 2000. Development and integration of winning technologies as key to competitive advantage. *Systems Engineering*, 3(4), pp.180–211.
- Suh, E.S., De Weck, O. & Chang, D., 2007. Flexible product platforms: framework and case study. *Research in Engineering Design*, 18(2), pp.67–89. Available at: <http://dx.doi.org/10.1007/s00163-007-0032-z>.
- Wang, T., 2005. Real options“ in” projects and systems design: identification of options and solutions for path dependency.
- Wang, T. & de Neufville, R., 2006. Identification of real options “in” projects. In *4th Conference on Systems Engineering Research, Los Angeles, CA*.
- De Weck, O.L., Roos, D. & Magee, C.L., 2011. *Engineering systems: Meeting human needs in a complex technological world*, The MIT Press.
- Woodruff, R., 1997. Customer value: The next source for competitive advantage. *Journal of the Academy of Marketing Science*, 25(2), pp.139–153. Available at: <http://dx.doi.org/10.1007/BF02894350>.

SUSTAINABLE DEVELOPMENT IN SERIOUS GAMES: RETHINKING GAME-BASED LEARNING STRATEGIES FOR MASTER'S DEGREE ENGINEERS

M. STANITSAS¹, E. VAREILLES², K. KIRYTOPOULOS³, M. ALDANONDO²

¹National Technical University of Athens - School of Mechanical Engineering - Greece
mariossta@central.ntua.gr

²IMT Mines Albi - Industrial Engineering Center - France
elise.vareilles@mines-albi.fr,
michel.aldanondo@mines-albi.fr

³University of South Australia - School of Natural and Built Environments - Australia
konstantinos.kirytopoulos@unisa.edu.au

ABSTRACT: *Sustainable development and sustainability assessment is one of the most important challenges of our time. On the other hand, game-based learning (GBL) is proven to increase soft skills, such as critical thinking, creative problem solving and teamwork, as well as to improve cognitive development, learning retention and social learning, which are of vital meaning for future sustainable aware researchers and professionals. Additionally, the literature review indicates that there is a complete lack of studies in integrating social and environmental sustainability into serious games. The purpose of the study is to integrate sustainability issues into educational project management using an existing serious game as educational tool and thus to create a new sustainable informed basis for serious games in master's degree engineering courses. The proposed scenario indicates that the focus is on both the deliverable/product of the project, such as the project specifications, requirements, deliverable, quality criteria and materials used as well as the delivery/process of the project, such as management planning and project sequencing. Through a case study all the game characteristics that can evolve in order to include sustainability concepts are presented.*

KEYWORDS: *Sustainability, Serious Games, Project Management, Game-Based Learning.*

1 INTRODUCTION

Experimenting and learning are increasingly considered as means to innovate governance approaches for pursuing a more sustainable society. This paper puts forward the case for the use of serious games to help support sustainable growth through the use of games for conventional skills and knowledge development related to project management. It presents all the elements of sustainability that can be applied to a serious game (SG) oriented course for master degree engineering (MEng) students, as well as a SG framework for industrial engineering and project management, which implements an extensible framework for different organizations to develop sustainability-oriented games. Moreover, it analyzes a content based on sustainability issues under the scope of game mechanics. This paper provides an overview of game-based learning and the state of SGs for sustainable management, offering insight into their potential as effective tools in facilitating project management education. Serious games have been shown to possess numerous qualities that have been connected with improved learning experiences and cognitive development (Madani et al., 2017).

The aim of the paper is to propose a SG environment that will enable students to become aware of sustainability topics. Our proposal can be thought as an application that introduces sustainable development (SD) principles

in an engaging way, as a precursor to designing with professional parametric design and environmental simulation processes. This new concept bears the name “Aeiforia¹ Serious Game” or “AmuSInG” and it introduces values concerning planning, management and SD.

The objectives of the study reported in this paper are to: (a) Identify all the elements of a SG that can sustainably evolved, (b) design and propose a SG prototype that aims to improve students' skills as engineers and leads them towards a sustainable project management philosophy.

The paper is organized as follows. Section 2 describes the sustainable way of thinking and SD. In Section 3, some selected definitions of SGs and game-based learning (GBL) are given. In Section 4, the design, the basic features and the class activities of the SG case study used are presented. The implementation of the sustainability feature into the case study SG and the “AmuSInG” concept is reported in Section 5. Finally, the paper ends with conclusions and future work.

¹ derives from the Greek word for sustainability

2 SUSTAINABILITY AND SUSTAINABLE DEVELOPMENT

According to Cairns and Martinet (2014), the term "sustainability" is inherently ambiguous. Sustainability can be understood either as the conservation and protection of ecosystems or as an improvement in living standards, under the aegis of economists. These different perspectives also refer to different interpretations, making it more difficult to understand the term sustainability. As proposed by Silvius and Schipper (2015) and later by Huemann and Silvius (2017), the most commonly accepted definition is that of the Brundtland Report: "*Sustainable development is a development that meets the needs of the present without jeopardizing the ability of future generations to meet their own needs*" (World Environment and Development Committee, 1987). Sustainability, according to this report, deals with three dimensions: economy (cost), society (people) and environment (planet).

2.1 Economic Dimension

Maximizing profits, reducing costs and increasing revenue are considered as some of the traditional business techniques (de Lange, 2017). The primary goal of an organization is to generate wealth for shareholders. Silvius (2017) recognize the importance of the economic dimension as it protects the shareholders' capital. Profits are reinvested in the organization to ensure that the organization achieves growth.

The strategic value of a project can be considered in terms of various economic dimensions that are featured in sustainability. Benefits, value, and value creation may imply different things to the project contractor and the customer and appear differently during the different phases of the project's and the product's life cycle thus making the specification of project goals very challenging (Kivilä et al., 2017).

2.2 Social Dimension

The social dimension refers to the communities in which organizations operate, as well as to employees of an organization. Employees are the ones who produce the results of the organization and should be rewarded by the organization itself. The results of the organization's work also depend on how the community supports the organization. Utilizing communities and workers for organizational success without exploiting them is the balance adopted by the notion of social dimension. Dempsey et al. (2012) proposed that organizations must "take care" of the individual communities in which they are active, making later Silvius (2017) totally agree with their point of view. Anything that adversely affects the society in which the organization is active may create problems for its own reputation (for example the BP Deepwater Horizon (1988) oil spill or Marikana) and is very likely to

reduce the available clientele, which customers will turn to more reputable organizations.

Organizations that recognize the importance of human factor often participate in corporate social responsibility initiatives. These initiatives are organizational actions that take place to improve the quality of life of workers as well as society in general, while ensuring economic growth (Bebbington et al., 2017). Social sustainability is about identifying and managing impacts, both positive and negative, on people.

2.3 Environmental Dimension

As Ludwig (1993) said, this dimension deals with the natural environment in which people live. The literature shows that the planet has been adversely affected by the activities of the human race Sha et al. (2017). It has already been reported since 1995 that the pursuit of economic goals has led and would continue to lead to the deterioration of the environment that supports humanity (Pretty, 1995). The environment is an important source of resources that must be maintained to ensure the continuity of natural processes (Hopwood et al., 2005).

Sustainability has been largely linked to environmental protection and mankind's failure to maintain natural wealth (Ferguson, 2016). Support provided by the natural environment is essential for the functioning of most organizations. Organizations are based on natural resources around the world. As these resources are reduced, organizations have difficulty in continuing their activities (Nawaz and Koç, 2018). Environmental non-sustainability has a negative impact on organizations and affects earnings, as operating costs are rising as supplies become costlier.

2.4 Combining Sustainability Dimensions

A balance is needed between the three dimensions of sustainability. It has been widely accepted that economic development, social prosperity and the prudent use of natural resources cannot be achieved without taking into account all dimensions and interactions between them (Elkington, 2006). The dimensions and their relationship can be seen in Fig. 1, which points out that only when there is a balance between the three dimensions, sustainability is achieved. If not, issues within the three spheres of sustainability emerge, depending on which dimension we are more focused on. Some examples of these issues can also be seen in Fig.1. Being more focused in social-environmental aspects provides good environmental management ethics towards society but lacks economic growth consideration. On the other hand, environmental-economic focus will assure energy efficiency by reducing costs and at the same time achieve resource efficiency however; the social dimension (concern for the values, norms, rules, and roles) would be absent. Finally, economic-social development neglects environmental

ethics and focuses on business ethics and worker's rights.

The Three Spheres of Sustainability

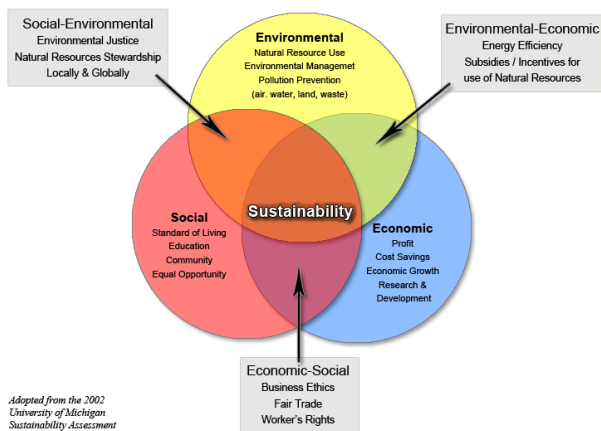


Figure 1: The three dimensions of sustainability, the relationship between them and the issues within (Sandra I. Rodriguez, 2012).

In addition to adopting policies and undertaking actions to improve their sustainability, corporations are also developing tools for monitoring, assessing, and reporting their environmental and sustainability performance. In his 1997 book "Cannibals with Forks", John Elkington proposed a triple bottom line against which corporations should measure the value they create and destroy. The triple bottom line (TBL) framework consists of three equally-weighted categories of impact: economic, social and environmental. Like definitions of sustainability, the TBL framework is built on the recognition that business sustainability depends upon positive stocks of economic, social and environmental resources (Sandra I. Rodriguez, 2012).

To promote dialogue between students and understanding of concepts related to sustainability and resource management, changes need to be made in the learning process and the use of the currently available tools. The use of a fun SG context to introduce SD principles, through the integration of sustainable issues in an existing SG, is what the main idea of this paper is.

3 SERIOUS GAMES AND GBL BACKGROUND

SGs are defined as games that educate, train, and inform. These games are designed for a primary purpose other than entertainment, enjoyment or fun. The initial intention of SGs is to combine the serious aspects (learning, instruction, etc.) with the playing aspect of digital games (game) (Khenissi et al., 2015).

There are numerous definitions for SGs, often varying depending on the different perspectives and purposes that they were developed to explain (Giessen, 2015). Some of the most used terms in this field are: "serious

games", "digital serious games", "edutainment", "Game Based Learning (GBL)", "Digital Game Based Learning" (Digital GBL) and "applied games" (Caballero-Hernández et al., 2017).

GBL is a pedagogical method of learning that utilizes role-plays, board games, card games or video games to promote retention of learned material and cognitive development. The game-based simulation is starting to be employed in several areas (Ricciardi and Paolis, 2014).

SGs that incorporate digital GBL are deemed by many to create the greatest potential for student learning over other media (Pereira et al., 2012). The creation of immersive and engaging environments in which players can explore and learn is more feasible than ever, especially with today's technology (Deshpande and Huang, 2011).

GBL has the potential to help engage students and enhance learning. Master's degree engineering students are receptive to the use of GBL in their courses, and educators have the opportunity to implement commercially available educational serious gaming tools like the SuPeRB² game, which is fully described in paragraph 4. Students will have to implement sustainable logic changes into the game itself in order to add extra value to the learning process. Through this process they will engage themselves into the SD philosophy and thus become aware of the sustainable way of thinking. Principles very close to the GBL and SGs philosophy.

4 CASE STUDY: THE SUPERB GAME

SuPeRB game is a typical serious game that incorporates GBL by giving students the opportunity to use a physical problem-solving tool by engaging themselves into programming activities.

Professors in the University of IMT Mines Albi in France were looking for ways to empower their master's degree engineering students' skills, turning their natural curiosity into creative exploration. They came up with the use LEGO® Mindstorms as an instant engagement tool and built upon it a whole instructional method. The instructional method of the specific course was inspired by the RobAFIS™ student competition. From systems engineering to the engineering of one system. The main objective of RobAFIS™ is to highlight the benefits of basing systems engineering education on a project life cycle realization: a full life cycle including the implementation of an operational system, deployed by a client, in a real environment.

So, in that way, students are reviewing the basic principles encountered in the planning and designing processes and at the same time they explore the latest ideas and

² deSign Programming Robot Brick

thinking around key project management topics. They end up finding key solutions in upcoming designing issues.

4.1 Context of the game

The SuPeRB game that has been developed in IMT Mines Albi and inspired by the RobAFIS™ program is a useful tool for the master's degree engineer students in order for them to achieve greater learning gains from playing SGs. It is taking place at the last semester of the fifth year of the graduate school of engineering where students have the opportunity to deal with design processes, create and operate their own system/device and program it using a Product Lifecycle Management (PLM) software, for the first time during their studies. Through this, they discover the difference between theory and reality and understand the importance of a system that works perfectly according to the initial design. The main idea is for the students to learn by themselves basic principles of system engineering, design and project management. The whole concept might have been inspired by the RobAFIS™ program but the use of a PLM software support gives the SuPeRB game a whole new dimension creating a big difference between the two programs.

In previous class activities, students used an ad-hoc robotic simulation package in order to improve their knowledge of the course subject (learning basic principles of system engineering, designing and project management); it clearly follows the importance of developing an innovative teaching method to improve the involvement of the students. As consequence, the class activities based on LEGO® Mindstorms were included in the curriculum activities of the students.

The activities have been attended every year by an average of 30 to 40, usually divided in teams of 7 to 9 in order to design, build and control a mobile robot able to compete in the execution of a predefined task. Each team clarifies predefined tasks among their members in order to complete their task more efficiently. Theoretically the goal is to have a team of 6 people, roughly with some specialization divided in two members responsible for the LEGO® bricks, two members to program the device and two members that organize the PLM activities. It is important while appointing the responsibilities to avoid having one skill mastered by only one student.

The LEGO® bricks give to the students the power to create and command their own robotic creatures, vehicles and machines built by the LEGO® bricks. By combining LEGO® elements with a programmable brick, motors and sensors, they can make their creations walk, talk, grab, think, shoot and do almost anything they can imagine. Each robot comes complete with its own unique features and a program that controls the robot's behavior. Once the task to be accomplished is assigned, each team works autonomously. Every year the task is

different. As typically few constraints are specified regarding the robot design (e.g. maximum size and number of motors and sensors), given the same resources each team usually develops solutions that differ a lot one from the others.

Every project has to produce something. "If you don't have any deliverables, you don't have a project" (Hamilton, 2002). That's why students in graduate school of engineering of IMT Mines Albi submit four main deliverables each year:

- Deliverable 1: In this deliverable the specifications and the planning process of the project given are specified. Identifying the risks in the systems engineering project is an important part too. Thus, students determine where to apply management resources and what to leave alone, as management resources are not unlimited. They can simplify many calculations and, more importantly, provide information with insight into the effects of uncertainty on project outcomes. The professors' feedback gives students the chance to improve their work and build more accurate self-assessment skills.
- Deliverable 2: This part is all about brainstorming, project drawings and evaluation of possible solutions. Students dig into design thinking. Design thinking is a methodology that has been used in product design, branding design, service design and other areas such as information systems design (Chou, 2018). It can be used to generate new and innovative products and systems. Students list all their basic ideas concerning possible solutions towards the requirements given and they create drawings to visualize their robot construction. They carry forward to a brief evaluation of the project and conclude to the final methodology.
- Deliverable 3: This part incorporates the bill of materials (BOM) and the assembly process. Students explain what, how, and where to find the required materials to build their robot and include all the necessary instructions on how to assemble their product/system from the various parts ordered.
- Deliverable 4: Examination day; where everything needs to be as planned. A personal bet for the students and their work.

4.2 Materials and software

The series of LEGO® Mindstorms Kits, with the programmable brick, a set of modular sensors and motors and standard LEGO® parts coming from the LEGO® Technics line, allow to design and implement small, customizable and programmable robots both from hardware and software point of view. The basic hardware/software element of Mindstorms Robotics Invention System kit is a programmable brick that derives from the one created at the MIT Media Lab. Normally the brick may be programmed by uploading a

program written using the bundled graphical software or one of the several available languages.

Different versions of LEGO® Mindstorms Kit have been released over the time, each of which is characterized by an increasing level of complexity, thus allowing the development of more sophisticated and complex robots both from the mechanical and the control software point of view.

A bill of materials (BOM) describes the different components that together create a product/system. Students have the chance to understand the importance of the right selection between big lists of LEGO® bricks in order to give solutions to upcoming design problems. A BOM for a bicycle, for example, consists of all the parts that make up the bicycle, including the number needed (quantities): one frame, one saddle, two wheels, and so on. In our case students are called to clarify the BOM for the LEGO® robot in order to achieve the goals that it is made for.

Students have the chance to use a PLM software, an information management system that can integrate data and processes. In the class activities, students utilize the Windchill PLM software that allows them to manage the information needed throughout the entire lifecycle of a product, efficiently and cost-effectively, from ideation, design and manufacture, through service and disposal. Students realize the importance of documentation through this process. After everything is done, students hand over the three deliverables mentioned in paragraph 4.1.

As for the delivery process, students dig deeper into appearing problems after the delivery of their robot where they understand the importance of accuracy during the designing process. They have to carefully inspect all the contents of their product/system. A delivery process is a special process describing a complete and integrated approach for performing a specific project type. It provides a complete end-to-end lifecycle and can be used as a reference for running projects with similar characteristics (Peñaloza, 2018). From the software point of view, in the class activities the students utilize the LEGO® Mindstorms EV3 software which is easy to use and free to download. Nevertheless, programming the brick is possible through another programming language (big list of some of the most popular third-party alternatives). In fact, by installing the proper plug-in, the developing environment allows not only to fully exploit all the potentials of the libraries, but also to directly download the compiled code on the brick by using USB or Bluetooth connection.

4.3 Course schedule

The lesson plan and the class activities of the game which takes place in IMT Mines Albi are as follows:

- Requirements and specifications are given regarding the robot design and the tasks it is called to fulfil.
- One lecture (3 hours) on how to build the already defined robot.
- One lecture (3 hours) dedicated to the pre-design process (BOM) and project management skills.
- Two lectures (3+3 hours) where the students try the LEGO® robot construction and programing.
- Exams session: 45 minutes outside the examination class to build and program the robot – 20 minutes to present their completed task inside the class and to explain the logic behind – 20 minutes dedicated to questions and feedback analysis.
- Final grades for the students are given as follows: 1/3 professors' evaluation – 1/3 other teams' evaluation (students) – 1/3 proper robot delivery (product design).
- The following link provides a wider scope of understanding towards the SuPeRB game: <https://www.youtube.com/watch?v=3Koxe6Nk mOE&feature=youtu.be>

4.4 The need for sustainability inclusion

The objective of this class, so far, was for engineering students to discover the complicity of product design, to understand the difficulty of scheduling a project (project management), to dig into the programming philosophy through a software, to comply with the user's requirements and cultivate their innovation skills.

Analyzing the current set-up of the game and at the same time recognizing the needs of modern societies, a big change towards the modernization of the game itself and its possibilities is required. The integration of sustainability into GBL is picking up momentum. Integrating sustainability stretches the game's boundaries and adds extra value.

After careful consideration and after analyzing the process of the conduction of the classes, the authors of this paper came up with conclusions on opportunities for inclusion of sustainability concepts. Table 1 outlines all the current elements of the game that can evolve in order to entail sustainable concepts and categorizes them according to the TBL; Economy, Society and Environment.

Integrating sustainability issues in the SuPeRB game/product will give the chance to students to expand their horizons towards a more sustainable way of thinking. The supplementary knowledge in terms of SD will create a new sustainable SG with great significance ("AmuSInG").

Dimension of Sustainability	SuPeRB Game
Economic	<ul style="list-style-type: none"> ▪ Location of construction of the parts of the game
Social	<ul style="list-style-type: none"> ▪ Working in teams ▪ Working during business hours ▪ Students learn basic principles of project management
Environmental	<ul style="list-style-type: none"> ▪ Rechargeable batteries ▪ Plastic bricks ▪ Working hours in relation to energy consumption

Table 1: SuPeRB game characteristics that can evolve in order to include sustainability concepts

5 INTEGRATING SUSTAINABILITY ISSUES INTO GBL

The continuing depletion of natural resources has become a major focus for the society at large. There is an increasing recognition of the need to sustain an ecologically-balanced environment, while, at the same time, exploring and exploiting the natural resources to satisfy the ever-increasing demands of the human race. A profound solution to this is the adoption of SD practices. Increasing the awareness towards a more sustainable future is thus critical (Moloney et al., 2017). To achieve this in the education sector, changes in the use of decision games called “serious games” need to be done. Serious games are gaining in popularity as tools that add entertainment to teaching and training.

5.1 Context

This paper explores how sustainable philosophy can be integrated in SGs with a view to facilitate the understanding of the issues around sustainability, to identify opportunities towards improving the feature-set of these games, and to enhance knowledge around SD strategies. To achieve this goal, the SuPeRB game was taken into account, and improvements that will foster the notion of sustainability are proposed that will add significant value to the game and thus create a new sustainable oriented version under the name of “AmuSInG”.

5.2 Material

Taking into account the basic elements of the SuPeRB game that can be changed towards a more sustainable philosophy (refer to Table 1), Table 2 provides solutions on how something like that is possible, introducing the “AmuSInG” approach. The dimensions of sustainability

summarized in Table 2 provide input for integrating sustainability requirements into the content related aspects of the game, such as the specifications and design of the deliverable, materials used, benefits to be achieved, quality and success criteria.

Starting with the economic dimension of the TBL, the proposed scenario suggests that a way to maximize profits and reduce costs is the construction of some parts of the game to be made on-site. Some of these parts can be made upon request in IMT Mines Albi, (bricks) using an existing 3D printer. As simple as this may sounds, a business plan has to be conducted in order to clarify the sustainable contribution behind this proposal. The sustainability value of this implementation has to be clarified, which constitutes a great suggestion for future research.

Continuing with the social dimension, an upgrade towards the current status of the game can be made by introducing basic GBL principles. Working in teams is not enough. The new proposal includes budgets, helpful pictures, virtual reality (VR) environments and a mobile app. Following the students’ progress, reward budgets will be given by the professors resulting to extra rewards like helpful designing related images that will guide them through the construction phase easier and faster. A VR app will immerse students in incredible 360-degree visuals of the game. In addition to all these, a mobile app will provide real-time insights of all the actions that take place during the class. Students have to complete more than one quest on the same topic; they earn more virtual points, a higher leader board standing and hopefully, better understanding of the material. By this way retention of learned material and cognitive development is promoted making the game more interesting to the students. Applying SD principles into the game, students get the chance to become aware of sustainability issues, increasing the social impact.

Finally, the environmental contribution to the TBL suggests the use of electricity instead of batteries (photovoltaics) for the movements of the robot and low-cost environmental materials for the bricks. The energy goes straight from the photovoltaic panel to the electrical circuits without the need of a battery. Working outside the class or during daylight to reduce electricity consumption to a minimum and the creation of environmental aware engineers are significant contribution towards ecological issues as well. During this process, the social dimension gains significant value as well.

The main question that this paper discusses is whether all the elements that can be upgraded in the game to include sustainability concepts (Table 1) have a positive or a negative impact to the additional value of the deliverable/product and the delivery/process of the project. The discussion of the seven defining characteristics presented on Table 2 represents the qualitative outcome of this study. Next to this analysis, the study also indicates how

these characteristics improve the TBL criteria of sustainability and thus create additional value to the game itself and the stakeholders.

Dimension of Sustainability	“AmuSInG”
Economic	<ul style="list-style-type: none"> ▪ Parts built on site upon request
Social	<ul style="list-style-type: none"> ▪ Working in teams with rewards (budies, best student of the week, etc.), integrating VR and mobile apps
Environmental	<ul style="list-style-type: none"> ▪ Electricity based game ▪ Low cost environmental materials ▪ Working outside during daylight

Table 2: Implementing sustainability issues into the SuPeRB game

5.3 Course schedule

The lesson plan and the class activities for the “AmuS-InG” approach game, along with the highlighted new features, could be as follows:

- Requirements are specified regarding the **sustainable design** and the tasks it is called to fulfil.
- One lecture (1hour) about **SD and environmental awareness (renewable energy sources – RES)**.
- One lecture (2 hours) on how to build the robot **following the TBL principles**.
- One lecture (2 hours) dedicated to the pre-design process (BOM) and **sustainable project management skills**.
- One lecture (1 hour) on how to **utilize a 3D printer and design 3D bricks on a computer**.
- One lecture (3 hours) in **informing students about GBL, VR, mobile app and rewards given (budies, helpful pictures, etc.) to teams that achieve high performance**.
- One lecture (3 hours) organized in teams, outside class (**field**), where the students try the robot construction, the programing part, **VR and the mobile app application**.
- Exams session: 40 minutes to build and program the robot – 20 minutes to explain the **sustainable logic** behind – 20 minutes dedicated to questions and feedback analysis.
- Final grades for the students are given as follows: 1/3 professors’ evaluation – 1/3 other teams’ evaluation (students) – 1/3 proper robot delivery (product design).

6 CONCLUSION & FUTURE RESEARCH

This work has been focused on the design and development of the first prototype of a SG that follows sustainable logic, following the TBL criteria for educating master’s degree engineers. To the best of authors’ knowledge, no previous research has been conducted towards the SD of a SG that incorporates awareness into sustainability topic for educational purposes. As far as the specific case study is concerned, the SG is intended to be integrated into the course "éléments de conception de produits et procédés" and to be used as a supplementary learning tool for students in industrial engineering. The game will be played in practical work sessions and it is expected to include more levels and more challenges that the students should gradually resolve. Some configurations of the game are too easy and others are more complex. Overall, the idea of a sustainable oriented game (“AmuSInG”) was well received and the students manifested an increased willingness to want to play it. Stimulate this kind of intrinsic motivation is what the game is aimed at and thus which has been achieved successfully.

This study opens an interesting prospect for research. Firstly, it is interesting to improve and extend the original game. Once the game is enriched with a long storyline and additional challenges following the sustainable path, it will be interesting to reassess. Another way which is also promising is to provide the system with an algorithm for collecting and storing the results of training sessions to assess learner’s skills.

Paragraph 5.2 provides a list of supplementary ideas that can be integrated into the sustainable logic of the game, following the Tables 1 and 2. As mentioned in previous paragraph the main question is whether all the elements that can be upgraded to include sustainability concepts in the game (Table 1) have a positive or a negative impact to the additional value of the deliverable/product and the delivery/process of the game. In paragraph 5.2 we can observe all the previously mentioned updates that “AmuSInG” approach has to offer and their gaining value according to the TBL philosophy.

One direction for future research should therefore be the specification of the gaining value for each implementation mentioned in Table 2. Positive and negative impact on TBL and sustainable philosophy can be mentioned. The project can be expanded even further, going one step ahead and taking into account not only the narrow borders of the game/product itself but also the sustainable impact of the delivery/process of the project. To incorporate sustainability into GBL projects and raise the level of capability, organizations and professors must reconsider the way that sustainability is incorporated.

REFERENCES

- BEBBINGTON, J., RUSSELL, S. & THOMSON, I. 2017. Accounting and sustainable development: Reflections and propositions. *Critical Perspectives on Accounting*, 48, 21-34.
- CABALLERO-HERNÁNDEZ, J. A., PALOMO-DUARTE, M. & DODERO, J. M. 2017. Skill assessment in learning experiences based on serious games: A Systematic Mapping Study. *Computers & Education*, 113, 42-60.
- CAIRNS, R. D. & MARTINET, V. 2014. An environmental-economic measure of sustainable development. *European Economic Review*, 69, 4-17.
- CHOU, D. C. 2018. Applying design thinking method to social entrepreneurship project. *Computer Standards & Interfaces*, 55, 73-79.
- DE LANGE, D. E. 2017. Start-up sustainability: An insurmountable cost or a life-giving investment? *Journal of Cleaner Production*, 156, 838-854.
- DEMPSEY, N., BROWN, C. & BRAMLEY, G. 2012. The key to sustainable urban development in UK cities? The influence of density on social sustainability. *Progress in Planning*, 77, 89-141.
- DESPANDE, A. A. & HUANG, S. H. 2011. Simulation games in engineering education: A state-of-the-art review. *Computer Applications in Engineering Education*, 19, 399-410.
- ELKINGTON, J. 2006. Governance for Sustainability*. *Corporate Governance: An International Review*, 14, 522-529.
- FERGUSON, P. 2016. Productivity growth as a barrier to a sustainability transition. *Environmental Innovation and Societal Transitions*, 20, 86-88.
- GIESSEN, H. W. 2015. Serious Games Effects: An Overview. *Procedia - Social and Behavioral Sciences*, 174, 2240-2244.
- HAMILTON, A. 2002. Considering value during early project development: A product case study. *International Journal of Project Management*, 20, 131-136.
- HOPWOOD, B., MELLOR, M. & O'BRIEN, G. 2005. Sustainable development: mapping different approaches. *Sustainable Development*, 13, 38-52.
- HUEMANN, M. & SILVIUS, G. 2017. Projects to create the future: Managing projects meets sustainable development. *International Journal of Project Management*, 35, 1066-1070.
- KHENISSI, M. A., ESSALMI, F. & JEMNI, M. 2015. Comparison Between Serious Games and Learning Version of Existing Games. *Procedia - Social and Behavioral Sciences*, 191, 487-494.
- KIVILÄ, J., MARTINSUO, M. & VUORINEN, L. 2017. Sustainable project management through project control in infrastructure projects. *International Journal of Project Management*, 35, 1167-1183.
- LUDWIG, D. 1993. Environmental Sustainability: Magic, Science, and Religion in Natural Resource Management. *Ecological Applications*, 3, 555-558.
- MADANI, K., PIERCE, T. W. & MIRCHI, A. 2017. Serious games on environmental management. *Sustainable Cities and Society*, 29, 1-11.
- MOLONEY, J., GLOBALA, A., WANG, R. & ROETZEL, A. 2017. Serious Games for Integral Sustainable Design: Level 1. *Procedia Engineering*, 180, 1744-1753.
- NAWAZ, W. & KOÇ, M. 2018. Development of a systematic framework for sustainability management of organizations. *Journal of Cleaner Production*, 171, 1255-1274.
- PEÑALOZA, L. 2018. Ethnic marketing practice and research at the intersection of market and social development: A macro study of the past and present, with a look to the future. *Journal of Business Research*, 82, 273-280.
- PEREIRA, G., BRISSON, A., PRADA, R., PAIVA, A., BELLOTTI, F., KRAVCIK, M. & KLAMMA, R. 2012. Serious Games for Personal and Social Learning & Ethics: Status and Trends. *Procedia Computer Science*, 15, 53-65.
- PRETTY, J. N. 1995. Participatory learning for sustainable agriculture. *World Development*, 23, 1247-1263.
- RICCIARDI, F. & PAOLIS, L. T. D. 2014. A comprehensive review of serious games in health professions. *Int. J. Comput. Games Technol.*, 2014, 9-9.
- SANDRA I. RODRIGUEZ, M. S. R., SAMANTHA C. STURHAHN, ELIZABETH H. TERRY 2012. Sustainability Assessment and Reporting for the University of Michigan's Ann Arbor Campus. In: SYSTEMS, C. F. S. (ed.) 2005 ed. Michigan USA: University of Michigan.
- SHA, L., JIANG, H., SEIDENKRANTZ, M.-S., LI, D., ANDRESEN, C. S., KNUDSEN, K. L., LIU, Y. & ZHAO, M. 2017. A record of Holocene sea-ice variability off West Greenland and its potential forcing factors. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 475, 115-124.
- SILVIUS, A. J. G. & SCHIPPER, R. 2015. A Conceptual Model for Exploring the Relationship Between Sustainability and Project Success. *Procedia Computer Science*, 64, 334-342.
- SILVIUS, G. 2017. Sustainability as a new school of thought in project management. *Journal of Cleaner Production*, 166, 1479-1493.

Système de systèmes dans les architectures PHM

Alexandre SARAZIN, Sébastien TRUPTIL, Aurélie MONTARNAL, Jacques LAMOTHE

Université de Toulouse, IMT Mines Albi
Campus Jarlard
81013 Albi cedex 09 - France
prenom.nom@mines-albi.fr

RÉSUMÉ : La complexité grandissante des systèmes industriels couplée à des normes et exigences de fiabilité toujours plus élevées posent un enjeu majeur pour de nombreuses industries comme les transports et l'énergie. Afin de rendre les systèmes plus sûrs, de nombreux travaux ont été menés sur le thème de la maintenance prédictive et sur le Prognostics and Health Management en particulier. Toutefois ces solutions souffrent encore d'une trop forte spécificité à un élément en particulier et il est nécessaire de se pencher sur l'application de ces technologies à l'échelle du système entier. Cet article propose un métamodèle PHM et pose les attentes auxquelles il doit afin de correspondre aux besoins d'adaptabilité et de performance industriels.

MOTS-CLÉS : *Prognostics and stics ealth Management (PHM), métamodèle, Système de systèmes*

1 INTRODUCTION

Le développement de nouvelles technologies ou concepts comme l'Industrie 4.0 (Hermann et al. 2016), les objets connectés ou la Smart factory devient une source d'enjeu majeur pour la compétitivité des entreprises. De fait, les industries sont à la recherche de fonctionnalités de plus en plus nombreuses ce qui a pour conséquence d'augmenter la complexité des systèmes.

Malgré cette complexité, ces systèmes doivent répondre à des exigences en termes de qualité et de fiabilité toujours plus fortes (Z. Wang et al. 2012). En effet, (K. L. Tsui et al. 2015) rappellent que les conséquences des défaillances peuvent engendrer un coût financier (dû par exemple aux retards) mais également des dommages humains et matériels inacceptables sans oublier l'éventuel impact sur l'image de l'entreprise.

Afin de limiter ces dangers et les risques économiques associés, de gros efforts ont été réalisés ces vingt dernières années par les chercheurs et les industriels afin de trouver

des méthodes de prévention et de prédiction des défaillances et de leurs conséquences. Ces méthodes ont été développées dans différents domaines d'application comme l'électronique (Bloch et Geitner 1983), la santé (K.-L. Tsui et al. 2008) ou encore l'aéronautique (J. Xu, Y. Wang et al. 2015) et peuvent être regroupées sous le terme Prognostics and Health Management (PHM).

Bien que l'approche PHM ait été développée depuis plus de 10 ans, (Lee, Wu et al. 2014) indique que ces approches sont essentiellement basées sur l'estimation de la durée de vie restante (Remaining Useful Life : RUL), qui est un aspect du PHM, mais pas le seul à considérer. (Lee, Jin et al. 2017) pointe le fait que ces systèmes sont encore très spécifiques à un seul équipement et parle de l'intérêt de développer un système PHM à l'échelle d'une flotte afin de regrouper les données d'appareils partageant des caractéristiques communes afin d'améliorer la performance de la prédiction.

Cet article propose une approche basée sur la notion de système de systèmes support

au PHM permettant de détecter les anomalies, de diagnostiquer les pannes et d'estimer la RUL tout en permettant d'adapter l'architecture à différents systèmes selon leur nomenclature. Cette approche repose notamment sur un métamodèle décrit dans ce papier

La première partie de cet article a pour but de poser la définition et les objectifs du PHM. La seconde partie détaillera les différentes étapes d'une structure PHM et un aperçu des solutions proposées par la littérature pour chacune d'entre elles. Enfin la troisième partie présentera le métamodèle proposé ainsi que les enjeux et attentes auxquels il doit répondre.

2 DÉFINITION ET OBJECTIFS DU PROGNOSTICS AND HEALTH MANAGEMENT

2.1 Définition

(Ahmad et Kamaruddin 2012) distingue 2 grands types de maintenance : la maintenance basée sur le temps (Time-Based Maintenance : TBM) et celle basée sur les conditions (Condition-Based Maintenance CBM). La TBM est une maintenance traditionnelle dont le principe est de réaliser les actions de remplacement ou réparation en se basant sur la durée moyenne entre les défaillances (Mean Time Between Failures : MTBF). La MTBF est dans ce cas déterminée au préalable par une étude statistique ou expérimentale avant de servir à déterminer l'intervalle de temps entre les actions de maintenance programmées. Ce type de maintenance est particulièrement pertinent pour les systèmes non critiques, dont la défaillance n'engendre pas de risque pour la sécurité des utilisateurs ou de conséquence économique importante comme l'arrêt d'une chaîne de production (Yam et al. 2001). La CBM est elle basée sur une détection des anomalies en temps réel par l'analyse des données émises par le système pendant son fonctionnement (Lee, Jin et al. 2017). D'après (Bloch et Geitner 1983), 99% des défaillances sont précédées par des signaux, conditions ou indications permettant de détecter qu'une panne va se produire. Cette méthode permet donc d'avoir une vision plus réaliste de l'état réel du système.

Avec les récentes avancées liées au développement des capteurs et des techniques d'analyse de données, de nombreux travaux se concentrent aujourd'hui sur la représentation en temps réel de l'état de santé des systèmes dans le but de prédire la durée de vie restante utile (Remaining Useful life : RUL) le plus précisément possible. (Si et al. 2011) reconnaît que la RUL peut se définir comme la période durant laquelle un bien ou une propriété est supposé être utilisable dans le cadre de l'utilisation pour laquelle il a été acquis.

Le Prognostics and Health Management (PHM) est une CBM définie par (Vichare et M. G. Pecht 2006) comme une méthode permettant d'évaluer la fiabilité d'un système dans les conditions réelles de son cycle de vie afin de prévenir l'apparition de pannes et réduire les risques. Elle consiste à observer et analyser les données envoyées par le système afin de détecter les anomalies, diagnostiquer les pannes et estimer la RUL des composants ou sous-systèmes en temps réel.

2.2 Objectifs

L'intérêt de mettre en place un système PHM est multiple. Tout d'abord, le déploiement d'une maintenance prédictive pour les systèmes dont la panne peut engendrer des risques pour la sécurité des utilisateurs est un enjeu essentiel. (K. L. Tsui et al. 2015) rappellent que le niveau d'exigence en terme de fiabilité des équipements et infrastructures est de plus en plus élevé malgré la complexité grandissante des systèmes. Pourtant (Vichare et M. G. Pecht 2006) indique que des accidents parfois mortels ont encore lieu à cause de défaillances techniques comme l'accident de métro de Washington en 2009 ou le crash de l'A330 dans l'océan Atlantique la même année.

Au delà de l'aspect sécuritaire, les enjeux économiques occupent également une place importante dans l'intérêt et l'essor de cette technologie. La détection des pannes est une piste majeure dans la diminution des actions de maintenance curative au profit d'actions planifiées. La planification est un facteur important de diminution des coûts et du temps d'intervention car elle permet de rassembler les ressources

humaines et matérielles de manière optimisée. L'augmentation de la disponibilité du système générée par la planification est en lui-même un enjeu économique primordial dans certains secteurs industriels comme l'aéronautique. D'après (Alter et al. 2014, les économies engendrées par la maintenance prédictive s'élèveraient à environ 12% sur les réparations non programmées et 30% sur l'activité de maintenance globale en diminuant le nombre de pannes de 70%.

3 ARCHITECTURES PHM DANS LA LITTÉRATURE

Comme le montre la figure 1, un système de PHM se décompose en 6 phases : la récupération des données, le pré-traitement des données observées, l'évaluation de l'état de santé du système, le diagnostic de l'anomalie, l'estimation de la RUL et l'interface utilisateur couplée à un outil d'aide à la décision.

3.1 Récupération des données

La première phase de récupération des données est, d'après (Campos 2009), le coeur des CBM. Durant cette étape, les signaux renvoyés par le système sont continuellement observés au moyen de différents types de capteurs ou d'indicateurs pertinents. Le choix et l'optimisation des capteurs peuvent s'avérer complexes. (Najjar et al. 2016) se sont penchés sur le problème de la captation des données d'un système d'échangeur thermique intégré à un système aérospatial et ont relevé trois problèmes majeurs. Premièrement, dans le cas de systèmes complexes et inter-connectés, la multitude des capteurs rend la sélection difficile. Deuxièmement, la fiabilité n'est pas assurée. Un capteur de température situé loin de la zone à mesurer ne renverra pas une valeur suffisamment fiable et précise pour être exploitée. Enfin, les conditions de mesure comme la présence de vibrations ou de bruits génère une incertitude qu'il faut prendre en compte dans l'analyse des données récoltées. (J. Xu, Y. Wang et al. 2015) proposent d'intégrer un système d'évaluation des capteurs en parallèle de l'architecture PHM pour résoudre ce problème de sélection.

3.2 Pré-traitement des données

La seconde étape consiste à pré-traiter les données reçues afin d'éliminer les valeurs absurdes, les bruits et les vides afin de limiter l'impact des mauvaises mesures sur l'évaluation de l'état de santé du système. Son impact est d'autant plus important qu'il impacte la performance globale du système et même l'apprentissage de la détection des anomalies lorsque celle-ci s'effectue grâce à du machine learning (ML).

3.3 Analyse de l'état de santé

La phase d'analyse de l'état de santé du système est l'étape qui va permettre de détecter les anomalies à partir des données recueillies et pré-traitées. Parmi les technologies utilisées, la plus simple est la création, pour chaque capteur, de seuils dont le franchissement signale une valeur anormale. Cette méthode, bien que facilement implémentable, ne tient pas compte de l'interaction et de la relation entre les données des différents capteurs (Orsagh et al. 2006). Une autre méthode, introduite par (M. Pecht et Jaai 2010), consiste à se baser sur la Physics of Failure, autrement dit, utiliser des modèles théoriques afin de représenter le système et détecter les dégradations par des techniques statistiques. Cette méthode a l'avantage de tenir compte de toutes les données dans l'analyse de l'état de santé, cependant elle requiert une connaissance approfondie du système et s'adapte mal aux situations non prévues par le modèle. Une troisième méthode consiste à utiliser les techniques de reconnaissance de modèle basées sur le ML (J. Xu et L. Xu 2011). Ce choix nécessite d'alimenter le système de détection avec un jeu de données d'apprentissage suffisamment complet et précis pour lui permettre de définir un fonctionnement nominal et de reconnaître les anomalies. Les avantages de cette méthode sont qu'elle ne requiert aucune connaissance spécifique du système et qu'elle s'adapte facilement aux nouvelles pannes grâce à sa capacité d'apprentissage.

3.4 Diagnostic de la panne

Le diagnostic permet de définir la panne rencontrée à partir des anomalies détectées. D'après (Lee, Jin et al. 2017), les technologies les plus utilisées pour le diagnostic dans le PHM sont les

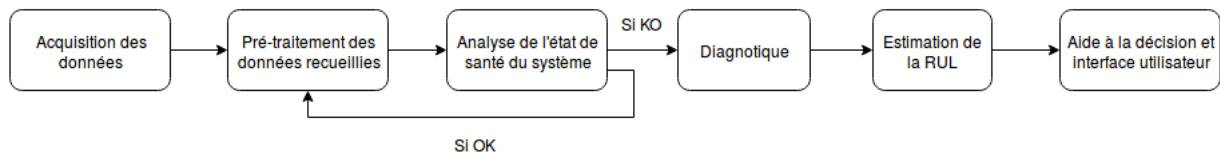


FIGURE 1 – Description des étapes principales du PHM

réseaux bayésiens et la classification basée sur du ML, en particulier les réseaux de neurones et le Support Vector Machine (SVM). L'avantage du ML est sa capacité à apprendre les relations entre les valeurs des paramètres mesurés et le type de panne diagnostiqué bien que cela requiert des jeux de données pour chaque type de panne afin d'entraîner l'algorithme. Les réseaux bayésiens modélisent également bien les relations entre les valeurs anormales mesurées et la catégorie de la panne mais (Yang et Lee 2012) pointe le fait que définir et configurer un réseau bayésien demande une connaissance très pointue du domaine.

3.5 Estimation de la RUL

Une fois la panne catégorisée, l'objectif est de déterminer la RUL du système afin de pouvoir faciliter la prise de décision concernant la planification de la maintenance. De nombreux travaux ont été réalisés ces dernières années afin d'améliorer la performance de cette estimation. (Gupta et Lawsirirat 2006) propose un modèle de détérioration multi-composants dans le but d'optimiser les coûts des actions de maintenance. (Si et al. 2011) fait un état de l'art assez complet des méthodes de calcul de RUL basées sur les modèles parmi lesquelles figurent les techniques de régression, les Hidden Markov Models (HMM) et Hidden Semi-Markov Models (HSMM). Les HMM et HSMM ont la particularité d'intégrer une approche stochastique dans l'estimation de la RUL. (Zhang et al. 2017) présente une technique nommée Multiobjective Deep Belief Network Ensemble (MODBNE) basée sur du ML mais ayant la particularité de pas se baser sur une saisie manuelle des paramètres d'entrée. Toutefois, les méthodes les plus prometteuses sont celles associant plusieurs approches, appelées fusion prognostics. (J. Xu et L. Xu 2011) propose par exemple de combiner le SVM, les Fuzzy Neural Networks (FNN) et l'Autoregressive Moving Average (ARMA). Un autre exemple est l'association de HSMM et de

SVM par (Xinxin et al. 2016).

4 SYSTÈME DE SYSTÈMES ET PHM

Il existe plusieurs définitions du concept de système de systèmes (SoS). (Luzeaux et Ruault 2008) propose une synthèse de ce concept par la définition suivante : « Un système de systèmes est un assemblage de systèmes pouvant potentiellement être acquis et/ou utilisés indépendamment, pour lequel le concepteur, l'acquéreur et/ou l'utilisateur cherche à maximiser la performance de la chaîne de valeur globale, à un instant donné et pour un ensemble d'assemblages envisageables ». On retrouve dans cette définition une forme très condensée des différences entre le concept de système de systèmes, et celui de système. Ces différences sont qualifiées par cinq critères discriminants proposés par (Maier 1996) :

1. indépendance opérationnelle des éléments (pouvant être des systèmes),
2. indépendance managériale des éléments,
3. développement évolutif,
4. comportement émergent,
5. distribution géographique.

En projetant ces critères au domaine de la maintenance, il est possible d'émettre les hypothèses suivantes :

- chaque sous composant d'un système complexe peut être assimilé à un système et par conséquent chaque assemblage de composant peut être assimilé à un système de systèmes.
- pour chaque système d'un système de systèmes, il est possible d'appliquer les approches de maintenance décrite en section 2 de ce papier.

Sur la base de ces deux hypothèses, il est possible de tirer profit

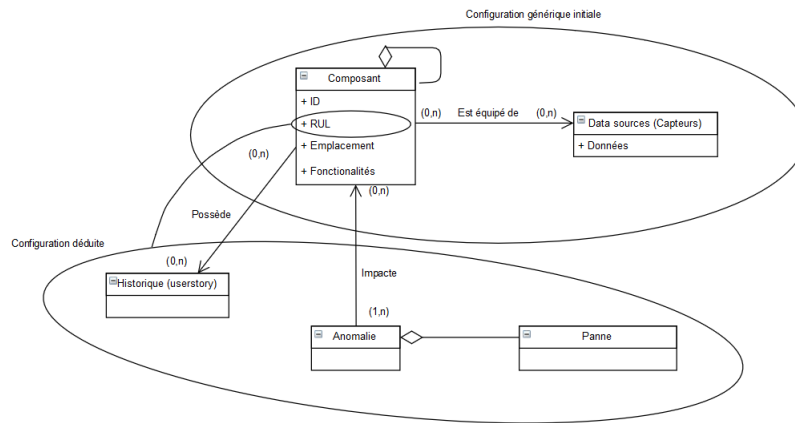


FIGURE 2 – Proposition de métamodèle pour le PHM

- des études sur chaque composant
- de gérer la granularité des données et des informations remontées
- de construire la composition du système de systèmes sur la base de la nomenclature du produit

Ce dernier point est essentiel. En effet, Chaque produit peut être vu comme un système de systèmes et chacun de ces sous-composants peut lui même être vu comme un système de systèmes ou au moins un système. Ainsi la notion de système de système permet de proposer une approche de maintenance indépendante de la complexité du produit.

Afin d'être applicable, cette approche nécessite de pouvoir décrire n'importe quel composant quelque soit sa nature ou ses fonctionnalités. Par conséquent, un méta-modèle a été créé de sorte à supporter cette approche puisqu'un méta-modèle est défini comme « un modèle qui définit le langage utilisé pour exprimer un modèle » (MOF 2002) ou encore « une description de tous les concepts d'un langage ». La méta-modélisation permet de définir directement les concepts manipulés, leur sémantique et la syntaxe du langage liée à l'utilisation de ces concepts » (Chapron 2006).

Dans le métamodèle représenté sur la figure 2, une distinction est faite entre la configuration générique initiale et la configuration déduite.

La configuration initiale décrit les éléments qui composent le système et dont l'agrégation définit la nomenclature. On y retrouve également les capteurs permettant de récupérer les données

de fonctionnement in-situ des composants et sous-systèmes de l'appareil. Cette configuration initiale a la particularité d'être basée sur les éléments physiques du système tout en étant assez générique pour permettre une adaptation à différentes nomenclatures et donc différents produits et/ou modèles.

La configuration déduite regroupe elle toutes les informations issues de l'analyse des données de capteurs, autrement dit les anomalies détectées, les pannes identifiées et la durée de vie restante (RUL). Cette configuration intègre également l'historique des actions de maintenance dont l'analyse peut contribuer à améliorer la performance des prédictions.

Le métamodèle intégrant ces 2 configurations répond ainsi au besoin d'adaptabilité de la maintenance prédictive à différents systèmes et contribue à améliorer la performance de la détection des anomalies, du diagnostic et de l'estimation de la RUL en rendant possible la mise en place d'un PHM à l'échelle d'une flotte (Lee, Jin et al. 2017).

5 CONCLUSION

L'augmentation du nombre de données disponibles doit permettre de construire une approche pro-active de la maintenance. En effet, une fois qu'une panne se produit, la réalisation d'un diagnostic doit empêcher le même type de défaillance de se reproduire. Cette prédiction, appelée pronostique, ne peut se baser uniquement sur les diagnostics car elle

doit également anticiper le cas de situations d'erreur inconnues. De plus, il faut répondre au problème de spécificité des architectures PHM par la définition d'un métamodèle permettant d'améliorer la performance et l'adaptabilité de ces architectures. L'enjeu est de pouvoir gérer plus facilement et plus efficacement des systèmes complexes et évolutifs.

De fait, ce papier propose un métamodèle différenciant la configuration générique initiale comprenant la nomenclature et les caractéristiques physiques du système et la configuration déduite englobant les résultats de la détection d'anomalie, de l'analyse de l'état de santé, du diagnostic de la panne et de l'estimation de la RUL.

RÉFÉRENCES

- Ahmad, R. et S. Kamaruddin (2012). « An overview of time-based and condition-based maintenance in industrial application ». In : *Computers & Industrial Engineering* 63.1, p. 135–149.
- Alter, A., P. Banerjee, P. Daugherty et W. Negm (2014). *Driving Unconventional Growth through the Industrial Internet of Things*. ed.
- Bloch, H. P. et F. K. Geitner (1983). *Practical machinery management for process plants. Volume 2: Machinery failure analysis and troubleshooting*. Gulf Publishing Co., Houston, TX.
- Campos, J. (2009). « Development in the application of ICT in condition monitoring and maintenance ». In : *Computers in Industry* 60.1, p. 1–20.
- Chapron, J. (2006). « Organizational Urbanism: methodology and decision making supports for the information system evolution management of the enterprise ». Theses. Ecole Nationale Supérieure des Mines de Saint-Etienne.
- Gupta, A. et C. Lawsirirat (2006). « Strategically optimum maintenance of monitoring-enabled multi-component systems using continuous-time jump deterioration models ». In : *Journal of Quality in Maintenance Engineering* 12.3, p. 306–329.
- Hermann, M., T. Pentek et B. Otto (2016). « Design Principles for Industrie 4.0 Scenarios ». In : *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2016 49th Hawaii International Conference on System Sciences (HICSS), p. 3928–3937.
- Lee, J., C. Jin, Z. Liu et H. D. Ardakani (2017). « Introduction to Data-Driven Methodologies for Prognostics and Health Management ». In : *Probabilistic Prognostics and Health Management of Energy Systems*. DOI: 10.1007/978-3-319-55852-3_2. Springer, Cham, p. 9–32.
- Lee, J., F. Wu, W. Zhao, M. Ghaffari, L. Liao et D. Siegel (2014). « Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications ». In : *Mechanical Systems and Signal Processing* 42.1, p. 314–334.
- Luzeaux, D. et J. Ruault (2008). « Systèmes de systèmes: concepts et illustrations pratiques ». In : *Lavoisier, Paris, France* 36.
- Maier, M. W. (1996). « Architecting principles for systems-of-systems ». In : *INCOSE International Symposium*. T. 6. Wiley Online Library, p. 565–573.
- MOF, O. (2002). *OMG Meta Object Facility (MOF) Specification v1. 4*. OMG Document formal/02-04-03 [Online]. Available: <http://www.omg.org/cgi-bin/apps/doc>.
- Najjar, N., S. Gupta, J. Hare, S. Kandil et R. Walthall (2016). « Optimal Sensor Selection and Fusion for Heat Exchanger Fouling Diagnosis in Aerospace Systems ». In : *IEEE Sensors Journal* 16.12, p. 4866–4881.
- Orsagh, R. F., D. W. Brown, P. W. Kalgren, C. S. Byington, A. J. Hess et T. Dabney (2006). « Prognostic health management for avionic systems ». In : *2006 IEEE Aerospace Conference*. 2006 IEEE Aerospace Conference, p. 1–7.
- Pecht, M. et R. Jaai (2010). « A prognostics and health management roadmap for information and electronics-rich systems ». In : *Microelectronics Reliability* 50.3, p. 317–323.
- Si, X.-S., W. Wang, C.-H. Hu et D.-H. Zhou (2011). « Remaining useful life estimation – A review on the statistical data driven approaches ». In : *European Journal of Operational Research* 213.1, p. 1–14.
- Tsui, K. L., N. Chen, Q. Zhou, Y. Hai et W. Wang (2015). *Prognostics and Health Management: A Review on Data Driven Approaches*. Mathematical Problems in Engineering. DOI: 10.1155/2015/793161. URL : <https://www.hindawi.com/journals/mpe/2015/793161/> (visité le 15/12/2017).

- Tsui, K.-L., W. Chiu, P. Gierlich, D. Goldsman, X. Liu et T. Maschek (2008). « A Review of Healthcare, Public Health, and Syndromic Surveillance ». In : *Quality Engineering* 20.4, p. 435–450.
- Vichare, N. M. et M. G. Pecht (2006). « Prognostics and health management of electronics ». In : *IEEE Transactions on Components and Packaging Technologies* 29.1, p. 222–229.
- Wang, Z., J. L. Zarader et S. Argentieri (2012). « A novel aircraft engine fault diagnostic and prognostic system based on SVM ». In : *2012 IEEE International Conference on Condition Monitoring and Diagnosis*. 2012 IEEE International Conference on Condition Monitoring and Diagnosis, p. 723–728.
- Xinxin, X., L. Qing et C. Nong (2016). « Remaining useful life prognostics of aircraft engine based on fusion algorithm ». In : *2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*. 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), p. 628–633.
- Xu, J., Y. Wang et L. Xu (2015). « PHM-Oriented Sensor Optimization Selection Based on Multiobjective Model for Aircraft Engines ». In : *IEEE Sensors Journal* 15.9, p. 4836–4844.
- Xu, J. et L. Xu (2011). « Health management based on fusion prognostics for avionics systems ». In : *Journal of Systems Engineering and Electronics* 22, p. 428–436.
- Yam, R. C. M., P. W. Tse, L. Li et P. Tu (2001). « Intelligent Predictive Decision Support System for Condition-Based Maintenance ». In : *The International Journal of Advanced Manufacturing Technology* 17.5, p. 383–391.
- Yang, L. et J. Lee (2012). « Bayesian Belief Network-based approach for diagnostics and prognostics of semiconductor manufacturing systems ». In : *Robotics and Computer-Integrated Manufacturing* 28.1, p. 66–74.
- Zhang, C., P. Lim, A. K. Qin et K. C. Tan (2017). « Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics ». In : *IEEE Transactions on Neural Networks and Learning Systems* 28.10, p. 2306–2318.

THE DIGITAL TWIN AT THE SERVICE OF STORAGE ZONE OPTIMISATION AND ASSEMBLY LINE SIMULATION

A. PIEL, M. BENAZOUZ

C. BRUNET

CEA, LIST, Executable Language Engineering
and Optimization Laboratory
91191 Gif-sur-Yvette, France
ariane.piel@cea.fr, mohamed.benazouz@cea.fr

Sherpa Engineering
R&D Department
92250 La Garenne Colombes, France
c.brunet@sherpa-eng.com

ABSTRACT: *This paper presents a complete optimisation methodology to improve costs linked to storage zones in assembly lines. It consists of coupled combinatorial optimisation and simulation techniques, relying on a Digital Twin UML model. The methodology is applied to a use-case provided by Groupe PSA, an automobile manufacturer. Asynchronisms necessarily appear in assembly lines because, generally, different types of products are produced on the same line, and because of hazards which result in a temporary stop of part of the line. The role of storage zones is to improve the flow of the line despite these desynchronisations. They however represent a financial cost, so a goal for industrialists is to minimise the size of these zones while still insuring the good performance of the line. At present, this optimisation process is done by experts who explore solutions by hand with a flow simulation tool in a trial and error approach. In this paper, a three step solution is proposed. A modelling framework is defined to create a Digital Twin of the factory, which serves as input to an optimisation process. A first combinatorial optimisation tool reduces the possible number of configurations, and then a simulation tool chooses the best configuration insuring the performance of the line.*

KEYWORDS: *Multi-model assembly line, Industry 4.0, Digital Twin, UML, Cyclo-static data-flow, Discrete simulation.*

1 INTRODUCTION

The current economical constraints in the industrial field are getting tighter, which leads to increased competitiveness and a need to produce more and quicker. The optimisation of production processes and their design therefore lies at the center of the considerations on the Factory of the Future. Optimisation needs are large and cover various scopes ranging from design to logistics and processes, with the objective of reducing time and costs while maintaining or even increasing the quality of products and services.

Optimisation and simulation tools need a comprehensive vision of the systems they study. However, a wide variety of software is used in the industrial field throughout the design and production processes. A lack of interoperability between these software results in what is commonly called digital discontinuity, which is both time-consuming (data has to be transferred manually between tools) and error-prone, and which prevents having an overall picture of the factory. The movement towards digitalising the Factory of the Future with Industry 4.0 and the Digital Twin, a general model of the factory, offers the perfect

framework for deploying optimisation and simulation strategies. The Digital Twin acts as a first step towards digital continuity since it allows uniting and mapping all concepts and data.

In this paper is presented a methodology of coupled optimisation and simulation techniques relying on a Digital Twin model of the industrial system. The aim is to:

- reduce the time necessary for the design of a production line to release the design office from tedious and repetitive tasks;
- reduce the costs of the production line.

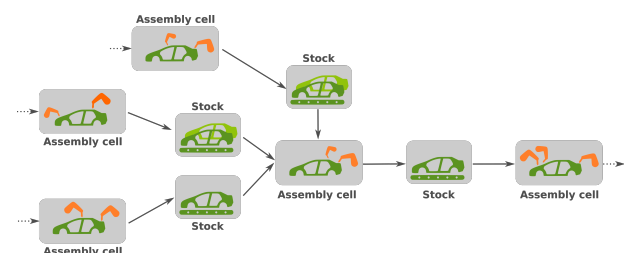


Figure 1 – Part of an assembly line, as studied in the Groupe PSA use-case of this paper.

In this context, Groupe PSA, a French multinational manufacturer of automobiles and motorcycles, has provided a use-case dealing with the optimisation of a body-in-white line (an assembly line in which a car body's sheet metal components are welded together). The line is composed of assembly cells and storage zones (stock) as illustrated in Fig. 1. Different models of cars may be assembled on the same line, which means that the processing time varies, depending on the variety which is being produced in each assembly cell. This phenomenon creates desynchronisations between cells, which impact on the line is smoothed by storage zones. Note that even on a perfectly balanced line, hazards (which result in the temporary stop of a cell) would lead to asynchronism and thus justify the necessity for storage zones. However, these storage zones represent a financial cost: the financial value of both the stored products and the occupied ground surface, as well as the cost of the buffer conveyor (eg. maintenance and power supply of a conveyor belt). The aim of the use-case is to minimise the size of these storage zones while ensuring a good performance of the line.

At present, the work to calculate the dimensions of these storage zones is done manually in a trial and error approach: a design engineer launches flow simulations on a tool and manually explores different configurations chosen using his expertise. Probability laws are integrated into the tool to model hazards and generate hazard scenarios. The simulation is then run with these scenarios for a sufficient length of time to be representative of the life-cycle of the line regarding hazards. The flow simulation tool is used to validate the final configuration. The validation criterion is the steadiness of the productivity of the line: a goal of a certain number of assembled products is fixed for each working day (targeted daily production rate), and the objective is that 90% of these working days reach their rate over the whole simulation. This approach is time-consuming, and the number of configurations rises exponentially with the size of the line, which is why a solution automating and optimising the process is needed. In order to be able to integrate a new solution in the current process, there is the constraint of keeping the same validation environment. The hazard scenario and the length of the simulation are therefore input data.

It must be mentioned that studied assembly lines actually have a higher capacity of production than the targeted daily production rate. The spare capacity of production is primarily intended to cope with strong desynchronisations unabsorbed by storage zones and to catch up with the initial production schedule in case of long breakdown days. The recovery should occur at the least within the day or two that follow which tends to guarantee a high level of production rate regularity.

State of the art The storage zones allocation problem that seeks minimising the total size of storage zones or maximising the production rate is known to be an NP-hard combinatorial optimisation problem (Dolgui, Ereemeev, Kovalyov & Sigaev 2013). In (Demir, Tunali & Eliiyi 2014) a thorough survey is provided on the different approaches developed to solve several variants of this problem. Nevertheless, none of the exposed techniques address the problem of achieving a certain production rate regularity while sizing storage zones as sought for in this paper.

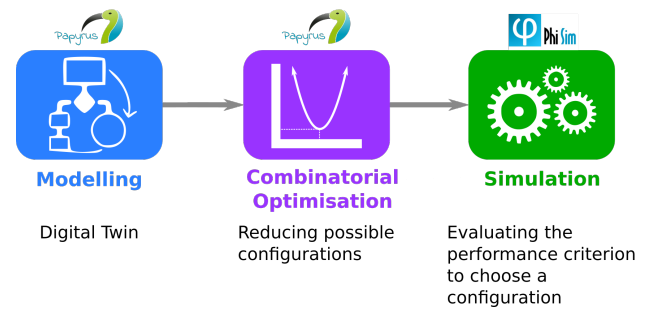


Figure 2 – Global methodology.

The contribution of this paper is to offer a methodology to automate and optimise the dimensioning of storage zones in a production line. This reduces both the time necessary for the design of the line and the cost of the line by proposing an optimised solution. The solution is in two steps, as illustrated in Fig. 2:

1. the production line and its processes are *modelled* in UML to offer a Digital Twin of the factory and provide a global view of the system as a first step towards digital continuity;
2. an algorithmically efficient *optimisation* process is then run on the Digital Twin:
 - (a) a first *combinatorial optimisation approach* allows to reduce the possible configurations to be explored;
 - (b) a *simulation* tool then chooses between these configurations the one which fulfils the production regularity criterion that 90% of working days reach their production rate (note that this criterion cannot be analytically evaluated and requires simulation).

In order to have a generic holistic solution, the simulation tool is conceived so that discrete and continuous simulation may be collectively run. Continuous processes are frequent in industry (for instance, the painting of the vehicle shell, which occurs after the welding of the body-in-white, is a continuous process), and may not always be representatively discretised.

This paper is structured as follows. Section 2 details the modelling approach which allows to create the Digital Twin of the factory. Section 3 presents the first part of the optimisation approach, where

the possible size configurations to be explored are reduced. The performance criterion is then evaluated for these selected configurations using a simulation tool exposed in Sect. 4. Finally, results on the use-case illustrating this paper are detailed in Sect. 5.

2 THE DIGITAL TWIN

2.1 Modelling approach

This Section details the modelling process to create the Digital Twin.

A single generic modelling language is not enough to cover the various concerns involved in a specific domain, but creating specific languages in each situation raises fragmentation issues and increases digital discontinuity. It is necessary that the various models be interfaced into a consistent integrated whole which can be verified, tested, and unambiguously understood.

To this end, a general-purpose modelling language called the Unified Modelling Language (UML) (UML 2007) together with its extension mechanism is used in this paper. UML, widely adopted in industry, is endorsed by the Object Management Group (OMG, www.omg.org) which is one of the principal international organisations promoting standards supporting the usage of model-based software and systems development.

UML provides the profile concept for deriving domain-specific modelling languages from its set of general language concepts. This allows reuse of existing UML tools and available UML expertise, and ensures that all domain-specific modelling languages derived from UML share a common semantic and syntactic foundation. The basic premise of profiles is that all domain-specific concepts are derived as extensions or refinements of existing UML concepts. These extensions are called stereotypes. A stereotype definition must be consistent with the abstract syntax and semantics of standard UML meta-classes it extends. The modelling tool used in this paper is Papyrus, the Eclipse graphical editing tool for UML2 (*Papyrus 2.0.X*). In accordance with its primary goal to implement the complete standard specification of UML2, Papyrus provides an extensive support for UML profiles and powerful tool customization capabilities.

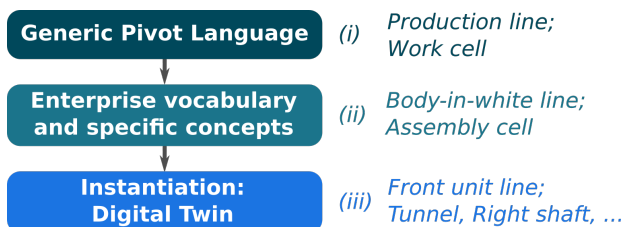


Figure 3 – Abstraction levels of the modelling approach and example on the Groupe PSA use-case.

The modelling methodology put forward in this paper and illustrated in Fig. 3 follows three steps which

correspond to three abstraction levels:

- (i) A *generic pivot language* is defined using UML profiles to introduce collective non enterprise-dependant concepts of the Digital Twin.
- (ii) The *enterprise vocabulary and specific concepts* are then introduced by extending and applying the generic pivot language profile.
- (iii) Finally, an *instantiation* corresponding to the studied use-case is built using the concepts introduced in Step (ii).

This final instantiated UML model constitutes the Digital Twin of the factory and its processes.

Note that the optimisation and simulation techniques require their own models. These models derive from the UML definition model of the Digital Twin which is the general reference.

2.2 A generic pivot language for industry

The generic pivot language proposed is based on IEC 62264 / ISA95, an international standard for enterprise control system integration that defines activity models, function models, and object models in the Manufacturing Operations Management (MOM) domain (Lu, Morris & Frechette 2016). This standard was chosen because it positions itself at the frontier of Manufacturing Operations & Control and Business Planning & Logistics (Levels 3 and 4 of the Purdue Enterprise Reference Architecture) (Williams 1994). Note that ISA95 defines how to express communication messages to exchange information between different parts of the enterprise, while the pivot language presented in this paper offers a description language to create a complete digital twin, based on ISA95 concepts. The language regroups notions distributed between different tools and models used by the enterprise, as a step towards digital continuity.

The pivot language is implemented through a UML profile which defines concepts identified by the standardisation committees in ISA95. The main concepts relevant to the use-case presented in this paper constitute two categories:

1. *Resources*, extending UML classes, which group equipment (from enterprise sites down to work cells, the equipment hierarchy is detailed as an example in Fig. 4), material (from raw material to the finished product), and operators;
2. *Processes*, extending UML opaque actions, which describe how products may be built, how products are scheduled to be built, and then what actually happens (including hazards for instance).

Resources are allocated to processes to create the schedule.

This generic pivot language brings together the key concepts needed to describe the production system. Its generic nature is established by the ISA95 norm on which it relies, and was proven through several use-cases from different fields of industry, one of which is presented in this paper.

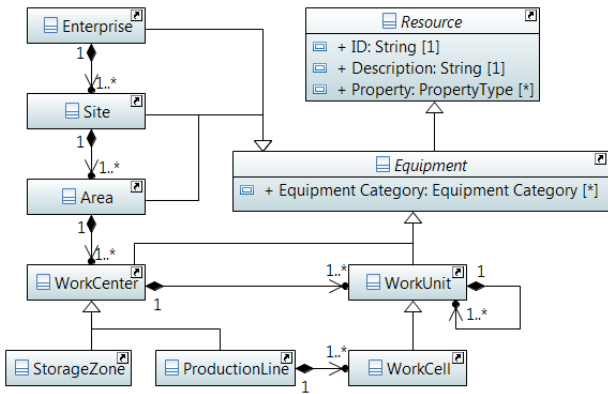


Figure 4 – Pivot Language Equipment hierarchy.

2.3 Introducing Groupe PSA specific vocabulary and concepts

The concepts of the pivot language now have to be specialised to introduce the enterprise specific vocabulary as well as particular non-generic parameters. For instance, as illustrated in Fig. 5, the Groupe PSA use-case deals with a specific type of production line, a body-in-white line. This type of line has two types of workcells, assembly cells where the metal fitting is done, and intermediate storage zones which are positioned between assembly cells. This constrained architecture must be defined in an extension of the pivot language.

There are two extension mechanisms to achieve this. Resources such as assembly cells and storage zones are represented by classes on which the corresponding stereotype from the Pivot Language is applied (in this example, the workcell stereotype). For processes, the pivot language profile is extended by a new profile to personalise process concepts which are extensions of UML opaque actions. Specific properties may be added through attributes in both cases.

This second modelling step adds an abstraction layer which corresponds to the mapping of the enterprise concepts to the Pivot Language.

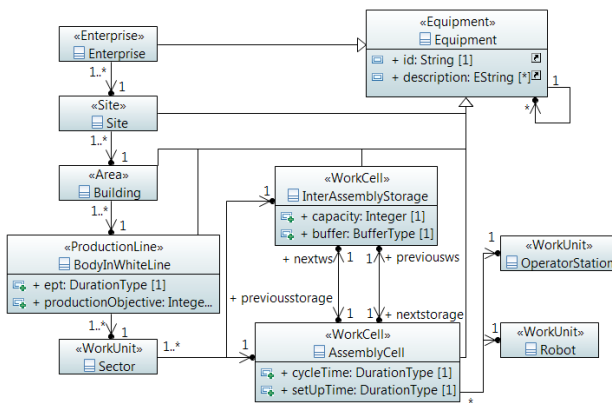


Figure 5 – Groupe PSA types of equipment.

2.4 An instantiation as a Digital Twin of the body-in-white line

All enterprise concepts are now defined. The last step is to instantiate these concepts to create the model of the studied use-case of Fig. 1. For instance, the body-in-white line contains five assembly cells with four intermediate storage zones. These nine elements are instantiations of the two concepts introduced in Sec. 2.3.

The architecture of the line and its properties are fully described. This instantiated model constitutes the Digital Twin of the studied production system. It contains the input data for the optimisation processes which aim is to determine the size of the storage zones and which is described in the following sections.

3 COMBINATORIAL SPACE REDUCTION

Given the exponential number of possible allocations of storage zones, exploring all of them blindly while checking by simulation whether they achieve the production rate regularity objective or not is a too much time consuming task even for small assembly lines. Thus there is a need for an efficient space reduction technique that should allow to quickly identify solutions that may potentially satisfy the objective. In this section, a two steps technique is proposed as follows.

In the first step, an optimised sizing is computed for each day of simulated production such that the production goal is achieved. However, if the assembly line as simulated during one of these days suffers a long cumulative breakdown period, no solution can be computed because the production objective cannot be achieved even if every storage zone is sized to infinity. Thus, such days are dropped from computation. By the end of the simulation, we are in possession of as much solutions as the number of days simulated, except those days of long outage times.

Afterwards, in a second step, these solutions are combined into a series of non-decreasing upper bounds on storage zone sizes. This series of candidate solutions is then checked through simulation in Section 4 in order to find the smallest in total storage size that achieves the production regularity objective.

3.1 Step 1 : Computation of an optimised solution per day

In order to model the variability of processing times of working cells in assembly lines, the Cyclo-Static DataFlow (CSDF) model was considered (Bilsen, Engels, Lauwereins & Peperstraete 1995). This model of computation is used in embedded systems and Electronics domains mainly to express the variability of data rate transfers (produced/consumed) between the different components of an application. A given process through subsequent executions might consume/produce different amounts of data. However, in this paper, we are only interested in the

ability of this model to represent the fluctuations of processing durations of a working cell (the production/consumption rate remains unchanged during production and is always equal to 1). In fact, a CSDF graph is expressive enough to model mixed-model assembly lines and the impact of production hazards on processing durations.

Let us denote by μ the period of time over which we would like to compute storage zone sizes. In the present case, μ is equal to the daily actual production time; time for rest breaks and briefings, for instance, are not taken into account in this part of the optimisation process. Let us also consider that the number of products to be delivered during this period of time is given by φ (*i.e.* daily production rate). In the following, it is shown how a CSDF graph is built to model production during the considered period.

3.1.1 Modelling using a CSDF graph

In a CSDF graph $\mathcal{G} = (T, A)$, the set of nodes T , called tasks, models the working cells of the assembly line and the set of arcs A corresponds to storage zones with a FiFo (*i.e.* First in First out) semantic. See Figure 6 for the CSDF graph of a simple assembly line composed of two working cells linked by a storage zone.

$$\ell_t = [60, 84, 70, 113]s \quad \ell_{t'} = [73, 70, 50, 152]s$$

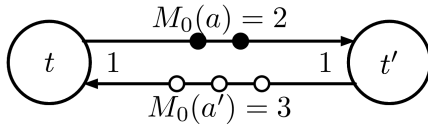


Figure 6 – A CSDF graph that models an assembly line manufacturing 4 products by means of two working cells (with different successive processing times) linked by a storage zone of capacity 5.

A task that models a working cell should go through as many phases as the number of products scheduled to be produced during the considered period. Thus, in the present case, every task $t \in T$ should be decomposed into $\varphi \in \mathbb{N} - \{0\}$ distinct phases (*i.e.* as much phases as the targeted number of products through the period μ). For every value $k \in \{1, \dots, \varphi\}$, the k^{th} phase of t is denoted by t_k . It has a constant duration $\ell_t(k) \in \mathbb{Q}^+$ that is the nominal processing time of the current variant produced to which a hazard duration must be added (retrieved from the simulated scenario) if eventually the working cell modelled by t goes down during the treatment of the k product.

In a CSDF graph, a storage zone from task t to task t' , denoted by $sz_{tt'}$, is modelled by a couple of arcs ($a = (t, t'), a' = (t', t) \in A^2$). Tokens on the forward arc a model the number of products that occupy the storage zone $sz_{tt'}$. Their initial number is denoted by $M_0(a)$. On the other hand, tokens on the backward arc a' of number $M_0(a')$ represent empty space. Therefore, the capacity of the storage zone $sz_{tt'}$ is

obviously given by $M_0(a) + M_0(a')$.

3.1.2 CSDF graph feasible schedules

A feasible schedule of a CSDF graph is a function s that associates, for every couple $(t, k) \in T \times \{1, \dots, \varphi\}$ the execution start time $s(t_k)$ of the phase t_k such that the number of tokens in every arc $a \in A$ remains non negative, *i.e.* no product is retrieved before it is available. In fact, a storage zone results in precedence constraints between the executions of the different phases of adjacent tasks. A feasible schedule is a schedule that satisfies these precedence constraints. Besides, all the phases of a task must be executed within the specified period of time μ such that $\forall t \in T, \mu \geq \ell_t \cdot \mathbf{1}$. Let us define the phases scheduling ζ_t that is used to determine the start times of the phases t_k ; $k \in \{1, \dots, \varphi\}$, with regards to the start time of t_1 , *i.e.* $\forall k \in \{1, \dots, \varphi\}, s(t_k) = s(t_1) + \zeta_t(k)$. A valid phases scheduling must ensure that two successive phases do not overlap, *i.e.* $\forall k \in \{2, \dots, \varphi\}, \zeta_t(k) \geq \zeta_t(k-1) + \ell_t(k-1)$, and the last phase ends before the end of the period μ , *i.e.* $\zeta_t(\varphi) + \ell_t(\varphi) \leq \mu$.

It has been proved in (Benazouz & Munier-Kordon 2013) that to ensure that no token (*i.e.* product) is consumed by t' before it is available a minimum delay must exist between $s(t_1)$ and $s(t'_1)$. The following theorem restricts this result to the specific case of assembly lines with production/consumption rate that is constant and equals 1.

Theorem 1. *Given the delays $\beta_a = \max_k f_a(k) + \max_{k'} g_a(k') - M_0(a) \frac{\mu}{\varphi}$ with $f_a(k) = \zeta_t(k) + \ell_t(k) - k \frac{\mu}{\varphi}$ and $g_a(k') = k' \frac{\mu}{\varphi} - \zeta_{t'}(k')$, then a schedule that verifies: $\forall a = (t, t') \in A, s(t'_1) - s(t_1) \geq \beta_a$ is feasible.*

3.1.3 Total Storage Zones Size Minimisation

The storage zone sizes minimisation problem consists of computing a feasible schedule s such that the total storage zones size $\sum_{sz_{tt'}=(a,a')} (M_0(a) + M_0(a'))$ is minimised and such that, for every task $t \in T$, all the φ phases are executed within a period of time μ . This guarantees achieving the daily production rate.

Based on Theorem 1, this optimisation problem can be formulated by the following Integer Linear Program (ILP):

$$\begin{aligned} & \min \sum_{sz_{tt'}=(a,a')} (M_0(a) + M_0(a')) \quad \text{subject to} \\ & \left\{ \begin{array}{l} \forall sz_{tt'}=(a=(t,t'), a'=(t',t)) \in A^2, \\ s(t'_1) - s(t_1) \geq \max_k f_a(k) + \max_{k'} g_a(k') - M_0(a) \frac{\mu}{\varphi} \\ s(t_1) - s(t'_1) \geq \max_{k'} f_{a'}(k') + \max_k g_{a'}(k) - M_0(a') \frac{\mu}{\varphi} \\ \forall t \in T, \quad s(t_1) \geq 0 \\ \forall a \in A, \quad M_0(a) \in \mathbb{N} \end{array} \right. \end{aligned}$$

It follows from this ILP that in order to minimise the total storage zones size, the sum $\max_k f_a(k) + \max_{k'} g_a(k') + \max_{k'} f_{a'}(k') + \max_k g_{a'}(k)$ must be minimised for each storage zone $sz_{tt'}$ (a detailed proof of this claim can be found in (Benazouz & Munier-Kordon 2013)). The value of this sum for a storage zone $sz_{tt'}$ is subject to the scheduling of the phases of both tasks t and t' .

For a task t , the following Min-Max Linear

Program (LP) computes optimised values of $\zeta_t(k)$, $k \in \{1, \dots, \varphi\}$ such that the sizes of storage zones that are adjacent to t are minimised (Benazouz & Munier-Kordon 2013):

$$\begin{aligned} & \min \sum_{sz_{tti}=(a_i, a'_i)} (F_{a_i} + G_{a'_i}) \quad \text{subject to} \\ & \left\{ \begin{array}{l} \forall sz_{tti} = (a_i = (t, t^i), a'_i = (t^i, t)) \in A^2, \forall k \in \{1, \dots, \varphi\}, \\ \quad f_{a_i}(k) = \zeta_t(k) + \ell_t(k) + k \frac{\mu}{\varphi} \leq F_{a_i} \\ \quad g_{a'_i}(k) = -\zeta_t(k) + k \frac{\mu}{\varphi} \leq G_{a'_i} \\ \forall k \in \{2, \dots, \varphi\}, \quad \zeta_t(k) - \zeta_t(k-1) \geq \ell_t(k-1) \\ \zeta_t(1) = 0 \text{ and } \zeta_t(\varphi) + \ell_t(\varphi) \leq \mu \end{array} \right. \end{aligned}$$

Fortunately, the phases scheduling optimisation of a task t can be performed independently from the other tasks of the assembly line. Therefore, Min-Max LPs for different tasks can be launched in parallel to reduce computation time.

3.2 Step 2 : candidate solutions generation

As a result of the first step of the optimisation part, thousands of solutions are obtained. Each of them is a potential solution that may achieve the targeted production rate regularity. However, such a number prevents us from checking all of them. Thus, a basic heuristic has been adopted in order to reduce the number of solutions to an acceptable one.

First, obtained solutions are sorted in a non-decreasing order according to their total storage size. Second, solutions with the same total storage size are substituted with a unique candidate solution. This solution is computed such that the size for each storage zones takes the highest value amongst those proposed by the replaced solutions. For instance, the following two solutions for a four storage zones assembly line (2,1,2,4) and (3,1,1,4), of the same total size 9, will be replaced with the candidate solution (3,1,2,4). While such a technique will result in an increase of the total size of the candidate solution in comparison to the solutions it replaces, it ensures that the candidate solution can cope with all the hazards covered by the substituted solutions.

Afterwards, as we advance in the computation of the following candidate solutions, we ensure in addition that, for each storage zone, the size is not smaller than the value proposed by the previously computed candidate solution. In such a manner, by the end of the algorithm, computed candidate solutions will form non-decreasing orders according to each of the storage zones. Considering the previously computed candidate solution (3,1,2,4) and the following solutions of size 10 to be substituted, (3, 2, 2, 3) and (4,2,1,3), the subsequent candidate solution should be (4,2,2,4) instead of (4,2,2,3) in order to ensure the non-decreasing order between candidate solutions on the size of the last storage zone.

As a result of this heuristic, tens of candidate solutions are produced instead of the thousands obtained initially. These solutions are then evaluated through simulation in order to determine the smallest solution that allows the achievement of the targeted production rate regularity. That is the subject of the next

section.

4 CRITERION EVALUATION AND DECISION

As mentioned above, the next step is to check which configurations among the solutions given by the combinatorial optimisation approach of Sect. 3 can reach the targeted regularity. The performance criterion is the percentage, over a significant operating time, of working days during which the production rate objective is reached. The evaluation of this criterion requires taking into account the operation of the system and all its perturbations over time. Dynamic simulations therefore have to be performed.

4.1 A need/supply simulation methodology

In this framework, a simulation tool was developed with the requirement of taking into account discrete events but also continuous phenomena. The underlying idea is, anticipating future needs, to be able to model in more details the operating systems and its control unit or to evaluate more criteria: for example power consumption. To this end, the PhiSim continuous simulation toolbox (*PhiSim* 2017.1), based on the Simulink software (*Simulink* R2013a), was extended with a library dedicated to discrete simulation. With this library, the model may be built by connecting simple components according to the architecture line. This modelling method is based on a need/supply approach of discrete objects. This approach was inspired by a continuous method already used for energy and material management (Mokukcu, Fiani, Chavanne, Taleb, Vlad, Godoy & Fauvel 2016), and may be viewed as a discretisation of it. This method allows the expression of the control logic in a structured and intuitive way. Here operations are driven by the production finality. The principle for discrete objects is as follows. A component sends to its upstream component(s) a need request for an object. If this request can be satisfied, the upstream component sends the object. The need and the supply may be expressed by internal states of the element called availability and acceptance:

- *Availability* characterises the capacity of a component to supply an object. If the object is available, it is ready to be supplied when a need is requested.
- *Acceptance* characterises the capacity of a component to receive an object.

With these concepts two basic components are built:

- a *transformer* component models a transformation taking a finite time, called its *cycle time*;
- a *storage* component accumulates objects until its *maximum capacity* is reached.

These blocks are designed to be generic : the transformer is parametrised with a cycle time, the storage is parametrised with a maximum capacity.

The production line model is then assembled as shown in Fig. 7 using transformation blocks, which corre-

spond to assembly cells, and storage blocks, which correspond to storage zones, according to the definition model of the Digital Twin.

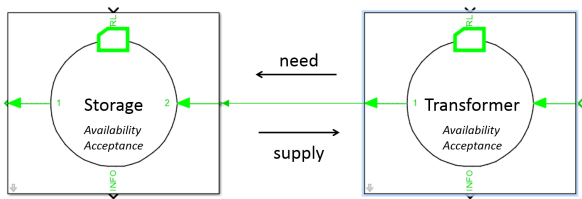


Figure 7 – Illustration of the transformer and storage blocks and the need/supply connection.

4.2 Simulation process

With this new simulation library that follows a need/supply methodology, simulations may be launched to evaluate the performance criterion for the configurations provided by the combinatorial optimisation process. In accordance with the combinatorial process, the architecture of the production line and the cycle time parameters of the transformers are constant and provided by the Digital Twin. The storage parameters vary according to the values of the configurations.

For each set of parameters, the simulation is launched with the same scenario to evaluate the impact of the different configurations on the performance criterion. The simulation scenario consists of the definition of the high-level processing of the line and the perturbations occurring during this operating process. Perturbations correspond to faults: for each transformer, a hazard scenario disables its transformation process according to the duration of each fault. For the operating process, the scheduling of the working days and the breaks of operators are taken into account. The line is in operation only when an operator is working on it. The same production rate is given for each working day, and quantified as a number of objects to produce. If the number is not reached, the number of unproduced objects are added to the next working day's production target rate but this next working day is still evaluated on the standard target, not including the extra-target.

The results of this simulation process are presented in the following section.

5 RESULTS

Running the simulation process of Sec. 4 on the configurations provided by the combinatorial optimisation approach described in Sec. 3 allows evaluating the performance criterion chosen by Groupe PSA. As explained earlier in this paper, this criterion is the production regularity which is expressed by the percentage, on the length of the simulation, of working days during which the targeted daily production rate is reached. The regularity objective on the use-case

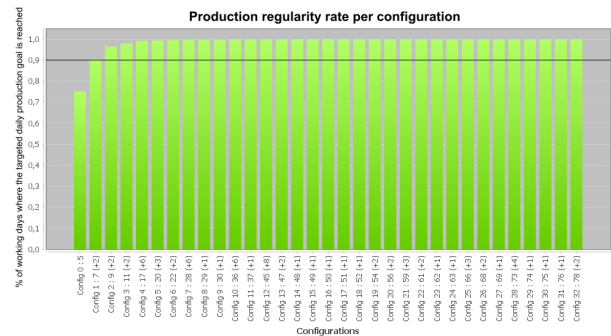


Figure 8 – Evaluation of the validation criterion for each configuration issued from the first optimisation approach.

illustrating this paper is 90%. The output of the simulation process for the use-case of this paper is presented in Fig. 8. Here, the second configuration provided by the optimisation process fulfils the performance criterion.

The configurations are sorted by increasing total size of storage zones and the results allow to take a balanced decision. Indeed, the expert in charge has several indicators to help him decide. Along with production regularity, the size of the storage zones and their cost are computed which allows to apprehend at which cost production regularity is increased and thus take an informed decision.

The adequacy of the combinatorial optimisation approach was tested with regard to the reference flow simulation tool used by Groupe PSA, on the same production line architecture but with varying initial parameters (varying cycle times for assembly cells, varying production rate objectives) which have a direct impact on the optimal dimensioning of storage zones. The result provided by the combinatorial optimisation process (candidate solution) is compared to the best result found by manually launching the reference flow simulation tool on possible configurations (optimal solution if the solution space is limited enough to be fully explored, better solution if not). For the initial scenario, the candidate solution corresponds to the second configuration of Fig. 8. This study, illustrated by Table 1, showed that the solution provided by the combinatorial optimisation approach was either the optimal solution or close to the optimal solution of a few units, so the solution provided by the tool may be refined if needed by exploring around the selected configuration. This witnesses the pertinence of the combinatorial optimisation approach which selects the configurations.

6 CONCLUSION AND PERSPECTIVES

This paper has presented a complete methodology for the optimisation of the sizing of storage zones in a production line : minimising their size while guaranteeing a performance criterion. A generic UML pivot

	Stock1	Stock2	Stock3	Stock4	Tot.
Candidate solution*	1	2	2	2	7
Optimal solution**	1	2	2	2	7

Initial scenario

	Stock1	Stock2	Stock3	Stock4	Tot.
Candidate solution*	9	11	10	10	40
Better solution**	9	9	11	10	39

More constrained scenario

	Stock1	Stock2	Stock3	Stock4	Tot.
Candidate solution*	4	3	6	3	16
Better solution**	3	3	6	3	15

Scenario with a higher production rate

* : Solutions provided by the combinatorial optimisation tool

** : Solutions obtained manually with the reference flow simulation tool

Table 1 – Pertinence of the results provided by the combinatorial optimisation process.

language for industry is proposed, deriving from the ISA95 norm. Based on this language, a modelling methodology is defined. It allows the description of a production line and its processes, and thus the creation of a Digital Twin of the factory. This global view of the studied system may then serve as input to an optimisation process in two phases to resolve the storage zone dimensioning problem. In a first phase, a combinatorial optimisation tool has been developed to reduce the possible number of configurations. The issuing configurations may then be simulated in order to evaluate the performance criterion and choose an optimised solution. To this end, a new library dedicated to discrete simulation but compatible with continuous phenomena has been developed.

In order to get more accurate solutions at the end of the first step of our combinatorial optimisation technique, the development of a phase scheduling algorithm taking into account a more detailed daily production calendar (operator breaks, briefings, ...) is considered. Besides, a more sophisticated heuristic than the one adopted in Subsec. 3.2 should be developed to propose more refined candidate solutions to the simulation part. Actually, candidate solutions as computed by our basic heuristic can be seen as upper and lower bounds of the optimal solution. Thus, an exploration within the range of these bounds should quickly lead to more accurate solutions.

ACKNOWLEDGMENTS

This work, as part of the SODALite project, has been funded by the FactoryLab¹ community, a French initiative towards the industry of the future.

The authors thank M.-L. Strady, C. Aubrit, and F. Derouet from Groupe PSA for their help and technical support on this application.

References

- Benazouz, M. & Munier-Kordon, A. (2013). Cyclo-static dataflow phases scheduling optimization for buffer sizes minimization, *16th Int. Workshop on Software & Compilers for Embedded Systems, M-SCOPES '13*, ACM, NY, USA, pp. 3–12.
- Bilsen, G., Engels, M., Lauwereins, R. & Peperstraete, J. A. (1995). Cyclo-static data flow, *1995 International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 3255–3258.
- Demir, L., Tunali, S. & Eliiyi, D. T. (2014). The state of the art on buffer allocation problem: a comprehensive survey, *Journal of Intelligent Manufacturing* **25**(3): 371–392.
- Dolgui, A., Ereemeev, A., Kovalyov, M. Y. & Sigaev, V. (2013). Complexity of buffer capacity allocation problems for production lines with unreliable machines, *Journal of Mathematical Modelling and Algorithms in Operations Research* **12**(2): 155–165.
- Lu, Y., Morris, K. C. & Frechette, S. (2016). Current standards landscape for smart manufacturing systems, *National Institute of Standards and Technology, NISTIR 8107*.
- Mokukcu, M., Fiani, P., Chavanne, S., Taleb, L. A., Vlad, C., Godoy, E. & Fauvel, C. (2016). A new concept of functional energetic modelling and simulation, *The 9th Eurosim Congress on Modelling and Simulation*, Oulu, Finland, pp. 536–541.
- Papyrus (2.0.X)*.
URL: www.eclipse.org/papyrus
- PhiSim (2017.1)*.
URL: www.sherpa-eng.com/produits/phisim/
- Simulink (R2013a)*.
URL: mathworks.com/products/simulink.html
- UML, O. (2007). Unified modeling language, *Infras-structure Specification, version 2*(1).
- Williams, T. J. (1994). "The Purdue enterprise reference architecture", *Computers in industry* **24**(2-3): 141–158.

¹<https://factorylab.fr/>

THE USE OF INTERNAL CROSS-DOCKING IN JUST-IN-TIME PLANTS

Mohammed Hichame BENBITOUR, Evren SAHIN

Laboratoire Génie Industriel, CentraleSupélec, 3 Rue Joliot Curie
91190 Gif-sur-Yvette
mohammed-hichame.benbitour@centralesupelec.fr, evren.sahin@centralesupelec.fr

ABSTRACT: We study the use of internal cross-docking in Just-in-Time plants. The considered plants consist of a cross-docking zone, a picking zone and assembly lines. The cross-docking zone is used to receive, store (temporarily) and ship pallets. Pallets are then broken into cases and stored in the picking zone which supplies the assembly lines. Cross-docking is used instead of mass storage in order to accelerate the physical flows of pallets. In this paper, we introduce different cross-docking policies (set of rules which specify how to store pallets) commonly used in practice. A cross-docking cost evaluation model is developed in order to compare the cross-docking policies considered in terms of required labour and surface costs.

KEYWORDS: in-house logistics, cross-docking, just-in-time, picking, facilities planning.

1 INTRODUCTION

Cross-docking is defined by Stephan and Boysen (2011) as a logistics concept, which integrates intermediate nodes into a transportation network. Within a cross-docking terminal (called cross-dock) incoming shipments delivered by inbound trucks are collected, sorted by destination, and moved across the terminal to be directly loaded onto outbound trucks. Cross-docks are high speed warehouses, said Bartholdi and Hackman (2008). Cross-docking (in distribution systems) has led to reduce significantly warehousing costs. However cross-docking success in distribution systems, it still remains not well known for manufacturing systems. In this paper, we propose a cross-docking model that can be applied to Just-in-Time (JIT) plants.

The model addressed in this paper is a cross-docking zone used to receive, store and ship pallets. The customer of the cross-docking zone represents the picking zone of the plant (also called the supermarket) which supplies the assembly lines (see Figure 1). The cross-docking zone contains inbound docks where the suppliers' trucks delivering assembly components are received. Each supplier has a given delivery frequency. The cross-docking storage area is composed of I-shape parallel lanes.

Each time slot (e.g., one hour), a pick-list is released by the picking zone and pallets are loaded from the cross-docking zone to the picking zone. Note that the main difference between cross-docking used in distribution systems and cross-docking used in JIT plants is the downstream part. In distribution systems cross-docking, there are many customers and outbound docks. Some practitioners use the term "Internal cross-dock" to describe cross-docking used in JIT plants.

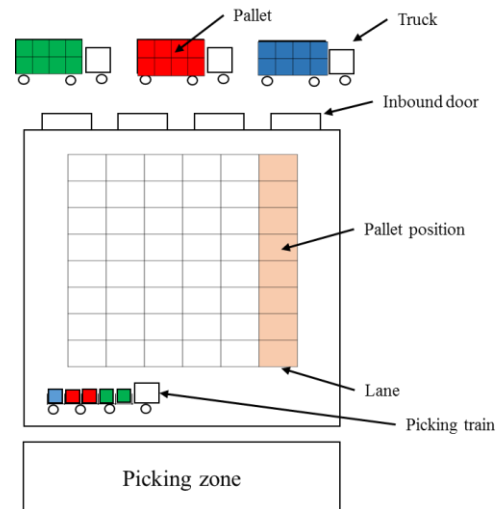


Figure 1 : Internal cross-docking in JIT plants

We call a cross-docking policy, a set of rules that specifies where to store pallets in the cross-docking zone. We start by presenting two cross-docking policies commonly used in practice: cross-docking per supplier and cross-docking per time slot. In cross-docking per supplier, the pallets of each supplier are stored in the same lane(s). In cross-docking per time slot, each lane contains all the pallets (from different suppliers) that will be consumed during a fixed time slot and hence, shipped to the picking zone during this time slot. The advantages and limits of each policy are presented. To cope with these limits, two new cross-docking policies are developed.

We develop a cross-docking cost calculation model in order to compare the different cross-docking policies considered in this paper. The model developed is applied to the case an auto parts supplier.

The literature review is presented in the next section. Section 3 describes the model and different cross-docking policies used in JIT plants. Section 4 presents the cost calculation model, section 5 presents the numerical example and section 6 concludes.

2 LITERATURE REVIEW

Our work is related to two research streams: 1) cross-docking operations and 2) in-house logistics.

The literature on cross-docking addresses decision problems which can be: i) strategic such as cross-dock location, cross-dock design, ii) tactical such as cross-dock network and iii) operational such as truck scheduling and storage assignment (see Van Belle et al (2012) for a literature review on cross-docking). Bartholdi and Gue (2000) proposed a model to construct efficient layout of cross-docking terminals. Their model seeks to minimize worker travel time and worker waiting time required in sorting and consolidating freight within the terminal. De Koster et al (2007) reviewed literature about order picking systems. They focused on layout design, storage assignment, zoning, batching and routing policies. In this paper we address a storage assignment problem which consists on assigning pallets to pallet locations before they can be picked to fulfil the picking zone orders. The cross-docking per supplier policy considered in this paper is somewhat similar to dedicated storage explained in De Koster et al (2007).

Vis and Roodbergen (2008) determined the temporary storage locations that minimize the total travel distance in the cross-dock. They gave a row-based storage assignment algorithm to resolve the problem of storage locations. Boysen and Fliedner (2010) proposed a classification of truck scheduling problems studied in literature. A truck scheduling problem consists on the assignment of trucks to dock doors and the sequence of all inbound and outbound trucks. Larbi et al (2011) addressed the scheduling of outbound trucks problem in a cross-docking centre where the sequence of arrival of inbound trucks can be fully known, partially known or unknown.

In a recent review, Ladier and Aplan (2016) compared the literature on cross-docking operations with cross-docking practices in industry. The different industry practices were mainly extracted from the visits and interviews of 8 different cross-docks in France. They provided several future research directions to answer industrial needs.

Concerning in-house logistics, Battini et al (2013) reviewed literature on JIT-supermarket concept commonly used in the automotive industry. In our paper, JIT-supermarket corresponds to the picking zone. Emde and Boysen (2012) studied the optimization of the number and the placement of supermarket-areas. Boysen and Emde (2014) addressed the scheduling problem of refill events in line-integrated supermarkets. Sali and Sahin

(2016) studied the problem of line feeding in JIT assembly lines. They compared the performance of three line feeding modes: line stocking, kitting and sequencing.

In our work, we show how cross-docking can be integrated in internal-logistics. We present different cross-docking policies commonly used in practice. An exact cost calculation model of cross-docking costs is proposed.

3 MODEL

We consider an internal cross-docking zone used to receive, store and ship pallets (i.e., unit loads) received from suppliers. Without loss of generality, we assume that each supplier delivers one component, i.e., all pallets delivered by the same supplier contain the same component. When a supplier shipment arrives (a shipment can be in one or in multiple trucks), pallets are first unloaded from trucks and stored temporarily (very short time) in the receiving area (Figure 2). Then, an operator transfers controlled pallets from the receiving area to the storage area by using a forklift. This operation is called staging (or positioning as described by Vis & Roodbergen (2008)). The storage area is a floor storage area composed of I-shape parallel lanes where each lane consists of several pallet positions. One pallet position can contain one or several pallets; the number of pallets per position is decided by the cross-dock managers.

Each time slot (a time interval defined by plant managers), a pick-list of pallets is released from the picking zone. A pick-list contains several pick-lines where each pick-line specifies what component pallets to pick and in what quantity. An operator transfers pallets requested in the pick-list from the storage area to a picking train by using a stacker (a stacker is smaller than a forklift and can handle narrower space). The picking train is unloaded in the picking zone; pallets that are transferred are broken into cases (boxes) and stored in flow racks. Afterwards, the picking zone feeds the assembly line in terms of quantity of components necessary for assembly operators. The routing problem of the picking train is not addressed in this paper.

Each day, the plant receives different shipments. Each supplier is assumed to have a constant delivery frequency, denoted by F , which represents the number of deliveries (or shipments) per day. For example, $F=2$ corresponds to a supplier making 2 deliveries per day while $F=0.1$ corresponds to a supplier making 1 delivery every 10 days (5 working days per week are considered). The number of pallets shipped in each delivery, denoted by Q , is assumed to be constant. In all the study, we use the hour as a time unit (which we denoted by t) and we assume that there are H working hours (time units) per day.

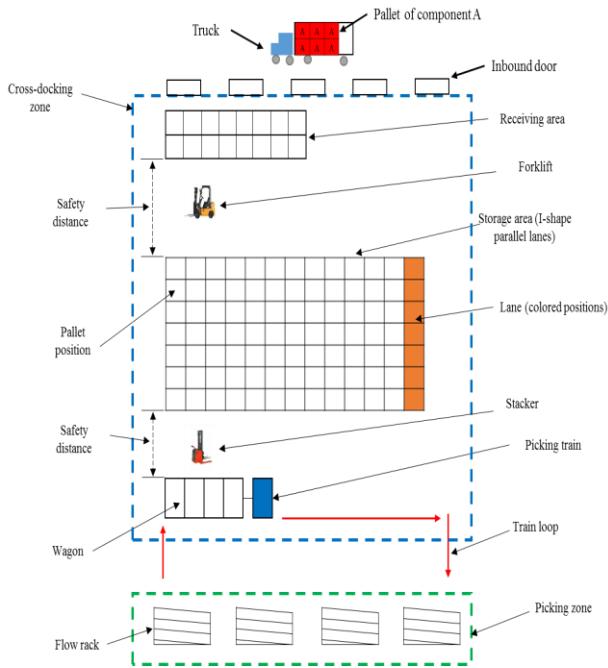


Figure 2. Layout of internal cross-docking zone in JIT plants

The below assumptions are introduced in the model developed:

- The receiving area consists of two parallel lanes where pallets are put temporarily before being positioned in the storage area.
- Each I-shape parallel lane of the storage area has the same number of pallet positions, i.e., the storage area is of rectangular shape. The number of pallet positions per parallel lane can be calculated or fixed by the cross-dock managers as it will be discussed after.
- The picking train contains w wagons. Each wagon can hold one pallet.
- Pallets positioned in the same position can belong to different suppliers. We remind that one pallet position can contain one or several pallets.
- There are two safety distances to facilitate movements for the forklift and the stacker: one is between the receiving and the storage areas and the other between the storage area and the picking train.

In this paper, we are interested in studying the performance of different staging policies. A staging policy, equivalently a cross-docking policy, is a set of rules used to tell operators where to put each pallet (from the receiving area) in the storage area and how to retrieve them. The objective of this paper is to compare different cross-docking policies in terms of required labour and surface costs. The way pallets are stored in the storage area impacts the surface of the storage area and the number of operators required to move pallets within the cross-docking area.

We start by presenting two cross-docking policies commonly used in practice:

3.1 Cross-docking per time slot

In this policy, each lane in the storage area contains the pallets that will be consumed during a fixed time slot. For illustration purposes, we consider the case of two suppliers delivering components A and B with daily delivery frequency ($F_A=F_B=1$). Supplier A (we call a supplier of component A, supplier A) delivers 8 pallets per shipment and supplier B delivers 4 pallets ($Q_A=8$ pallets and $Q_B=4$ pallets). The number of time units per day H equals 8 hours and the duration of a one time slot, denoted by STS, is equal to 2 hours (hence, there are 4 time slots per day). We also assume that one pallet position can contain one pallet. The pick-lists are known in advance and are presented in Table 1 (column STS 1 represents the pick-list released at the beginning of time slot STS 1).

	STS 1	STS 2	STS 3	STS 4
Supplier A	2	2	2	2
Supplier B	1	1	1	1

Table 1. Pick-lists in cross-docking per time slot (numbers represent the number of pallets)

Suppliers' shipments are assumed to be received at the beginning of the day and a pick-list is released at the beginning of each time slot. A top view of the storage area at the beginning of each time slot (before the stacker retrieves pallets) would look as represented in Figure 3. In this example, the number of lanes in the storage area is 4 and the number of positions per lane is 3.

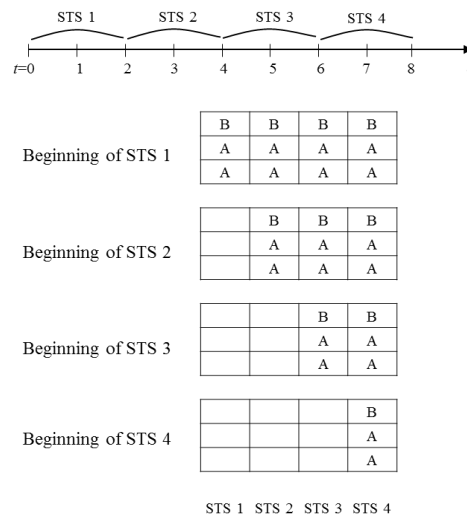


Figure 3. Top view of the storage area in a cross-docking per time slot policy

As illustrated by the example, cross-docking per time slot has many characteristics:

First, the number of positions per lane depends on the maximum number of pallets per pick-list and the number of pallets per position. Second, the number of lanes in

the storage area depends on the time slot duration and on the minimum delivery frequency per supplier.

The advantage of using cross-docking per time slot is facilitating the loading of picking train for stacker operators. Because each lane represents one time slot, operators have to pick all the pallets of the same lane without looking for pallets in other lanes. The disadvantages of cross-docking per time slot are:

- The number of lanes increases with decreasing the minimum delivery frequency. For instance, if the time slot duration is 1 hour, $H=8$ hours and the minimum delivery frequency is 0.1 (one delivery every 10 days), the number of lanes in the storage area is 80 (a very big number).
- Staging of pallets towards the storage area is complicated for the forklift operator because he is obliged to go to different lanes and to know how many pallets to put in each lane.

In the example above, the pick-list for each time slot is assumed to be given. If the pick-list is not defined in advance, it has to be built. Let's take another example to show how pick-lists can be generated.

In this example, there are 4 time slot per day, and one pallet position can contain one pallet: Supplier A delivers 8 pallets per day ($F_A=1$) which will be consumed during the 4 time slots. Under the assumption that customer demand is regular, $8/4=2$ pallets are consumed during each time slot. Hence, 2 pallets will be put per lane.

In addition to supplier A, we consider a second supplier, i.e., supplier B which delivers 10 pallets per day. Each lane will contain at least 2 pallets ($\lfloor 10/4 \rfloor = 2$) of supplier B and there still remains 2 pallets to be positioned. Where will these 2 pallets be stored? For example, we can put 1 pallet in the first lane (in STS 1) and the other one in the third lane (in STS 3). The pick-lists will thus be as given in Table 2.

	STS 1	STS 2	STS 3	STS 4
Supplier A	2	2	2	2
Supplier B	3	2	3	2

Table 2. Pick-lists with suppliers A and B

In addition to suppliers A and B, we consider supplier C which delivers 10 pallets per day ($F_C=1$). Similar to supplier B, each lane will contain at least 2 pallets of supplier C and there still remains 2 pallets to be positioned. If we put 1 pallet in the first lane and the other one in the third lane, there will be 8 pallets in lanes 1 and 3 and 6 pallets in lanes 2 and 4. The storage area is not optimally used because 2 pallet positions will always be empty in lanes 2 and 4. That is why, to reduce storage area, one would put one pallet of supplier C in the second lane and the other one in the fourth lane as shown in Table 3.

	STS 1	STS 2	STS 3	STS 4
Supplier A	2	2	2	2
Supplier B	3	2	3	2
Supplier C	2	3	2	3

Table 3. Pick-lists with suppliers A, B and C

Under the assumption that customer demand is regular, the generation of pick-lists can not be done randomly and a calculation algorithm is required to smooth the arriving pallets over the lanes. When customer demand is not regular, the generation of pick-lists should follow the non-regularity of demand. For example, a component demand is high in the first time slot and low in the other time slots. The pattern of this component demand should be respected in the generated pick-lists. In our paper, we consider only the case of regular customer demand.

3.2 Cross-docking per supplier

In this policy, the pallets of the same supplier have to be put in the same lane(s). Let's consider the same explanatory example taken in cross-docking per time slot. We add the following precision: the maximum number of positions per lane is fixed to 5 (in cross-docking per supplier, the number of positions per lane is defined in advance). The pick-lists presented in Table 1 are used. A top view of the storage area at the beginning of each time slot would look as represented in Figure 4.

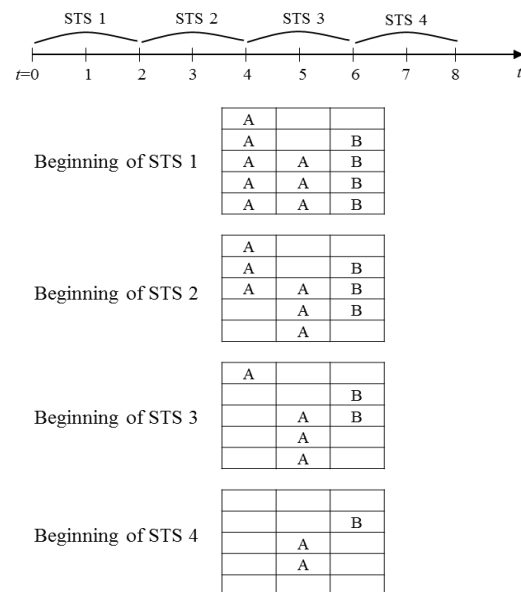


Figure 4. Top view of the storage area in a cross-docking per supplier policy

There is a specified picking rule in cross-docking per supplier: if a supplier's pallets are stored in many lanes (such as supplier A pallets stored in 2 lanes in Figure 4), then the pallets of each lane have to be picked totally before beginning picking from other lanes. For example,

supplier A pallets are to be picked from the first lane before passing to the second lane.

As said before, in cross-docking per supplier, the number of positions per lane is fixed by the cross-dock managers and is not calculated as in cross-docking per time slot. The advantage of using cross-docking per supplier is that the staging of pallets is easy for operators since all the pallets of the same supplier are put in the same lane(s) without dispatching them according to time slots. Another advantage of this policy concerns the number of lanes which depends on the number of positions per lane (lane capacity) and the number of pallets per shipment Q . For example, if the number of positions per lane is 10 and a supplier delivers 25 pallets per shipment; as a result, 3 lanes are required to store the pallets of this supplier (we assume that one pallet position can contain one pallet). Unlike cross-docking per time slot, the number of lanes in cross-docking per supplier does not depend on the delivery frequency of suppliers. This advantage would make cross-docking per supplier interesting for suppliers with low delivery frequency. The disadvantages of using this policy are:

- Loading the picking train would be complicated for stacker operators because they are obliged to constitute the pallets requested by the pick-list.
- The storage area is not optimally used because lanes are permanently assigned to suppliers (e.g., lanes 1 and 2 are always assigned to supplier A and can not be used to store another supplier's pallets even when they are partially used).

To cope with the limits of cross-docking per time slot and cross-docking per supplier, new cross-docking policies based on these two policies are developed:

3.3 Cross-docking per supplier/time slot

In this policy, suppliers are divided into two classes: suppliers with high delivery frequency and suppliers with low delivery frequency. A threshold level, denoted by γ , is used to separate the two classes where a supplier having a delivery frequency $F \geq \gamma$ is considered to be a high delivery frequency supplier. γ is often fixed by cross-docking managers. Cross-docking per time slot is used for the staging of pallets of suppliers with high delivery frequency and cross-docking per supplier is used for the staging of pallets of suppliers with low delivery frequency. The example below illustrates this:

Suppliers A, B and C have respectively delivery frequencies equal to $F_A=1$, $F_B=1$ and $F_C=1/3$ (1 delivery every 3 days) with delivered total quantities $Q_A=12$, $Q_B=6$ and $Q_C=18$. There are 8 time units per day ($H=8$ hours) and the time slot duration is set equal to 2 hours. The threshold level γ is set equal to 0.5. Therefore, supplier C is considered as a supplier having a low delivery frequency and its pallets are staged according to a cross-docking per supplier. The pick-lists of this cross-dock for 3 days (the time interval covered by one delivery of supplier C) are generated using the smoothing algorithm explained in subsection 3.1 and are presented in Table 4.

	Day 1				Day 2				Day 3			
	STS 1	STS 2	STS 3	STS 4	STS 1	STS 2	STS 3	STS 4	STS 1	STS 2	STS 3	STS 4
Supplier A	3	3	3	3	3	3	3	3	3	3	3	3
Supplier B	2	1	2	1	2	1	2	1	2	1	2	1
Supplier C	1	2	1	2	1	2	1	2	1	2	1	2

Table 4. Example of pick-lists in cross-docking per supplier/time slot

B		B		C	C	C	
B	B	B	B	C	C	C	
A	A	A	A	C	C	C	C
A	A	A	A	C	C	C	C
A	A	A	A	C	C	C	C

STS 1 STS 2 STS 3 STS 4

Figure 5. Top view of the storage area in a cross-docking per supplier/time slot policy

Figure 5 depicts a top view of the storage area at the beginning of the first day where the number of positions per lane is set equal to 5. The pallets put in the storage area are picked according to the pick-lists given in Table 4. The number of lanes required in this policy is equal to 8. If cross-docking per time slot were used in this policy, then the number of lanes would be equal to 12 lanes. In cross-docking per supplier, this number would be equal

to 9. Cross-docking per supplier/time slot policy may be limited because lanes where suppliers with low delivery frequency are staged are not fully used and may stay empty for a long time (inefficient use of surface).

3.4 Cross-docking per big/small time slots

Similar to cross-docking per supplier/time slot policy, two supplier classes are considered: suppliers with high delivery frequency and suppliers with low delivery frequency. In addition, we consider two types of pick-lists: 1) a pick-list released each small time slot STS and 2) a pick-list released each big time slot, denoted by BTS. The first type is similar to pick-lists used in the previous policies and the second type has the same format as the first type but instead of being released each time slot STS, it is released, less frequently, each time slot BTS. This policy has the same logic as cross-docking per time

slot presented in subsection 3.1 where each lane contains the pallets that will be picked during a time slot (STS or BTS). In this new policy, there are two types of lanes: those which contain pallets that are picked each STS and those which contain pallets that are picked each BTS. Logically, the STS lanes are used to store the pallets of suppliers with high delivery frequency and the BTS lanes are used to store the pallets of suppliers with low delivery frequency. The reason behind the use of BTS lanes is the fact that the number of STS lanes required to store the pallets of suppliers with low delivery frequency would be very high. In this policy, the smoothing algo-

rithm used to generate pick-lists start by generating the pick lists corresponding to STS lanes, then the pick-lists corresponding to BTS lanes are generated independently of STS lanes. For the example considered in section 3.3, the pick-lists in cross-docking per small/big time slots are presented in Table 5 where the duration of big time slot is set equal to 1 day (the duration of STS is 2 hours and the duration of BTS is 8 hours). The top view of the storage area using this policy is presented in Figure 6. The number of lanes required in this policy is equal to 7 (less than the other policies).

	Day 1					Day 2					Day 3				
	STS 1	STS 2	STS 3	STS 4	BTS 1	STS 1	STS 2	STS 3	STS 4	BTS 1	STS 1	STS 2	STS 3	STS 4	BTS 1
Supplier A	3	3	3	3	0	3	3	3	3	0	3	3	3	3	0
Supplier B	2	1	2	1	0	2	1	2	1	0	2	1	2	1	0
Supplier C	0	0	0	0	6	0	0	0	0	6	0	0	0	0	6

Table 5. Example of pick-lists in cross-docking per big/small time slots

				C	C	C
B		B		C	C	C
B	B	B	B	C	C	C
A	A	A	A	C	C	C
A	A	A	A	C	C	C
A	A	A	A	C	C	C
STS 1	STS 2	STS 3	STS 4	BTS 1	BTS 2	BTS 3

Figure 6. Top view of the storage area in a cross-docking per small/big time slots policy

We remind that the objective of this paper is to compare the four cross-docking policies presented in this section: cross-docking per time slot, cross-docking per supplier, cross-docking per supplier/time slot and cross-docking per small/big time slots. The comparison is done in terms of required surface and labour costs. In the next section, we present the method used to evaluate the average labour and surface costs per day.

4 CROSS-DOCKING COSTS EVALUATION

Two types of cost are considered: surface and labor costs. We mean by surface cost, the cost incurred by using the storage area. This cost is the product of the storage area surface and the generic surface cost. Labour cost represents the wages of person-hours spent to move the pallets from the receiving area to the storage area and from the storage area to the picking train. The average total cost per day of a cross-docking policy, denoted by C , is calculated using the following expression:

$$C = l * E[O] + s * E[S] \quad (1)$$

where:

$E[O]$: average number of person-hours spent per day

$E[S]$: average storage area surface per day

l : unit labour cost per person-hour (€/person hour)

s : unit surface cost per square meter per day (€/m²/day)

The average storage area surface $E[S]$ is the product of the number of cross-dock lanes and the number of pallet positions per lane.

4.1 Calculation of average person-hours per day $E[O]$

To calculate the average number of person-hours spent per day, we need to know where operators (persons) are exactly involved in the cross-docking zone. Operators mainly take action in the following activities:

- Pallets receiving: this activity consists on unloading suppliers' trucks and storing the load temporarily in the receiving area. We do not include this activity in the evaluation of labour cost because it does not depend on cross-docking policy and its cost is the same whatever is the cross-docking policy.
- Pallets staging: this activity is directly impacted by the cross-docking policy used. It can be divided into 4 elementary activities:
 - Loading pallet(s) from the receiving area in the forklift.
 - Moving the forklift (filled) from the receiving area to the storage area.
 - Unloading the forklift in the pallet position.
 - Returning of the forklift to the next pallet(s) to be moved (if there still remains pallets to move) or to its initial position (the initial position of the forklift is set by the cross-dock managers).
- Pallets loading in the picking train: this activity is also impacted by the cross-docking policy used. Similar to pallets staging, this activity can be divided into 4 elementary activities:
 - Loading pallet(s) from the storage area in the stacker.
 - Moving the stacker from the storage area to the picking train.

- Unloading the stacker and loading the picking train.
- Retuning of the stacker to the next pallet position or to its initial position (the initial position of the stacker is set by the cross-dock managers).
- Driving the picking train to the picking zone: when the picking train is totally filled (w wagons are occupied) or when all the pallets in the pick-list have been loaded in the picking train, then the train goes to the picking zone where pallets are unloaded, and broken. Component cases are stored in the flow racks of the picking zone. This activity is not directly impacted by cross-docking policies and its labour cost is not included in the calculation.

To summarize, the average number of person-hours spent per day is the average number of person-hours spent to fill cross-dock lanes (pallets staging) and to retrieve pallets (pallets loading in the picking train). Let's explain how $E[O]$ can be calculated:

A working day can be viewed as a set of "pallets staging" activities and "pallets loading in the picking train" activities. We associate a coordinate system (O,X,Y) to the cross-docking zone as shown in Figure 7. One pallet staging activity consists on moving the pallet(s) from a position (x_1,y_1) in the receiving area to a position (x_2,y_2) in the storage area, then going back to another position (x_3,y_3) in the receiving area (this position is either the next pallets to stage or the initial position of the forklift). The distance travelled in one pallets staging activity is $d_f = |x_1 - x_2| + |y_1 - y_2| + |x_2 - x_3| + |y_2 - y_3|$.

The number of person-hours spent in a pallets staging activity is calculated using the following expression:

$$\text{Pallet staging time} = lt_f + \frac{d_f}{v_f} + ut_f \quad (2)$$

where:

lt_f : loading time used by the forklift in terms of person-hours

v_f : velocity of the forklift

ut_f : unloading time used by the forklift in terms of person-hours

Similarly, the time required in a pallets loading in the picking train activity is calculated using the following expression:

Similarly, the time required in a pallets loading in the picking train activity is calculated using the following expression:

$$\text{Pallet loading in picking train time} = lt_s + \frac{d_s}{v_s} + ut_s \quad (3)$$

where:

d_s : distance travelled by stacker in one "pallets loading in the picking train" activity

lt_s : loading time used by the stacker in terms of person-hours

v_s : velocity of the stacker

ut_s : unloading time used by the stacker in terms of person-hours

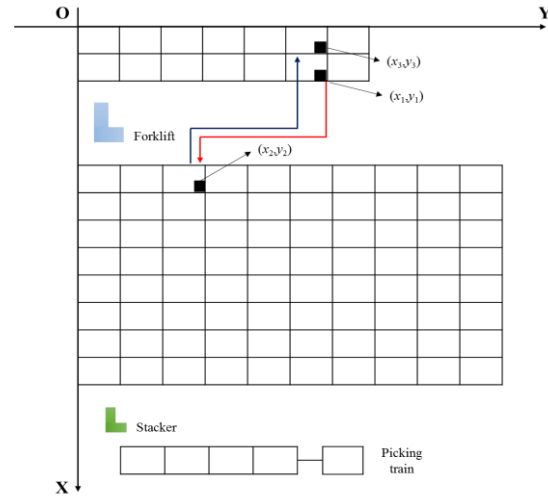


Figure 7. Cross-docking associated coordinate system

We calculate the average number of person-hours spent per day, $E[O]$, using a Discrete Event Simulation over n days where n is the least common multiple of time intervals that separate two consecutive shipments of each supplier (inverse of delivery frequency). n is set equal to the least common multiple because it is the first number where all delivered pallets are completely picked. We calculate average cost C using the following algorithm:



Algorithm 1. Algorithm of calculation of cross-docking cost

5 NUMERICAL EXAMPLE

We apply the tool developed to the case of an auto parts supplier plant in France. The characteristics of the cross-docking zone in this plant are: the unit labour cost per person-hour l is set equal to 12 €/person-hour and the unit surface cost per square meter per day s is set equal to 0.33 €/meter²/day. The dimensions of all pallets are set equal to 1*1 meter². Using the coordinates system given in Figure 7, the initial position of the forklift is set in $(x=2, y=0)$, i.e., near to the receiving area. The initial position of the stacker is $(x=2+\text{safety distance receiving/storage} + \text{number of positions per lane} + \text{safety distance storage/picking train}, y=0)$.

- Number of positions per lane (fixed for cross-docking per supplier policy)=30 positions. One pallet position can contain two pallets.
- Safety distance receiving/storage=4 meters
- Safety distance storage/picking train=3 meters
- STS duration=1 hour, BTS duration=1 day
- $H=8$ hours, threshold level is set equal to 1
- $lt_f=0.2$ min, $v_f=0.006$ min/meter, $ut_f=0.27$ min, $lt_s=0.27$ min, $v_s=0.008$ min/meter, $ut_s=0.27$ min.

Results	Cross-docking policy			
	Per supplier	Per time slot	Per small/big time slots	Per supplier/time slot
Number of lanes	41	80	18	42
Number of positions per lane	30	34	55	25
E[S]	1230	2720	990	1050
E[O]	8.87	11.85	10.50	9.10
Average cost (€)	512.34	1039.82	452.66	455.75

Table 6. Comparison between cross-docking policies in case study

In Table 6, we compare the four cross-docking policies when applied to the plant considered. It can be seen that cross-docking per small/big time slots is the best policy in terms of average cost per day.

6 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we study the use of cross-docking in Just-in-Time plants. Two cross-docking policies commonly used in practice are first presented: cross-docking per time slot and cross-docking per supplier. We showed the advantages and limits of each policy and proposed new policies to cope with these limits. We also developed a cross-docking cost calculation model in order to compare the different cross-docking policies considered in this paper. The cost model allows to calculate the average surface and labour costs. We developed an Excel VBA Script which implements the cost model. The tool was applied to the case of an auto parts supplier.

For future research, it would be of great value to develop efficient approximate analytical expressions to evaluate the labour and surface costs of each cross-docking policy. Such an approximate model would be helpful to avoid computer programming difficulties (hundreds of lines of code) used in the exact cost calculation model. Another interesting direction would be to study a coordinated organization of picking and cross-docking operations.

REFERENCES

- Bartholdi III, J. J., & Gue, K. R. (2000). Reducing labor costs in an LTL crossdocking terminal. *Operations Research*, 48(6), 823–832.
- Bartholdi, J. J., & Hackman, S. T. (2008). *Warehouse & Distribution Science: Release 0.89*. Supply Chain and Logistics Institute.
- Battini, D., Boysen, N., & Emde, S. (2013). Just-in-Time supermarkets for part supply in the automobile industry. *Journal of Management Control*, 24(2), 209–217.
- Boysen, N., & Emde, S. (2014). Scheduling the part supply of mixed-model assembly lines in line-integrated supermarkets. *European Journal of Operational Research*, 239(3), 820–829.
- Boysen, N., & Flidner, M. (2010). Cross dock scheduling: Classification, literature review and research agenda. *Omega*, 38(6), 413–422.
- De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2), 481–501.
- Emde, S., & Boysen, N. (2012). Optimally locating in-house logistics areas to facilitate JIT-supply of mixed-model assembly lines. *International Journal of Production Economics*, 135(1), 393–402.
- Ladier, A.-L., & Alpan, G. (2016). Cross-docking operations: Current research versus industry practice. *Omega*, 62, 145–162.
- Larbi, R., Alpan, G., Baptiste, P., & Penz, B. (2011). Scheduling cross docking operations under full, partial and no information on inbound arrivals. *Computers & Operations Research*, 38(6), 889–900.
- Sali, M., & Sahin, E. (2016). Line feeding optimization for Just in Time assembly lines: An application to the automotive industry. *International Journal of Production Economics*, 174, 54–67.
- Stephan, K., & Boysen, N. (2011). Cross-docking. *Journal of Management Control*, 22(1), 129.
- Van Belle, J., Valckenaers, P., & Cattrysse, D. (2012). Cross-docking: State of the art. *Omega*, 40(6), 827–846.
- Vis, I. F., & Roodbergen, K. J. (2008). Positioning of goods in a cross-docking environment. *Computers & Industrial Engineering*, 54(3), 677–689.

TOWARDS A NEW METHODOLOGY TO IMPROVE MAINTENANCE PERFORMANCE BASED ON THE PROCESS APPROACH AND LEAN THINKING

B. FRAKSO, A. MOKHLIS S. ELFEZAZI, A. EL HACHADI

Cadi Ayyad University
Higher School Of Technology
Safi , Morocco

ABSTRACT: *The increasing demand from stakeholders besides the fierce competition and technology expansion in the industrial world have made it necessary for manufacturing companies to ensure maximum availability, reliability and maintainability of their production facilities. Therefore, seeking for continuous improvement of the maintenance function has become nowadays an ineluctable challenge for companies to stay competitive. However, the success of any continuous improvement process relies undeniably on the implementation of performance measurement system with well defined performance indicators. In this perspective, the paper aims to propose a simple methodology to evaluate and improve maintenance performance based on the process approach, which will help to outline the dysfunction areas of the maintenance system. Once detected and analysed, an improvement approach, in the frame of lean thinking adopted into maintenance activities, will be introduced. The application of our method presents an occasion for managers to appraise and provide a clear picture of the current status of the maintenance service organization, and moreover, it is an opportunity to promote the culture of continuous improvement aiming at the excellence of the maintenance function.*

KEYWORDS: *Maintenance Performance, Process Approach, Assessment, Lean Maintenance, , Indicators*

1 INTRODUCTION

The increasing expansion of technology, the competition among industries, the supremacy of safety and the implication of industry in the environmental issues are among many factors that have brought companies to reconsider their strategies. This concern, which has been conveyed to the maintenance function, evolving from a “necessary evil” to a “value adding” activity (Adriaan Van Horenbeek and Liliane Pintelon, 2013), and on the basis of this finding, has led several companies to move towards the measurement of the performance of their production facilities by adopting maintenance performance measurement. Moreover, research evidences show that companies that are managed using an integrated performance measurement systems outperform than to those not undertaking the performance measurement (Aditaya Parida *et al.*, 2015).

In fact, nothing is more penalizing for a company than to see its production plant unavailable while time is lacking to meet the deadlines and losses pile up: Less productivity, bad quality, rising of costs etc. Furthermore, it was proved that maintenance practices hold a significant percentage of total production costs estimated between 15 and 70 % (Kim Fraser, 2014). This percentage is likely to increase in the future due to the advent of technology and automation within in the fourth industry revolution. So, in order to avoid such as situation and in order to sustain and survive in such a challenging environment, companies

must assess regularly the performance of the maintenance function. However, in practice, and even though most companies have recognized the important impact of the maintenance performance on their global performance, few of them have managed to adapt performance measurement system since measuring performance of the maintenance function is very complex as it involves various indicators and requires a systematic approach to measure, monitor, and continuously improve performance (Parida and Chattopadhyay, 2007; Simoes *et al.*, 2011).

Maintenance performance is an issue that has been widely analysed in the literature. Nevertheless, there are a few of the published researches that suggest a comprehensive and simple model which portrays the process of actually designing the measurement systems in practice (Aditaya Parida *et al.*, 2015). This paper will represent a comprehensive and simple methodology which offers a solution to this problem. The methodology aims to facilitate the complexity of the maintenance function using a decomposition of the macro-process into sub-processes, in turn, decomposed into key activities and tasks based on the process approach. Each activity will be evaluated from the necessary tasks that ensure the right and good running of the whole process. This decomposition will make it easy for practitioners to first identify the areas that need improvement through and second hold suitable improvement actions in the frame of lean maintenance.

The paper is structured as follow: Section 2 provides a brief review on maintenance performance and the main

and the most cited maintenance performance measurement approaches in the literature. Then the proposed methodology will be explained in section 3. The proposal will be illustrated by applying the methodology to one the operational processes of the maintenance function in section 4. In the conclusion, we will state the progress of our research.

2 STATE OF THE ART

2.1 Maintenance performance

Traditionally, maintenance has been considered as a cost generator for companies. As a matter of fact, managers had in mind one first concern: it was to minimize its expenses instead of trying to detect the root cause of the dysfunction and improve their operating mechanisms in order to avoid factors that affect not only maintenance costs but also the quality of the product. Nowadays, and following the progress of the automation and mechanization of production systems, maintenance is no longer seen as an additional expense to the companies but as a strategic element of revenue generation for organizations (Mostafa *et al.*, 2015). As stated by Muchiri (2010), maintenance is vital for a sustainable performance of a production plant. Moreover, it is seen as a direct contributor to the organization's profit (Davies and Greenough, 2010). Therefore, it is necessary to assess and improve maintenance activities regularly to ensure its effectiveness and efficiency.

Maintenance can be controlled and improved based on its performance. According to Dwight (1999), performance is defined as the level to which the aims and objectives are attained. Thus, maintenance performance translates the capability of the maintenance function to ensure continuous production of good quality product with acceptable cost. Moreover, it has been acknowledged by many authors that maintenance performance is a major contributor to the performance and profitability of manufacturing systems.

Maintenance performance, and therefore its measurement, has become a very recurrent topic. In fact, measuring performance is seen as a fundamental management principle. Like other manufacturing functions, performance measurement is very important in managing the maintenance function (Muchiri *et al.*, 2011). In fact, maintenance performance measurement gives a quantitative value to the maintenance performance (Dwight *et al.*, 1999). By means of this tangible value, managers could identify the gap between the current situation and the desired one, and also glean the maintenance performance levels and ensure a continuous improvement plan (Arts *et al.*, 1998). Besides, maintenance performance measurement is perceived as an essential function to ensure a sustainable performance of any industry. So, in order to fulfil this aim, managers need to follow up the performance of the maintenance activities (Parida and Chattopadhyay 2007), which can be reached, in turn, by defining a rigorous performance measurement system (Adriaan Van

Horenbeek et Liliane Pintelon, 2014). This is reflected and supported by several maintenance performance measurement approaches proposed in the literature which will be the subject of the following section.

2.2 Review on the main maintenance performance measurement approaches and indicators

The issue of maintenance performance measurement has stirred up much interest and attention of many researchers and practitioners in the field. With various factors affecting the efficiency of maintenance, there are also various types of maintenance performance measurement methods discussed in the literature and practiced in industry (Hasnida Abdul Samat *et al.*, 2011). For the sake of being precise and concise, we will limit our review to the main and the most recurrent approaches in the literature. For instance, Tsang proposes a review on four approaches to measure maintenance performance, one of them being the audit system, defining it as "A comprehensive review of the various dimensions in the maintenance system, such as organization, personnel training, planning and scheduling, data collection and analysis, control mechanisms, measurement and reward systems, etc."; thus, the aim is to be able to suggest improvement actions, determine the priorities for the recommended activities and setup a plan of action (Tsang *et al.*, 1999). Meanwhile, Dwight (1999) proposed the use of a system audit approach in measuring the performance of maintenance system contribution to organizational success called value-based performance measurement. It takes into account the impact of maintenance activities on the future value of the organization. Another most popular approach is the balanced scorecard approach, developed by Kaplan and Norton (1992), has received an increasing acceptance among the industry (Tsang *et al.*, 1999). It is based on four perspectives: (1) financial, (2) customers, (3) internal processes, and (4) innovation and learning that covers financial and non-financial aspects of the business process. This has made from the approach a very successful and useful method for measuring organization's performance. For instance, the BSC was applied by Tsang 1998; he presents the BSC as a framework for translating organization's strategy into operational measures. The framework was used to evaluate the impact of maintenance activities on the future value of organization. Also, Alsayouf (2006) suggests an extended BSC framework to assess the contribution of maintenance to strategic business objectives, based on 7 perspectives, while kumar (2011) proposes to audit the degree of the fulfilment of objectives and the degree of satisfaction obtained from each of the four perspectives of the balanced scorecard.

From literature, different classifications of maintenance measures can be identified. For instance, Campbell (1995) classifies the commonly used measures of maintenance performance into three categories: equipment performance (e.g. availability etc.), cost performance (e.g. maintenance, materiel costs etc.) and process performance

(e.g. schedule compliance, etc.). Arts uses the time horizon to classify maintenance control and performance indicators into three levels namely strategic, tactical and operational (Arts *et al.*, 1998). Maintenance measures have also been utilized to support control at each level. A good example is the multi-criteria hierarchical framework for maintenance performance measurement proposed by Parida and Chattopadhyay (2007) which consist of multi-criteria indicators for each level of management (strategic, tactical and operational).

The total productive maintenance (TPM) concept introduced the well known quantitative metric “overall equipment effectiveness (OEE)”. It is a widely used measure within industries as it identifies and measures losses of important aspects of manufacturing namely availability, performance/speed and quality rate. This supports the improvement of equipment effectiveness and thereby its productivity (Muchiri *et al.*, 2010). Also, mean time between failure (MTBF) and mean time to repair (MTTR) are very popular metrics used in performance analysis in companies. Both have a direct impact on the equipment availability. Komonen (2002) presents a hierarchical system of maintenance performance indicators and classifies indicators into three main dimensions of maintenance performance (OEE, production costs and production quality). The objective of the system is to explain the purpose and significance of various indicators at different hierarchies.

The other commonly used classification is leading and lagging indicators. Leading indicators monitor if the tasks are being performed that will ‘lead’ to results. On the other hand, lagging indicators monitor whether the results or outcomes that have been achieved. Both leading and lagging indicators are therefore important for managing the performance of the maintenance function. Moreover, the leading indicators are even more important than lagging indicators because they have the potential to avoid unfavorable situations from occurring (Muchiri *et al.*, 2010).

Various are the approaches for maintenance performance measurement and classification of indicators proposed in literature. These approaches and classifications are different one from another according to the point of view and interests of researchers and practitioners.

3 RESEARCH METHODOLOGY

As seen in the last section, many studies and researches have dressed up the issue of maintenance performance measurement. Nevertheless, few of them have presented a rigorous and clear methodology to analyse, assess and improve its performance, which remains a real challenge for managers. To figure out this problem, this paper proposes a simple methodology to improve maintenance performance based on the process approach.

The research methodology is divided into 6 steps as follows:

- Step 1: identification of maintenance objectives

The aim of this step is to make clear maintenance objectives and the expected outcomes. These objective should be in compliance with the organization goals. Moreover, this is a preliminary step to the identification of maintenance processes.

- Step 2: Definition of maintenance main processes

The definition of the maintenance main processes ,that allow the achievement of the set objectives, is the first step toward the process approach . For instance: The preventive maintenance process can reduce the probability of equipment failure and therefore meet the objective of improving functional and technical aspects.

- Step 3: Decomposition of the maintenance processes using the process approach

This step is necessary to ensure an accurate assessment of the performance. Also, the decomposition of processes into activities is a crucial step in the approach since this decomposition will facilitate the extraction of the required information to carry out a detailed analysis of the dysfunction causes that could counterfeit the expected result of a process, based on the fact that the activities are the support of the performance approach.

- Step 4: Elaboration of the audit referential related to the key activities

In order to evaluate maintenance function, it is necessary to go into the deepest level of details. In other words, the core of the evaluation will be the assessment of the activities. Indeed, an activity is a set of elementary tasks using specific know-how and consuming resources to achieve a common goal. These elementary tasks will constitute the basis of the evaluation referential. It is then necessary to first formulate a questionnaire to understand the reality of execution of each task and then compare this reality to the reference, define the statement of conformity and analyze the gaps.

- Step 5: Improvement plan

In this phase, we will first implement improvement actions based on the selection of adequate lean maintenance tools to the spotted wastes in dysfunction areas and then propose performance indicators which, beyond of being logically related to the defined objective, should absolutely allow to appreciate the improvement of the target situation and to measure effectively the required performance.

- Step 6 : Industrial application

The last phase will be devoted to the development of a real case study to test the validity of our approach in an industrial company (in progress).

Figure1 represents a summary diagram of the proposed model.

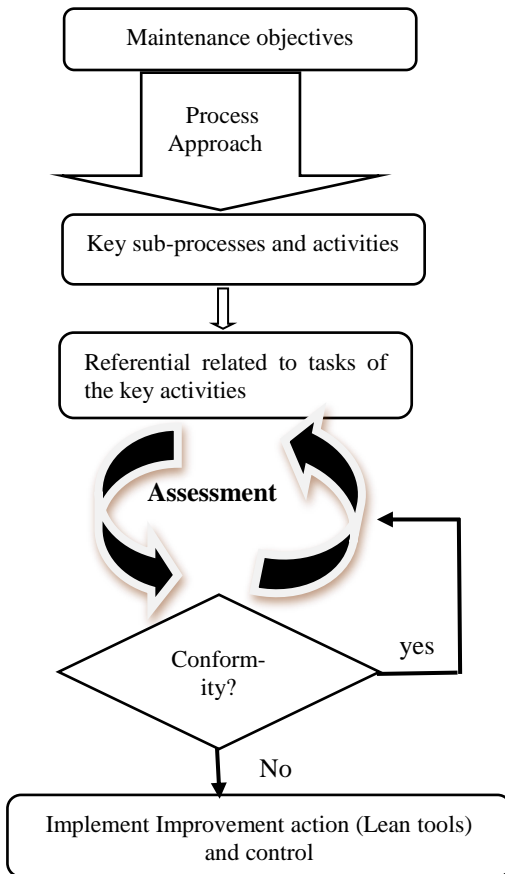


Figure 2: The proposed model

4 PROCESS APPROACH: A NEW TRACK TOWARD PERFORMANCE

4.1 Maintenance objectives

The maintenance function is referred to in the company as a support process composed of a set of activities with clearly and consensually identified objectives within the organization and which will consume resources in order to achieve these goals. In fact, Maintenance objectives should be consistent with and subordinated to manufacturing goals which, in turns, should be aligned with the corporate strategy (Muchiri *et al.*, 2010). Moreover, Coetzee (1998) emphasizes that “The main objective of the maintenance function is to support the production process with adequate levels of availability, reliability, operability and safety at an acceptable cost.”. According to this definition and to other authors experiences, these objectives are the most relevant and the most common among all manufacturing companies (Van Horenbeek & Pintelon, 2013):

- ✓ Improve functional and technical aspects: availability, reliability, maintainability, overall equipment efficiency, productivity, quality of production and quality of maintenance.

- ✓ Optimize the maintenance budget: maintenance costs.
- ✓ Increase the plant life: capital replacement decisions and life-cycle optimization.
- ✓ Support: inventory of spare parts, logistics and personnel management.
- ✓ Improve working conditions and environment: environmental impact, safety / risk / health.

4.2 Maintenance processes

According to the standard “the process approach refers to the application of a process system within an organization, as well as the identification of interactions and management of these processes in order to achieve the desired result”. Also, the process approach integrates processes into a complete system to achieve strategic and operational goals. Moreover, the identification and decomposition of the sub processes into activities is a crucial step in the approach since this decomposition will facilitate the extraction of the required information to conduct a detailed analysis of the dysfunction causes that could counterfeited the expected results and in the other hand, it is an occasion for the maintenance managers to reconsider the whole organization of the maintenance service. Parida (2007) states that process mapping is a critical step to understand the existing flow of various work processes incorporating operation and maintenance process work practices in the plant.

For any organization and for any macro process, there are generally 3 types of processes: management processes, Operational processes, and the third type, concerns the support processes. The three types of processes are then divided into a set of maintenance processes and sub-processes that formalize the main activities to be implemented in order to meet the predetermined objectives. For instance, the sub process corrective maintenance (CM) aims to reduce downtime of the equipment and this corresponds to the objective set above: improving equipment availability, also preventive maintenance process meets with increasing the equipment life objective etc.

Maintenance macro-process can be seen as a set of sub processes as described in Figure 2 which constitute the first step toward a total decomposition of the maintenance process.

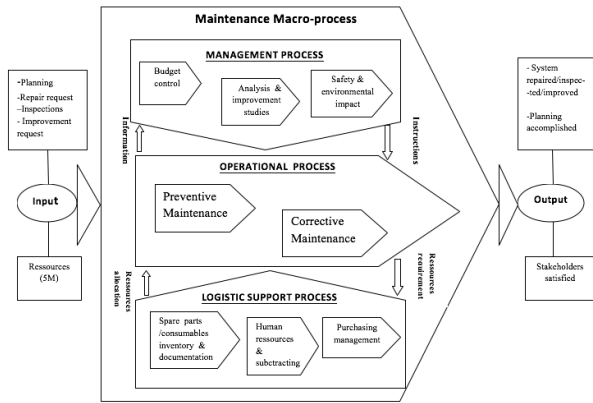


Figure 2: Maintenance macro-process mapping

4.3 Decomposition of the maintenance processes into sub-processes and activities

After defining the maintenance function main processes, the next step is the identification of the key sub-processes and activities of each process. Considering the consistency of the operational processes referring to En-nhaili (2016), and other similar studies that demonstrate that the operational processes are the processes that influence the most the overall performance of the organization, we decided to limit our study in this paper to one of the operational processes: Corrective maintenance.

Figure 3 represents the mains sub-processes and activities of the corrective maintenance process. We took as reference the standard EN 17007:2016, and the decomposition elaborated by Meddaoui and Bouami (2014).

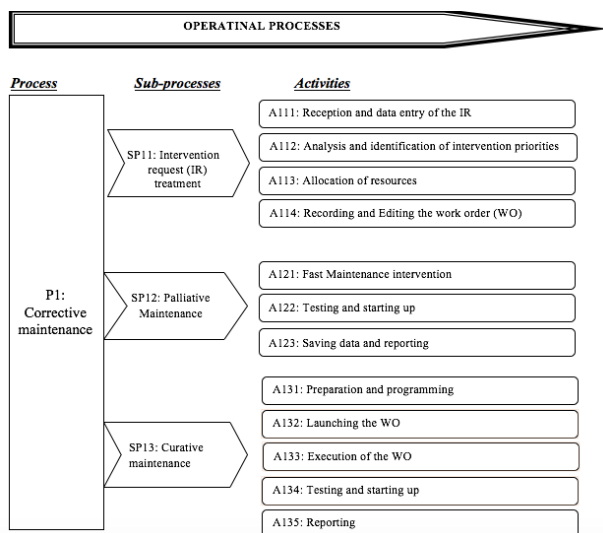


Figure 3: Decomposition of the corrective maintenance process

4.4 Elaboration of the audit referential

The main goal of this approach is to detain a referential through the established tasks of the main activities that

form, according to our investigation and based on the literature reviewed on the maintenance organization, a detailed and an exhaustive representation of the logical functioning of the maintenance function. In fact, the audit will allow us to have an accurate insight on how the activities are performed and also to detect where exactly the process has weaknesses and losses and hence identify opportunities for rapid improvement actions. So, the idea is to first improve the dysfunction areas detected by auditing the tasks related to each activity and then to generalise the improvement on the other activities as a part of a continuous improvement process.

Evaluation methodology:

The evaluation grid is composed of the tasks related to the activities of the sub- process. Each task is expressed in the grid as an affirmation. To each affirmation will be assigned a response: True, Quit True, False or Quite False, and each response is consecutively linked to a rate (100%, 70%, 30%, 0%) reflecting the level of veracity of the execution of the task. The average of the rates related to the veracity of the tasks performance reflects the level of the execution of the activity according to predetermined intervals. These intervals of the average of the execution rate are modifiable according to the objectives and the expectations of the company. Here after the explanatory table:

level of veracity	Veracity	Veracity rate	Minimal average rate	Maximal average rate	Execution of the activity/sub-process	Level of execution
Level 1 : The task is not executed	False	0%	0%	25%	Very insufficient	Level 1 : it is necessary to revise the tasks not executed
Level 2 : the task is sometimes executed	Quit False	30%	26%	59%	Insufficient	Level 2 : it is necessary to make sure that the tasks of the activity are executed correctly
Level 3 : the task is executed in a convincing way.	Quit True	70%	60%	89%	Convincing	Level 3 : The activity/sub-process should be traced and improved
Level 4 : the task is executed ,traced and improved	True	100%	90%	100%	Very convincing	Level 4 : The activity/sub-process is well controlled

Table 1 : Evaluation explanatory table

The overall evaluation of the sub-process, and eventually the maintenance function is obtained by calculating the average of the execution rates of the activities/sub-processes

Ref.	Sub-processes/activities/tasks	Evaluation	Rate (%)	Execution level
P1	Corrective maintenance	---	---	---
SP1	Intervention request treatment	Insufficient	55	Level 2
A111	Reception and data entry of the intervention request	Convincing	66	Level 3
	Once the IR is received, a confirmation is sent to the requester(via FMSS or receipt)	V	100	
	The IR contains all the necessary data (date, time, equipment reference etc.)	PV	70	
	Data are saved and gathered for treatment	PF	30	
A112	Analysis and Identification of intervention priorities	Insufficient	57	Level 2
	After consulting the file history of the concerned equipment, a technical report is edited	T	100	
	The IR are classified regarding to the equipment criticality	QT	70	
	An ordered list of interventions is established and continuously updated	QF	30	
	A preliminary schedule of the corrective maintenance interventions based on the received IR is established	QF	30	
A113	Resources Allocation	Insufficient	25	Level 1
	A list of the required resources is addressed to the preparation agent	QF	30	
	Qualified technicians are affected to execute the intervention	QT	70	
	Information on the requested spare parts is provided (availability ,delivery time etc. .)	F	0	
	Provision of any tools facilitating the progress of the intervention (scaffolding, technical documentation etc.)	F	0	
A114	Editing work order	Convincing	70	Level 3
	Once the IR is approved by the manager, a work order is transmitted to the designed team work.	QT	70	
	The work order provide all instructions related to the intervention (tasks, spare parts, security etc.)	QT	70	

Table 2: Example of an extract from the evaluation grid of the corrective maintenance process

Once the evaluation is over, we can summarize the results in the table

Processes	Rate	Corresponding level
P1-Corrective maintenance		
P2-Preventive maintenance		
P3-Spare part and documentation		
P4-Human resources and training		
.....		
P9-Safety and environmental impact		
Maintenance function evaluation		

Table 3 : Evaluation results

in order to better visualize the weak points and to give a comparative impact, the results obtained can be presented as a polar graph of the sub process and then of the whole function. This representation will allow tracing the profile of each maintenance process and compare it to the company objective.

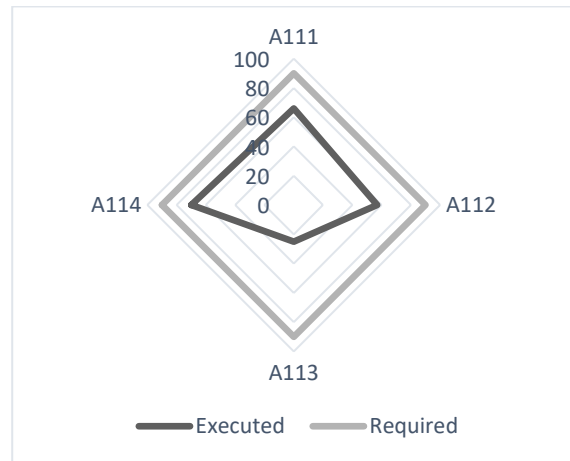


Figure 4: “Intervention request treatment” sub-process profile

4.5 Lean maintenance to improve maintenance processes

4.5.1 Identification of maintenance wastes

The aim of this phase is to discover the sources of waste of each activity by analyzing the results of the assessment and focusing more on the weak points that need improvements.

The principal 7 types of wastes of the maintenance process can be summarized as follows (Mostafa *et al.*, 2015):

- W1: Wastes of waiting (i.e. waiting for resources)
- W2: Wastes of unnecessary motions and transportation (i.e. poor ergonomics of the workplace due to centralised maintenance)
- W3: Wastes of processing (i.e. lack of procedures and maintenance methods)
- W4: Wastes of human potential (i.e. lack of training, under-utilisation of maintenance crew)
- W5: Wastes of inventory (i.e. unbalanced stock of spare parts)
- W6: Wastes of defects (i.e. errors during interventions)
- W7: Wastes of inappropriate systems (i.e. poor information)

These wastes identification method involves systematically an analysis of maintenance processes in order to identify activities that are source of waste as shown in Table 4. Actually, this step represents a starting point to eliminate waste and subsequently helps to implement appropriate and effective improvements.

		WASTES						
		W1	W2	W3	W4	W5	W6	W7
ACTIVITIES	A111	✓		✓				✓
	A112			✓				✓
	A113	✓	✓	✓	✓	✓		✓
	A114	✓		✓	✓			
SP11		✓	✓	✓	✓	✓		✓

Table 4: Link between wastes and “Intervention request treatment “sub-process

4.5.2 Lean tools selection

The core concept of lean is to eliminate source of waste in the processes by adopting appropriate lean tools/practises into wasteful activities. These tools are designed to prepare for, plan and execute the function using lean thinking. when referring to the maintenance function, these lean tools will allow lean thinking to be applied with the end result that maintenance excellence is attained (Tendayi and Fourie,2013)

In this section, and so as to identify area of lean tools application, we will propose and link appropriate lean maintenance tools to each source of 7 sources of waste highlighted previously. According to Mostafa (2015) we can identify 26 lean tools/Practises gathered into 4 bundles: just-in-time (JIT), total quality management (TQM), human resource management (HRM), and TPM (figure 5). Each tool will provide much benefits within the maintenance processes and as a result, the performance of a whole maintenance process can be improved.

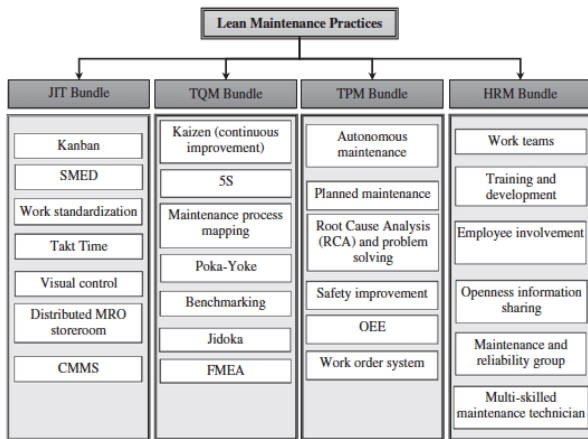


Figure 5: Lean maintenance tools (Mostafa *et al.*, 2015)

Selecting relevant lean maintenance tools is the next step after identifying and discussing the types of waste and lean tools. The success of this selection consists on associating suitable tools to eliminate spotted waste types of the maintenance activities. For the purpose of achieving this end, we opt to develop the association between the lean tools and type of waste using the relationship matrix of the House of quality (HoQ) decision making method that will be more detailed in future articles.

4.5.3 control

As we stated before, any process of continuous improvement of performance relies, in principle, on the establishment of relevant indicators essential for the sustainability of the improvement actions. So, the aim will be to propose an set of measures to assess the implementation of lean tools (For example, introduction of 5S practices into maintenance processes shortens the time of repair (one of the indicators most commonly used to assess the effectiveness of the work of maintenance is MTTR - Mean

Time To Repair), which affects the efficiency of production equipment, and consequently increases OEE (Małgorzata Jasiulewicz and Anna Saniuk, 2017) .

5 CONCLUSION AND FUTUR WORK

This basic contribution of this paper is to present a straightforward and simple method for maintenance performance measurement and improvements based on the process approach and lean thinking adapted into maintenance activities. This method is around 3 main axes: the first one is to map the maintenance function in order to identify the key processes and activities based on the process approach, the second is to elaborate a referential to evaluate the performance of each activity in the process in order to highlight the dysfunction areas and finally to propose an adequate improvement actions in the frame of lean thinking to the weak point of the process with relevant indicators that allow to measure the efficiency and effectiveness of the improvement actions.

With a successful accomplishment of the proposed method, the performance of the maintenance function could create more improvement opportunities over time to reach the maintenance excellence status. Nevertheless, spreading out the implementation of the lean practices into the whole maintenance activities is not that easy and requires the commitment and direct managerial implication. Therefore, future research will focus further more on clarification of the lean maintenance concept and also will concentrate on the development of measures to control the implementation of lean tools .

Case study will be undertaken within industrial companies to make proof of the validity and reliability of the proposed method.

REFERENCES

Aditaya Parida Udaya Kumar Diego Galar Christer Strenstrom ,2015. Performance measurement and management for maintenance: a literature review. Journal of quality in Maintenance engineering , vol.21

Adriaan Van Horenbeek ,Liliane Pintelon ,2014. Development of a maintenance performance measurement framework-using the ANP for maintenance performance indicator selection. Omega vol. 42, pages 33-46.

Alsyouf Imad, 2006. Measuring maintenance performance using a balanced scorecard approach,. Journal of Quality in Maintenance Engineering, Vol. 12 Iss 2 pp. 133 – 149.

Campbell, J.D., 2015. Uptime: Strategies for Excellence in Maintenance Management. Productivity Press, Portland.

Coetzee JL., 1998. Reliability-centred maintenance. Hat field: Republic of South Africa: Maintenance Publishers.

- Davies, C., & Greenough, R. M. (2010). Measuring the effectiveness of lean thinking activities within maintenance. Retrieved June 24, 2013,
- Dwight, R., 1999 : Frameworks for Measuring the Performance of the Maintenance System in a Capital Intensive Organization , University of Wollongong, Australia, Doctoral Report .
- En-Nhaili A., A. Meddaoui, D. Bouami , 2016. 'Effectiveness improvement approach basing on OEE and lean maintenance tools', *Int. J. Process Management and Benchmarking*, Vol. 6, No. 2, pp.147–169.
- Galar, D, Parida, A, Kumar, U, Stenström, C and Berger, L. 2011. Maintenance metrics: A hierarchical model of balanced scorecard. *IEEE International Conference on Quality and Reliability (IEEEQR)*, Piscataway, NJ, pp. 67-74.
- Hasnida Abdul Samat, Shahrul Kamaruddin and Ishak Abdul Azid, 2011. Maintenance Performance Measurement: A Review. *Pertanika J. Sci. & Technol.* 19 (2): 199– 211.
- Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard—measures that drive performance *Harvard Business Review* 70(5) pp :71–79.
- Komonen, K. 2002. A cost model of industrial maintenance for profitability analysis and benchmarking, *International Journal of production Economics* 79 (1), pp :15–31
- Małgorzata Jasiulewicz-Kaczmarek and Anna Saniuk, 2017. How to Make Maintenance Processes More Efficient Using Lean Tools? *International conference on social and occupational Ergonomics, USA.*
- Meddaoui A. and Bouami D., 2013. Cost modelling in maintenance using time-driven activity based costing. *Int. J. Productivity and Quality management Vol 12*
- Milton P. Dentch , 2017 . *The ISO 9001: 2015, Implementation handbook; Using the process approach to build Quality Management System.* ASQ quality Press
- Mostafa S. , Sang-Heon L., J. Dumrak, N. Chileshe & H. Soltan , 2015. Lean thinking for a maintenance process. *Production & Manufacturing Research*, 3 : 1, pp : 236-272.
- NF 17007: 2016. Maintenance Process And Associated Indicators . AFNOR
- Parida, A., Chattopadhyaya G, 2017. Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM). *Journal of Quality in Maintenance Engineering* 13 (3), pp:241–258.
- Peter N. Muchiri a , Liliane Pintelon a , Harry Martin b & Anne-Marie De Meyer, 2010 . Empirical analysis of maintenance performance measurement in Belgian industries. *International Journal of Production Research*, Vol : 48, pp5905-5924.
- Peter N. Muchiri., L. Pintelon, L. Gelders. 2011: Development of a maintenance performance measurement framework and indicators. *Int. Journal of Production Economics*. 131 pp 295-302
- Rødseth H, Skarlo T and Schjølberg P (2015) "Profit loss indicator: a novel maintenance indicator applied for integrated planning" *Advances in Manufacturing* 3(2):139-150.
- Tendayi, T. G., & Fourie, C. J. (2013). The combined AHP-QFD approach and its use in lean maintenance. Paper presented at the 25th Annual Southern African Institute of Industrial Engineering Conference, Stellenbosch.
- Tsang, A.H.C. , 1998. A Strategic Approach to Managing Maintenance Performance. *Journal of Quality in Maintenance Engineering*, Vol. 4 No. 2, pp. 87-94.
- Tsang, A.H.C., 1999. Measuring maintenance performance: a holistic approach. *International Journal of Operations and Production management* 19 (7), pp : 691–715.

Towards the design of a distributed aircraft flight control system connected to simulation components

Jean-Baptiste Chaudron

Complex systems engineering department
ISAE-SUPAERO, University of Toulouse
10 Avenue Edouard Belin, Toulouse, France
jean-baptiste.chaudron@isae-supaero.fr

David Saussie

Department of Electrical Engineering
École Polytechnique de Montréal
C.P. 6079, Montréal (QC), H3C 3A7, Canada
d.saussie@polymtl.ca

ABSTRACT: *The design and the implementation of Flight Control Systems (FCS) still remains a key element of modern avionic systems. During the development process and before flight tests, aeronautical standards require performing piloted simulations. Based on our background in distributed real-time aircraft simulation, we developed a distributed flight control system connected to a simulation environment. This complex distributed architecture is composed of several avionic entities (e.g., primary flight control systems, autopilot) interconnected to simulated components (e.g., aircraft flight dynamics, primary and secondary control surfaces, sensors). Based on a detailed bibliography, we present in this paper the building bricks of this special architecture and the design characteristics of its implementation. In particular, we are introducing architectural fault tolerance aspects and present results to assess the global behaviour of the system.*

KEYWORDS: *Aircraft simulation, Hardware-in-the-loop Simulation, Flight Control Systems, Distributed Systems*

1 INTRODUCTION

Simulation consists in imitating the operation of a real-world system over time, and requires a mathematical model of the system behaviour (J. Banks et al., 2009). It is a well-known design technique, widely used and accepted among research institutes and aerospace industry. Compared to the real flight environment, simulation offers tight control and easy tuning of the different parameters, and one can swiftly investigate many different scenarios and operating conditions. Over the years, the use of simulation has contributed to the improvement of aircraft, such as aerodynamic efficiency or advancements in avionic systems. In particular, to assess the maturity of certain types of equipment or hardware components, the use of a dedicated type of simulation called Hardware-In-the-Loop (HIL) simulation is essential. HIL simulation combines hardware components (to be implanted on the real system) interacting with a simulation environment, reproducing as close as possible the behaviour of the system to be controlled, as depicted in Fig. 1.

The first digital Flight Control Systems (FCS) were studied in the early 70s (J.P. Sutherland, 1968). As an example, in 1972, the F-8C military aircraft became the first aircraft equipped with digital FCS to operate without a mechanical backup system. Since

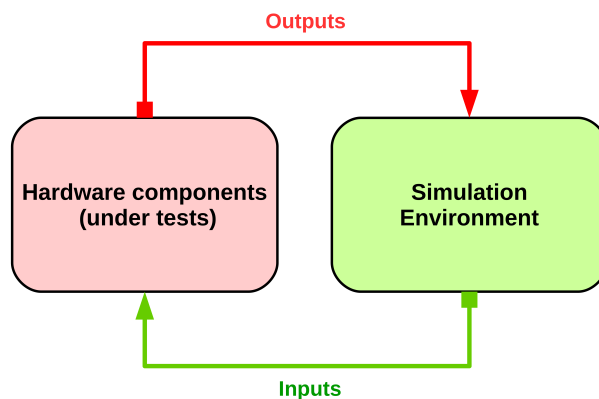


Figure 1 – HIL Simulation illustration

then, the industry and research institutes have been pursuing their effort in this area. Over the years, aircraft FCS have then become more and more complex to ensure multiple functionalities and are currently composed of several redundant components as on-board computers or fly-by-wire units (I.Moir, 2006). The Airbus A320 aircraft is a well-known example and was the first commercial aircraft to contain a full digital complex FCS architecture with autopilot and fly-by-wire functionalities (D. Briere and P. Traverse, 1993). The development of an aircraft FCS typically involves the use of a HIL simulation environment where the flight control system components

have to be tested in certain scenarios prior to the first flight (E.L. Duke, 1989). To allow this, the FCS architecture under study has to be plugged to a relevant and accurate simulation environment modelling the realistic behaviour of the aircraft with mathematical models of flight dynamics, actuators, engines and sensors. The whole system composed by the FCS connected to a dedicated simulation environment is depicted as in Fig. 2.

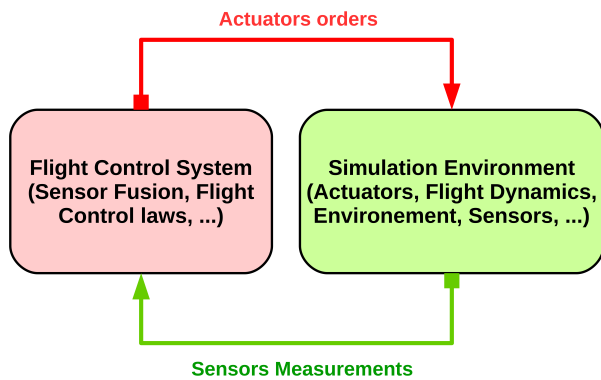


Figure 2 – HIL Simulation for Flight Control System

In the context of the PRISE project¹ (Research Platform for Embedded Systems Engineering), we implemented and tested an aircraft distributed simulation compliant with the High Level Architecture (HLA) standard and composed of several distributed simulators (J-B. Chaudron et al., 2014). Based on this background, we decided to further investigate avionic system architectures and started a new project called Simulation Modules for Avionics Real-Time Embedded Systems (SMARTIES). Based on literature and assumptions, we designed and implemented from scratch our own aircraft flight control system which is connected to the simulation architecture. The remainder of this paper introduces the main building blocks of the global system and is structured as follows:

- Section 2 provides a general overview of the whole architecture;
- Section 3 describes the different implementation details and design choices;
- Section 5 outlines some fault-tolerance aspects of our system;
- Section 6 concludes and presents the perspectives for future works.

2 ARCHITECTURE DESCRIPTION

By analogy with the illustration given in Fig. 2, our HIL simulation architecture, depicted in Fig. 3, is

¹French acronym for *Plate-forme pour la Recherche en Ingénierie des Systèmes Embarqués*.

also composed of two main blocks: the simulation environment which is described in Section 2.1 and the distributed flight control system architecture detailed in Section 2.2.

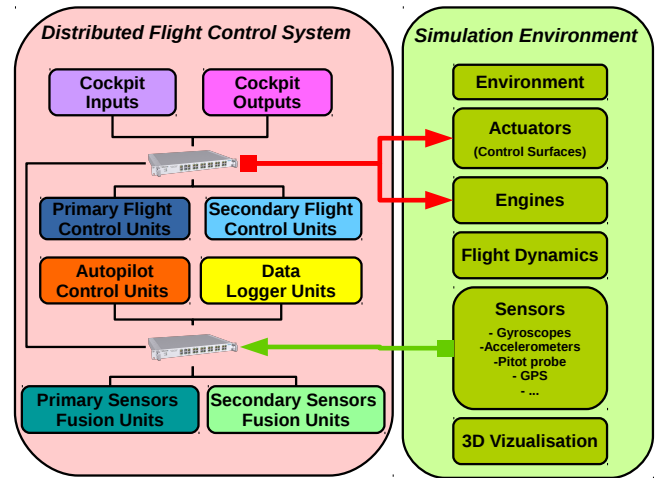


Figure 3 – Our HIL architecture overview

2.1 Simulation Environment

We decided to implement the model of a Boeing 747-100 (C.R. Hanke, 1970) (C.R. Hanke and D.R. Nordwall, 1970). The information provided in these two references are largely sufficient to reproduce faithfully the aircraft behaviour. The documents have been released publicly, and the accuracy and the relevance of the model have been demonstrated by NASA with careful comparisons to temporal simulations provided by Boeing; this motivated our choice to implement this model². To be more explicit, the flight simulation environment is subdivided into several components:

- The **Environment** component reproduces realistic weather conditions. It models the US Standard Atmosphere 1976 (NOAA and NASA, 1976) and, depending on the altitude, it calculates the corresponding atmospheric variables such as temperature, pressure or air density. It also integrates different types of wind models, such as wind shears and gusts, as well as turbulence models like Dryden (H.L. Dryden and I.H. Abbott, 1949) and Von Karman ones (T. Von Karman, 1948).
- The **Actuators** component gathers all the control surfaces whose deflections change the aerodynamic forces, and thus influence the aircraft motion. We consider here left and right ailerons, left and right inboard/outboard elevators, stabilizer, upper/lower rudder, slats, flaps and gears.

²Note that the B747-100 aircraft was not equipped with a current generation FCS system, we are using the flight dynamics model because of its relevance and its accuracy.

Each control surface is modeled by a second-order system with position and rate saturations. Actuators defaults such as delay, bias, and hysteresis phenomena are also considered in our model. As described in Section 4, aircraft actuator systems are redundant (D. R. Ryder, 1973.) and four groups of actuators denoted *Actuators Group (AG)* are considered.

- The **Engines** component simulates four high bypass ratio Pratt & Whitney JT9D-3 jet engines whose behaviour changes with the atmospheric conditions (i.e., temperature) and the aircraft Mach number. An accurate model of these jet engines is also described in the B747-100 reference documents (C.R. Hanke, 1970) (C.R. Hanke and D.R. Nordwall, 1970) .
- The **Flight Dynamics** component (or Flight Dynamics Model, FDM) is the heart of the simulation model as it computes the equations of motion described, in their simplest form, by a set of twelve first-order ordinary differential equations. Under the action of aerodynamic, gravity, and propulsion forces, the aircraft state evolves accordingly. This part of the simulator is usually the most demanding one depending on the sought fidelity. Fortunately, the reference documents (C.R. Hanke, 1970) (C.R. Hanke and D.R. Nordwall, 1970) provide a sufficiently accurate description of the aerodynamic coefficients.
- The **Sensors** component simulates different sensors available in an aircraft. The navigation system is connected to numerous sensors and warning systems and they record data and transmit them to the flight control system. The various sensors included in our simulation are Inertial Navigation System (INS), Global Positioning System (GPS), Pitot Static System, Total Air Temperature (TAT) probe, Angle of Attack (AOA) and Side Slip Angle (SSA). We also consider sensors measuring the engines and control surfaces states. As each type of sensor has its own imperfection, we added phenomena such as delay, bias, drift, and noise. The role of these components is essential as they feed the FCS with the inputs from the simulation environment, the details on the hardware and software implementation will be discussed in Section 3. As described in Section 4, aircraft sensors systems are redundant and we are considering four groups of Sensors noted Sensors Group (SG).
- The **3D Visualization** part is the display of actual aircraft position, attitude, control surfaces states and so on, in existing virtual 3D environments such as FlightGear (<http://www.flightgear.org/>) or X-plane (<http://www.x-plane.com/>). We can in-

stantiate many instances of these flight simulators to be able to reproduce an accurate graphical environment to the user. For instance, we are using several instances to get a full cockpit windows display with multiples screens and, also, we are using some other screens to visualize the aircraft from an external point of view.

2.2 Flight Control System

As mentioned in Section 1, our distributed flight control system tends to reproduce the behaviour of a real avionic network. Aircraft FCS are very complex systems composed by several entities with different purposes such as the autopilot functions or sensors fusion. For the time being, the different entities considered in our distributed FCS architecture are as follows:

- The **Cockpit Inputs** components are in charge of relaying the pilot/co-pilot inputs coming from the cockpit elements such as ailerons/elevators-stick, throttle-stick, flaps-stick, slats-stick and gears/stick devices in order to send commands to the aircraft actuators and engines (generally through the implemented fly-by-wire system). We also have a Flight Control Unit (FCU) interface which can be used by the pilot/co-pilot to set the parameters for the autopilot units. For example, a reference heading or a reference altitude can be selected via this interface and, if the autopilot is enabled, the FCS system will move the aircraft in this predefined reference state.
- The **Cockpit Outputs elements** are mainly pilot/co-pilot graphical interfaces such as in a real aircraft cockpit. As of now, our architecture Primary Flight Display (PFD), Navigation Display (ND) and Electronic Centralized Aircraft Monitor (ECAM). These interfaces provide visual and sound cues to the pilot/co-pilot from the information sent by the different units of the distributed system.
- The **Primary Flight Control Units (PFCU)** units implement the different algorithms to control the aircraft and to maintain it in a safe flight envelope (following a set of configurable rules) depending on the inputs from the pilot on the side-stick or throttle stick or the autopilot (via APCU). There are four PFCU units integrated in the system to ensure some fault-tolerant properties as described in Section. 4. These flight control laws are composed of stability and control augmentation system (CSAS) loops that allow the pilot to control the normal acceleration in the longitudinal motion, and the roll rate in the lateral motion to make the aircraft turn. Moreover, one usually finds protection system of the angle of attack and speed to avoid stall.

- The *Secondary Flight Control Units (SFCU)* also implement some algorithms to control the aircraft but with degraded features. There are two SFCU units which are only used in case of failure of the PFCU units, the reconfiguration of the architecture from PFCU to SFCU is not discussed in this paper.
- The *Autopilot Units Control Units (APCU)* units are in charge of the autopilot calculation based on the inputs from the pilot or the co-pilot on dedicated interfaces to select for example a reference altitude or a reference heading. There are two APCU units: one unit computes the autopilot functions based on the inputs from the pilot and the other one computes it based on the inputs from the co-pilot.
- The *Data Logger Units (DLU)* units are in charge of gathering and storing all the data coming from the whole FCS System. There is only one unit currently in our architecture.
- The *Primary Sensor Fusion Units (PSFU)* implement some algorithms for sensor fusion based on the inputs from the simulated sensors. These PSFU units, similar to Air-Data and Inertial Reference Units (ADIRU) (M.L. Sheffels, 1992), must provide the most possible accurate values to the control units for each measured parameter. Currently, the parameters that are estimated by our algorithms are attitude (Euler angles and quaternions), position (latitude, longitude and altitude), airspeed parameter (such as true airspeed), angle of attack, and so on. An Extended Kalman Filter (EKF), which is an estimation algorithm for non-linear systems, derived from the original Kalman filter (R.E. Kalman, 1960), has been implemented as a baseline for the algorithm of these units. There are four PSFU units, as described in Section 4, and we are illustrating the correctness of our algorithm by presenting some results in Section 4.3.
- The *Secondary Sensor Fusion Units (SSFU)* also implement Kalman filter-based algorithms to fuse the data from sensors but with degraded features. As for SFCU units, there are two SSFU units which are only used in case of failure of the PSFU units; however the reconfiguration of the architecture from PSFU to SSFU is not discussed in this paper.

3 DESIGN AND IMPLEMENTATION

3.1 Simulation Interfaces and Execution

In an HIL environment, the way to interface the hardware under test and the simulation is paramount be-

cause it must use the same interfaces as the ones found in the real system. In our case, we are considering a FCS system based on real-time Ethernet network to connect all the FCS computing devices (see Section 3.2). To allow the connection between the FCS system and the simulation environment, we used some specific devices called *Rackmount*, which is very similar to a computer with multiple Ethernet interfaces. As illustrated in Fig. 5, the Rackmount device is running Engines, Actuators and Sensors components, which are sending and receiving data from the real-time Ethernet network (via Real-Time Ethernet Switches i.e. RTES).

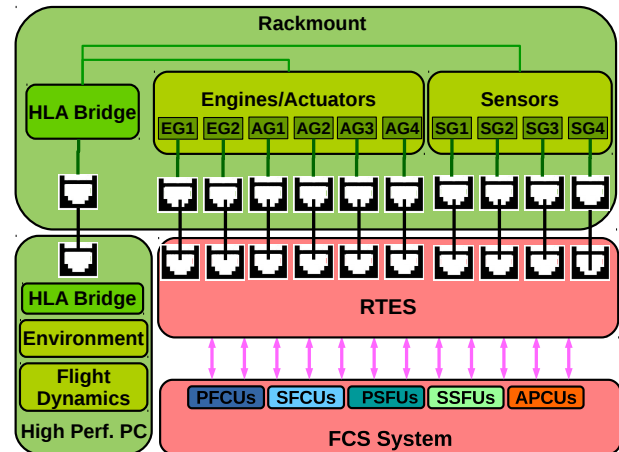


Figure 4 – Simulation/FCS interfaces illustration

The simulation part must respect the real-time constraints inherent to the components under test (such as the operating frequencies) and, therefore, our simulation environment has to be executed fairly quickly to behave according to real-time requirements imposed by the FCS system. To ensure a good execution of the simulation and allow enough computing power to each simulation component, we distributed the flight dynamics and the environment simulation components to another high performance machine. As depicted in Fig. 5, this distribution has been done using IEEE HLA standard and especially the open-source middleware CERTI (E. Noulard et al, 2009). In particular, HLA and CERTI mechanisms provide time management algorithms, guaranteeing a consistent global logical time throughout the whole simulation, which are one of the main benefits of this simulation standard. Our simulation models are based on Ordinary Differential Equations (ODE) and the use of HLA time management ensures the proper schedule of the computations and the communications between the distributed ODE based simulators (J-B. Chaudron et al., 2016). Thus, the global simulation provides results that will always be fair and relevant according to simulation model semantics.

From the software point of view, the whole simulation environment has been home made and written

in C/C++. The whole simulation environment is configurable through a set of eXtensible Markup Language (XML) files. As an example, the user can select the numerical integration method and the integration step to solve the corresponding ODE.

3.2 FCS Interfaces and Execution

As described in Fig. 5, the current version of our FCS system is composed of 2 standard devices (Cockpit Inputs/Outputs and Data Logger) and 14 embedded devices (2 APCUs, 4 PFCUs, 2 SFCUs, 4 PSFUs and 2 SSFUs). The whole system use Commercial Off-The-Shelf (COTS) technologies and in particular an Ethernet based network. Thus, we use Real-Time Ethernet Switches (RTES), which are compliant with standard Ethernet protocols, as well as the ARINC 664 part 7 standard (A664p.7) (ARINC Industry Activities, 2006) and Time-Triggered Ethernet (TTEthernet) standard (SAE International, 2011).

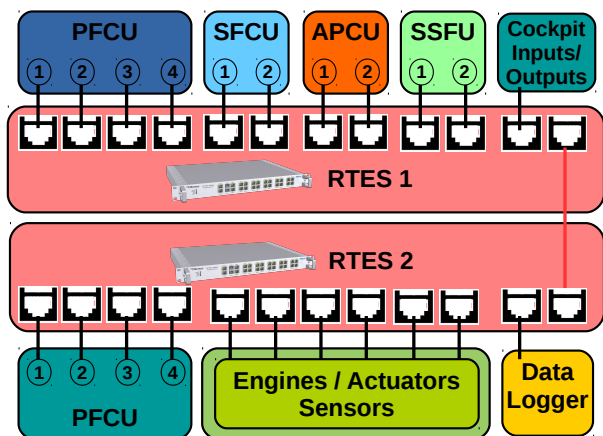


Figure 5 – FCS network illustration

At the processing unit level, the Cockpit and the Data Logger units are running on high performance machine with a Linux Fedora 24 (64 bits) installed. For the embedded units, we decided to use Beagleboards X15 cards (Gerald Coley, 2016) on which are installed Xenomai 3 real-time OS (Jan Kiszka, 2016) (one card per unit). The current version of the communication stack works with standard UDP Ethernet protocol and an home-made A664p.7 software-based compliant stack (light version).

As for the simulation, the whole software design and implementation for each has been home made and written in C/C++. The APCU, PFCU and PSFU units have been designed to work at 50Hz or 100Hz and the SFCU and SSFU units have been designed to work at 20Hz or 50Hz for the moment. The real-time scheduling analysis for the different algorithms implemented in the FCS devices is out of the scope of this paper.

3.3 Discussion

We have described in this section the implementation details of our global architecture (the simulation part and the FCS part). We have made some choices based on our current knowledge and experience. As an example, we decided to use the HLA standard, instead of other existing standard as the Functional Mockup Interface (FMI), we have a complete knowledge of HLA CERTI software which we are using for years now. Also, the HLA synchronization mechanisms using time management services, which are not available in other simulation standards, are extremely useful to ensure proper schedule and consistency of the simulation (J-B. Chaudron et al., 2016). The whole software (for both simulation and FCS components) has been done per us and we have a complete control on the implementation and the tuning of every part.

As a recall, the purpose of this position paper is to introduce our new project SMARTIES, we have described what we have done and how it is related with the different research areas. The overall architecture is complex and has required the combination of a lot of concepts, algorithms and details and it is not possible to describe every aspects within one paper. Thus, we have decided to illustrate, in the next section 4, the fault-tolerance and redundancy features implanted and tested on our platform.

4 FAULT TOLERANCE AND REDUNDANCY

4.1 Quadruplex design

The occurrence of a fault in an aircraft flight control can lead to catastrophic events with significant costs, both economically and in terms of human life. Therefore, such systems are said to be safety critical systems and have to be designed to ensure fault tolerant properties to continue operating properly even if a fault occurs on one of its components (computers systems, actuators, sensors, etc.). The fault tolerance analysis is a wide research domain (D.K. PradhanK. 1986) which goes way beyond the scope of this paper. We introduce in this section the impact on the design of our architecture.

In avionic systems, the fault tolerance and re-configuration properties can be considered in the implemented algorithms themselves (adaptive control/command laws) as well as in the global system architecture design (computers, backbone network, etc.). From the architectural point of view, fault-tolerance and reconfiguration techniques are usually built upon per the usage of redundant, specific and heterogeneous computer based components. In aircraft flight control systems, fault tolerance is usually achieved through redundancy, i.e., the comput-

ing units are duplicated, triplicated or quadruplicated (R.P.G. Collinson, 2003). The redundancy ensures that the overall cumulative failure probability required for the whole global system (example, 10^{-9} failures/hour for civilian transport aircraft) is smaller than the failure probability of one of its standalone sub-system (example, 10^{-7} failures/hour for one processing unit). Therefore, if one of the redundant components fails, the architecture can ignore the faulty component, or switch to a spare one depending on the implanted fault detection and recovery mechanism.

Currently, the primary level of our distributed FCS is based on a quadruplex architecture (D.B. Mulcare et al., 1988) with multi-cast communication flows as illustrated in Fig. 6.

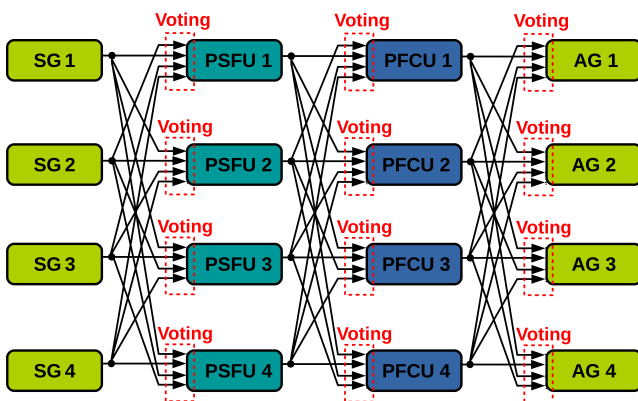


Figure 6 – Communication flows illustration

4.2 Voting algorithms

In addition, due to redundancy and communication flows (see Fig. 6), a voting technique must be integrated in order to elect the *best*³ value to be processed. There are many varieties of voting techniques, which mainly rely on and depend on the architecture of the system. As detailed in Section. 3.2, the current version of our demonstrator only implements asynchronous communication protocols (standard UDP and light version of A664 p7) and, therefore, we have implemented a voting algorithm that is compliant with this specificity (G.J. Davis and Ames Research Center, 1987). Currently, the implemented algorithm is a fault-tolerant median algorithm which is composed per 4 successive phases and works as follows.

1. **Collection phase:** Collect the incoming data from each channel (collection duration is a configurable amount of time).
2. **Consistency phase:** Compare the incoming values two by two and ensure that they remain within a pre-defined interval.

³With respect to the logic implanted in the voting algorithm

3. **Selection phase:** Sort the incoming data and remove the smallest and the biggest inputs (only if 3 or 4 inputs are available).
4. **Fusion phase:** The remaining values are averaged (in case of 4 inputs) or the median value is kept as it is (in case of 3 inputs).

4.3 Simulation Results

We will illustrate in this sub-section the behaviour of the algorithm implanted in the redundant PSFU units. We have considered a scenario with 100Hz IMU sensors with 5% of white noise (for gyroscope and accelerometers) and 10Hz GPS sensors with ± 10 meters error in position (worst case) and ± 0.006 meters/second error in speed. Note that the error model considered here is worst than the navigation grade requirements for civilian transport aircraft sensors. Figures 7 and 8 are showing the simulation results obtained from redundant computations of the roll angle ϕ by the PSFU units. The purple line (the one from the Flight Dynamics Model i.e. FDM) represents the "real" value given per the simulation environment. The others lines show the output of the different PSFU units (the value estimated per our EKF algorithm implanted in PSFU units) as well as the voted value (i.e. the median value).

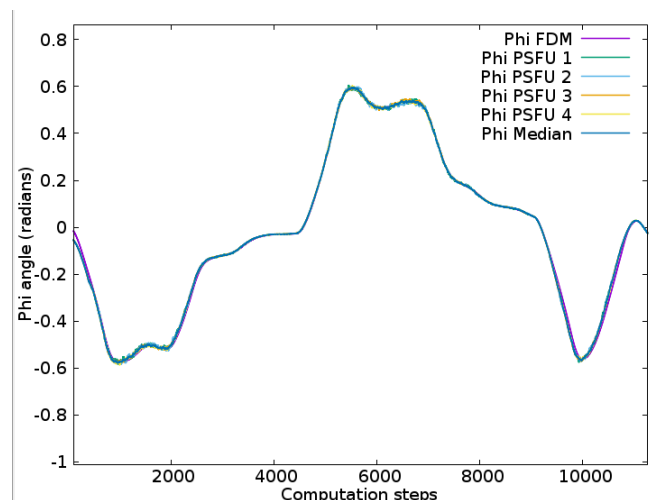
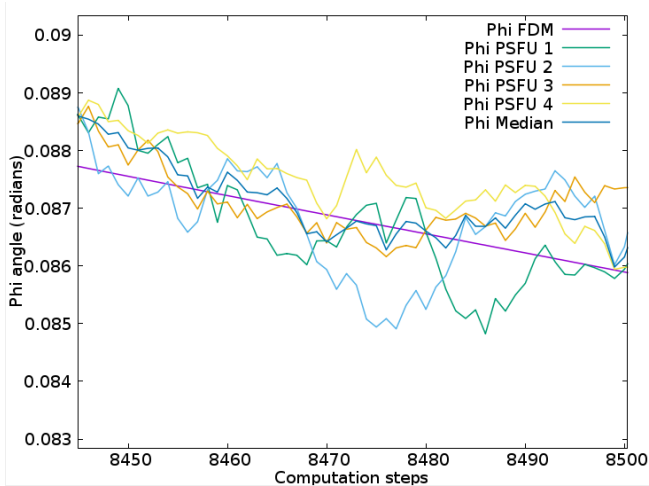
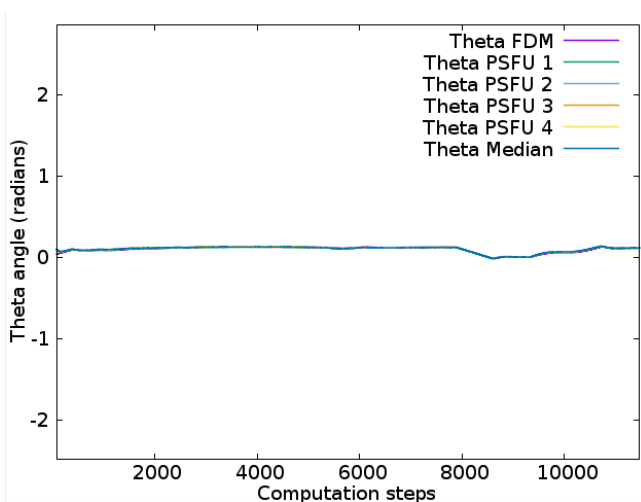
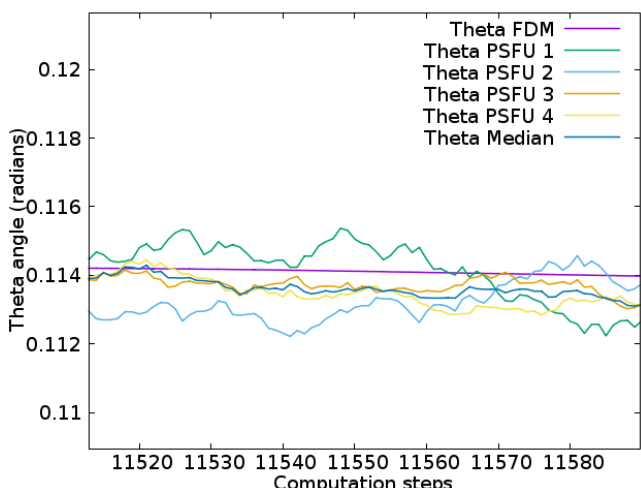


Figure 7 – Roll angle ϕ redundant calc. (Global)

Figures 9 and 10 show the simulation results obtained from redundant computations of the pitch angle θ by the PSFU units.

5 CONCLUSIONS AND PERSPECTIVES

In this position paper, we have presented an overview of our new SMARTIES architecture. To our knowledge, this is the first academic COTS based aircraft FCS architecture which tends to capture every aspects of real FCS systems, in particular it includes

Figure 8 – Roll angle ϕ redundant calc. (Zoom)Figure 9 – Pitch angle θ redundant calc. (Global)Figure 10 – Pitch angle θ redundant calc. (Zoom)

complex sensors fusion algorithms. This complex project has required the mastering of many aspects: from the realistic design of close-to-real avionics sys-

tem unit running on dedicated target to the implementation of an efficient simulation environment compliant with real-time constraints. We have described here the current status of our architecture, and we have introduced fault-tolerant aspects and presented some first results from redundant computation and voting algorithms from sensor fusion units.

The project is ongoing, and many parts still have to be implemented, tested and optimised. On the APCU unit level, we are currently working on some extension to integrate high level control loops to mimic the behaviour of aircraft Flight Management System (FMS) which allows pilot and co-pilot to define flight plan according to weather, fuel consumption and so on. We are also currently planning the integration of new units that aim to reproduce simple Full Authority Digital Engines Control Systems (FADEC) (L. Padon, 1988) to control the engines. On the system level, we are investigating the feasibility of the migration to software based time triggered architecture.

ACKNOWLEDGMENTS

The authors would like to thank TTTech Computertechnik for their courtesy using their switches hardware components. The authors would also like to thank Mr. Vashan Srinath Kemthoor for his great work on the sensor fusion algorithms during his master internship at ISAE-SUPAERO.

REFERENCES

- J. Banks, J. Carson, B. Nelson. D. Nicol, 2009. Discrete-Event System Simulation (5th edition). Prentice Hall.
- J.P. Sutherland, 1968. Fly-By-Wire Flight Control Systems. *Joint Meeting of Flight Mechanics and Guidance and Control Panels of AGARD*, Oslo, Norway, page 1.
- D. Briere and P. Traverse, 1993. AIRBUS A320/A330/A340 electrical flight controls - A family of fault-tolerant systems. *FTCS-23 The Twenty-Third International Symposium on Fault-Tolerant Computing*, Toulouse, France.
- I. Moir, 2006. Civil Avionics Systems (Chapter 9: Flight Control Systems). *Aerospace Series (PEP)*, John Wiley & Sons Ltd.
- E.L. Duke, 1989. V&V of flight and mission-critical software, *IEEE Software*, vol. 6, no. 3, pp. 39-45.
- J-B. Chaudron, D. Saussie, P. Siron, M. Adelantado, 2014. Real-time distributed simulations in an HLA framework: Application to aircraft simulation, *Simulation Journal*, Vol.90, Issue 6, p. 627-643

- C.R. Hanke, 1970. The simulation of a jumbo jet transport aircraft - Volume I: Mathematical Model. Prepared by Boeing Company Wishita Division, Kansas for National Aeronautics and Space Administration, Ames Research Center, Moffet Field California.
- C.R. Hanke and D.R. Nordwall, 1970. The simulation of a jumbo jet transport aircraft - Volume II: Modeling Data. Prepared by Boeing Company Wishita Division, Kansas for National Aeronautics and Space Administration, Ames Research Center, Moffet Field California.
- D. R. Ryder, 1973. Redundant Actuator Development Study. Report N74-21655. Prepared by Boeing Commercial Airplane Company Seattle, Washington for National Aeronautics and Space Administration, Ames Research Center, Moffet Field California.
- NOAA and NASA, 1976. Standard US atmosphere. *Washington DC: Government US Printing Office*, N77-16482.
- H.L. Dryden and I.H. Abbott, 1949. The Design of Low-Turbulence Wind Tunnels. Technical Report 940, National Advisory Committee For Aeronautics, Washington, D.C, USA.
- T. Von Karman, 1948. Progress in the Statistical Theory of Turbulence. *National Academy of Science Journal*, Vol. 34.
- L. Paddon, 1988. Active-control Engines, *Flight International*.
- M.L. Sheffels, 1992. A fault-tolerant air data/inertial reference unit. *Proceedings of 11th Digital Avionics Systems Conference*, Seattle, MA, USA.
- R.E. Kalman, 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, Vol. 82.
- E. Noulard, J-Y. Rousselot, and P. Siron, 2009. CERTI, an Open Source RTI, why and how. *Proceedings of the 2009 SISO Spring Simulation Interoperability Workshop*, San Diego-Mission Valley, United States, p. 23-27.
- J-B. Chaudron, David Saussié, Pierre Siron, Martin Adelantado, 2016. How to solve ODEs in real-time HLA distributed simulation. *Proceedings of the SISO Spring Simulation Interoperability Workshop*, Orlando, United States.
- Gerald Coley, 2016. Beagleboard X15 System Reference Manual (Revision B1). SRM_X15, Released on July 22, 2016.
- Jan Kiszka, 2016. Xenomai 3 - An Overview of the Real-Time Framework for Linux. *Embedded Linux Conference*, San Diego, California, USA,
- ARINC Industry Activities, 2006. Aircraft Data Network Part 7 Avionics Full Duplex Switched Ethernet (AFDX) Network. *ARINC Standard Document*.
- SAE International, 2011. Time-Triggered. *SAE Standard Document*, AS6802.
- R.P.G. Collinson, 2003. Introduction. In: Introduction to Avionics Systems. Springer, Boston, MA, USA.
- D.K. PradhanK. 1986. Fault-tolerant computing: theory and techniques. *Prentice-Hall, Inc.*, Vol. 1.
- D.B. Mulcare, L.E. Downing and M.K. Smith, 1988. Quadruplex Digital Flight Control System Assessment, *National Aeronautics and Space Administration, Ames Research Center*, Moffett Field, California.
- G.J. Davis and Ames Research Center, 1987. An analysis of redundancy management algorithms for asynchronous fault tolerant control systems, *National Aeronautics and Space Administration, Ames Research Center*, Moffett Field, California.

Une approche centrée usage pour accompagner l'introduction d'un outil numérique pour la prise en charge à domicile

Christophe MERLO^{1,2}, Katarzyna BORGIEL^{1,2}

¹ESTIA - ESTIA Recherche - 92, allée Théodore
Monod, Technopole Izarbel - 64210 Bidart - France
{c.merlo|c.borgiel}@estia.fr

²IMS UMR 5218 - Université de Bordeaux – 351 cours
de la libération - 33405 Talence cedex - France

RÉSUMÉ : *Les outils du système d'information sont un support aux activités des acteurs de l'organisation. L'amélioration de la performance, de la qualité, etc., dans les organisations conduisent parfois à remettre en cause certains de ces outils et à les faire évoluer ou à les remplacer. Néanmoins, l'implémentation d'un nouvel outil numérique est un processus d'adaptation mutuelle entre l'organisation et la technologie, qui comporte des changements sur plusieurs dimensions et à différents niveaux de granularité. L'objectif de nos travaux est de contribuer à accompagner ces changements dans le but de favoriser l'innovation organisationnelle tout en facilitant l'implémentation de nouveaux outils. Nous proposons d'appréhender la diversité des évolutions dans l'organisation et des adaptations de l'outil à travers la diversité des transformations « potentielles » et « effectives » des usages. Nous proposons un modèle d'usage permettant de caractériser les usages et leur diversité. À partir de ce modèle, nous définissons des méthodes d'exploration graduelle de la diversité des usages et de la diversité des transformations des usages, orientées organisation et outil. Nous illustrons l'application des méthodes proposées à travers le cas d'étude d'informatisation du dossier de soins dans une structure de prise en charge à domicile.*

MOTS-CLÉS : *home care, système d'information, transformation bidirectionnelle technologie-organisation, innovation organisationnelle, approche centrée usage.*

1 INTRODUCTION

Les organisations de santé à la recherche d'améliorations de leur performance, de la qualité, parfois sous l'impulsion de la réglementation ou encore sous l'effet de pressions extérieures, sont amenées à réévaluer leur système d'information (SI). Elles mettent alors en place des démarches longues et complexes afin de faire évoluer certains de leurs outils informatiques voire les remplacer, et/ou introduire de nouveaux outils. Néanmoins, l'implémentation d'un nouvel outil numérique est un processus d'adaptation mutuelle entre l'organisation et la technologie (Leonard-Barton, 1988 ; Berg, 2001), qui comporte des changements sur plusieurs dimensions et à différents niveaux de granularité. La technologie, qui est un moyen de transformer l'organisation, se retrouve elle-même transformée, ce qui peut à son tour mener à de nouveaux changements pour l'organisation.

Nous nous plaçons dans un contexte de « Home Care » (Bricon-Souf et al., 2005), pour lequel la structure de santé qui nous sollicite, souhaite transformer son système d'information en remplaçant le support papier utilisé au domicile des patients par un dispositif numérique. L'objectif de nos travaux est de contribuer à accompagner ces changements, en étudiant simultanément les transformations de l'organisation provoquées par l'introduction de ce type d'outil, et les transformations que l'organisation peut espérer sur l'outil lui-même.

Dans la section 2 nous développons ce contexte industriel et précisons les questions portées par la structure de

santé, avant de construire notre problématique de recherche. En section 3 nous proposons un modèle pour l'analyse et la caractérisation des usages, puis en section 4 plusieurs méthodes exploitant ce modèle en vue d'accompagner la structure de santé durant son projet de remplacement d'outil. La section 4 rappelle également les résultats de ce projet et nous effectuons un retour d'expérience avant de conclure.

2 PROBLÉMATIQUE

L'hospitalisation à domicile (HAD) est un mode de prise en charge qui relie la médecine hospitalière et la médecine de ville. Elle implique par définition un contexte collaboratif et distribué, où le patient est au cœur des préoccupations d'une multitude d'acteurs : les familles, les médecins, les infirmiers, les aides-soignants, les pharmaciens, etc. La qualité des soins prodigués est assurée notamment grâce à un bon niveau de coordination entre les interventions, asynchrones, de tous les acteurs. Cette coordination est facilitée, elle, par la présence d'outils partagés, comme par exemple le dossier de soins au chevet du malade.

2.1 Contexte industriel

La structure de santé qui est le partenaire industriel de nos travaux de recherche est une association privée à but non lucratif, fondée en 1968, qui assure des services complets de prise en charge à domicile. Elle est la seule structure de ce type sur son territoire d'interventions,

malgré la présence de différents types de structure plus spécialisées, publiques ou libérales. Elle gère environ 500 patients par jour et ses activités se répartissent à la fois sur de l'Hospitalisation à Domicile (HAD) pour 20%, et sur du Service de Soins Infirmiers à Domicile (SSIAD) pour 80%.

Si l'HAD a pour but d'éviter ou d'écourter un séjour à l'hôpital et de traiter des malades atteints de pathologies graves, aiguës ou chroniques, le SSIAD est une alternative à la maison de retraite ou à l'établissement spécialisé, pour les personnes en perte d'autonomie (âgées ou handicapées) nécessitant des soins infirmiers et d'hygiène générale. La structure de santé supervise aussi dans ce cadre des patients atteints de la maladie d'Alzheimer.

Elle emploie plus de 250 salariés : personnel médical, paramédical, administratif, etc., répartis sur plus de 20 métiers différents. Malgré cela, elle n'est pas en capacité d'apporter au patient à son domicile tous les services nécessaires au bon déroulement de sa prise en charge. Le concept même de l'HAD exige qu'une structure s'appuie dans son fonctionnement sur le réseau d'établissements et de professionnels du territoire, par exemple les médecins hospitaliers et généralistes, les pharmaciens, les laboratoires et plusieurs professions libérales.

Notre structure de santé pratique donc une prise en charge globale au travers de processus métiers et d'activités complexes, multiples, et distribués géographiquement. Dans ce contexte, le défi de la coordination n'est pas seulement de gérer le planning des passages au domicile et des déplacements à l'hôpital ou en ville, mais aussi et surtout d'assurer une continuité et une cohérence entre les différentes activités de soins liées à un patient donné, et ceci malgré le caractère asynchrone de la plupart des interventions, et les aléas de présence de personnes remplissant les fonctions nécessaires auprès du patient. Le dossier de soins, situé au domicile de chaque malade, est un outil crucial de traçabilité et de transmission d'informations entre différents intervenants.

Le rôle majeur des TIC dans l'amélioration des pratiques professionnelles dans le monde de la santé se manifeste par un nombre important d'initiatives nationales visant à accompagner les structures et les éditeurs dans l'évolution des systèmes d'information hospitaliers (SIH). Le Programme Hôpital Numérique (DGOS, 2014), par exemple, définit la stratégie nationale relative aux SIH pour les années 2012-2017, et établit un palier de maturité, permettant le partage et l'échange d'informations entre les différents acteurs du territoire, et l'amélioration de la qualité, de la sécurité et de la performance au service du patient. Le Livre Blanc de la (FNEHAD, 2009) met en avant des principes et recommandations pour le SI au sein des structures d'HAD.

Consciente des avantages que peuvent lui apporter les TIC, et en vue d'intégrer les évolutions nationales relatives aux systèmes d'information hospitaliers, notre structure de santé a pour projet de remplacer le dossier de soins papier par un outil numérique. Elle y voit une

opportunité pour améliorer la façon dont elle réalise ses missions.

Concrètement ce projet a pour but essentiel de :

- diminuer voire éliminer la ressaisie d'informations ;
- diminuer le temps passé sur la saisie d'informations ;
- donner l'accès à distance aux informations ;
- améliorer la lisibilité et l'exhaustivité de l'information, et par là la prise de décision ;
- assurer la « Prescription électronique alimentant le plan de soins ».

Ce sont les trois derniers objectifs qui caractérisent réellement les objectifs stratégiques de la structure de santé et sur lesquels sera évaluée la solution déployée.

Le dossier de soins actuel est constitué de plus de 49 documents, regroupés dans 4 sections au sein d'un classeur, constituant ainsi le référentiel unique pour la structure de santé et l'ensemble des acteurs qu'elle coordonne, qu'ils soient salariés ou des professionnels extérieurs.

Considérant le contexte que nous venons de décrire, l'outil numérique choisi se doit d'être distribué, en permettant d'accéder au dossier de chaque patient depuis son domicile, et aussi aux dossiers de plusieurs patients depuis le siège de notre structure de santé et des locaux de ses partenaires. Il se doit aussi d'être collaboratif par des mécanismes de communication et d'alerte visant à transmettre rapidement des informations critiques entre soignants/coordonateurs. Il se doit enfin d'être interopérable a minima avec le SI métier qui gère les dossiers administratifs des patients. L'outil est composé de deux applicatifs : un « Portail Web » et un applicatif « Tablette » pouvant fonctionner en mode « déconnecté », c'est-à-dire sans nécessiter une connexion continue au réseau Internet. L'éditeur est une entreprise récente et les applicatifs sont encore en cours de développement, ce qui permet à la structure de santé d'instaurer un partenariat étroit pour permettre des améliorations de l'outil suite aux tests sur le terrain.

Le grand nombre de documents composant le dossier papier, ainsi que le grand nombre de personnes l'utilisant d'une part, et le caractère adaptable de l'outil choisi d'autre part, sont la raison pour laquelle la structure s'interroge sur les moyens à mettre en œuvre pour réussir cette informatisation. Nos travaux se sont déroulés dans un contexte de thèse CIFRE (Borgiel, 2017) à partir des questionnements de la structure de santé :

- Quelle démarche adopter pour informatiser le dossier de soins au chevet du patient afin d'améliorer la traçabilité et la transmission d'informations par et entre les différents acteurs de prise en charge ?

Et par conséquent :

- Comment l'outil numérique « dossier de soins informatisé » va impacter l'organisation : son système d'information, ses acteurs et ses processus métier ?

Et réciproquement :

- Comment l'organisation peut-elle traduire ses besoins, sa façon de travailler, pour adapter l'outil numérique « dossier de soins informatisé » ?

Nous étudions dans la sous-section suivante ces questionnements en nous appuyant sur la littérature dans les différents domaines impliqués pour élaborer notre problématique de recherche et poser les bases de notre contribution.

2.2 Étude de la problématique de recherche

Selon (Le Moigne, 1990) l'organisation est composée de trois systèmes, et le SI joue le rôle crucial de l'intermédiaire entre le système opérant et le système de décision : « *le système d'information de l'organisation enregistre les représentations – sous forme symbolique – des opérations du système opérant (le comportement du système complexe), les mémorise, et les met à disposition, sous forme en général interactive, du système de décision. Lequel, après avoir élaboré ses décisions d'action (les commandes), les fait également enregistrer et mémoriser par le S.I., en les transmettant « pour action » au système opérant.* ».

Par ailleurs, les informations sont des faits bruts qui ont été mis en une forme compréhensible et utile aux gens (Laudon et Laudon, 2012). L'extraction et l'interprétation des informations à partir des données est, elle, possible grâce aux connaissances, apprises et maîtrisées par les individus (Reix et al. 2016). Le SI est un ensemble d'informations circulant dans l'organisation, associé aux moyens de les gérer : il s'appuie sur le système informatique (Morley et al., 2005). Selon (Reix et al., 2016) l'objectif opérationnel du SI « *est d'apporter un soutien aux processus de travail dans l'organisation, selon trois modalités principales (pouvant être combinées) : fournir ou partager l'information ; assister ou coordonner le travail humain ; automatiser ou intégrer le travail* ».

Nous voyons les outils du SI en accord avec (Jonassen, 1992) en tant que des extensions de l'être humain, matériels ou cognitifs. Les outils cognitifs - les dispositifs mentaux ou computationnels - soutiennent, guident et augmentent le processus de pensée de leurs usagers (Derry, 1990).

Pour étudier les impacts entre organisation et outil numérique, nous choisissons de nous intéresser au concept de l'usage, et plus particulièrement à celui de la diversité des usages d'un outil du système d'information. Dans cet article nous ne nous intéressons pas aux technologies utilisées pour construire ces outils.

Nous adoptons la définition de Proulx (2005) : « *utilisation d'un objet, naturel ou symbolique, à des fins particulières* ». Le concept d'usage matérialise le lien bidirectionnel entre l'outil et l'activité dont il est support. Ce même concept partage aussi les autres caractéristiques de ce lien : plusieurs dimensions peuvent être prises en compte pour décrire l'usage (Ammenwerth et al., 2006) ; il existe un lien particulier entre les individus et un col-

lectif utilisant un système (Burton-Jones et Gallivan, 2007) ; et les usages d'un outil se construisent dans le temps (Brangier, 2004), (Fidock et Carroll, 2012). Le projet de remplacement d'un outil numérique par un autre consiste alors à réaliser un changement « *d'une ancienne façon de faire les choses vers une nouvelle façon de les faire* » (Lorenzi et Riley, 2003). L'analyse des usages vise donc à analyser les pratiques des utilisateurs, « *ce que les gens font effectivement avec les objets et dispositifs techniques* » (Proulx, 2005), (Buisine et Roussel, 2008).

En outre, comme le soulignent (Burton-Jones et Gallivan, 2007) :

- différentes personnes du collectif peuvent employer des fonctionnalités différentes du système (technologique) ;
- différentes personnes du collectif peuvent employer le système pour des tâches différentes ;
- et plus globalement : différentes personnes du collectif peuvent employer des fonctionnalités différentes pour des tâches différentes.

Par conséquent le support qu'apporte un outil du système d'information se manifeste dans la diversité des usages qui en sont faits dans l'organisation.

Notre première proposition consiste donc dans la définition d'un **modèle d'usage pour rendre compte de la multiplicité des liens entre l'organisation et la technologie**. Ce modèle sera décrit dans la section 3 suivante.

Le changement dans le système d'information couvre la génération, l'implémentation, et l'adoption de nouveaux éléments, dans les sous-systèmes social et technique de l'organisation, qui stockent, transfèrent, manipulent, traitent et utilisent les informations (Lyytinen et Newman, 2008). La technologie est un vecteur de changement dans le système d'information, qui à son tour est un vecteur de changement dans l'organisation (Leonard-Barton, 1988).

Notre seconde proposition porte donc sur l'exploitation du modèle d'usage en vue de comprendre ces changements bidirectionnels, tout en se projetant des usages « présents » vers les usages « futurs ». Nous définissons ainsi plusieurs **méthodes pour identifier et caractériser les transformations des usages** :

- caractériser la diversité des usages présents ;
- caractériser la diversité des transformations des usages ;
- prendre des décisions sur les transformations des usages.

Ces méthodes permettent, de même que les approches issues du génie industriel de réingénierie des processus (BPM), de transformation des systèmes d'information, ou de modélisation d'entreprise (Merlo et al., 2014), d'étudier l'existant, de le confronter à une projection dans le futur pour aider à prendre des décisions qui vont impacter le niveau de performance atteint à l'issue d'un tel projet.

Ces deux propositions constituent notre contribution en vue de **comprendre et maîtriser la transformation bidirectionnelle de l'organisation et d'un nouvel outil numérique introduit dans son système d'information**.

3 MODÉLISATION DES USAGES

La définition du modèle d'usage repose sur l'identification d'un ensemble de variables qualitatives nécessaires à la caractérisation de l'usage d'un outil du SI. Ce modèle vise ainsi la description des usages observés dans un contexte organisationnel suivant les différentes modalités que peut prendre chacune des variables. Nous choisissons d'illustrer la construction du modèle d'usage par un cas d'étude extrait de nos observations : « prescrire les médicaments ». Ce cas d'étude intervient dès les premières activités du processus de soins mené par la structure de santé pour assurer la prise en charge du patient de son entrée à son départ.

3.1 Proposition d'un modèle

Pour définir les variables d'usage constituant le modèle conceptuel de l'usage, nous nous basons sur la méthode d'analyse QQQCCP, qui sert de guide général pour recueillir des informations complètes sur une situation ou sur une problématique (Equixor, 2016). Cette méthode de questionnement s'appuie sur sept questions « Qui ? Quoi ? Où ? Quand ? Comment ? Combien ? Pourquoi ? » qui peuvent être appliquées à une variété de contextes : définition des processus, élaboration des diagnostics, management de projet.

En nous appuyant sur notre analyse de la bibliographie et le contexte industriel, nous retenons dans un premier temps quatre variables que nous associons aux descripteurs suivants :

- Qui ? **Individu** (celui qui agit)
- Quoi ? **Outil du SI** (objet de l'action)
- Pour quoi ? **Activité** (l'action menée)
- Où ? **Lieu** (localise l'activité)

Nous désignons ce premier groupe de 4 variables comme la caractérisation d'un **cas d'usage**.

Dans le cas d'étude « prescrire les médicaments », la situation synthétisée dans la Figure 1 suivante permet d'identifier l'Individu « Dr Ospital » comme l'Individu concerné (Qui).

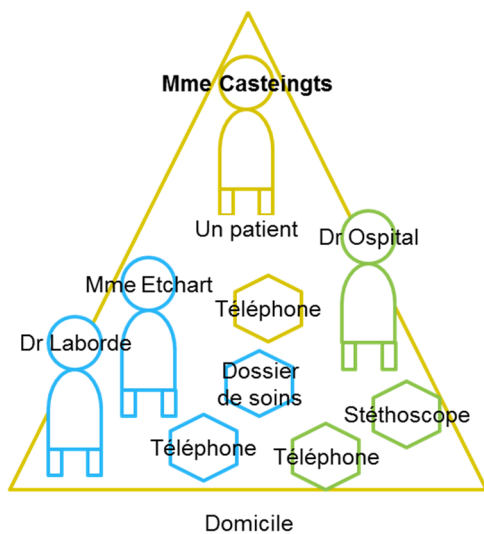


Figure 1 : Situation décrivant le cas d'étude

Le médecin « Dr Ospital » s'appuie sur le « Dossier de soins » (Outil du SI actuel, papier – Quoi) pour pouvoir « Prescrire les médicaments » (Activité – Pour quoi) au « Domicile » du patient (Lieu – Où).

La Figure 2 traduit ce cas d'usage à l'aide d'un formalisme graphique que nous avons défini spécifiquement pour faciliter la traçabilité puis les analyses de différents usages.

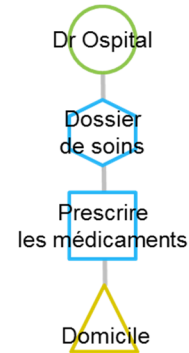


Figure 2 : Cas d'usage « prescrire les médicaments »

Nous complétons cette description par une cinquième variable, Quand, qui représente la **dimension temporelle**, et permet de décrire l'enchaînement de plusieurs activités. Nous désignons par **scénario d'usage** cet enchaînement d'activités. Ce scénario est une extension du cas d'usage, soit pour le situer dans un processus métier, soit pour le détailler davantage.

La Figure 3 illustre les actions du Dr Ospital pour réaliser sa prescription. Le cas d'usage de la Figure 2 est représenté et situé dans le contexte d'activités qui le précèdent.

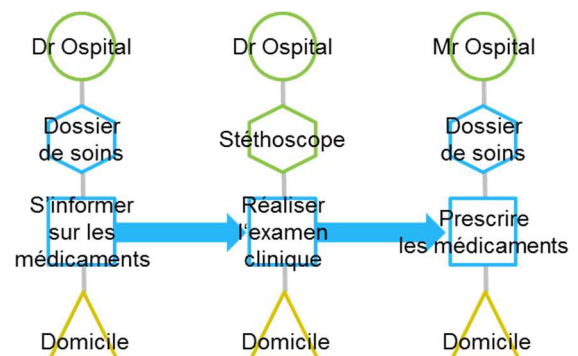


Figure 3 : Scénario d'usage

Les 5 variables identifiées constituent le modèle d'usage que nous proposons. Elles permettent de définir un cas d'usage et le(s) scénario(s) d'usage qui en découlent. Ensemble elles permettent donc de comprendre un usage complètement et comment celui-ci est réalisé.

Elles permettent également de caractériser et d'étudier la diversité des usages. Les observations que nous avons menées sur le terrain ont permis de constater qu'à partir d'un premier cas d'usage, d'autres cas d'usage associés pouvaient exister avec des valeurs différentes pour une (ou plusieurs) variable(s), ou bien que plusieurs scénarios

rios d’usage pouvaient correspondre au même cas d’usage initial.

Pour exemple, l’activité « Prescrire les médicaments » ne s’effectue pas toujours au « Domicile » du patient : lors de la réception des résultats des analyses médicales, le médecin peut modifier le traitement à distance depuis son « Cabinet », comme le montre la Figure 4. De plus, dans ce cas, un autre outil SI est employé.

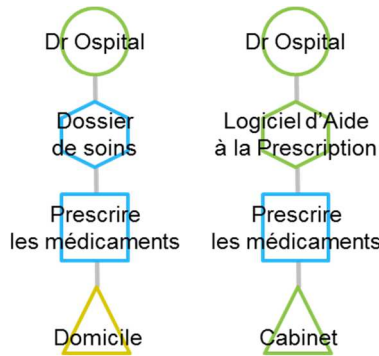


Figure 4 : Diversité du cas d’usage « prescrire les médicaments »

Enfin, de nombreux travaux issus du génie industriel sur les usages ou sur la modélisation des organisations et des processus métiers montrent l’importance de distinguer différents niveaux de granularité lors de la description d’une activité notamment. Nous avons synthétisé ces travaux pour définir 4 niveaux de granularité pour chacune des 4 variables décrivant un cas d’usage :

- Qui ? Individu - Fonction/rôle - Groupe - Collectif métier
- Quoi ? Information/donnée - Composant - Outil SI - SI métier
- Pour quoi ? Action/opération - Activité - Processus métier - Fonction de l’organisation
- Où ? Poste - Pièce - Site - Espace métier

Ces différents niveaux permettent de décrire de façon plus précise un cas d’usage général, ou de décrire un scénario d’usage de façon plus détaillée que le cas d’usage auquel il fait référence.

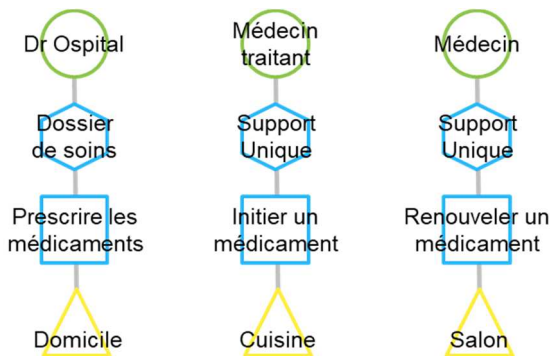


Figure 5 : Différents niveaux de granularité des variables d’usage

La Figure 5 montre par exemple que le « Dr Ospital » en tant qu’Individu pourrait aussi être représenté de façon

plus générique par la Fonction/rôle de « Médecin-traitant », ou encore par la Fonction/rôle « Médecin ». De même l’emploi du « Dossier de soins », Outil du SI, peut être décrit de façon plus précise par « le support unique » un Composant, qui est LE document à remplir pour formaliser la prescription, et qui est contenu dans le dossier de soins. De la même façon l’activité « prescrire les médicaments » peut être précisée : « initier un médicament », ou « renouveler un médicament », etc ; ou encore le « domicile » : la « cuisine » ou le « salon ».

Ainsi, il n’est pas nécessaire que toutes les variables décrivant un cas d’usage soient au même niveau de granularité.

À l’aide de ces différents niveaux de granularité, nos travaux sur le terrain ont montré qu’il était ainsi possible d’établir des corrélations entre les cas d’usage, dont les activités sont souvent décrites à un niveau de granularité détaillé, avec des activités métiers et donc les processus métiers de l’organisation.

Nous allons à présent décrire les méthodes permettant de mettre en œuvre ce modèle, et définies pour accompagner l’innovation organisationnelle portée par un nouvel outil numérique introduit dans le SI pour remplacer un outil ancien, sous forme papier dans notre cas d’étude.

4 MÉTHODES D’EXPLORATION DES USAGES

Nous considérons que le nouvel outil numérique, à travers les usages qu’il permet, est porteur de transformations pour l’entreprise qui souhaite l’intégrer dans son système d’information.

4.1 Contexte méthodologique

Nous désignons les usages présents que nous observons comme étant les usages présents « effectifs ». Les usages futurs « effectifs » ne pouvant pas être observés qu’une fois que l’outil numérique est implémenté, il n’est pas possible d’étudier directement les transformations « effectives » des usages avant l’implémentation et l’appropriation de l’outil. Nous introduisons donc le concept de transformation « potentielle » des usages pour caractériser les usages « potentiels » visés. Il nous est alors possible d’étudier les usages présents « effectifs », puis de nous projeter sur des usages futurs « potentiels », soit par des évolutions de l’outil du SI étudié, soit par des évolutions de l’organisation (processus, activités ou actions/opérations). La Figure 6 récapitule ces hypothèses.

Comme vu à la section 2, il nous faut caractériser les transformations des usages, et conformément au périmètre des transformations (Figure 6), il est nécessaire d’étudier :

- la diversité des usages présents, afin de comprendre la situation actuelle de l’organisation ;
- la diversité des usages futurs « potentiels », en se projetant dans le futur ;
- prendre des décisions sur les transformations désirées des usages.

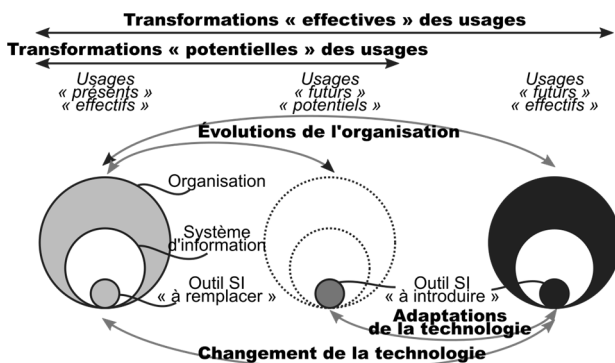


Figure 6 : Périmètre des transformations des usages

Nous définissons ainsi 3 méthodes :

- Méthode 1 : Exploration de la diversité des usages « présents » « effectifs » ;
- Méthode 2 : Exploration de la diversité des usages « futurs » « potentiels » ;
- Méthode 3 : Exploration de la diversité des transformations « potentielles » des usages.

Dans le cadre de cet article, nous nous focalisons sur la Méthode 1, de façon à illustrer le principe général de ces méthodes.

4.2 Caractérisation de la Méthode 1

L'exploration de la diversité des usages « présents » « effectifs » d'un outil vise à identifier différents cas d'usage (représentation statique avec 4 variables d'usage) et de différents scénarios d'usage (représentation dynamique avec 5 variables d'usage) par lesquels est concerné l'outil du SI « à remplacer ».

Nous proposons une méthode en 3 étapes :

1. Définition des observables.
2. Exploration de la diversité des usages présents à partir de la variable « Quoi ? ».
3. Exploration de la diversité des usages présents à partir de la variable « Pour quoi ? ».

L'étape 1 a pour but de caractériser le périmètre à étudier au sein de l'organisation. Il s'agit de définir les modalités pour les 4 variables initiales considérées indépendamment les unes des autres : quels sont les différents usagers à envisager, les activités à analyser, les outils du SI à examiner et les différents lieux possibles.

Les étapes 2 et 3 s'appuient sur les résultats de cette 1^{ère} étape pour caractériser les cas et scénarios d'usage qui concernent l'outil SI « à remplacer ». Nous proposons de considérer l'exploration comme graduelle, où une seule (nouvelle) variable est explorée à la fois. Ainsi l'étape 2 vise à explorer la diversité des cas d'usage à partir d'une analyse détaillée de l'outil SI dans l'usage – ses Composants, les Informations/données supportées (variable « Quoi ?). À partir de là est définie la diversité des usages pour l'ensemble des Informations/données supportées par l'Outil SI, avec les différentes Actions réalisées par les Individus, et ceci depuis plusieurs Lieux.

L'exploration de la diversité des usages à partir de la variable « Pour quoi ? », centrée sur la Fonction de l'organisation, démarre par l'analyse d'un processus métier, et de différentes activités, actions et opérations le

composant. Cette étape 3 permet de poser un cadre pour les Actions/opérations réalisées avec l'aide de l'outil, en les inscrivant dans le contexte plus global du Processus métier composé de différentes Activités, supportées ou pas par l'Outil SI étudié.

Nous identifions plusieurs techniques de recueil de données compatibles avec les objectifs de cette méthode, par exemple :

- analyse du contenu de l'Outil SI « à remplacer » avant l'usage et dans l'usage ;
- entretiens individuels et/ou collectifs avec les usagers ;
- questionnaires ou enquêtes auprès des usagers ;
- observations des usagers ;
- analyse des documents de l'entreprise (p.ex. procédures, protocoles).

4.3 Synthèse et résultats

Dans un projet de remplacement d'un outil du SI par un autre, l'étude des transformations « potentielles » des usages se base sur la comparaison de la diversité des usages « présents » « effectifs » (Méthode 1), supportés par l'Outil SI « à remplacer », et de la diversité des usages « futurs » « potentiels » (Méthode 2), supportés par l'Outil SI « à introduire ». Bien entendu, l'Outil du SI « à introduire » doit être préalablement identifié afin de pouvoir l'étudier. L'exploration des transformations « potentielles » des usages (Méthode 3) s'appuie sur la mise en correspondance de cas d'usage et de scénarios d'usage « présents » « effectifs » et « futurs » « potentiels ». Il s'agit ensuite pour l'organisation de choisir les cas d'usage et les scénarios d'usage « présents » qui doivent être conservés, ceux « futurs » qui doivent être introduits (potentiels) ou encore ceux qui doivent être combinés, par exemple.

Dans le cadre de nos travaux, nous avons pu suivre l'ensemble d'étapes du projet d'évolution et de déploiement du dossier de soins informatisé auprès du personnel de la structure de santé, des professionnels libéraux qu'elle coordonne, et de plusieurs de ses patients. Nous avons participé à l'élaboration des cahiers de charges, des tests et de recettes tout au long de différentes versions du logiciel ou de ses fonctionnalités. Nous avons réalisé le déploiement pilote auprès de 6 patients et le déploiement généralisé auprès plus de 70 patients. Nous avons formé plus de 200 soignants salariés et externes. Nous avons participé à la création du guide utilisateur et de plus d'une dizaine de procédures internes. Tout au long de notre présence sur le terrain, nous avons cherché à capturer l'usage du dossier de soins à travers : des observations d'utilisation du classeur par une dizaine d'infirmiers et d'aides-soignants lors de leurs tournées ; des dizaines d'entretiens individuels avec des patients, des médecins traitants et des pharmaciens d'officine, réalisés à l'occasion du déploiement de l'outil numérique ; la mise en place d'un groupe de travail dédié à la reconception d'un menu dédié à la gestion des médicaments ; ou encore l'analyse du contenu de documents

papier dans les archives ou de menus pour les patients pris en charge avec un dossier - tablette.

Les expérimentations constituent un point fondamental de notre démarche projet, en particulier pour étudier les usages « futurs » « potentiels ». En effet, une fois les cas et les scénarios d'usage « futurs » « potentiels » identifiés, il s'agit de les évaluer par rapport aux transformations dont ils sont porteurs pour l'organisation. Cette évaluation, autant quantitative que qualitative, peut mener aussi bien à des adaptations de l'outil (changements des profils utilisateurs, des informations/données supportées, des fonctionnalités, etc.), qu'à des adaptations de l'organisation (changements des usagers, des processus métier supportés). Par exemple, même si l'outil dossier de soins informatisé permettait au « Médecin traitant » de « Prescrire les médicaments » depuis le « Domicile » à l'aide d'un stylet, la structure de santé a demandé à l'éditeur de permettre une saisie d'ordonnances au clavier pour améliorer la lisibilité de la liste du traitement. En même temps, l'introduction par l'éditeur d'une fonctionnalité non prévue de saisie de prescriptions via des photos des ordonnances papier a amené la structure à revoir ses flux informationnels et décisionnels, et à se prononcer sur les profils des futurs-usagers prescripteurs. Aujourd'hui, l'outil numérique est devenu le nouveau dossier de soins des patients relevant du service d'hospitalisation à domicile. Ce dossier de soins informatisé s'accompagne d'un classeur papier, afin de pallier aux aléas des pannes du matériel et des problématiques de gestion des accès des intervenants. Il a pu être déployé auprès d'environ 20% de malades pris en charge quotidiennement par la structure. Les séjours de 80% de malades restant sont gérés, pour des raisons principalement économiques (coût important du matériel informatique), toujours avec l'ancien outil papier.

La mise en œuvre du projet d'informatisation a été avant tout une aventure humaine, où chaque personne a réagi au changement suivant ses propres expériences et valeurs. Par exemple, pour certains patients et leurs proches, l'informatisation de données médicales à domicile était le symbole de leur sécurisation grâce à l'accès par identifiants personnels, pendant que pour d'autres la mise à disposition des données à distance était vue comme négative. De même, certains soignants percevaient l'outil comme fastidieux, pendant que d'autres comme ludique et intuitif.

4.4 Discussion

Le modèle des usages proposé a pour but de formaliser l'ensemble des informations caractérisant les usages d'un outil du SI dans le but de les identifier de façon systématique, visant une meilleure exhaustivité, et de les analyser plus facilement.

Ce modèle a été utilisé avec la structure de santé avec laquelle nous avons travaillé et son utilisation s'est particulièrement révélée utile pour cadrer les usages « présents » « effectifs » de l'outil à remplacer et les usages « futurs » « potentiels » de l'outil du SI à déployer. L'identification des écarts entre les usages de ces deux

outils, ainsi que l'évaluation des usages à travers les protocoles et scénarios d'expérimentation, ont permis de faciliter la prise de décisions par la structure de santé sur les évolutions dans son organisation et sur les adaptations du dispositif numérique.

La diversité des usages d'un outil du SI se manifeste par une multitude d'informations et de données supportées, par un grand nombre de profils de ses usagers, par de nombreux processus pour lesquels il est support, et/ou par l'hétérogénéité des lieux où il est utilisé. Elle se caractérise également par différents niveaux de granularité de sa description.

Les trois méthodes proposées sont destinées à proposer une exploitation directe du modèle d'usage, dans le but de comprendre l'existant et de permettre de se projeter vers un fonctionnement futur. Ces méthodes ne constituent pas une méthodologie complète qui permettrait de piloter le projet de déploiement d'un nouvel outil du SI. Elles ont vocation à s'intégrer à des démarches existantes de réingénierie des processus ou d'évolution du système d'information, afin de compléter et améliorer certaines étapes : analyse de l'existant, diagnostic, élaboration de spécifications fonctionnelles, voire techniques, etc. Si ces démarches partent généralement du global vers le particulier, elles ne descendent pas jusqu'au niveau de l'individu et beaucoup de solutions rencontrent des difficultés lors de leur déploiement. Notre positionnement est inverse afin de concilier la prise en compte de l'individu, à partir de représentations basées sur l'usage, pour enrichir et consolider les solutions étudiées et impacter les processus métiers en les rendant plus efficaces. Cette intégration via les trois méthodes proposées est encore à l'étude et fera l'objet de prochains travaux.

Par ailleurs, si nous avons pu étudier l'innovation organisationnelle et les interactions bidirectionnelles entre technologie et organisation à un niveau de granularité très détaillé, nous n'avons pas pour mission de remettre en cause de façon importante le fonctionnement de l'organisation, donc ses processus métiers ou ses mécanismes de prise de décision. Cela est dû en partie au fait que la structure de santé a déjà mené les années précédentes un travail important avec ses salariés et les professionnels de santé pour faire évoluer son référentiel (dossier patient, dossier de soins, procédures, etc).

L'étude exhaustive de cette diversité a nécessité un travail très important que nous avons pu mener à bien par une grande présence sur le terrain durant ces 3 années. Notre modèle et nos méthodes ne sont pas directement transférables par exemple à des consultants du fait de cette charge de travail importante.

5 CONCLUSION

Notre recherche est centrée sur l'accompagnement des transformations d'une structure de prise en charge à domicile dans le cadre de son projet d'informatisation du dossier de soins au chevet des malades. Nous avons proposé d'appréhender la multitude des évolutions dans l'organisation et des adaptations de l'outil à travers l'étude de la diversité des transformations « poten-

tielles » et « effectives » des usages. Afin d'identifier cette diversité nous avons proposé un modèle d'usage, basé sur cinq variables et quatre niveaux de granularité, pour tenir compte de la diversité des usages. À partir de ce modèle, nous avons défini trois méthodes d'exploration graduelle de la diversité des usages et de la diversité des transformations des usages, orientées organisation et outil.

Une fois identifiées, les transformations « potentielles » peuvent être maîtrisées à travers leur évaluation, menant à des décisions sur les usages à délaissés, à conserver, à adopter, jusqu'à l'émergence de nouveaux usages. Dans nos travaux futurs, nous prévoyons d'étudier plus en détail les différentes façons d'évaluer les transformations des usages, et ceci à différents moments d'implémentation d'un nouvel outil.

Les travaux présentés ont exploré le lien entre technologie et organisation à un niveau de granularité fin, centré sur les usagers et leurs activités. Nos travaux à venir se pencheront sur l'impact des transformations au niveau de l'organisation et de ses processus métiers.

Notre approche a permis à la structure de santé avec laquelle nous avons travaillé de déployer complètement l'outil du SI choisi, tout en obtenant l'adhésion de son personnel, des professionnels de santé qu'elle coordonne, ainsi que des patients et de leur entourage.

REMERCIEMENTS

Nous tenons à remercier les responsables et l'ensemble des salariés de la structure de santé avec laquelle nous avons travaillé sur ce projet.

REFERENCES

- Ammenwerth, E., Iller, C., and C.Mahler, 2006. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. *BMC Medical Informatics and Decision Making*, 6(3).
- Borgiel, K., 2017. Proposition d'une approche centrée usage pour l'accompagnement de l'innovation organisationnelle liée à l'introduction d'un outil numérique collaboratif dans un système d'information : application au contexte du « home care ». Thèse de doctorat, Université de Bordeaux.
- Brangier, E., 2004. La boucle usage - adaptation - reconception : l'usage comme intégration des points de vue de l'utilisation et de la conception. *Congrès SELF 2004*, p. 535-543.
- Bricon-Souf, N., F.Anceaux, N.Bennani, E.Dufresne, and L.Watbled, 2005. A distributed coordination platform for home care: analysis, framework and prototype. *International Journal of Medical Informatics*, 74, p. 809-825.
- Buisine, S. and B.Roussel, 2008. Analyse de l'usage. In: B. Yannou, H. Christofol, D. Jolly, N. Troussier (Eds.), *La conception industrielle de produits, vol. 3: Ingénierie de l'évaluation et de la décision*, Chapitre 2, p. 43-60, Paris : Lavoisier Hermès Science.
- Burton-Jones, A. and M.J.Gallivan, 2007. Toward a Deeper Understanding of System Usage in Organizations: a Multilevel Perspective. *MIS Quarterly*, 31(4), p. 657-579.
- Derry, S. J., 1990. Flexible cognitive tools for problem solving instruction. *Annual meeting of the American Educational Research Association*, Boston, MA, April, p.16-20.
- DGOS, 2014. Programme Hôpital Numérique. *La politique nationale relative aux systèmes d'information hospitaliers*. Direction Générale de l'Offre de Soins.
- Equixor, I., 2016. La méthode QOOQCP, un outil d'analyse. *Techniques de l'Ingénieur : Fiche pratique* 0447.
- Fidock, J., and J.Carroll, 2012. Theorizing about the Life Cycle of IT Use: An appropriation perspective. In: D.N. Hart, S.D. Gregor (Eds) *Information Systems Foundations: Theory Building in Information Systems*, p. 79-112. ANU E Press.
- FNEHAD, 2009. *Livre Blanc des Systèmes d'Information en Hospitalisation à Domicile*. 1ère édition - Juin 2009. Fédération Nationale des Etablissements d'Hospitalisation à Domicile.
- Jonassen, D.H., 1992. What are Cognitive Tools? In: Kommers, P.A.M., Jonassen, D.H., Mayes, J.T., Ferreira, A. (Eds) *Cognitive tools for learning*, p. 1-6. Springer.
- Laudon, K.C., and J.P.Laudon, 2012. *Management Information Systems. Managing the Digital Firm*. Pearson Prentice Hall.
- Le Moigne, J.-L., 1990. *La modélisation des systèmes complexes*. Paris, Dunod.
- Leonard-Barton, D., 1988. Implementation as mutual adaptation of technology and organization. *Research Policy*, 17(5), p. 251-267.
- Lorenzi, N.M., and R.T.Riley, 2003. *International Journal of Medical Informatics*, 69, p. 197-203.
- Lyytinen, K., and M.Newman, 2008. Explaining information systems change: a punctuated socio-technical change model. *European Journal of Information Systems*, 17, p. 589-613.
- Merlo C., Vicien G., and Y.Ducq, 2014. Interoperability Modelling Methodology for Product Design Organisations. *Int. Journal of Production Research*, Taylor & Francis, 2014, Vol.52, Issue 15, p. 100-120.
- Morley, C., J.Hugues, B.Lebanc, and O.Hugues, 2005. *Processus métiers et S.I. Évaluation, modélisation, mise en œuvre*. Paris, Dunod.
- Proulx, S., 2005. Penser les usages des technologies de l'information et de la communication aujourd'hui : enjeux - modèles - tendances. In: L. Vieira, N. Pinède (eds.) *Enjeux et usages des TIC : aspects sociaux et culturels*, tome 1, Presses Universitaires de Bordeaux (Bordeaux), p. 7-20.
- Reix, R., B.Fallery, M.Kalika, and F.Rowe, 2016. *Systèmes d'Information et Management* (7ème éd.) Vuibert.

Urban Freight by Rail: An Algorithm Genetic for Optimizing the Transport of Goods

A. Sahli, S. Berraf

D. Uritu

W. Behiri

Université Paris-Est, ESIEE Paris,
Laboratoire d'Informatique Gaspard-Monge
UMR 8049, 93162 Noisy-le-Grand France
abderrahim.sahli@esiee.fr, sana.berraf@esiee.fr

Scuola di Ingegneria e Architettura
dell'Università di Bologna
40131 Bologna BO, Italy
doina.uritu@studio.unibo.it

Laboratoire Génie Industriel, CentraleSupélec
Université Paris-Saclay
Gif-sur-Yvette, France
walid.behiri@centralesupelec.fr

ABSTRACT: *This paper provides a genetic algorithm in order to solve the Freight Rail Transport Scheduling Problem (FRTSP) where goods have to be transported using the urban rail network. The FRTSP considers a single line with a known set of loading/unloading stations and a predefined trains timetable. The customer demand is also supposed to be known in advance such that each demand corresponds to one or several standard boxes to be transported from one departure station to a destination station. The FRTSP defines the loading times for each demand to a given train such that the total waiting time of goods at their departure stations is minimized. This problem has been proven to be NP-hard since it is an extension of the generalized assignment problem. The proposed genetic algorithm (GA) is developed to tackle big instances since the MILP reached its limits for 140 demands. The obtained computational results confirm its effectiveness.*

KEYWORDS: *Urban mobility, urban freight, rail transport, genetic algorithm.*

1 INTRODUCTION

The current development of e-commerce in European cities, changes in consumption modes and densification of cities population lead to a very important increase of urban goods transport demand. A direct consequence is an explosion of the corresponding generated kilometers which are expected to be tripled by 2050 (Van Audenhove et al., 2015). The most used mean of conveyance to transport goods within cities is the truck; in fact, urban freight transport by trucks represents 10% of the traffic, but it is responsible of 40% of pollutant emission (Behiri et al., 2016). In addition to the negative environmental impact, road vehicles are liable of causing road accidents, noise, and city land occupation. Facing with this situation, local authorities have reacted by introducing more restrictive transport policies concerning polluting vehicles, the size and the weight of road transportation vehicles, the time access to specific areas and so on. Also, European Union leaders have responded with more restrictive policies such as drastically lowering pollutant emissions linked to urban freight transport (Behiri et al., 2016).

Regarding this situation, sustainable alternatives to transport goods in urban areas should be found. Many alternatives are suggested in the literature. In (MacHaris et al., 2007), the authors have proposed to develop Electro-Mobility for goods deliveries. Some researches as (Browne et al., 2014) and (Dessouky and

Leachman, 1995) have proposed the use of rail network and infrastructure to transport freight as one of potential sustainable urban solution to decrease the greenhouse effect. However, rail network could not completely substitute to road mode since the issue of the last kilometer cannot be solved with this alternative. Nevertheless, the main idea is to reduce road pressure by shifting as much road traffic as possible toward rail and improve the use of the currently under-used passenger trains' space. In particular, this work investigate the feasibility of introducing freight flow into the passenger one. The "Grand Paris" project, which aims to shift toward resilient and smart city with innovating services and infrastructure, is one of the possible space where such system could be experimented. This project promotes sustainable development with the ambition to study the sharing of its resources in order to insert and absorb a part of the freight transport demand. Sharing the passenger rail network means the utilization of one or several physical resources (trains, stations and railways) either in the same operating time interval or at different moments. Indeed, in (Behiri et al., 2016), the authors have identified and classified all different resource levels by determining their respective critical complexity. They have considered the most critical level where all resources are shared which corresponds to the case where freight would be transported in the same train than passenger. In this context, several decision problems are identified and classified into strategic, tactical and operational problems. The

FRTSP is identified as one of the operational issues that has to be addressed. This problem is central and critical since the trains are cadenced to absorb the passenger flow and an additional flow which is freight without disturbing the passenger transport service. Moreover, the transport company would not accept to deteriorate its principal service but would be interested to increase its revenue once the insertion of freight flow could be done in hidden way for passengers.

This paper is an extension of the works of (Behiri et al., 2016), which have proposed a generic framework to model and optimize the sharing of urban rail network between freight and passengers. A Mixed Integer Linear Programming (MILP) has been proposed for the central problem which is the FRTSP. In the present paper, a metaheuristic based on genetic algorithm is proposed to complete the resolution schema by improving the limit in terms of instance size since the MILP reaches its limit for 140 demand.

The remainder of this paper is organized as follows. Section 2 is dedicated to the literature review where similar experimentations are reported and analyzed. Section 3 is devoted to the formulation of the FRTSP and its assumptions. Section 4 deals with the proposed GA and its adaptation to the FRTSP. Section 5 provides numerical results of GA. Finally, some perspectives of future works and conclusions are presented.

2 LITERATURE REVIEW

Goods flow corresponds to the delivery from upstream supply chain actors toward downstream ones (Nuzzolo and Comi, 2014). In France, between 85% and 90% of this traffic is ensured by road which represents 20% of urban traffic and 1/3 of pollutant emissions (Mairie de Paris, 2009). To cope with this situation many alternatives are studied, some of them are still at the design step and others are experimented at real scale.

2.1 Underground future solutions for freight rail transport

In the following, the works that are reported aims to improve the underground space for freight transport.

In the nineteen, a Japanese team aimed to develop a dedicated underground freight distribution system based on passenger rail network to cope with environmental issues in Tokyo (Kashima et al., 1993). CargoCap is another underground solution where electric vehicles could be used on rails for this issue (Cargo Cap, 2002). "Cargo Sous Terrain" (CST) is an underground freight transport system which is considered as a feasible option from technical and economic

points of view it could come into operations in 2030.

The Parisian project called "Chapelle international" currently under built, is a global urban logistic solution where urban rail shuttle would be used for the multi-modal logistics hub. The idea is to connect a big warehouse in the north of Paris with this hub by railway. Goods will be transported by trains and each train would transport 60 mobile boxes. After, trucks using NGV fuel would carry products to stores. This project could save approximatively 10 000 trucks per year.

2.2 Experimentations mixing freight and passengers transport

The idea of inserting freight into passenger flow using urban transportation system has been considered in (Trentini and Malhene, 2012). The transportation system is passenger buses where a space is dedicated to absorb a part of the urban freight. In (Masson et al., 2017), the authors have proposed a mathematical model to ensure goods deliveries from bus stops to final customers, using near-zero-emissions vehicles.

In Italy, the authors of (Nuzzolo and Comi, 2014) have studied the possibility of delivering goods using passenger trains, between Naples and Sorrento. They have proposed to share railways and stations. They have modified old passenger trains by removing seats and inside doors, to free more space for freight. Goods are transported during off-peak periods to avoid disturbances.

Another successful real experimentation is the one of MONORPIX (Dablanc, 2009): MONORPIX is a French supermarket chain and thanks to the collaboration with SNCF Group (Société Nationale des Chemins de Fer français), a part of its goods are delivered by rail to some supermarkets in Paris. The MONORPIX warehouse which is situated at Lieusaint is the departure point of goods toward bercy station using the commuter line D (abr. RER D).

Another experimentation concerns the subway waste collection in New York City through stations using modified subway cars. The collect is done overnight, and as the New York subway runs 24/7, it is mixed between freight (waste) transport and passenger one. This solution enables the collect of 14 000 tons of waste per year, with 11 dedicated subway trains, covering 359 stations with 567 stops.

Another interesting case in Dresden uses a portion of 4 km tram line to transport automotive parts, from Volkswagen warehouse toward the factory which is in the city center. This solution copes with limited storage space in the warehouse. This line can deliver 300 000 tons of products per year with 10 daily trips.

In (Dotoli, Epicoco and Seatzu, 2015), a train load planning technique has been proposed for intermodal rail-road terminals. The proposed method aims at maximizing the train commercial value while respecting priority, physical, financial, and prosecution constraints (i.e., taking into account containers that prosecute their trip after the first destination). In (Dotoli, Epicoco, Falagario, Angelico and Vinciullo, 2015), the authors have focused on the optimization of containers pre- and end-haulage by road in intermodal terminals, which is one of the most critical factors in door to door transport effectiveness and profit. In (Dotoli and Epicoco, 2016), the authors have proposed a technique to solve in an exact and optimal way the pick-up and delivery problem under the typical assumptions of intermodal transportation: full truck load, split delivery, clustered backhauls, and time windows.

3 PROBLEM DESCRIPTION

The Freight Rail Transport Scheduling Problem (FRTSP) deals with the physical and temporal assignment of client demands to trains subject to several technical and temporal constraints while the total waiting time of goods at their departure stations is minimized. The idea behind this objective function is to maximize the turnover which is directly proportional to the generated cash flow. In fact, since the space in stations is generally very limited and in order to maximize the transport company benefit, it is obvious that it should transport the most important goods volume. Nevertheless, the economical model should be precised in order to adjust the formulation of the objective function. At this stage of the study, this formulation is relevant to evaluate the behaviour of this transport alternative for goods deliveries.

The FRTSP considers a single line from the rail network. This line consists of S stations in which one can proceed to the loading and the unloading of goods. The goods are packed in boxes with standard dimensions. All trains are considered able to transport goods with a pre-dimensioned capacity which is expressed by a maximal number of standard boxes. Goods are transported to different clients, requiring different quantities of goods, with independent departure ready times. The assumptions of the problem are the following:

- There are J client demands per day and they are known in advance;
- A departure ready time r_j is attributed to every single demand j ;
- The number of boxes required for a client demand j is Q_j . In fact, clients can demand various product in different quantities, so standard boxes are required to pack the goods;

- The transport of goods take place during the off-peak hours of the operating period;
- During the operating period, the schedule of train is known and T trains are available to serve S stations;
- The departure time of trains from the first station is given by l_t ;
- The travel time between two successive stations is tt ;
- The maximum capacity that trains can dedicate to goods is $MaxCap$;
- Any station can be used as loading/unloading platform for goods, so for every demand j there are a departure station Dep_j and an arrival station Arr_j ;
- The minimum waiting time of trains at stations is $wait_{min}$ and the maximum waiting time is $wait_{max}$;

The objective in the FRTSP is to determine for each client demand j a loading time R_j and a train x_k on which j is transported. Hence, the objective function is expressed as follows: $Minimize \sum_{j=1}^J (R_j - r_j)$

4 GENETIC ALGORITHM

The FRTSP belongs to the class of NP-hard problems. Considering its complexity, an efficient solving method must combine both quickness and efficiency even on large size problems. In this case, the adoption of metaheuristics is the most suitable choice to implement. Metaheuristics enclose a set of general methods that can be adapted to different problems. For this purpose, Genetic Algorithm (GA) has been chosen, because this metaheuristic is able to explore a good part of the search space in a reasonable amount of time.

The Genetic Algorithm was first described in (Holland, 1992). The basic idea of GA is to simulate the process of biological evolution, natural selection and survival of the fittest in the population. The GA starts with the generation of an initial population, made up of a certain number of individuals, generated randomly or by using heuristics. Individuals are identified by a set of feasible solutions to the problem. The population is evaluated according to a fitness function, which judges the performance of every individual. Some parents are selected among this population and are subjected to a crossover operator. This operator will create new feasible solutions (the children). A mutation operator will then be applied to the children with a low probability. At some point of the algorithm, the renewal of the population consists in selecting some individuals from the

current population and to replace them with newly-created children. This whole process is identified as the creation of a new generation, and is repeated until reaching a stopping criterion. A general scheme of GA is provided in Figure 1.

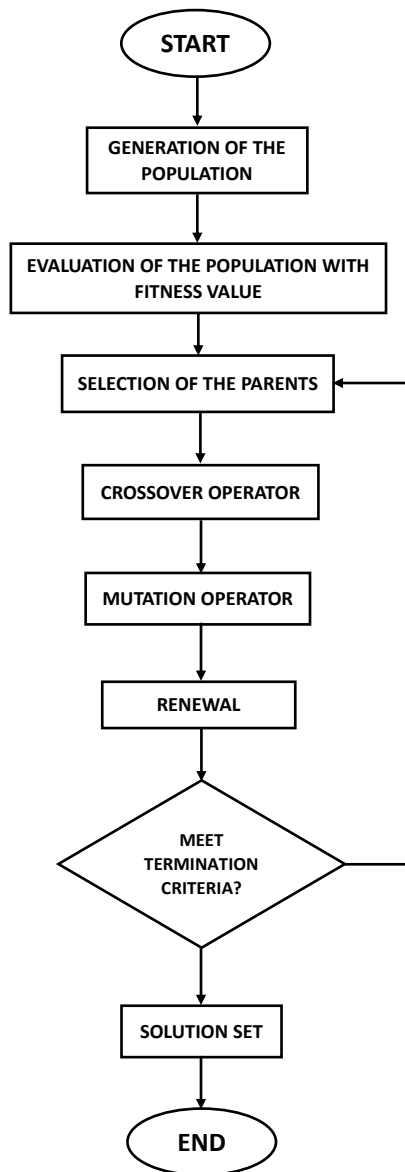


Figure 1 – General scheme of a GA

As previously mentioned, a metaheuristic is a powerful framework of general methods. So, every component of the GA must be adapted to the considered problem, starting with the coding of the individuals.

4.1 Coding

The coding is the first step of GA. This process permits to represent feasible solutions into individuals that can be handled by the algorithm. In the FRTSP, the objective is to assign each demand to a train. A unique number is attributed to each demand. Then

an individual is coded as a list of demand numbers. The demands are ordered in a decreasing way according to their priority. Each demand is considered before the demands that have a lower priority than it. So, the most priority demand is the first one to be assigned to a train. To have a clearer idea of the meaning of coding, an example is given below.

Consider an individual I made up of 10 demands (see Figure 2). Every demand is identified by a unique number: 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. If the individual is coded as (0, 9, 2, 3, 4, 8, 6, 7, 5, 1), this means that demand 0 is the first demand to be assigned to a train, and then demand 9, etc. We denote by $L_I(j)$ the value of the gene at the position j , where $j \in \{0, 1, \dots, J - 1\}$.

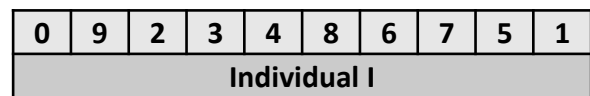


Figure 2 – Coding representation

This type of coding is named indirect, because it does not represent the direct solution (assignment of demands to trains). To obtain an assignment, the individual must be decoded. Figure 3 illustrates the steps to build a demand-train assignment. Each demand $L_I(j)$ is affected to the first train t which respects all the constraints of the problem. Then the next demand $L_I(j + 1)$ is considered. We denote by $L_{I'}(j)$ the value of the gene at position j in the decoded individual I' . The constraints which must be respected by each train, when a new demand is affected to it, are the following:

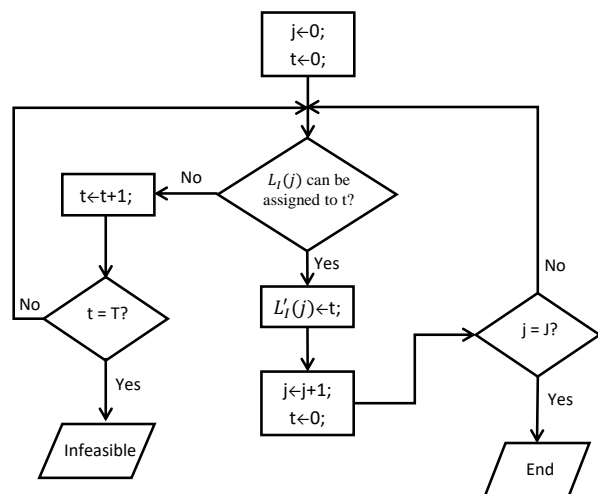


Figure 3 – Decoding algorithm

- It must always respect the minimum and maximum waiting times at each station.
- The quantity of goods transported by it does not exceed its maximum capacity at each station.
- It must arrive at the departure station of the new demand after the departure ready time of the demand.

If for some demand no train is found, then the individual does not represent a feasible solution. We say that an individual is feasible if it represents a feasible solution. Figure 4 provides an example of decoding. According to this example, demand 0 is assigned to train 1, demand 9 is assigned to train 4, etc.

Coded Individual									
0	9	2	3	4	8	6	7	5	1
Decoded Individual									
1	4	5	1	0	5	5	0	9	6

Figure 4 – Example of decoding an individual

4.2 Evaluation

The evaluation step consists in determining the fitness of each individual. The fitness represents the performance of the considered individual related to the optimization criterion. In our GA, the evaluation step will be the calculation of the total waiting time of demands at their departure station induced by an individual.

Beforehand, the individual has to be decoded. So, the sequence of genes must be translated into a feasible solution. In biology, this step corresponds to the translation of the genotype into the phenotype.

Let us consider an individual I , coded following the previous coding. $L_I(j)$ is the value of the gene at position j where $j = 0, \dots, J-1$. We denote by $L_{I'}(j)$ the value of the gene at the same position in the decoded individual I' (I' is a list of trains rather than a list of demands). In fact, each demand $L_I(j)$ is assigned to train $L_{I'}(j)$. By considering the decoded individual, the waiting time C_{ts} of each train t at each station s can be computed. After that, we can compute the loading time of each demand j as follows:

$$R_j = l_t + tt \times dep_j + \sum_{s=0}^{dep_j-1} C_{ts} \quad \forall j \in \{0, 1, \dots, J-1\}$$

where $t = L_{I'}(i)$ and $j = L_I(i)$.

The last step is to compute the total waiting time of demands at their departure station associated with the solution. We denote by $RT(I)$ the fitness value of individual I :

$$RT(I) = \sum_{j=0}^{J-1} (R_j - r_j)$$

4.3 Generation of the Initial Population

A population in this algorithm consists of N feasible individuals. Several methods can be used to generate the initial population. The individuals usually are generated randomly or using simple heuristics. In our method, they are first generated randomly. Then six simple heuristics are used to improve the efficiency of the algorithm. The implemented heuristics are the following:

- FIFO (First In First Out): the main factor of this heuristic is to respect the departure ready time of demands. It means that, the demands are sequenced in nondecreasing of their availability date. So, the first ready demand is the first one to be assigned to a train.
- MIN-DISTANCE: the idea of this heuristic is based on the distance that a client demand must cover. The parameter is calculated as the difference between the arrival and the departure demand stations. In this case, the most urgent demand to be assigned to a train is the one with the shortest distance.
- MAX-DISTANCE: this heuristic is the opposite of the previous one. The most priority demand to be assigned to a train is the one that presents the longest distance between its arrival and departure stations.
- MIN-BOXES: this heuristic takes into account the number of boxes of each client demand. So, the demand which has the lowest number of boxes is the most priority demand.
- MAX-BOXES: this heuristic is the opposite case of the last one, which means that the most priority demand is the one with the largest number of boxes.

To ensure that the initial population is formed by feasible individuals, all the genes must be decoded and respect the above-mentioned constraints.

4.4 Crossover

The crossover operator is the process in which individuals from the population exchange genetic information, hopefully to create a better individual.

Before the crossover operation, a number of parents have to be selected. In our GA, each individual in the population is considered for crossover. In fact, a crossover rate parameter, that determines the frequency in which crossover is applied, is used. By comparing the crossover rate to a random number, we can decide whether the individual should have crossover applied to it, or whether it should be added into the next population unaffected by crossover. If an individual is selected for crossover then a second parent must be selected. To find the second parent, we have used the Roulette Wheel Selection method.

The Roulette Wheel Selection (also known as fitness proportionate selection) is a selection method which uses the analogy of a roulette wheel to select individuals. The idea is that individuals from the population are placed on a metaphoric roulette wheel depending on their fitness value. The higher the fitness of the individual, the more space it is allocated on the roulette wheel. Then a random selection is made similar to how the roulette wheel is rotated. In this way, an individual with a better fitness value has a better chance to be selected as a second parent.

For the crossover, we need a crossover operator. There are many operators. In our algorithm, we have adapted the 1X Crossover Operator which was initially proposed by (Lee et al., 1998). The principle of this operator is easy. Each parent is divided into two parts, and the children are made with one part from each parent (see Figure 5).

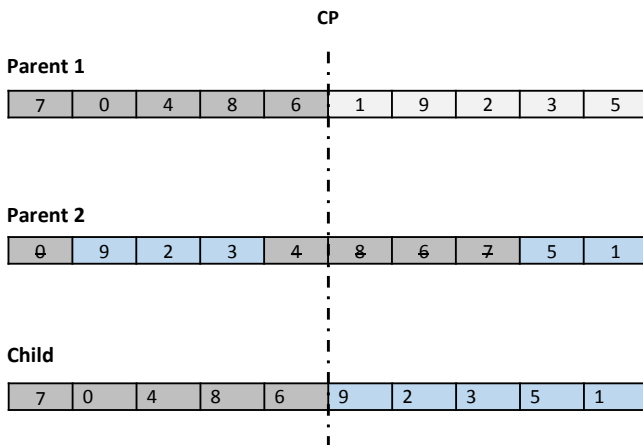


Figure 5 – 1X Crossover Operator

A cross point (CP) is randomly selected between position $A = 0.25 * J$ and position $B = 0.75 * J$. These positions ensure that the crossover provides children different from their parents. Then the following three steps are executed:

- **Step 1:** The first part of parent 1 is copied into

child.

- **Step 2:** All genes appearing in the first part of parent 1 are deleted from parent 2.
- **Step 3:** The remaining genes are copied from parent 2 to child.

4.5 Mutation

Once crossover is finished, a mutation is applied on some children. The objective of mutation is to explore new solutions that may not be reached by crossing existing solutions. For each child, the probability of being mutated is 25%. The Precedence Preserving Shift mutation (PPS) has been chosen as a mutation operator.

The PPS was designed by (Lee et al., 1998). In this operator, a position is randomly generated. Then, the corresponding gene is swapped with another gene, randomly chosen between the immediate successor and predecessor of the demand corresponding to the drawn position (see Figure 6).

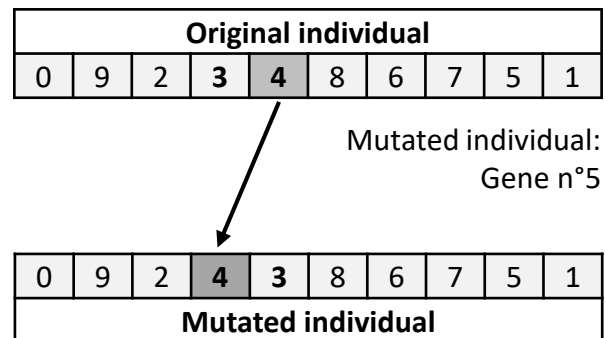


Figure 6 – PPS Mutation Operator

4.6 Renewal

Once the crossover and mutation are finished, a number of children can be inserted into the population in order to create a new generation. In our method, the size of the population is kept constant. Therefore, some individuals must be discarded. For this purpose, the individuals of the population are sorted in nondecreasing order of their fitness values. Let $r(I)$ be the rank of individual I in this order. For each individual I , a random number is generated between 0 and N , where N is the population size. If this number is less than $r(I)$, then I is deleted and replaced by a child. So, the probability for an individual I to be deleted is:

$$P_{del}(I) = \frac{r(I)}{N}$$

In this way, individuals with a better fitness value have less chances to be deleted. The algorithm repeats this procedure until all children are inserted into the new population.

4.7 Stopping Criterion

GA can employ different conditions to stop the search. In our method, the number of generations is set as a stopping criterion. So, the algorithm stops when it reaches a maximum number of generations. Then, it gives the best solution appearing in all the generations.

5 COMPUTATIONAL RESULTS

Our genetic algorithm has been implemented in java and tested on Intel Xeon CPU E3-1245 v3 at 3.40 Ghz. It has been trained on 250 instances randomly generated by (Behiri et al., 2016). These instances are classified into 10 categories according to the number of demands. Each instance considers a typical line of 10 stations indexed from 1 to 10. The goods are transported from the lowest-indexed station to the highest one. There are 30 trains which travel through the line with a frequency defined as one train every 10 minutes. The capacity allocated to freight in each train is 15 boxes. The travel time between two successive stations is 5 minutes. The minimum waiting time of each train at each station is 30 seconds and the maximum waiting time is 1 minute. The time required to load or unload each box is 10 seconds. The other parameters have been randomly generated.

Table 1 – Computational results.

Number of demands	GA(25)		GA(50)		GA(100)	
	Gap	CPU	Gap	CPU	Gap	CPU
10	0.0	0.2	0.0	0.3	0.0	0.6
20	0.0	0.3	0.0	0.6	0.0	1.2
30	0.0	0.6	0.0	1.0	0.0	2.0
40	0.0	0.8	0.0	1.6	0.0	3.0
50	0.0	1.3	0.0	2.5	0.0	5.0
60	0.2	1.9	0.1	3.6	0.0	7.2
70	0.3	2.6	0.2	4.9	0.0	9.6
80	0.6	3.4	0.2	6.5	0.0	12.6
90	1.5	4.0	0.5	7.9	0.1	15.3
100	3.4	5.0	1.4	9.5	0.5	19.4

In our computational studies, we have fixed the crossover rate to 0.9 and the maximum number of generations to 500. To set the most appropriate population size, we have tested GA with different values. We denote $GA(x)$ the genetic algorithm when the population size is set to $x \in \{25, 50, 100\}$. Table 1 reports the computational results for each category of instances. In this table, we use the following abbreviations:

- *Gap*: the average deviation in percent from the optimal solution value.

- *CPU*: the maximum CPU time required for each instance of each category, in seconds.

As we can see, our GA is very efficient on the instances with up to 50 client demands. In fact, all these instances are solved by our algorithm for each population size $x \in \{25, 50, 100\}$. On the instances with up to 100 client demands, we observe a direct correlation between the population size and the efficiency of the algorithm. In fact, $GA(100)$ yields the tightest average deviation *Gap*, but it is too costly in computation time. $GA(50)$ provides the second best average deviation and it is faster than $GA(100)$. $GA(25)$ provides a good performance while being impressively fast. Therefore, we deduce that with a larger population size, the search space is explored more than with a smaller population size. This will result in finding more efficient solutions in globally optimal areas of the search space. A small population size will often result in finding less arcuate solutions, however they need less computational resources per generation.

All these instances are solved by the MILP model proposed in (Behiri et al., 2016). Our GA is faster than this model. In fact, some instances have been solved by the MILP in more than 90 seconds. We have also tested these two approaches on 25 instances with 150 client demands. The MILP is not able to find the optimal solutions in less than 30 minutes. Moreover, our GA provides the best average deviation *Gap* in less than 3 minutes.

6 CONCLUSION

In this paper, we have dealt with the Freight Rail Transport Scheduling Problem (FRTSP) where goods have to be transported using the urban rail network. The FRTSP considers a single line with a known set of loading/unloading stations and a predefined timetable of trains. The customer demand is also supposed to be known in advance. The FRTSP defines the loading time of each demand to a given train such that the total waiting time of goods at their departure stations is minimized.

We have proposed a genetic algorithm GA to solve this problem. We have introduced dedicated heuristics to generate the initial population and a basic selection operator to pick the parents in the reproduction process. We have also adapted the one-point crossover operator which was initially proposed by (Lee et al., 1998). The proposed algorithm proved its efficiency by the quality of provided solutions. Perspectives related to this work are various. One of them is the improvement of the coding and the different proposed operators to reduce the computational time. Another perspective is to explore the possibility to develop other solution approaches, such as simu-

lated annealing, tabu, or fast heuristics, to tackle the problem.

References

- Behiri, W., Ozturk, O. and Bemokhtar-Berraf, S. (2016). Urban freight by rail: A milp modeling for optimizing the transport of goods, *6th International Conference on Information Systems, Logistics and Supply Chain, ILS 2016*.
- Browne, M., Allen, J., Woodburn, A. and Piotrowska, M. (2014). The potential for non-road modes to support environmentally friendly urban logistics, *Procedia-Social and Behavioral Sciences* **151**: 29–36.
- Cargo Cap (2002). Automated underground transportation of cargo, *The 5th transportation alternative for the transport of goods in congested urban areas*.
- Dabanc, L. (2009). Freight transport for development toolkit: Urban freight, *The World Bank, Waszyngton*.
- Dessouky, M. M. and Leachman, R. C. (1995). A simulation modeling methodology for analyzing large complex rail networks, *Simulation* **65**(2): 131–142.
- Dotoli, M. and Epicoco, N. (2016). A technique for the optimal management of containers' drayage at intermodal terminals, *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, IEEE, pp. 000566–000571.
- Dotoli, M., Epicoco, N., Falagario, M., Angelico, B. and Vinciullo, A. (2015). A two-step optimization model for the pre-and end-haulage of containers at intermodal freight terminals, *Control Conference (ECC), 2015 European*, IEEE, pp. 3472–3477.
- Dotoli, M., Epicoco, N. and Seatzu, C. (2015). An improved technique for train load planning at intermodal rail-road terminals, *Emerging Technologies & Factory Automation (ETFA), 2015 IEEE 20th Conference on*, IEEE, pp. 1–4.
- Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, MIT Press, Cambridge, MA, USA.
- Kashima, S., Nakamura, R., Matano, M., Taguchi, T. and Shigenaga, T. (1993). Study of an underground physical distribution system in a high-density, built-up area, *Tunnelling and Underground Space Technology* **8**(1): 53 – 59.
- Lee, K.-M., Yamakawa, T. and Lee, K.-M. (1998). A genetic algorithm for general machine scheduling problems, *Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES'98. 1998 Second International Conference on*, Vol. 2, pp. 60–66.
- MacHaris, C., Van Mierlo, J. and Van Den Bossche, P. (2007). Combining intermodal transport with electric vehicles: Towards more sustainable solutions, *Transportation Planning and Technology* **30**(2-3): 311–323.
- Mairie de Paris (2009). Le transport de marchandises dans la ville de paris, *Paris: Direction de la Voirie & des Déplacements, Agence de la Mobilité - Usagers & Marchandises*.
- Masson, R., Trentini, A., Lehuédé, F., Malhéné, N., Péton, O. and Tlahig, H. (2017). Optimization of a city logistics transportation system with mixed passengers and goods, *EURO Journal on Transportation and Logistics* **6**(1): 81–109.
- Nuzzolo, A. and Comi, A. (2014). Urban freight demand forecasting: a mixed quantity/delivery/vehicle-based model, *Transportation Research Part E: Logistics and Transportation Review* **65**: 84–98.
- Trentini, A. and Malhene, N. (2012). Flow management of passengers and goods coexisting in the urban environment: Conceptual and operational points of view, *Procedia-Social and Behavioral Sciences* **39**: 807–817.
- Van Audenhove, F., De Jongh, S. and Durance, M. (2015). Urban logistics - how to unlock value from last mile delivery for cities, transporters and retailers, *Arthur D. Little Future of Urban Mobility*.

VALIDATION DE LA FIABILITÉ DE COMPOSANTS AU STADE DE LA CONCEPTION

Daoud Ait-Kadi
Professeur titulaire
Département de génie mécanique
Université Laval, Sainte-Foy
Québec, Canada, G1K 7P4
Daoud.aitkadi@gmc.ulaval.ca

Abdelhak EL Khaïr EL Idrissi
et Nathalie Gouget
Direction de la recherche
Renault. France.

Résumé

La prise en compte de la fiabilité des composants au stade de la conception est une des préoccupations importantes des entreprises autant sur le plan purement économique que stratégique. L'analyse du processus de validation de la fiabilité d'un composant mécanique est à l'origine de cette réflexion. L'approche contrainte/résistance a été adoptée pour estimer la fiabilité de composants sujets à des sollicitations dont l'occurrence et l'intensité sont aléatoires et dont la résistance est également aléatoire. Cette étude propose une méthodologie de mise en œuvre de l'approche contrainte/résistance en procédant à partir d'une vision intégrée des différents concepts, des besoins exprimés par les divers intervenants et des outils de conception et d'analyse disponibles. Plusieurs indications sont fournies pour rationaliser les diverses actions aussi bien au niveau de la conception, de la fabrication que des essais de validation. Grâce aux nouvelles technologies de l'information, il est désormais possible de créer un environnement qui favoriserait les échanges entre les divers intervenants pour exploiter efficacement le potentiel de l'approche contrainte/résistance.

I-Introduction

La prise en compte de la fiabilité au stade de la conception est une des préoccupations importantes des entreprises qui visent à améliorer leurs avantages compétitifs sur le plan mondial.

Ces préoccupations, à la fois stratégiques et économiques, ont conduit plusieurs entreprises à revoir leurs produits, leurs procédés et leur organisation. Cet exercice a permis de mettre au point de nouvelles méthodes de travail qui tentent, grosso modo, de capitaliser sur le savoir faire de l'entreprise, sur la mise en place de nouvelles méthodes de travail et sur l'intégration d'outils efficaces d'aide à la décision. Ce processus de réingénierie des produits et des procédés a relancé l'intérêt pour certaines méthodes d'analyse jugées autrefois très onéreuses à mettre en œuvre et qui s'avèrent actuellement, grâce aux nouvelles technologies de l'information et aux nouveaux outils de conception, parfaitement adaptés au contexte réel d'exploitation des systèmes. Il s'agit, plus explicitement, de l'approche contrainte/résistance qui permet d'évaluer la fiabilité de composants sujets à des contraintes dont l'occurrence et l'intensité sont aléatoires et dont la résistance est également une variable aléatoire.

Les fondements de cette approche ont été introduits par plusieurs ouvrages de référence (voir Shooman[12], Lewis[7], Marcovici et Ligeron[8], Ireson et al.[5], Villemeur[13], Kapur et Lamberson[6], Chuech[2], Rao[11] et Carter[1]). Plusieurs contributions ont été publiées récemment sur le sujet (voir Xue et Yang[15], English et al.[4], Wang et Liu[14]). El Khaïr et Gouget [3] se sont penchés sur l'application de l'approche contrainte/résistance pour réévaluer les procédures d'essais de validation de la fiabilité de composants mécaniques sujets à des phénomènes de fatigue. Cette recherche s'inscrit dans le cadre des travaux élaborés par EL Khaïr et Gouget [3] qui visent essentiellement à développer une stratégie d'implantation de l'approche contrainte/résistance en procédant à partir d'une version globale de ses différents concepts, des besoins des divers intervenants, des outils analytiques et informatiques disponibles et enfin du savoir faire aussi bien à l'interne qu'à l'externe.

Pour aborder les différentes questions relatives à la mise en œuvre de l'approche contrainte/résistance, à la lumière des préoccupations manifestées par les divers intervenants dans le processus de conception d'une part et de validation de la fiabilité d'autre part, la structure suivante a été adoptée.

Les concepts de base de l'approche sont introduits à la section II. Pour mettre en évidence les actions qui pourraient découler d'une analyse du modèle de fiabilité reliant la contrainte et la résistance, la section III examine différents scénarios et indique la façon de faire pour apporter les correctifs jugés nécessaires pour une conception au juste nécessaire. Très peu de travaux publics ont considéré l'influence de la dégradation du composant sur la fiabilité. La section IV aborde cette question et fournit les principaux modèles qui permettent d'évaluer la fiabilité en tenant compte du processus de dégradation dû à la corrosion, à la fatigue, au vieillissement et autres. La section V est consacrée aux essais de validation. Elle reprend les objectifs et les enjeux techniques et économiques inhérents aux essais de validation. Cette section propose également un schéma général pour la validation de la fiabilité d'un composant. Une alternative aux essais de validation sur prototypes est proposée à la section VI. Une exploitation des méthodes de simulation sur prototypes virtuels est présentée. Un exemple d'application pour valider les différents concepts est traité. La conclusion de cette étude (section VII) résume le potentiel de cette approche et fournit des indications relatives à sa mise en œuvre. Une liste de références pertinentes est fournie à la section IX.

II- Approche contrainte/résistance. Concepts de base

Cette approche fait appel à la mécanique de la rupture et définit la fiabilité R d'un composant soumis à des contraintes l dont l'occurrence et l'intensité sont aléatoires et dont la résistance r est également une variable aléatoire. La fiabilité d'un tel composant est donnée par :

$$R = \Pr \{ l < r \} \quad (1)$$

La défaillance surviendra aussitôt que la contrainte l dépassera la résistance r du composant. La probabilité F qu'un tel événement se

$$F = \Pr\{l \geq r\} \tag{2}$$

Si $f_l(\bullet)$ désigne la fonction de densité associée à la contrainte et $f_r(\bullet)$ celle associée à la résistance, alors, l'expression de la fiabilité R est donnée par :

$$R = \int_0^\infty f_l(l) \left[\int_l^\infty f_r(r) dr \right] dl \tag{3}$$

ou encore
$$R = \int_0^\infty f_r(r) \left[\int_0^r f_l(l) dl \right] dr \tag{4}$$

La probabilité F de défaillance est donnée par :

$$F = \int_0^\infty [1 - F_l(r)] f_r(r) dr \tag{5}$$

où $F_l(r) = \int_0^r f_l(x) dx$.

Étant donné que le composant ne peut être qu'en opération ou hors d'usage, on peut écrire :

$$R + F = 1 \tag{6}$$

Connaissant les fonctions de densité $f_l(\bullet)$ et $f_r(\bullet)$, il est possible de calculer la fiabilité R du composant en utilisant soit la relation (4) ou (5) et (6).

Dans le cas général, on fait appel à des procédures numériques pour calculer R et F . des résultats numériques ont été obtenus par Kapur et Lamberson[6], pour différentes distributions de probabilités généralement utilisées en fiabilité à savoir : exponentielle, Gamma, Weibull, Normale, lognormale et valeurs extrêmes.

Pour un composant donné, le choix de la distribution $F_l(\bullet)$ associée aux contraintes est déduite à partir d'une analyse de son profil de mission. La distribution $F_r(\bullet)$ de la résistance est en revanche, obtenue à partir de résultats d'essais effectués sur bancs ou par simulation. Cette seconde façon de faire sera examinée à la section VI.

Il importe de souligner que les relations (4) et (5) sont établies pour un mode de défaillance donné. Une analyse des modes de défaillance permet d'identifier le mode dominant qui sera reproduit au niveau des essais.

Si la défaillance du composant est attribuable à plusieurs causes indépendantes ayant chacune une distribution de probabilité connue, les conditions du théorème central limite pourraient alors être valides pour toutes les combinaisons de distributions, ce qui permettrait d'adopter la distribution normale pour les résistances. Plusieurs résultats expérimentaux corroborent cette hypothèse.

Sans perte de généralité, on considérera le cas particulier où les contraintes l suivent une distribution normale de moyenne μ_l et de variance σ_l^2 ($l \sim N(\mu_l, \sigma_l)$) et les résistances r suivent une distribution normale de moyenne μ_r et de variance σ_r^2 ($r \sim N(\mu_r, \sigma_r)$). Ce cas particulier, largement couvert dans la littérature, est repris dans ce texte pour mettre en évidence certains éléments que le bureau d'études pourrait exploiter pour établir les actions à entreprendre en vue d'optimiser la conception du produit.

III- Exploitation du modèle contrainte/résistance dans le cas particulier où $l \sim N(\mu_l, \sigma_l)$ et $r \sim N(\mu_r, \sigma_r)$

Pour ce cas particulier, l'expression de la fiabilité (4) s'écrit :

$$R = \frac{1}{\sqrt{2\pi}} \int_\beta^{+\infty} \exp\left(-\frac{z^2}{2}\right) dz \tag{7}$$

où
$$\beta = -\frac{\mu_r - \mu_l}{\sqrt{\sigma_r^2 + \sigma_l^2}} \tag{8}$$

La figure 1 présente une illustration du modèle.

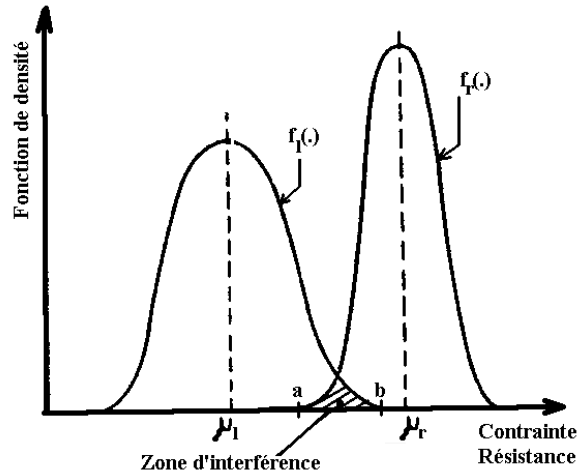


Figure 1- Illustration du modèle

L'aire de la zone de chevauchement des deux courbes $f_l(\bullet)$ et $f_r(\bullet)$ (zone d'interférence) correspond au risque de défaillance du composant. En effet, dans cette zone la probabilité F , que la contrainte soit supérieure à la résistance, est non nulle.

Une bonne connaissance du profil de mission du composant en clientèle et des moyens devant servir à sa réalisation permettraient de positionner correctement les deux distributions et d'évaluer la fiabilité R du composant. Si R^* représente le seuil de fiabilité spécifié dans le cahier des charges, alors les scénarios suivants peuvent être envisagés.

a) $R < R^*$

Les actions à entreprendre doivent conduire à une réduction de l'aire de la zone d'interférence pour la ramener à un seuil : $F^* \leq 1 - R^*$.

Le bureau d'études devra alors agir sur la valeur de la borne d'intégration β définie par la relation (8).

Le tableau 1 résume les actions envisageables pour réduire la valeur du paramètre β .

Paramètres	Actions envisageables														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
μ_r	+	0	+	0	+	0	+	0	+	0	+	0	+	0	+
σ_r	0	-	-	0	0	-	-	0	0	-	-	0	0	-	-
μ_l	0	0	0	-	-	-	-	0	0	0	0	-	-	-	-
σ_l	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-

Tableau 1 – Actions envisageables pour améliorer la fiabilité d'un composant

+ : augmentation du paramètre
 - : réduction du paramètre
 0 : paramètre inchangé

À titre indicatif, l'action 1 suggère d'améliorer la résistance moyenne sans agir sur les autres paramètres (σ_r, μ, σ). L'action 15 agit sur les quatre paramètres simultanément.

En pratique il n'est pas toujours facile d'agir sur les paramètres μ et σ . En effet ces derniers sont directement influencés par le comportement du client. Seules les actions 1,2 et 3 demeurent alors envisageables dans le contexte où aucune restriction sur l'usage du composant ne peut être imposée au client.

Mentionnons qu'une amélioration de μ_r peut être obtenue en jouant sur la qualité des matériaux utilisés, sur la forme, sur l'état de surface, sur les tolérances, sur le procédé de fabrication et d'assemblage, sur le choix du facteur de sécurité, sur le renforcement des points faibles, etc.

Une bonne maîtrise du procédé permet de garantir des produits homogènes et par conséquent un σ_r très faible.

Pour une décision rationnelle, il importe d'examiner l'influence des différents paramètres sur la fiabilité R du composant. Pour ce faire, on calcule les variations relatives de R par rapport à chacun des paramètres μ_r, σ_r, μ et σ . Ces variations sont données par les équations (9), (10), (11) et (12).

$$\frac{\partial R}{\partial \mu_r} = \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}} \times \frac{1}{\sqrt{\sigma_r^2 + \sigma_l^2}} = \psi \quad (9)$$

$$\frac{\partial R}{\partial \sigma_r} = -\psi \times \frac{\mu_r - \mu_l}{\sqrt{\sigma_r^2 + \sigma_l^2}} \times \sigma_r \quad (10)$$

$$\frac{\partial R}{\partial \sigma_l} = -\psi \times \frac{\mu_r - \mu_l}{\sqrt{\sigma_r^2 + \sigma_l^2}} \times \sigma_l \quad (11)$$

$$\frac{\partial R}{\partial \mu_l} = -\psi \quad (12)$$

Il en résulte qu'une réduction de μ_l produirait le même effet qu'une augmentation de μ_r . En revanche, les équations (10) et (11) suggèrent d'agir sur les causes qui engendrent la plus forte variation.

Pour tenir compte des différents facteurs qui peuvent influencer la tenue d'un composant (forme, matériaux, contraintes résiduelles, état de surface ..., etc.) un coefficient de sécurité ν est généralement utilisé. Dans un contexte déterministe, le facteur de sécurité ν_d , est donné par :

$$\nu_d = \frac{r_{\max}}{l_{\max}} \quad (13)$$

Dans un contexte probabiliste, le facteur de sécurité ν_p est donné par :

$$\nu_p = \frac{\mu_r}{\mu_l} \quad (14)$$

Une borne inférieure du facteur de sécurité ν_p a été établie pour garantir un niveau de fiabilité R donné (voir Kapur et Lamberson [6])

$$\nu_p \geq \frac{1}{(1 + \eta_r^2) - \sqrt{\eta_r^2 + \eta_l^2} \sqrt{\frac{R}{1-R}}} \quad (15)$$

$$\text{où } \eta_r = \frac{\sigma_r}{\mu_r} \text{ et } \eta_l = \frac{\sigma_l}{\mu_l} .$$

Un coefficient élevé conduirait à un surdimensionnement du composant et par conséquent à une augmentation de son coût.

b) $R > R^*$

Pour une conception au juste nécessaire, une analyse de l'écart $E=R-R^*$ s'impose pour tenter de réajuster, au besoin, les différents paramètres de conception, de fabrication et d'exploitation. Traditionnellement, lorsque la condition $b)$ est satisfaite, le produit est jugé acceptable. Aucun réajustement n'est prévu.

Il ressort de ce qui précède qu'il est possible d'évaluer la fiabilité d'un composant connaissant la distribution de probabilité des contraintes et des résistances. Le modèle (4) suppose, que la résistance du composant est indépendante du temps. En pratique, les phénomènes de fatigue, d'usure et de vieillissement couplés à de multiples autres tensions causés par le procédé de fabrication, altèrent la résistance du composant. La prise en compte de l'âge du composant dans l'approche contrainte/résistance est présentée à la section IV.

IV- Fiabilité d'un composant sujet à un processus de dégradation

Peu de travaux publiés sur l'approche contrainte/résistance ont abordé en profondeur la question relative au processus de dégradation de la résistance d'un composant en fonction du temps ou de tout autre unité d'usage. Lewis [7] et Shooman [12] exposent clairement le problème et proposent des modèles d'estimation de la fiabilité qui tiennent compte à la fois des contraintes, de la résistance et de l'unité d'usage du composant (le temps).

Xue et Yang [15] ont établi les expressions des bornes inférieure et supérieure de la fiabilité dans le cas où la résistance du composant se dégrade avec l'usage et ce, de manière aléatoire.

Si $R(t/r)$ désigne la probabilité conditionnelle de survie du composant au temps t , étant donné que sa résistance à cet instant est égale à r , alors, la fiabilité du composant au temps t (voir Lewis[7]) s'écrit :

$$R(t) = \int_0^\infty f_r(r) R(t/r) dr \quad (16)$$

Dans le cas où le taux de panne du composant est indépendant du temps mais varie avec sa résistance r , on obtient :

$$R(t/r) = e^{-\lambda(r)t} \quad (17)$$

Xue et Yang [15] indiquent trois façons de faire pour tenir compte du processus de dégradation d'un composant en fonction de son âge. La loi de dégradation $r(t)$ peut être :

a- Déterministe, auquel cas la résistance du composant est connue à chaque instant.

b- Aléatoire : Deux situations peuvent être envisagées :

b1) La loi de dégradation est connue et ses paramètres sont considérés comme des variables aléatoires.

b2) La fonction $r(t)$ est considérée comme une variable aléatoire dont les paramètres dépendent du temps.

Il va sans dire que $r(t)$ est une fonction décroissante du temps.

Si $r(t) \sim \mathbf{N}(\mu_r(t), \sigma_r(t))$; pour $t \in \mathfrak{R}$ alors :

pour un modèle de dégradation linéaire, on aura :

$$\mu_r(t) = \mu_0 + \alpha \cdot t \quad (18)$$

et
$$\sigma_r(t) = \sigma_0 + \xi \cdot t \quad (19)$$

Les paramètres μ_0, α, σ_0 , et ξ peuvent être obtenus à partir de valeurs observées de $\mu(t)$ et $\sigma(t)$ aux instants t_1, t_2, \dots, t_n .

Pour un modèle déterministe de dégradation, si les sollicitations se manifestent d'une manière aléatoire selon un processus de Poisson d'intensité γ alors, la fiabilité du composant est donnée par (Xue et Yang[15]) :

$$R(t) = \exp[-\gamma \int_0^t R_l(r(t)) dt] \quad (20)$$

où $R_l(r(t))$ est la probabilité de défaillance lorsque le composant subit une sollicitation d'intensité l au temps t .

Le taux de panne est donné par :

$$\lambda(r(t)) = \gamma R_l(r(t)) \quad (21)$$

Il ressort de cet exposé que la mise en œuvre de l'approche contrainte/résistance repose sur une bonne connaissance des conditions de réalisation, d'exploitation et de dégradation du composant. De plus en plus d'entreprises tentent de consolider leur expertise pour maîtriser les différents paramètres qui influencent le coût, la qualité, la fiabilité, la sécurité et la disponibilité de leurs produits.

Dans l'état actuel des choses, les entreprises sont tenues de démontrer les performances de leurs produits en procédant à des essais, normalisés ou non. Les contraintes de coûts et de délais font en sorte qu'on doit agir sur les facteurs environnementaux pour accélérer le processus de dégradation et obtenir ainsi des données fiables à partir d'un nombre limité de prototypes du produit.

Un examen sommaire de cette question est présenté à la section V.

V- Essais de validation

Ces essais demeurent incontournables surtout pour les composants dont la défaillance peut affecter la sécurité de l'utilisateur. Il s'agit de démontrer, par des essais, que la fiabilité R d'un composant est, pour une mission donnée, supérieure à une valeur R^* spécifiée dans le cahier des charges.

Ces essais sont généralement effectués sur un échantillon de taille n extrait d'un lot homogène de N composants ($N \gg n$). Ils tentent de reproduire, sur banc ou virtuellement, les conditions d'utilisation du composant pour apprécier non seulement ses capacités à supporter les diverses sollicitations mais aussi à valider les hypothèses et valeurs adoptées au niveau du bureau d'études.

Il n'est pas toujours facile ni utile de reproduire fidèlement toutes les conditions d'utilisation normale du composant. Une analyse des modes de défaillance (AMDEC) permet généralement de d'identifier les modes dominants qui méritent une investigation poussée.

Une fois cette étape franchie, on doit spécifier la taille n de l'échantillon qui permettra d'affirmer, avec un niveau de confiance (C) donné, que l'objectif de fiabilité spécifié R^* est satisfait.

Si un échantillon constitué de n composants extrait d'une population de taille N ($N \gg n$) est soumis à l'essai et si les k défaillances ont été observées à la fin de l'essai, alors (voir EL Khaïr et Gouget[3]), la probabilité que la fiabilité d'un composant extrait de la population, soit supérieure à une valeur R est donnée par :

$$\Pr(\text{fiabilité} \geq R) = 1 - (n+1) c_k^n \sum_{J=0}^{J=k} (-1)^J \cdot \frac{R^{n-k+J+1}}{n-k+J+1} \quad (22)$$

$$\text{où} \quad c_k^n = \frac{n!}{k!(n-k)!} \quad (23)$$

Si aucune défaillance ne s'est produite durant l'essai ($k=0$), on obtient :

$$\Pr(\text{fiabilité} \geq R) = 1 - R^{n+1} \quad (24)$$

Ainsi, dans le cas où aucune défaillance n'a été observée à l'issue de l'essai de n composants, on peut écrire (relation 24) :

$$1 - R^{n+1} \geq C \quad (25)$$

Il en résulte que, pour assurer un niveau de confiance C , la taille n^* de l'échantillon à soumettre à l'essai, doit satisfaire la condition :

$$n^* \geq \frac{\ln(1-C)}{\ln R} - 1 \quad (26)$$

Pour un niveau de fiabilité R donné, la taille n de l'échantillon à soumettre à l'essai augmente avec le niveau de confiance C désiré. Par ailleurs, plus le niveau de fiabilité à valider est grand et plus la taille de l'échantillon à tester devient importante (pour un niveau de confiance C donné).

Ces constatations sont particulièrement importantes pour la planification des essais où le nombre de spécimens à soumettre au test est souvent imposé par des contraintes techniques, économiques ou autres.

Si la forme de la loi de fiabilité $R(t)$ est connue, alors, le principe de l'essai de validation repose sur l'affirmation suivante :

« Si pour une mission t_1 , on a $\Pr(\text{fiabilité} \geq R(t_1)) \geq C$ alors, pour toute mission t_2 du composant ($t_2 \neq t_1$), on aura $\Pr(\text{fiabilité} \geq R(t_2)) \geq C$. »

La fonction de fiabilité étant une fonction décroissante, alors : pour tout $t_1, t_2 \mid t_1 < t_2$, on a $R(t_1) > R(t_2)$. (Voir figure 2)

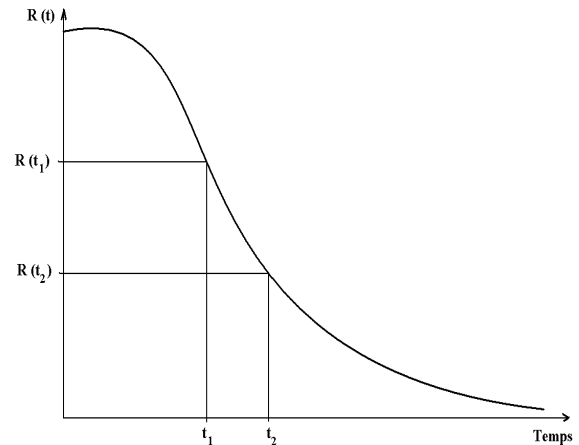


Figure 2 – Allure de la courbe de fiabilité $R(t)$

Ainsi, pour un niveau de confiance C donné, le nombre n_1 de composants requis pour valider $R(t_1)$ est supérieur au nombre n_2 de composants pour valider $R(t_2)$ (voir relation 26).

Dans le cas où les essais doivent être complétés dans un délai prédéterminé t_e et que les autres paramètres n et c sont fixés, le niveau de fiabilité $R_B(t)$ à valider au niveau de l'essai est obtenu à partir soit de la relation (25), dans le cas où aucune défaillance ne doit se produire durant l'essai, ou de la relation générale (22).

Si t_B désigne la durée de mission correspondant au niveau de fiabilité $R_B(t)$, deux situations peuvent se présenter :

- 1- La durée de mission est inférieure au délai t_e imposé pour compléter les essais.
- 2- $t_B > t_e$

Le premier cas n'impose aucune restriction particulière. La durée de l'essai est tout simplement égale à t_B . Le second cas, en revanche, est plus restrictif. Pour discuter de la façon de faire, dans le contexte de l'approche contrainte/résistance, désignons par : $F_B(t_B) = 1 - R_B(t_B)$, le risque de défaillance correspondant à une durée de mission de t_B . $F_B(t_B)$ n'est autre que l'aire de la zone d'interférence indiquée à la figure 3.

La durée de mission t_B étant supérieure à t_e , il en résulte que : $F_B(t_B) > F_B(t_e)$.

Pour les fins de discussion, et sans perte de généralité, on suppose que : $l \sim N(\mu, \sigma)$ et $r \sim N(\mu, \sigma)$.

Pour faire passer l'aire de la zone d'interférence de $F_B(t_e)$ à $F_B(t_B)$, on peut procéder par sévrisation des conditions de l'essai (augmentation de μ et/ou σ), par vieillissement (diminution de μ , et/ou augmentation de σ) ou par sévrisation et vieillissement tel qu'illustré à la figure 3.

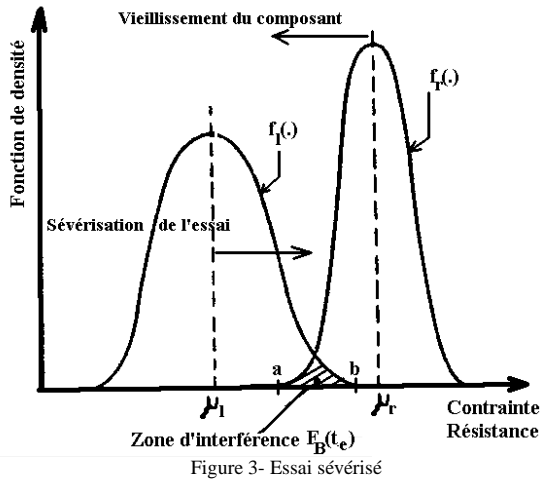


Figure 3- Essai sévéré

Le degré de sévérisation K de l'essai doit être tel que :

$$t_B = K t_e \tag{27}$$

Si D_r désigne l'endommagement résultant d'une exploitation du composant dans un environnement pris comme référence, et si D_s représente l'endommagement dans le cas d'un essai sévéré, alors, le facteur de sévérisation K n'est autre que :

$$K = \frac{D_s}{D_r} \tag{28}$$

Ainsi, si t_s est la durée de vie du composant dans un environnement caractérisé par son degré de sévérisation K , alors, sa durée t_r dans l'environnement de référence serait :

$$t_r = K t_s \tag{29}$$

Cette approche suppose que l'endommagement est proportionnel au degré de sévérisation de l'environnement. Le choix du facteur K est particulièrement important dans la conception et le pilotage des essais. Ce facteur influence non seulement la durée de l'essai (accélération du processus de dégradation) mais aussi la taille de l'échantillon.

Si $E(n,t,K)$ désigne l'essai caractérisé par la taille n de l'échantillon à tester, la durée t et son degré de sévérisation K , alors pour tout essai $E'(n',t',K')$ tel que $K' > K$, la taille n' de l'échantillon est inférieure à n .

Pour le cas particulier où $R(t)$ est une Weibull à deux paramètres

$$(\eta, \beta) \rightarrow R(t) = \exp\left(-\frac{t}{\eta}\right)^\beta, \text{ on obtient :}$$

$$n' = \left(\frac{K}{K'}\right)^\beta \times n \tag{30}$$

Pour $K' > K \Rightarrow n' < n$.

Une bonne connaissance des différents mécanismes de dégradation permettrait de faire un choix éclairé des moyens à utiliser pour atteindre le degré de sévérisation désiré tout en évitant d'accélérer des modes de défaillance dont la probabilité d'occurrence, en exploitation normale, serait très faible.

Une sévérisation mal dosée peut conduire soit à de nouvelles investigations souvent trop coûteuses (essai trop sévéré) ou à la validation d'un composant dont les objectifs de fiabilité visés ne seraient pas nécessairement satisfaits. Une concertation étroite entre le bureau d'études et les centres d'essais s'impose. La disqualification d'un composant par des essais sévérés, peut, dans le cas où les essais ne sont pas normalisés, remettre en question l'organisation et le choix des paramètres de l'essai. Dans un souci de conception au juste nécessaire, on assiste de plus en plus à un effort d'intégration des différentes

expertises disponibles, pour réaliser des produits de qualité, dans un court délai et à coût minimum.

L'organigramme présenté à la figure 4 décrit brièvement la procédure d'évaluation de la fiabilité en utilisant l'approche contrainte/résistance.

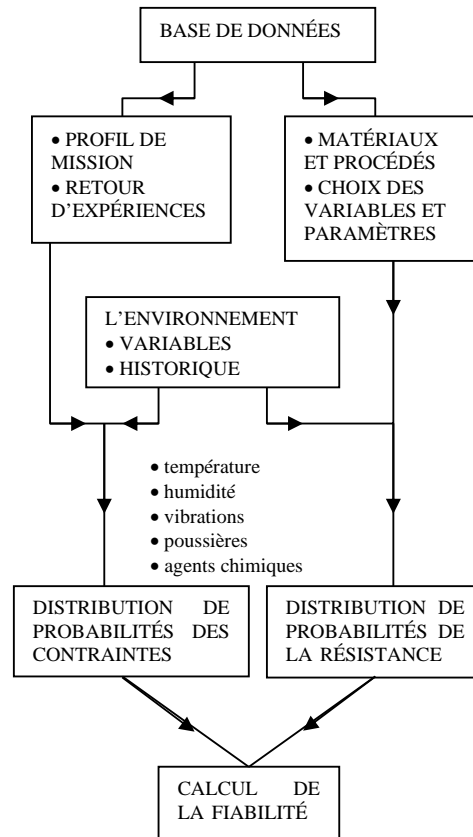


Figure 4- Procédure d'évaluation de la fiabilité à l'aide de l'approche contrainte/résistance

Dans cette optique, on tente d'exploiter la puissance de l'outil informatique pour permettre aux divers intervenants, non seulement d'accéder à de nouvelles connaissances, mais surtout d'interagir et de minimiser les délais occasionnés, entre autres, par les multiples rétroactions imposées par le processus séquentiel classique de développement et de mise en marché de nouveaux produits.

Les logiciels de conception et de fabrication assistées par ordinateur proposent maintenant de nouveaux modules qui permettent d'effectuer des simulations dynamiques sur des prototypes virtuels. Un aperçu du potentiel d'exploitation de ces outils pour la mise en œuvre de l'approche contrainte/résistance est donné à la section VI.

VI- Exploitation de la simulation sur prototypes virtuels

La quasi totalité des logiciels industriels de conception et de fabrication assistées par ordinateur offrent des modules qui permettent de simuler le comportement d'un composant dans un environnement virtuel. Une fois le dessin du composant complété, le logiciel permet de visualiser le composant en trois dimensions, de l'orienter dans l'espace, de lui appliquer des charges, de déceler ses points faibles et d'effectuer les améliorations jugées pertinentes. Jusqu'à très récemment cet exercice exigeait des connaissances approfondies en conception, en modélisation, en éléments finis et en analyse des résultats. Les nouvelles générations de logiciels offrent à l'utilisateur une assistance à tous les niveaux du processus de conception et de validation. Grâce à la flexibilité introduite par le paramétrage, il devient possible d'effectuer rapidement toutes les modifications désirées et d'évaluer l'impact de chacune d'elles sur la tenue du composant.

Ce ne sont là que quelques unes des fonctions qui pourraient servir à la mise en œuvre de l'approche contrainte/résistance. Il s'agit en fait de trouver un moyen d'obtenir, à partir de la simulation dynamique, les différentes fonctions de densité $f_i(\bullet)$, $f_r(\bullet)$ et $f_f(\bullet)$. Pour ce faire, on génère un lot de n composants. À partir de la liste de tous les défauts potentiels recensés, on génère, aléatoirement, pour chaque composant du lot, les valeurs des paramètres (dimensions, module d'élasticité, coefficient de Poisson, forme, tolérances, etc.) susceptibles d'influencer la tenue du composant. Une fois le lot défini, les contraintes sont générées d'une manière aléatoire. Pour chaque contrainte générée, on détermine la proportion de pièces ayant survécu à l'essai. L'analyse des observations recueillies est ensuite effectuée à l'aide d'un logiciel qui permet d'obtenir la distribution de probabilité $F_r(\bullet)$ qui ajuste au mieux les données.

Une simulation préalable du profil de l'utilisateur type du composant permet d'établir la distribution des contraintes.

Pour valider l'approche décrite succinctement ci-dessus, on s'est servi du logiciel Pro/ENGINEER et du module Pro/MECHANICA [9]. La figure 5 présente les principaux modules ayant servi aux divers traitements effectués.

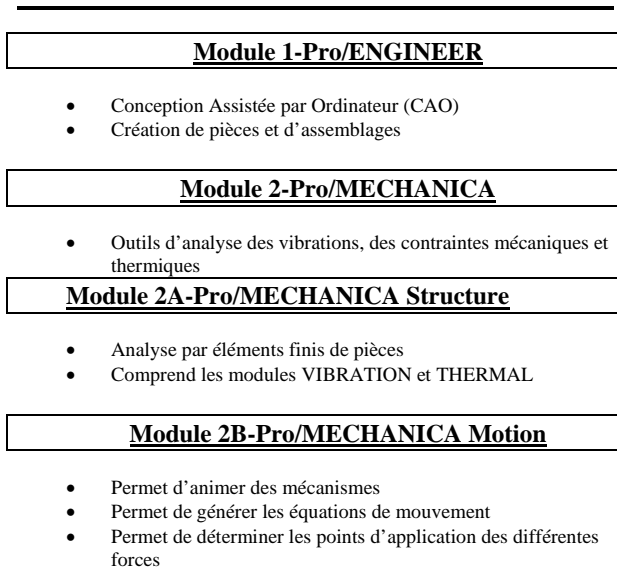


Figure 5 – Modules utilisés dans le logiciel Pro/ENGINEER

La figure 6 résume la procédure utilisée pour générer, analyser et modifier un composant.

La figure 7 présente un aperçu des différentes planches générées par le logiciel Pro/ENGINEER. Ces figures correspondent à l'exemple d'application traité dans le guide PROGUIDE fourni lors de l'achat de Pro/ENGINEER. Des détails plus spécifiques au module *Structure* sont disponibles dans le document de référence [9].

Il ressort de cette brève démonstration qu'à l'aide des outils de CAO, DAO et FAO actuellement disponibles sur le marché, il est possible d'effectuer des simulations dynamiques sur des prototypes virtuels et dans un environnement virtuel. Il est aussi possible d'agir aussi bien sur le prototype que sur l'environnement et ce, de manière relativement simple et dans un délai très court.

Plusieurs constructeurs automobiles [Renault, Chrysler...] exploitent déjà le potentiel de la simulation sur ordinateur ou sur banc d'essai pour développer de nouveaux produits. De plus, une récente publicité de Pro/ENGINEER mentionne que depuis 1995, quatre principaux manufacturiers d'automobiles ont standardisé le développement de leur groupe propulseur avec ce logiciel [10].

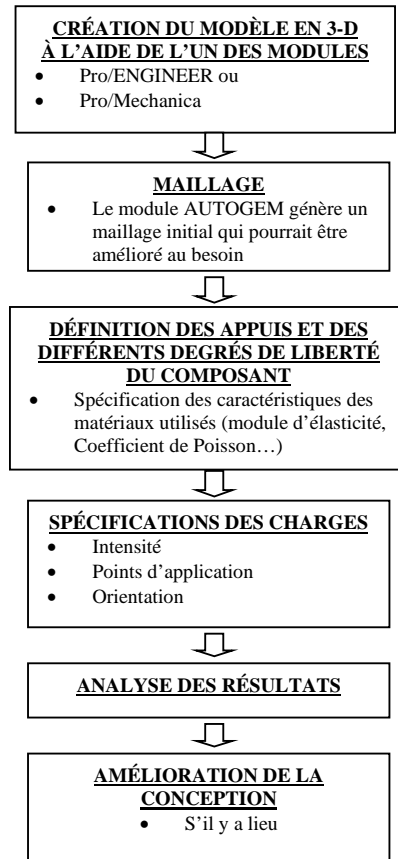


Figure 6 – Génération, analyse et modification d'un composant

VII-Conclusion

Cette étude a permis d'examiner les différents aspects de l'approche contrainte/résistance pour estimer la fiabilité de composants sujets à des contraintes dont l'occurrence et l'intensité sont aléatoires et dont la résistance est également aléatoire. Plusieurs travaux ont été consacrés aux divers aspects analytiques de l'approche sans trop se préoccuper de sa mise en œuvre. Les questions abordées dans cette recherche sont issues de considérations pratiques. L'objectif visé est de proposer une méthodologie de mise en œuvre qui procède à partir d'une vision intégrée des différents concepts et qui tente d'exploiter les outils disponibles pour raffiner les méthodes existantes ou introduire de nouvelles façons de faire.

La mise en œuvre de cette approche repose sur une étroite collaboration entre les différents intervenants pour mieux définir et modéliser les besoins du client et pour exploiter efficacement les expertises disponibles. Les nouvelles technologies de l'information répondent de plus en plus à cet impératif de collaboration. Par ailleurs, l'évolution des logiciels de conception et de fabrication assistées par ordinateur et la maîtrise de la simulation offrent des alternatives intéressantes pour le développement rapide de nouveaux produits de qualité et ce à coût minimum.

L'accent a été mis dans cette étude sur la façon d'exploiter le potentiel de l'approche contrainte/résistance aussi bien au niveau du bureau d'études que dans les centres d'essais. En intégrant à la fois les besoins du client et les capacités des procédés de fabrication et de validation, l'approche contrainte/résistance se révèle particulièrement intéressante pour les entreprises qui cherchent à améliorer leur compétitivité sur le plan mondial.

Parmi les orientations indiquées dans cette étude, on mentionne le potentiel de la simulation sur prototypes virtuels comme complément intéressant aux campagnes d'essais qui sont souvent onéreuses et très coûteuses. Grâce à la flexibilité des outils disponibles, il est possible d'agir virtuellement aussi bien sur le produit que sur l'environnement dans lequel il est appelé à évoluer.

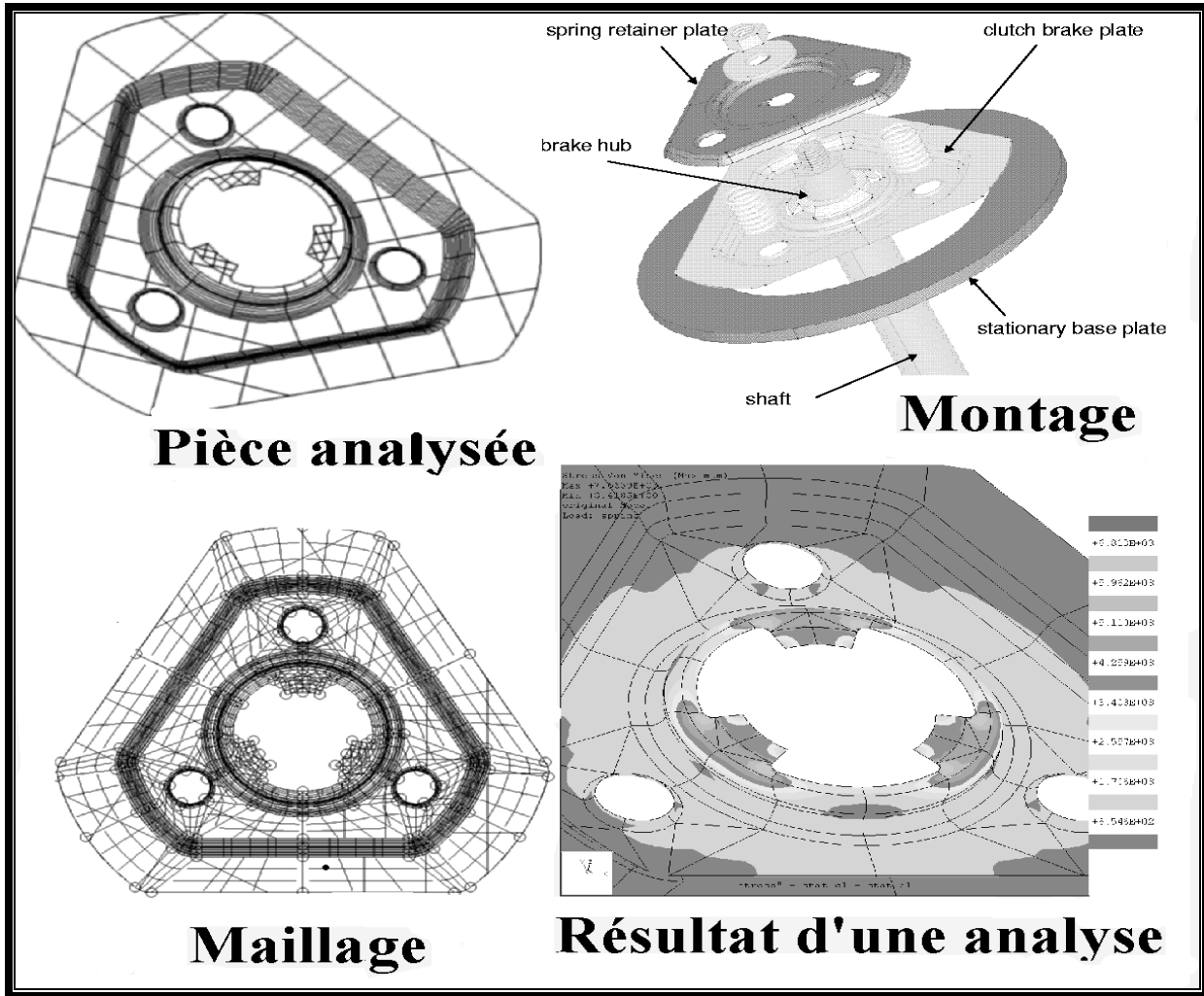


Figure 7- Planches générées avec Pro/ENGINEER et le module Structure de Pro/MECHANICA

VIII-Remerciements

Cette recherche a été financée par la direction de la recherche de Renault (France) et par le conseil de la recherche en sciences naturelles et en Génie du Canada (CRSNG).

Les auteurs tiennent à remercier tous les responsables de la direction de la recherche et des centres d'essais de Renault pour leur disponibilité et l'appui inconditionnel accordé aux auteurs.

IX-RÉFÉRENCES

[1] A.D.S. Carter, *Mechanical Reliability*, 1986; Macmillan.
 [2] J.D. Chuech, B. Harries, "The estimation of reliability from stress-strength relationships", *Techmetrics*, vol 12, num 1, 1970 Feb.
 [3] A. EL Khaïr et N. Gouget, *Validation de la fiabilité des systèmes innovants à l'aide de tests accélérés*, Rapport technique, direction de la recherche, Renault, 1997.
 [4] J. English, T. Sargent, T. Landers, "A discretizing approach for stress/strength analysis", *IEEE Trans. Reliability*, vol 45, 1996 Mar., pp 84-89.

[5] W.G. Ireson, C.F. Coombs, Jr. and R.Y. Moss, *Handbook of reliability engineering and management*, second edition, 1995, McGraw-Hill.
 [6] K.C. Kapur, L.R. Lamberson, *Reliability in Engineering Design*, 1977; John Wiley & Sons.
 [7] E.E. Lewis, H. Chen, "Load-capacity interference and the bathtub curve", *IEEE Trans. Reliability*, vol 43, 1994 Sep., pp 470-475.
 [8] Cl. Marcovici, and J.Cl. Ligeron, *Utilisation des techniques de fiabilité mécanique*, dernière édition, 1974, Technique & Documentation.
 [9] Parametric Technology Corporation, *Pro/MECHANICA Using Structure with Pro/ENGINEER*®, Release 18, 1997.
 [10] Parametric Technology Corporation, Publicité Aller plus vite., *Revue PLAN*, mars 1998, Ordre des ingénieurs du Québec, p.10.
 [11] S.S. Rao, *Reliability Based Design*, 1992; McGraw-Hill.
 [12] M.L. Shooman, *Probabilistic reliability : an engineering approach*, second edition, 1990, Robert E. Publishing company, Inc.
 [13] A. Villemeur, *Reliability Availability Maintainability and safety Assessment*, 1992; John Wiley & Sons.
 [14] J.D. Wang, T.S. Liu, "A discrete stress-strength interference model for unreliability bounds", *Reliability Engineering and System Safety*, vol 44, 1994, pp 125-130.
 [15] J. Xue and K. Yang, "Upper & Lower Bounds of Stress-Strength Interference Reliability with Random Strenght Degradation", *IEEE Trans. Reliability*, vol 46, no.1, 1997 Mar. pp 142-145

Vers une cartographie de processus explicite pour le modèle Demand Driven Adaptive Enterprise

Guillaume Martin¹, Pierre Baptiste², Jacques Lamothe³, Romain Miclo⁴, Matthieu Laurus⁵

¹ Centre Génie Industriel, Université de Toulouse
Mines Albi, Albi, France

² Polytechnique Montréal, Montréal, Canada

³ Centre Génie Industriel, Université de Toulouse
Mines Albi, Albi, France

⁴ AGILEA, Toulouse, France

⁵ Centre Génie Industriel, Université de Toulouse
Mines Albi, Albi, France

{gmartin@mines-albi.fr, pierre.baptiste@polymtl.ca, lamothe@mines-albi.fr, romain.miclo@agilea.fr,
matthieu.laurus@mines-albi.fr}

Abstract - Cet article présente une approche de cartographie de processus appliquée à la méthodologie Demand Driven Adaptive Enterprise. On constate, lors des déploiements réalisés par les professionnels, que les cas complexes de déploiement nécessitent d'adapter la méthode à divers degrés. Il faut notamment prendre en compte la réalité physique des ateliers, ainsi que les règles de gestion spécifiques. En croisant les principes issus des publications existantes et les retours fournis par des experts, une cartographie complète de la méthode Demand Driven Adaptive Enterprise a été établie. Elle fait apparaître les activités et les fonctions nécessaires et constitue une base pour de futures implantations, mais également pour la définition d'agendas de recherche partagés. On remarque en particulier que les possibilités d'interprétation de la cartographie sont nombreuses. Cela implique que chaque variation de la méthode doit être à son tour cartographiée et testée, afin de statuer sur son efficacité au cas par cas.

Keywords - Demand-Driven Adaptive Enterprise, modélisation de processus, Demand-Driven Materials Requirement Planning, DDMRP

1. Introduction

Quelle analyse peut-on avoir aujourd'hui de la gestion des nouveaux défis logistiques ? Quels sont, en prenant du recul, les nouveaux enjeux d'un système de production efficace ? Voici les questions auxquelles font face quotidiennement de multiples industries. Pendant le siècle dernier, scientifiques et industriels ont travaillé à mettre au point des méthodes de gestion afin d'assurer que les produits arrivent à la bonne date et dans les bonnes quantités (Orlicky 1975; Sugimori et al. 1977). Mais on entend de plus en plus souvent des plaintes concernant les limites des modèles existants.

C'est principalement parce que, dans le contexte actuel, ces méthodologies ne sont plus adaptées. Elles ne peuvent plus absorber les multiples sources de variation qui proviennent de l'environnement. Elles ne sont plus capables de gérer des gammes complexes, réparties sur un maillage d'acteurs parfois très différents. Cette diversité de produits, de procédés et de clients pousse à revoir l'approche des chaînes logistiques, et c'est ce que propose le modèle Demand Driven MRP (C. Ptak et Smith 2011).

Ce nouveau modèle est basé sur une meilleure flexibilité pour s'adapter aux fluctuations de la demande, mais également sur une plus grande résilience afin d'absorber la variabilité du système. Les publications récentes ont montré que le modèle Demand Driven était à même de tenir ses promesses d'agilité et de résilience (Romain Miclo 2016; Klein 2015; Ihme et Stratton 2015), et les acteurs industriels sont de plus en plus nombreux à se lancer dans des essais sur cette méthodologie. Cependant, si la capacité d'optimisation qu'offre le modèle Demand Driven a été prouvée pour des cas simples, le modèle n'a pas encore été appliqué à des systèmes complexes.

Le déploiement du modèle Demand Driven dans un système complexe va demander une grande capacité d'interprétation et d'adaptation de la théorie à la réalité physique de chaque cas. Mais cette adaptabilité est aujourd'hui rendue difficile par le manque d'outils permettant de modéliser et de simuler le fonctionnement d'un système piloté par la demande.

L'objectif est donc ici combler ce manque par l'établissement d'une cartographie des processus utilisés par la méthodologie Demand Driven. Cette cartographie sera basée à la fois sur les publications existantes, mais aussi sur une série d'entretiens avec des experts dans le déploiement de cette méthode. Il sera ainsi possible de déterminer la nature exacte des activités, mais aussi les possibilités d'interprétation du modèle, afin qu'il se conforme à la réalité des futurs chantiers. L'objectif final est d'utiliser cette cartographie pour établir les futures questions de recherche à traiter, mais également de rendre plus facile le déploiement de la méthode pour les industriels.

2. Un modèle Demand-Driven qui prend de l'ampleur

Pour répondre le plus efficacement possible aux défis que propose le nouveau monde de la logistique, le modèle Demand-Driven s'est construit sur la base de principes éprouvés, appelés dans la méthode les « piliers du modèle » (C. A. Ptak et Smith 2016). Ces principes intègrent les techniques de la gestion MRP (Orlicky 1975), de la distribution grâce au DRP (Martin 1995), du Lean (Womack et Jones 1997) et du Six-Sigma (Pande, Neuman, et Cavanagh 2000) et enfin de la théorie des contraintes (Goldratt 1999). A ces piliers, viennent s'ajouter deux principes qui sont au cœur de la méthodologie Demand-Driven : le découplage et l'adaptation dynamique qui apportent respectivement la résilience et l'agilité nécessaire.

L'objectif du principe de découplage est d'absorber et d'empêcher la transmission des incertitudes le long de la chaîne. Comme montré dans la figure 1 ci-dessous, la chaîne schématisée par deux étapes d'usinage et une étape d'assemblage encadrée par deux buffers physiques est sujet à la fois à des variations amont et aval. Les deux buffers physiques présentés sur la figure sont dimensionnés de manière à toujours pouvoir alimenter en matière première la chaîne, mais aussi à toujours disposer de produits pour les clients. Une fois protégée par ces deux buffers, la chaîne est donc découplée des sources de variation. On définit alors un nouveau temps de cycle, le *Decoupled Lead Time* (DLT), qui correspond au plus long temps non protégé entre deux buffers physiques (C. A. Ptak et Smith 2016).

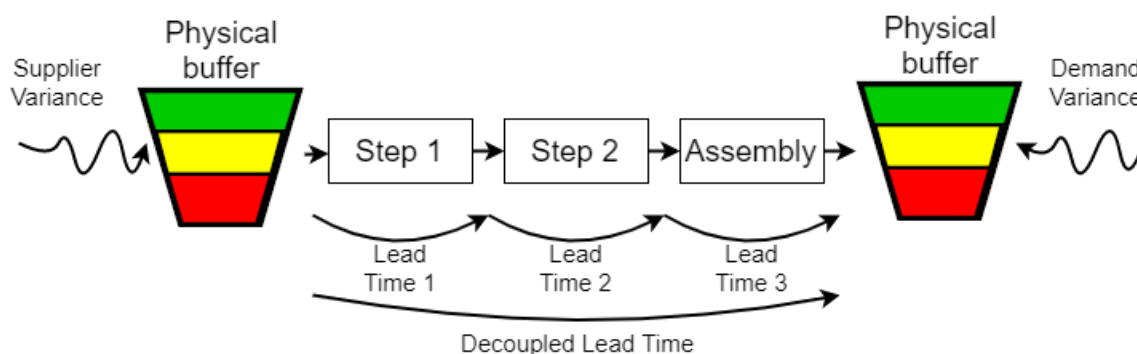


Figure 1 - Schéma de découplage des buffers physiques

L'adaptation dynamique, quant à elle, découle de la capacité à redimensionner les zones des buffers, telles que représentées en figure 1, en rouge, jaune et vert. En effet, lors de l'initialisation de la méthode DDMRP, la taille des zones de chaque buffer est calculée en fonction de la demande pour la référence concernée, ainsi que des incertitudes qui la caractérisent (transport, temps de cycle, ...) (C. A. Ptak et Smith 2016). Cependant, comme la demande ou les incertitudes pour cette référence évoluent avec le temps, il est impératif que les tailles de zones en fassent de même, assurant par-là l'agilité du modèle.

Chaque buffer est ainsi divisé en trois zones : (i) la zone rouge qui représente la plage de sécurité du buffer, (ii) la zone jaune qui représente le cœur de la couverture de la demande et (iii) la zone verte, qui couvre la génération des réapprovisionnements (C. A. Ptak et Smith 2016). Le niveau de chaque buffer est ainsi comparé quotidiennement aux demandes, stocks et en-cours au travers de la Net Flow Equation (NFE). Le résultat de la NFE est ensuite utilisé pour déterminer la priorité de réapprovisionnement de chaque référence gérée par buffer (C. A. Ptak et Smith 2016).

Le déroulement de la méthodologie DDMRP peut se présenter, pour sa part, en cinq étapes majeures :

- Le positionnement des stocks stratégiques, représentés par des buffers (physiques, de temps ou de capacité en fonction des cas)
- La détermination des profils d’incertitudes pour chaque référence bufferisée ainsi des niveaux des zones de buffers
- La mise en place d’une stratégie d’ajustements dynamiques, afin que les zones soient redimensionnées suivant les évolutions des références
- La planification des ordres en flux tiré par la demande
- Enfin, la mise en place d’un processus d’exécution basé sur un pilotage visuel et des règles de priorité.

Depuis les premières publications sur le modèle, le Demand Driven Institute a proposé une évolution du modèle (C. Smith, Ptak, et Ling 2017). Cette révision a pour objectif d’intégrer l’entreprise toute entière dans une démarche pilotée par la demande. Aux principes opérationnels tels qu’ils sont décrits plus haut dans ce paragraphe, le nouveau modèle, nommé Demand Driven Adaptive Enterprise, vient superposer une couche tactique, le Demand-Driven Sales and Operations Planning, ainsi qu’une couche stratégique, l’Adaptive Sales and Operations Planning. Il est donné ci-dessous en figure 2, tel qu’il apparaît dans la publication de référence sur le sujet (C. Smith, Ptak, et Ling 2017). Chaque « brique » du modèle est dorénavant reliée aux autres par un ou plusieurs liens, chacun contenant un ou plusieurs types de données

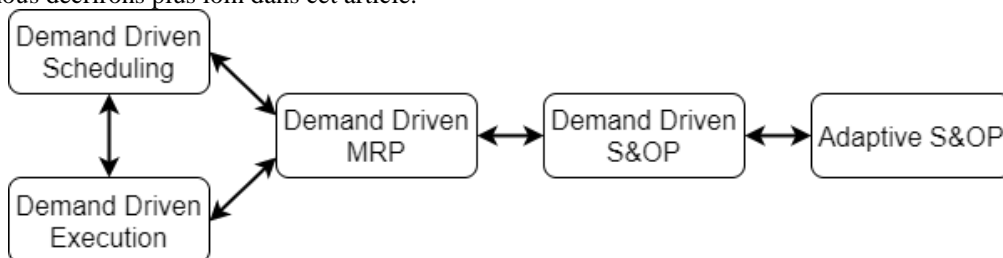


Figure 2-Schéma du modèle Demand Driven Adaptive Enterprise

Depuis la publication de cette méthode, plusieurs travaux de recherche ont été entrepris sur le sujet. Ils sont représentés ci-dessous dans le tableau 1 et forment une bibliographie commentée.

Orlicky's Material Requirements Planning, Third Edition (C. Ptak et Smith 2011)	Revue complète de tous les principes de la première version du modèle Demand Driven ainsi que de sa logique d’utilisation
Staying demand driven part III (C. Smith 2013)	Utilisation des indicateurs pertinents pour la méthodologie DDMRP
Effective production control in an automotive industry: MRP vs. demand-driven MRP (Shofa et Widyarto 2017)	Evaluation et comparaison de performances entre MRP et DDMRP en termes de stocks, par simulation d’une usine automobile
Evaluating Demand Driven MRP: a case based simulated study (Ihme et Stratton 2015)	Evaluation des performances de DDMRP dans la simulation d’une usine de production d’encre
Strategic Inventory Positioning in BOM with Multiple Parents Using ASR Lead Time (Jiang et Rim 2016)	Modèle pour déterminer les positions et quantités d’un buffer dans une nomenclature et complexe
MRP vs. Demand-Driven MRP: Towards an Objective Comparison (Romain Miclo et al. 2015)	Comparaison quantitative et qualitative entre MRP et DDMRP sur une étude de cas
An empirical study of Demand-Driven MRP (Romain Miclo et al. 2016)	Intérêts du DDMRP par rapport au MRP grâce à une simulation d’évènements discrets rapportée à une étude de cas
Challenging the “Demand Driven MRP” Promises: A Discrete Event Simulation Approach (Romain Miclo 2016)	Etude quantitative des promesses de DDMRP suivant plusieurs simulations sur deux cas d’études, centrée sur le taux de service et le niveau d’en-cours

Interpreting and applying Demand Driven MRP A case study (Ihme 2015)	Comparaison entre MRP et DDMRP sur un cas d'étude utilisant une simulation et montrant l'intérêt pour la réduction des stocks
Evaluation et simulation de l'implémentation d'une méthode de réapprovisionnement basée sur la consommation réelle au sein d'un environnement de production projet (Klein 2015)	Description et analyse du déploiement de la méthodologie DDMRP dans un contexte de production orienté projet. Revue des limites de la méthode.
The Demand Driven Adaptive Enterprise (C. Smith, Ptak, et Ling 2017)	Document de référence pour la méthode Demand Driven Adaptive Enterprise, comprenant les principales fonctions, activités et flux de données tels que définis par le Demand Driven Institute
Demand Driven Performance – Using Smart Metrics (D. Smith et Smith 2013)	Implémentation d'indicateurs dans des environnements gérés par DDMRP

Tableau 1- Bibliographie commentée des publications concernant le DDMRP

La bibliographie sur le sujet est assez concise et, en résumant le contenu de chaque publication, on s'aperçoit que les travaux portent principalement sur la validation des deux promesses centrales de la méthodologie DDMRP : l'amélioration de la résilience du système face aux incertitudes et la réduction des niveaux moyens de stocks.

Si l'on compare la littérature existante et la cartographie proposée par le Demand Driven Institute, on constate que nombre de nouveaux ajouts n'ont pas été validés scientifiquement. On n'y trouve notamment pas les fréquences recommandées pour mener les opérations, ou l'attribution des responsabilités par poste. Qui plus est, l'enchaînement des étapes au sein de ces « briques » est absent ou incomplet. Afin de préciser en détail quelles étapes doivent être réalisées dans chaque processus, par quel acteur et suivant quelles informations, il pourrait être utile d'appliquer aux informations disponibles une méthode de description de processus.

3. Cartographie du processus Demand-Driven Adaptive Enterprise : méthodes et résultats

Afin de proposer une modélisation précise de ce modèle complexe, une méthode en trois étapes a été utilisée : (i) la réalisation d'un questionnaire standard, permettant de remettre en question les enchaînements des activités au sein du modèle et les informations qu'il utilise, (ii) l'utilisation du questionnaire à la fois sur les modèles fournis par le Demand-Driven Institute mais également de récupérer l'expérience de déploiement de la méthode DDMRP par des experts et, (iii) l'agrégation de toutes les informations recueillies en une seule cartographie complète. Le terme « activité » défini ci-dessus correspond à la définition donnée dans le dictionnaire APICS (Cox et Blackstone 2002).

Le questionnaire a été appliqué à chaque activité au sein du modèle et comprend cinq questions :

- Quel est l'objectif de l'activité ? Il s'agit ici d'être certain que les entrées et sorties sont connus et utilisés
- Quelles sont les étapes de l'activité ? Il faut ici déterminer si l'activité peut être divisée en sous-activités spécifiques
- Quels sont les indicateurs pertinents pour cette activité ? Cette question sert à récupérer tous les indicateurs préconisés ou utilisés dans les déploiements de la méthode. Si des indicateurs existent, toutes les données d'entrée supplémentaires sont alors prises en compte.
- Quels sont les outils pour l'activité ? Il faut ici prendre tous les outils informatiques utilisés, ainsi que les appels aux fonctions des éventuels systèmes informatiques concernés.
- Quels sont les problèmes déjà rencontrés sur cette activité ? Cette question a pour but de collecter tous les points durs et manques de chaque activité, particulièrement avec les avis d'experts basés sur les expériences déjà réalisées auprès de diverses industries.

Les retours du questionnaire ont été traités de deux manières différentes. Concernant l'analyse des publications existantes, chaque réponse aux questions a été compilée dans une première version de la cartographie, formant ainsi une base de travail. La collecte des données pour cette première partie s'est donc faite grâce à une revue de littérature ainsi que des formations fournies par le Demand Driven Institute.

Dans un second temps, les experts dans le déploiement de la méthodologie DDMRP ont été consultés. Ceux-ci ont tous participé à plusieurs projets de mise en place des processus Demand-Driven et sont également certifiés par le Demand Driven Institute. Deux experts du panel sont également formateurs certifiés par cet organisme. De façon plus transversale, les experts bénéficient tous d'au moins quinze ans d'expérience dans la gestion de production.

Les experts ont reçu la première version de la cartographie, ainsi que le questionnaire. Ils ont pu, à leur tour, venir ajouter les écarts qu'ils constataient, ainsi que les manques à combler. Avec l'aide des experts, une seconde cartographie a donc vu le jour et se retrouve présentée dans cet article.

3.1. Un modèle de représentation des processus

La modélisation IDEF0 (Kim et Jang 2002), très proche de la modélisation SADT, a été retenue car elle offrait un bon compromis entre la représentation des activités physiques à réaliser et les échanges d'information entre ces activités. Le but de la représentation est de mettre en lumière les processus transverses au modèle Demand-Driven Adaptive Enterprise, les activités qu'ils intègrent mais aussi les acteurs nécessaires et les flux d'information mis en jeu. En parallèle, des notions comme l'horizon, la fréquence, la granularité, la nature des informations et des décisions ont été ajoutées au modèle. Celles-ci sont dérivées de la méthode GRAI (G. Doumeingts 1984).

Les concepts utilisés pour dresser la cartographie, et issus de la philosophie GRAI (Guy Doumeingts 1984) sont les suivants :

- La fonction (« fonction ») est définie comme une boîte possédant une frontière fixe, laquelle est traversée par des macro-processus. Il s'agit d'un rôle définissant les responsabilités, horizons, fréquences et granularités pour chaque tâche qui lui incombe.
- L'horizon représente la période de temps prise en compte par l'activité en question
- La fréquence (« frequency ») représente la répétition d'une activité au sein d'un horizon
- La granularité (« granularity ») est la taille des biens gérés par l'activité en question (référence spécifique, référence bufferisée, famille de produit)
- Les activités (« activities ») sont des tâches élémentaires, généralement regroupées en processus ou macro-processus pour plus de clarté
- Les informations et les décisions sont définies suivant la méthode GRAI. Les décisions sont prises en suivant des objectifs, des critères d'appréciation, des variables et des contraintes.

Une représentation type est donnée ci-dessous en figure 3. Celle-ci permet de comprendre comment les notions définies plus haut s'articulent entre elles et seront utilisées dans la cartographie finale.

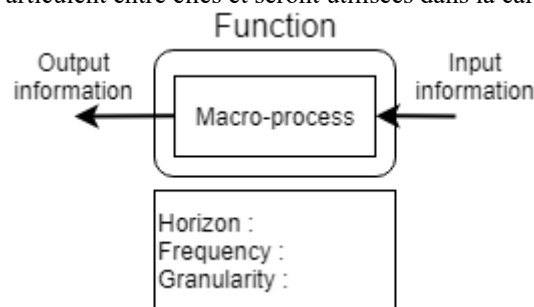


Figure 3 – représentation des fonctions et processus

La cartographie complète a été réalisée en trois parties, chacune suivant les macro-processus principaux, avant d'être assemblée. Ces macro-processus sont (i) le processus de définition du modèle opératoire, dont le but est de constamment adapter l'ensemble des buffers à la fois aux prévisions et à la demande réelle, (ii) le processus de production, qui se charge de lancer les bons ordres, de les planifier et de les prioriser au mieux et (iii) le processus d'amélioration continue, qui recueille tous les écarts constatés par le système et suit les améliorations.

3.2. Le processus de définition du modèle : réconcilier objectifs stratégiques et contraintes physiques

L'extrait réalisé du processus de définition du modèle est donné en figure 4 ci-dessous. Le principal objectif de ce processus est de réconcilier stratégie et incertitudes. Les volontés de développement stratégiques de la direction sont définies par la fonction Adaptive S&OP, notamment à travers le business plan, alors que l'ensemble des incertitudes enregistrées remontent en tant qu'améliorations possibles (« improvements »). Le processus de définition du modèle, pour cette version, fait intervenir trois fonctions : l'Adaptive S&OP, habituellement formée par la direction et le commerce, le Demand-Driven S&OP, qui se compose d'experts logistiques et managers, ainsi que le Système d'Information, qui joue ici le rôle de stockage de données. Ce processus possède deux cycles de décision : un cycle fonctionnant par trimestre pour la définition et la mise à jour du business plan et un cycle mensuel pour l'envoi des prévisions d'exceptions et l'intégration des projets d'amélioration continue. La granularité change également lorsqu'on accède aux couches plus basses du modèle. Si on démarre en effet avec le portefeuille complet des familles de produits, on prend en compte plus bas seulement les références gérées par buffer (généralement les produits finis, pièces détachées, matières premières et parfois articles achetés).

Au départ du processus, les personnes en charge de la fonction Adaptive S&OP vont devoir accorder tous les objectifs stratégiques (produits, finances, capacité, offre et demande) afin de rédiger et publier un business plan pertinent et chiffré. Ce business plan est alors transmis à la fonction Demand-Driven S&OP. Celle-ci va se charger de décomposer tous les besoins issus du business plan et impactant les références gérées sous buffer uniquement. Ces besoins, une fois agrégés, forment un plan de demande. Il peut éventuellement être nécessaire d'amener le plan de demande avec les données venant du commerce, telles que les campagnes spéciales, les promotions ou les pics de demande déjà connus sur certaines références. Ce plan de demande permet ainsi de définir une prévision générale pour les références gérées par buffer et donc considérées comme d'intérêt stratégique).

Le plan de demande est ensuite comparé avec les éventuels chantiers d'amélioration en cours, les écarts notés sur la période précédente ainsi qu'avec la gamme de production existante. Le modèle opératoire de la chaîne est alors créé à partir de cet ensemble d'informations, ou simplement mis à jour s'il existait déjà. Le modèle opératoire comprend l'emplacement des buffers le long de la gamme, leur profil (lié aux facteurs d'incertitudes qu'ils doivent absorber (C. A. Ptak et Smith 2016)), les tailles des zones et éventuellement des règles de gestion supplémentaires (par exemple, le lotissement, le groupement d'ordres sur certaines étapes). Ce modèle opératoire, ou modèle de configuration (« configuration model »), est essentiel et, une fois validé, se trouve stocké dans le système d'information.

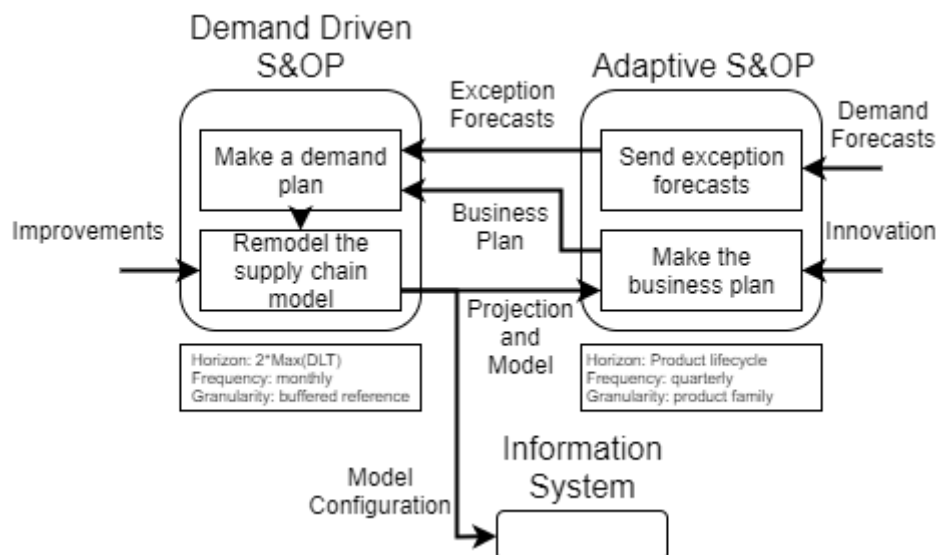


Figure 4 – Processus de définition du modèle

Les entretiens avec les experts en gestion de chaîne logistique ont fait apparaître plusieurs points sur ce macro-processus particulier. En premier lieu, on remarque que les activités présentées ici font rarement partie des essais en milieu industriel. Les experts jugent en effet que le sujet manque encore de clarté pour convaincre et être facile à déployer. Les experts notent également que peu d'informations sont données

concernant l'établissement des indicateurs à ce niveau de hiérarchie. On note cependant que la volonté des experts est d'utiliser les indicateurs proposés par le Demand Driven Institute (D. Smith et Smith 2013), basés sur la pertinence de l'information dans la chaîne. On retrouve principalement la marge contributive, le besoin en fonds de roulement et le suivi de la clientèle (C. Smith, Ptak, et Ling 2017).

Enfin, le point le plus difficile à mettre en œuvre au vu des retours fait par les experts reste le cycle spécifique de mise à jour du modèle opératoire en fonction des changements dans la demande. Cela nécessite en effet un effort constant de maintenance du modèle, ce qui passe par la remise en question régulière de plusieurs aspects du modèle. Il faut, en effet, réévaluer la position stratégique des buffers en fonction des nouveaux objectifs, mais aussi leur type (les buffers pouvant être non seulement physiques, mais aussi temporels ou capacitaires, selon le problème qu'ils viennent régler). Il faut, en parallèle, être capable d'attribuer le bon profil de buffer à chaque référence, ce qui suppose de détecter les incohérences.

3.3. Le processus de production : adapter les ordres de fabrication à la demande réelle

Ce processus est donné ci-dessous en figure 5. Il intègre quatre fonctions : (i) le système d'information, chargé d'effectuer tous les calculs et les opérations de logique simple, (ii) la fonction Demand-Driven MRP, où les décisions plus complexes sont prises et dont l'objectif est d'interpréter les alertes reçues du système d'information en ordres éventuels, (iii) la planification Demand-Driven (« Demand-Driven Scheduling ») qui vérifie au niveau des points de contrôle (C. A. Ptak et Smith 2016) que les ordres pourront être terminés à temps et avec la bonne quantité et (iv) la fonction Exécution, où se fait la priorisation des ordres de fabrication devant chaque machine et qui assure la production des biens.

Dans ce macro-processus, on rencontre à nouveau deux cycles de décision : le cycle quotidien pour la fonction Demand-Driven MRP et le système d'information, et le cycle horaire pour la planification et l'exécution. La granularité est elle aussi double dans ce processus : on s'intéressera uniquement aux références bufferisées pour les ordres, alors qu'entre les buffers, il faudra gérer toutes les références.

Le processus de production prend en compte deux données d'entrée essentielles que sont le modèle opératoire et la demande quotidienne réelle. Ces données d'entrées sont traitées par le système d'information dans la version proposée ici. Elles sont utilisées pour calculer le résultat de la « Net Flow Equation » (NFE), ceci pour chaque référence gérée par un buffer physique. L'utilisation de cette équation pour la création des ordres est la raison pour laquelle le système est capable de s'adapter rapidement aux variations de demande et évite d'accorder trop de poids aux prévisions. Les résultats de la NFE sont transformés en alertes si les niveaux de réapprovisionnement sont dépassés et ces alertes sont alors transmises à la fonction Demand-Driven MRP.

La fonction Demand-Driven doit alors analyser toutes les alertes du jour et décider s'il est nécessaire ou non de créer un ordre. Elle peut s'appuyer pour cela sur les recommandations de création d'ordre de la méthodologie DDMRP (commander dès que les références entament la zone jaune de leur buffer associé) mais également sur les règles de gestion optionnelles spécifiques de l'environnement (le groupement des ordres par exemple). Trois types d'ordres peuvent être créés, chacun suivant un processus particulier au sein du macro-processus. Les ordres d'achat et de transfert de stocks peuvent être transférés directement à l'Exécution, alors que les ordres de fabrication doivent d'abord passer par la Planification.

L'étape de planification est centrée sur la prise de décision en fonction des informations fournies par les points de contrôles placés tout au long de la chaîne. Les activités spécifiques aux points de contrôle sont centrées sur la vérification des capacités, la planification en avance ou le regroupement des ordres si nécessaire. C'est également ici que l'on peut s'assurer du bon avancement des activités de production et que l'on s'apercevra des retards pris par certains ordres de fabrication. Dans cette version, la planification gère des ordres ayant une date et une quantité connues à l'avance, contrairement à ce qui est préconisé par la méthodologie DDMRP. La fonction Exécution reçoit ensuite tous les ordres et doit décider des priorités à appliquer en fonction de l'état des buffers concernés, toujours suivant les préconisations de la méthodologie DDMRP.

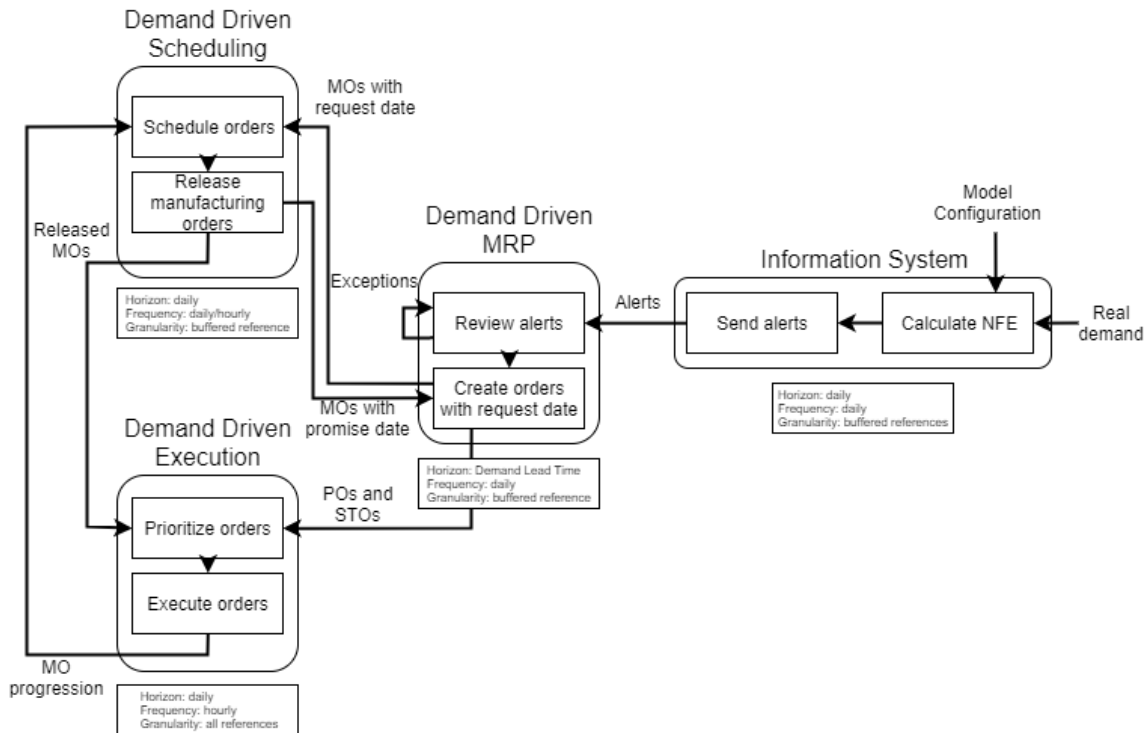


Figure 5 – Processus de production

On note, dans ce macro-processus, trois variations principales par rapport au modèle proposé par le Demand Driven Institute. Premièrement, l'utilisation d'un système d'information comme calculateur central représente une obligation pour les implémentations de la méthodologie DDMRP actuellement et, cependant, ne fait aucunement partie des sujets abordés dans les publications du Demand Driven Institute. Or, les experts font état d'un grand nombre de difficultés centrées sur le manque d'outils versatiles et performants qui pourraient venir soutenir leurs déploiements.

Dans un second temps, les experts ont exprimé un besoin supplémentaire de traitement des exceptions, représenté par une boucle spécifique à l'activité de revue d'alertes. Ce besoin est lié à la nécessité de devoir adapter chaque déploiement au contexte spécifique des industriels, que ce soit lié à l'organisation physique du flux de matières, ou à des règles spécifiques de gestion des ordres.

Le dernier point important à noter dans l'utilisation de la méthode DDMRP par les experts est l'utilisation de dates et de quantités fixes lors de la création des ordres. Ceci va à l'encontre de la méthodologie classique dans laquelle les ordres de production sont sensés remplir le carnet devant chaque machine sans date, ni quantité prédéfinie.

3.4. Le processus d'amélioration continue : comprendre les écarts aux prévisions

Le macro-processus est inclus ci-dessous en figure 6. Il fait principalement appel à la fonction Demand-Driven S&OP. Afin de donner une meilleure visibilité concernant la provenance des données d'entrées, les activités d'où ces données proviennent ont été ajoutées à la figure. Deux cycles sont présents : un cycle quotidien pour la collecte des données et un cycle d'une longueur maximale égale à deux fois le plus long DLT de la chaîne, dont le but est de suivre les chantiers d'amélioration tout en conservant une bonne agilité. Les entrées sont multiples mais elles partagent une caractéristique commune : ce sont des enregistrements d'écarts par rapport à un comportement attendu. Il peut s'agir, par exemple, d'une variation dans la demande reçue quotidiennement, d'une variance de planification due à une indisponibilité de la capacité ou d'une variation au cours de l'exécution, comme un cas de non-qualité problématique.

Les trois principales activités réalisées par la fonction Demand-Driven S&OP nécessitent : (i) l'analyse de la variance de toutes les autres activités, pour identifier les écarts problématiques et tenter de les expliquer, (ii) l'exploitation tactique, d'où partent les décisions d'amélioration basées sur des scénarios et (iii) le suivi des chantiers d'amélioration continue.

Puisque les améliorations proposées peuvent parfois avoir des conséquences sur le modèle opératoire, elles sont considérées comme une donnée de sortie pour ce processus. On peut en effet être amené à revoir les

facteurs d'incertitudes ou les tailles des buffers suites à ces chantiers d'amélioration continue, ce qui doit ensuite être reporté sur le modèle opératoire.

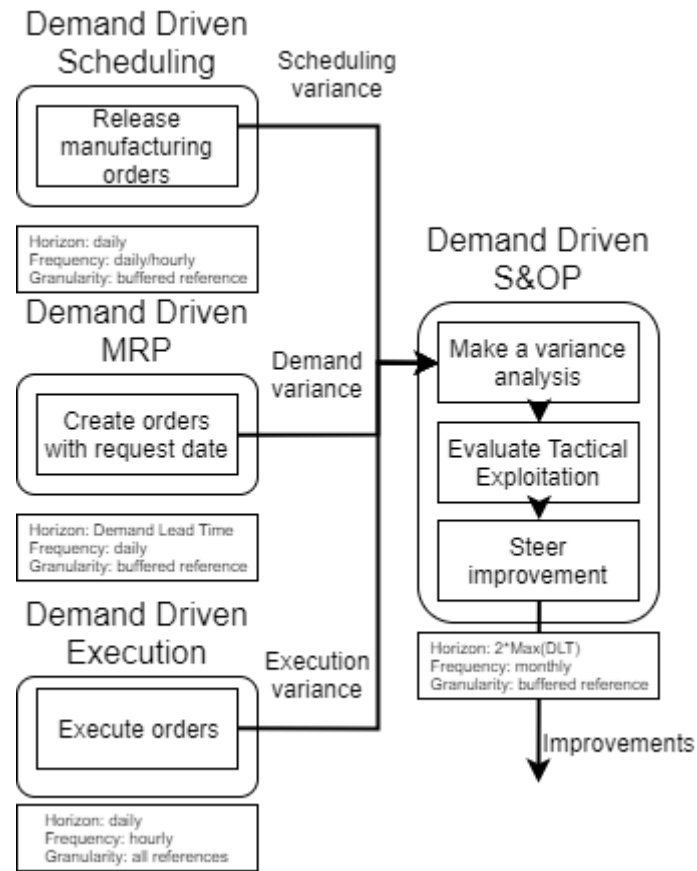


Figure 6 – Processus d'amélioration continue

Sur ce macro-processus particulier, on note deux points d'attention de la part des experts. Tout d'abord, les préconisations concernant les indicateurs à utiliser leur semblent insuffisantes. Ils ont noté que, pour offrir la meilleure analyse des écarts dans les données, le jeu d'indicateurs pourrait être enrichi. Deuxièmement, l'activité d'exploitation tactique (« Evaluate Tactical Exploitation ») manque encore de définition et pourrait être décomposée plus finement. Ceci permettrait de mieux comprendre les enjeux de cette activité, mais aussi l'utilisation à faire des données d'entrée-sortie à l'intérieur.

3.5. Cartographie complète : une vue d'ensemble du processus

La cartographie complète, résultat de l'agrégation de l'ensemble des macro-processus, est donnée ci-dessous en figure 7. On y retrouve les cinq fonctions telles qu'elles sont décrites par la méthodologie Demand Driven Adaptive Enterprise, ainsi que le système d'information. Celui-ci a été rajouté du fait de son utilisation centrale lors des déploiements par les experts. La cartographie présente également les horizons, fréquences et granularité pour chaque fonction.

Les fonctions peuvent être regroupées en trois horizons principaux : (i) l'horizon stratégique, qui comprend l'Adaptive S&OP, (ii) l'horizon tactique, qui comprend le Demand Driven S&OP et (iii) l'horizon opérationnel, où se retrouvent les fonctions incluses dans les premières publications sur le DDMRP (Demand Driven MRP, Demand Driven Scheduling et Execution).

A travers les paragraphes précédents, nous avons vu que des désaccords existaient entre la méthodologie telle qu'elle est présentée par le Demand Driven Institute et l'utilisation qui en est faite par les experts lors des déploiements. Ces désaccords portent principalement sur les variations qui existent autour de la cartographie initiale fournie dans la publication de référence (C. Smith, Ptak, et Ling 2017). Ceci prêche donc

à penser que la cartographie proposée ci-dessous aura besoin d'évoluer pour proposer la meilleure réponse possible à la réalité de l'utilisation de la méthode DDMRP.

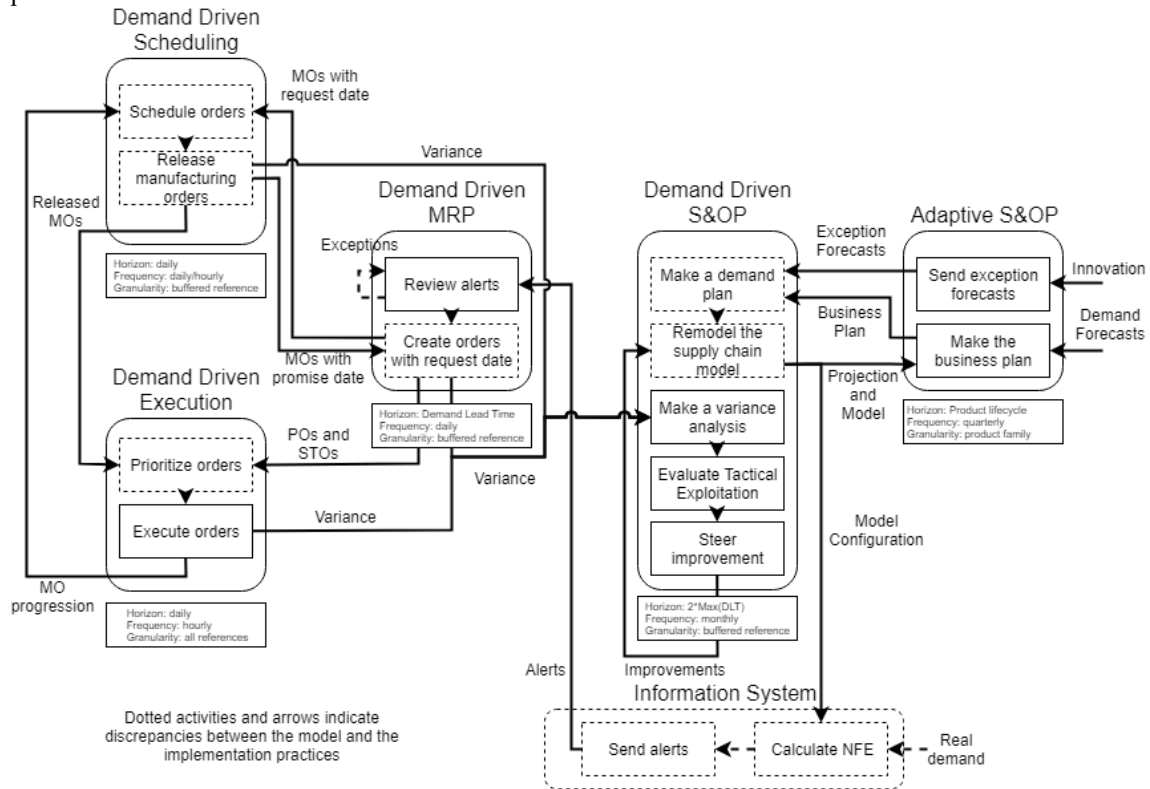


Figure 7-Cartographie complète du modèle Demand Driven Adaptive Enterprise

Les activités et les flux d'informations en pointillés représentent les écarts majeurs entre l'expérience de déploiement des experts et la méthode telle qu'elle est présentée dans les publications du Demand Driven Institute. On peut noter deux catégories principales :

- Les écarts dus à l'adaptation de la méthode à la réalité du terrain lors du déploiement, qui nécessitent souvent de revoir les méthodes de gestion préconisées, et qui se retrouvent principalement dans l'étage opérationnel
- Les écarts liés à un manque d'information et à une difficulté technologique, telle que la mise à jour fréquente des niveaux de buffers ou la gestion précise des alertes

4. Evolution et utilisation future de la cartographie

Un système de production en flux tiré peut répondre selon les auteurs à différentes caractéristiques qui sont de fait portées par des activités différentes du modèle DDAE :

- un système pour compléter un stock aval suite à la consommation de ce stock (Spearman, Woodruff, et Hopp 1990). Le travail est coordonné suite à la remontée d'un signal de consommation d'un buffer aval. Il peut remonter de machine en machine (kanban), ou de la source de consommation vers toutes les machines (EKCS, base stock) (Dallery et Liberopoulos 2000). Dans le modèle Demand Driven Adaptive Enterprise, on se reportera au processus de génération des ordres de fabrication de la fonction DDMRP.
- des boucles de contrôles servent à limiter l'encours et permettent de maîtriser le temps de cycle du système (Spearman, Woodruff, et Hopp 1990; Gaury 2000). C'est cette définition qui a aussi donné naissance aux méthodes CONWIP (Spearman, Woodruff, et Hopp 1990), COPACABANA et POLCA (Prakash et Chin 2014). Ceci est partiellement vrai pour les systèmes Kanban dans la mesure où, en environnement à forte diversité, la multiplicité des kanbans peut générer un engorgement ou un désamorçage de l'atelier (Askin et Krishnan 2009). Dans le modèle Demand Driven, ceci correspond au processus Demand Driven Scheduling qui doit garantir une gestion efficace des flux sur les points de contrôle dans le but de maîtriser l'encours de production.

- le flux tiré utilise des signaux visuels pour synchroniser les flux de travail avec la capacité des processus. Le choix du début d'une opération est ainsi déclenché par la fin d'une autre (Turner et al. 2012). Il s'agit d'une vision fortement décentralisée de la prise de décision, où l'opérateur sur ligne doit pouvoir prendre la décision du déclenchement des opérations. Dans le modèle Demand Driven Adaptive Enterprise, cela correspond à la fonction Demand Driven Execution, qui vise à fournir aux opérateurs des indicateurs pour prioriser en temps réel les activités à réaliser.

En distinguant les trois fonctions de l'horizon opérationnel (MRP, Scheduling et Execution), le modèle Demand Driven permet de reposer la question des processus à piloter de manière centralisée ou décentralisée. On peut parfaitement imaginer un modèle fortement décentralisé pour lequel, sur la base d'alertes, des opérateurs sur ligne responsables du réapprovisionnement de quelques buffers, génèrent des ordres de fabrication, les priorisent sur des points de contrôles et déclenchent leur exécution. Néanmoins, les cas d'études relatés dans la bibliographie font plutôt état d'un processus DDMRP centralisé, dans lequel un centre de décision génère les ordres pour toute l'usine et d'une exécution, par contre, décentralisée. La gestion des points de contrôle, quant à elle, reste souvent peu détaillée.

Les systèmes en flux tiré ont été associés à des productions en ligne et à des processus d'assemblage (Khojasteh et Sato 2015) et peu, voire pas, à des environnements en job shop. Une des particularités du modèle Demand Driven Adaptive Enterprise est de ne faire aucune hypothèse sur la structure du système de production si ce n'est la maîtrise des temps de cycle, supposés constants. Dans ces conditions, l'efficacité de la méthodologie Demand Driven à des unités de production avec des gammes multiples et variées demande à être validée. Il est possible de souligner, par exemple, un risque de nervosité du système dans certaines configurations physiques. Alors que, d'après ses auteurs, DDMRP offre des aptitudes d'amortissement des variabilités au niveau des buffers.

A l'horizon tactique, une des originalités de la fonction Demand Driven S&OP est de gérer des paramétrages sur des familles de produits afin de faire face à une diversité importante. De l'aveu même des auteurs, le paramétrage du dimensionnement des buffers, l'usage de prévisions pour adapter le modèle, le transfert de produits entre familles ainsi que le lissage de charge sur les ressources critiques restent basés sur des grands principes déjà présents dans les autres méthodes, et ne sont pas validés par des modèles dédiés. (C. Smith, Ptak, et Ling 2017) mettent, pour leur part, plus l'accent sur l'agilité de la fonction Demand Driven S&OP du fait du processus d'amélioration continue que sur la précision des modèles. Des outils et modèles restent donc à développer pour supporter ce processus. On peut notamment penser à améliorer la précision de la méthodologie concernant la meilleure façon d'intégrer les prévisions pour cet horizon.

Sur le niveau long terme, le processus Adaptive S&OP reste proche d'un modèle S&OP classique développé dans la démarche dite de *Manufacturing Resources Planning* (MRPII). Néanmoins, il intègre une question stratégique sur le positionnement des buffers qui peut ainsi être revu régulièrement, adapté à l'évolution de la demande ou des capacités. La littérature n'aborde cette question pour les systèmes en flux tirés avec des hypothèses importantes (ligne de production, système d'assemblage en ligne). Par comparaison, la fonction Adaptive S&OP offre un cadre pour aborder de façon beaucoup plus générique cette question.

5. Conclusion

Pour la première fois depuis la publication du modèle Demand Driven Adaptive Enterprise par le Demand Driven Institute, une cartographie détaillée a été proposée. On peut y retrouver, de manière explicite, chaque activité à réaliser, ainsi que les fonctions nécessaires à la réalisation de celle-ci. Cependant, les résultats montrent également que l'interprétation de cette cartographie peut mener à des interprétations très différentes. Au vu des écarts relevés entre les chantiers suivis par les experts et la méthodologie Demand Driven officielle, il serait bénéfique de continuer à faire évoluer cette cartographie.

En s'intéressant à l'adaptation de la méthodologie Demand Driven à la réalité physique des ateliers, il serait possible de faire un pas vers une cartographie réellement généralisable. L'utilisation de plusieurs cas d'études, extraits des projets réalisés par des experts, constituerait une base solide pour faire progresser les connaissances sur le modèle Demand Driven Adaptive Enterprise. Dans ce même esprit de généralisation, les hypothèses liées aux méthodes employées par les experts lors de leurs déploiements pourraient être éprouvées en détails.

Dans un horizon plus large, la cartographie présentée ici pourrait également servir de base pour la définition d'agendas de recherche partagés. La vision processus aiderait à délimiter clairement les activités, fonctions ou flux d'informations impactés par les travaux de recherches effectués par chacun.

6. References TBU

- Askin, Ronald G., et Shravan Krishnan. 2009. « Defining inventory control points in multiproduct stochastic pull systems ». *International Journal of Production Economics* 120 (2): 418–429.
- Cox, J. F., et J. H. Blackstone. 2002. « APICS Dictionary, 10th Edn (Alexandria, VA: APICS) ».
- Dallery, Yves, et George Liberopoulos. 2000. « Extended kanban control system: combining kanban and base stock ». *IIE transactions* 32 (4): 369–386.
- Doumeings, G. 1984. « La méthode GRAI ». *University of Bordeaux, Bordeaux*.
- Doumeings, Guy. 1984. « Méthode GRAI, Méthode de Conception et de Spécification des Systèmes de Productique ».
- Gaury, Eric Georges Antoine. 2000. « Designing pull production control systems: customization and robustness ». Tilburg University, School of Economics and Management.
- Goldratt, Eliyahu. 1999. *Theory of constraints*. North River Press, Great.
- Ihme, Mathias. 2015. « Interpreting and applying demand driven MRP: a case study ». PhD Thesis, Nottingham Trent University.
- Ihme, Mathias, et R. Stratton. 2015. « Evaluating demand driven MRP: a case based simulated study ».
- Jiang, Jingjing, et Suk-Chul Rim. 2016. « Strategic Inventory Positioning in BOM with Multiple Parents Using ASR Lead Time ». *Mathematical Problems in Engineering* 2016.
- Khojasteh, Yacob, et Ryo Sato. 2015. « Selection of a pull production control system in multi-stage production processes ». *International Journal of Production Research* 53 (14): 4363–4379.
- Kim, Soung-Hie, et Ki-Jin Jang. 2002. « Designing performance analysis and IDEF0 for enterprise modelling in BPR ». *International Journal of production economics* 76 (2): 121–133.
- Klein, Philippe. 2015. « Evaluation et simulation de l'implémentation d'une méthode de réapprovisionnement basée sur la consommation réelle au sein d'un environnement de production orienté projet. »
- Martin, Andre J. 1995. *DRP: distribution resource planning: the gateway to true quick response and continuous replenishment*. John Wiley & Sons.
- Miclo, R., F. Fontanili, M. Lauras, J. Lamothe, et B. Milian. 2016. « An empirical comparison of MRPII and Demand-Driven MRP ». *IFAC-PapersOnLine* 49 (12): 1725–1730.
- Miclo, Romain. 2016. « Challenging the " Demand Driven MRP" Promises: a Discrete Event Simulation Approach ». PhD Thesis, Ecole des Mines d'Albi-Carmaux.
- Miclo, Romain, Franck Fontanili, Matthieu Lauras, Jacques Lamothe, et Bernard Milian. 2015. « MRP vs. demand-driven MRP: Towards an objective comparison ». In *Industrial Engineering and Systems Management (IESM), 2015 International Conference on*, 1072–1080. IEEE.
- . 2016. « An empirical study of Demand-Driven MRP ». In *ILS 2016-6th International Conference on Information Systems, Logistics and Supply Chain*.
- Orlicky, Joseph. 1975. *Material requirements planning*.
- Pande, Peter S., Robert P. Neuman, et Roland R. Cavanagh. 2000. *The six sigma way*. McGraw-Hill.
- Prakash, J., et J. F. Chin. 2014. « Comparison between production controls in multi-stage multi-product manufacturing environments: two case studies ». *Production & Manufacturing Research* 2 (1): 477–500.
- Ptak, C., et C. Smith. 2011. « Orlicky's Material Requirements Planning McGraw-Hill Education-Europe ».
- Ptak, Carol A., et Chad Smith. 2016. *Demand Driven Material Requirements Planning (DDMRP)*. Industrial Press, Incorporated.
- Shofa, Mohamad Jihan, et Wahyu Oktri Widyarto. 2017. « Effective production control in an automotive industry: MRP vs. demand-driven MRP ». In *AIP Conference Proceedings*, 1855:020004. AIP Publishing.
- Smith, Chad. 2013. « Staying Demand Driven ». *Strategic Finance* 95 (6): 43.
- Smith, Chad, C. Ptak, et Dick Ling. 2017. « The Demand Driven Adaptive Enterprise Model ». Demand-Driven Institute.
- Smith, Debra, et Chad Smith. 2013. *Demand Driven Performance*. McGraw Hill Professional.

- Spearman, Mark L., David L. Woodruff, et Wallace J. Hopp. 1990. « CONWIP: a pull alternative to kanban ». *The International Journal of Production Research* 28 (5): 879–894.
- Sugimori, Y., K. Kusunoki, F. Cho, et S. Uchikawa. 1977. « Toyota production system and kanban system materialization of just-in-time and respect-for-human system ». *The International Journal of Production Research* 15 (6): 553–564.
- Turner, Richard, Dan Ingold, Jo Ann Lane, Ray Madachy, et David Anderson. 2012. « Effectiveness of kanban approaches in systems engineering within rapid response environments ». *Procedia Computer Science* 8: 309–314.
- Womack, James P., et Daniel T. Jones. 1997. « Lean thinking—banish waste and create wealth in your corporation ». *Journal of the Operational Research Society* 48 (11): 1148–1148.

VERS UNE INNOVATION DIRIGÉE PAR LES MODÈLES - APPLICATION EN TÉLÉMÉDECINE

Yosra CHAHER, Jean-Pierre BELAUD

Hervé PINGAUD

Université de Toulouse, Laboratoire de Génie Chimique,
UMR 5503 CNRS, ENSIACET INPT, 4, Allée Emile
Monso, BP 84234- 31432 Toulouse Cedex 4, France
yosra.chaher@ensiacet.fr, jeanpierre.belaud@ensiacet.fr

Laboratoire de Génie Chimique, CNRS
Institut National Universitaire Champollion, Ecole ISIS
Université de Toulouse, INPT, Castres, France
herve.pingaud@univ-jfc.fr

RÉSUMÉ : *Beaucoup d'organisations sont aujourd'hui dans l'obligation d'innover sous l'effet conjoint d'une pression externe (attentes des consommateurs et évolution de l'offre sur les marchés) et d'une pression interne (modes de travail évolutifs, limitation des moyens et besoins récurrents d'adaptation). L'offre de services de santé est concernée par ces effets de pression. L'innovation est concrétisée par un foisonnement d'initiatives jalonnant l'évolution future du système de santé. Dès lors, dans ce secteur comme dans tous les secteurs socio-économiques, le processus d'innovation doit être défini et placé sous contrôle. L'objectif de cet article est de montrer l'apport pour l'innovation d'une approche formelle relevant d'une démarche d'ingénierie dirigée par les modèles. Nous proposons de capturer les connaissances sur le concept d'innovation afin de les utiliser pour alimenter la modélisation d'un processus d'innovation. Nous présentons donc une modélisation conceptuelle de l'innovation et son instanciation sur une étude de cas en télémédecine.*

MOTS-CLÉS : *Ingénierie de l'innovation, Ingénierie d'entreprise, Ingénierie système, Modélisation conceptuelle, Cadre d'ingénierie, Santé connectée*

1 CADRE DES TRAVAUX

Pour beaucoup d'organisations de production de biens ou de services, des évolutions de l'environnement, comme par exemple l'ouverture internationale des marchés ou la transformation digitale de l'économie, sont perçues comme des signaux amenant à relever des nouveaux défis. Elles doivent alors imaginer des réponses pour maintenir la satisfaction de clients de plus en plus exigeants, pérenniser les savoir-faire à valeur ajoutée tout en consacrant une part raisonnable de leurs moyens à l'adaptation de leur outil de production et de leur offre. La question de l'innovation prend corps dans ce contexte et s'installe durablement dans le paysage socio-économique. (Xuereb 1991) affirme que la pérennité d'une entreprise repose sur sa capacité à générer un flux constant d'innovations. Innover est de plus en plus complexe et se fait dans un contexte de plus en plus incertain. C'est pourquoi nous visons de structurer le management de l'innovation par un cadre méthodologique qui guide les organisations et les aide à innover d'une manière efficace. Pour ce faire, nous adoptons une approche d'ingénierie système basée sur les modèles (ISBM). C'est une méthodologie qui peut être caractérisée comme un ensemble de processus, méthodes et outils reliés pour supporter l'ingénierie système dans un contexte «basé

sur les modèles» ou «dirigé par les modèles», (Estefan et al. 2007). En particulier, l'ingénierie système se base sur des préceptes liés au cycle de vie d'un système complexe, liés aux processus de sa mise en oeuvre et liés à des cadres de modélisation pour guider le raisonnement de modélisation d'un système complexe, (Castro 2010). Comme l'innovation est à la fois un résultat (un produit, un service, un procédé...) et le processus suivi pour parvenir à ce résultat (un projet d'innovation), nous pouvons considérer un système d'innovation à faire et un système d'innovation pour faire. La définition de modèles de processus est à la fois guidée et contrainte par un méta-modèle en amont qui définit les concepts dont les différents acteurs d'un processus d'ingénierie ont besoin. De plus, un méta-modèle permet de définir une trame générale pouvant être réutilisable dans une organisation en définissant des modèles propres à elle ou à chacun de ses projets. Les méta-modèles ont ainsi un rôle structurant très important. Ils permettent de faciliter la communication entre les différents acteurs impliqués. Ils aident à gérer les connaissances et le savoir-faire d'un système par la capitalisation et le retour d'expérience, au sein d'une démarche de qualité et d'amélioration continue du système. Ils peuvent aussi servir à structurer une base de connaissances pour stocker des expériences passées et qui peuvent servir dans une mécanique de *Case Based Reasoning*.

Ces bénéfices justifient l'adoption de cette approche dans le but de rationaliser les pratiques de management de l'innovation.

Les organisations du monde sanitaire et social sont impactées par ce mouvement d'innovation. Secteur fortement régulé par la sphère publique, il n'en fait pas moins l'objet de rapports de force entre offre de soins et demande de prise en charge. Récemment, l'évolution du contexte est marquée par deux caractéristiques : un vieillissement de la population et une augmentation de la demande sur des maladies chroniques, d'une part, et une raréfaction des praticiens de santé qui provoque une inégalité territoriale pour l'accès aux soins des populations et demande de modifier les pratiques et les conditions d'exercice du métier pour ceux qui sont en poste, d'autre part. Il faut trouver des réponses et la puissance publique incite à innover pour faire face à ces défis sociétaux. Cette innovation n'est pas que technologique, elle a d'autres dimensions. Un changement conséquent en matière d'organisations des soins a ainsi été initié. Il vise à plus de continuité à travers la notion de parcours de soins coordonné. La coordination passe par des relations accrues entre de nombreux acteurs (établissements hospitaliers, maisons de santé, praticiens libéraux,...) concourant ensemble à une prise en charge optimale (au sens de la santé du patient et des moyens mis en oeuvre pour le soigner). Dans cette optique, les technologies de l'information et de la communication doivent être mobilisées et contribuer à l'atteinte de cet objectif d'optimalité. La santé connectée est un qualificatif qui rassemble tous les changements impulsés par la transformation numérique sur les pratiques des professionnels et des citoyens, dans leur rapport à la santé. Le monde de la santé est complexe dans sa structure, et fortement réglementé dans son comportement. Innover dans ce secteur oblige à trouver le périmètre du système dans lequel l'innovation va se développer, à faire la preuve de sa valeur ajoutée sur le plan médical, à maîtriser tout effet secondaire indésirable et à valoriser la nouveauté par des gains financiers fruits de pilotes expérimentaux sur lesquels reposera la décision de rembourser par l'assurance maladie. Dans cet article, nous allons développer un cas d'études en télémédecine. Il s'agit de pratiquer des consultations médicales à distance pour apporter une forme de réponse à l'éloignement des citoyens des lieux de soins. Si la télémédecine a été reconnue par le code de la santé publique depuis 2008, les premiers décrets de mise en oeuvre de cette pratique en santé connectée ont été publiés en 2016. C'est une preuve s'il en est des résistances que l'innovation peut rencontrer dans ce secteur. Apprendre des expériences passées et pouvoir bénéficier de ces apprentissages pour être plus efficaces dans le futur nécessitent de capitaliser par au moyen d'une formalisation du concept et du processus d'innovation et à en faire un élé-

ment central du management de l'innovation. C'est l'objectif de notre travail de thèse. Il s'agit de créer ce cadre général à partir de représentations par des modèles. Nous pouvons parler d'un but de faire de l'innovation dirigée par les modèles. C'est l'esprit qui a présidé à la conception des cadres d'ingénierie d'entreprise connus que nous proposons d'adopter. Ces approches très souvent liées à des courants de normalisation, ont déjà prouvé leur utilité par le passé. Sans être exhaustifs, nous citerons les retombées constatées sur des sujets de première importance, comme la chaîne logistique (Stewart 1997), la maîtrise intégrée des risques et des processus (ISO 2009), cadre BPRIM (Sienou 2009), l'architecture des réseaux d'organisations (Zaidat, Boucher & Vincent 2005), et l'architecture d'entreprise (TOGAF 1998). Dans cet article, nous présentons une première composante de ce cadre et les premiers résultats de notre travail de capitalisation de connaissances. Il s'agit du modèle conceptuel qui décrit l'innovation.

La section 2 est une revue bibliographique sur les définitions de l'innovation et sur les premières tentatives de modélisation conceptuelle. La section 3 présente le modèle conceptuel de l'innovation qui sera instancié par une étude de cas en télémédecine dans la section 4. Enfin, dans la section 5, nous concluons sur les principaux résultats de notre recherche et sur les perspectives de nos travaux.

2 ANALYSE BIBLIOGRAPHIQUE

Dans la dernière décennie, plusieurs chercheurs ont travaillé au développement de modèles afin de mieux cerner l'innovation, considérée comme objet d'études. Beaucoup d'entre eux, se sont focalisés sur la notion de processus, souvent avec des périmètres d'études variés et des finalités différentes. Par une approche systémique, (Tomala, Senechal & Tahon 2001) proposent un modèle à base de processus qui permet de décrire les mécanismes d'élaboration d'une innovation et de comprendre son évolution. Ces auteurs affirment avec raison qu'une telle approche permettrait de mieux piloter l'innovation à condition de tenir compte des spécificités relatives à l'objet de l'innovation. (Castro 2010) propose un cadre de modélisation pour un système d'innovation visant un nouveau produit. Dans ces travaux, cet auteur choisit de considérer le modèle du cycle de vie d'un produit, en particulier le modèle en V afin de structurer le processus d'innovation qui l'accompagne. (Penide, Gourc, Pingaud & Peillon 2013) a proposé de généraliser la démarche en considérant un modèle d'innovation comme un processus de transformation des connaissances, fondé sur une cartographie guidant vers plusieurs niveaux de détail. Ce processus y est défini comme non linéaire et ponctué par des jalons qui permettent des boucles de rétroaction. Pour l'auteur, une démarche d'alignement entre les proces-

sus métiers et le processus d'innovation s'impose. Il suggère de positionner un ensemble de bonnes pratiques sur le modèle de processus pour faciliter le management de l'innovation. Il existe aussi dans la littérature des initiatives de cadre d'ingénierie pour systématiser la création d'écosystèmes d'innovation. Nous citons les travaux de (Rabelo & Bernus 2015) qui ont pour objectif de créer une vue holistique et de dépasser les limites dues à la complexité inhérente à de tels systèmes. (Rabelo & Bernus 2015) soulignent qu'il y a un manque d'une vision globale sur ce sujet de l'écosystème d'innovation. Dans un registre différent, celui des systèmes d'information, (Achi & Salinesi 2015) développent un modèle de l'innovation selon plusieurs dimensions: l'organisation autour du processus d'innovation, la veille technologique et métier, la gestion des connaissances, le capital humain, et la culture d'innovation. Ce cadre est un point d'entrée pour estimer la maturité d'une démarche d'innovation. Dans son ouvrage qui fait référence, (Boly 2004) promeut une démarche d'ingénierie de l'innovation. Ayant compris le caractère polysémique du terme, il propose de définir l'innovation selon 6 points de vue complémentaires : vision de l'économiste, vision opératoire, vision du cognitif, vision du sociologue, vision systémique et vision évolutionniste. N'ayant pas trouvé de référence dans la littérature sur la formalisation de telles connaissances, mis à part le modèle IpOp (Innovation par les Opportunités) (Cohen 2016). C'est un guide générique qui structure la pensée et la réflexion relative à l'innovation, mais n'intègre que la partie idée et faisabilité du processus d'innovation. Nous avons donc décidé de créer un modèle conceptuel de l'innovation réunissant en un tout ces différents aspects. Ce choix dépasse la simple capture de ces connaissances sur l'innovation. Après examen des 6 visions proposées par (Boly 2004), nous nous sommes aperçus que la vision systémique et la vision évolutionniste sont supportées par le développement naturel d'un système industriel et n'apportent donc pas de nouveaux concepts. Ainsi, nous avons jugé dans notre démarche que les 4 premières visions sont suffisantes. Nous avons été amenés ensuite à trouver les liens qui existent entre les différents angles et à unifier l'ensemble. Situé en amont de notre travail sur le processus d'innovation, les construits de l'innovation qui sont inclus dans ce modèle conceptuel serviront à décrire ensuite les flux d'innovation, c'est-à-dire les intrants et extrants des activités composantes du processus d'innovation.

3 MODÉLISATION CONCEPTUELLE

Les concepts, ou construits, sont des artefacts permettant de définir l'innovation. Notre démarche consiste à identifier des termes clés à partir des définitions littérales. Dans cette section, nous recensons ces

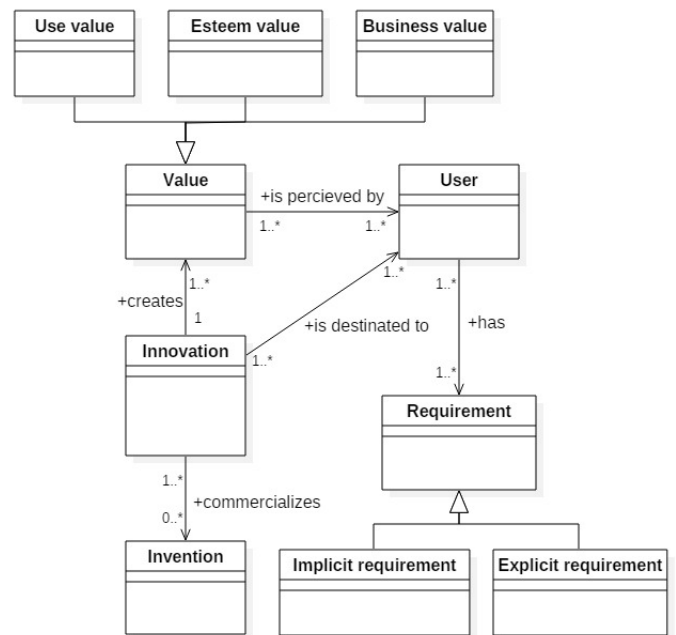


Figure 1 – Les concepts de modélisation selon la vision de l'économiste

concepts, les accompagnons de définitions consignées dans un tableau, et tissons les liens entre eux. Nous en faisons ensuite une synthèse sous forme d'un méta-modèle par un diagramme de classes en langage UML, les concepts deviennent des classes. Une classe conceptualise et généralise un groupement d'objets similaires appelés instances. Cette synthèse est représentée par une figure à mettre en vis-à-vis du tableau des concepts. L'innovation est une classe qui apparaît dans chacune des quatre parties, composant le traitement du problème. Chacun des diagrammes de classes deviendra un paquetage de classes pour former le méta-modèle de l'ensemble.

3.1 La vision de l'économiste

La vision de l'économiste considère que l'innovation est une *invention* qu'on a réussi à commercialiser. Ceci génère une *valeur économique* pour l'entreprise. L'*utilisateur*, étant le destinataire de l'objet innové (bien ou service), a un rôle important dans l'analyse de sa *valeur*. En effet, la satisfaction de l'*utilisateur* concerne l'ensemble des qualités d'usage (*valeur d'usage*) et d'estime (*valeur d'estime*) de l'objet innové. Parfois, l'innovation n'est pas définie pour répondre à une *exigence* exprimée par l'*utilisateur* mais pour la susciter afin d'améliorer sa qualité de vie (nous parlons d'*exigences explicites* et d'*exigences implicites*). Dans le tableau 1, nous définissons ces concepts. La traduction sous forme de méta-modèle est réalisée sur la figure 1.

Concept	Définition
Interrelation	Relation entre des éléments différents d'un même système.
Acteurs	Ce sont les parties prenantes qui constituent l'écosystème de l'innovation.
Membres du projet	Ce sont les personnes impliquées dans la réalisation d'un projet d'innovation.
Chercheurs	Les chercheurs se consacrent à l'élaboration de connaissances fondamentales. Dans le domaine de la technologie, les ingénieurs appliquent les connaissances des chercheurs pour concevoir des produits ou des services.
Connaissance	Ensemble d'informations structurées, basées sur le raisonnement et susceptibles de vérification, selon la Norme AFNOR.
Connaissance tacite	C'est un attribut personnel, c'est la connaissance produite par les expériences passées, les valeurs, les perspectives, l'intuition et d'autres éléments dans un contexte particulier et spécifique. (Cortes Robles 2006)
Connaissance explicite	C'est la connaissance qui peut être clairement formalisée, structurée et définie de façon systématique par le biais du langage, d'objets, de schémas, d'images, etc. (Cortes Robles 2006)
Information	L'information (élément stocké dans un système de connaissances) est le facteur, l'élément ou le moyen permettant de découvrir et de produire la connaissance. (Cortes Robles 2006)
Compétence	La mobilisation des connaissances dans le but d'accomplir une mission.
Capacité d'innovation	La capacité d'innovation renvoie aux connaissances et aux compétences nécessaires à l'entreprise pour utiliser efficacement, maîtriser et améliorer les technologies existantes et pour en créer de nouvelles. (Lall 1992)

Tableau 2 – Les concepts de modélisation selon la vision du sociologue

ressources qui peuvent être internes ou externes. Cette définition s'intègre dans le cadre d'une organisation par projets. En effet, dans la réalisation d'un projet il y a une innovation, puisqu'un projet est par définition une organisation temporaire décidée dans le but de créer une nouveauté. Et tout projet peut être

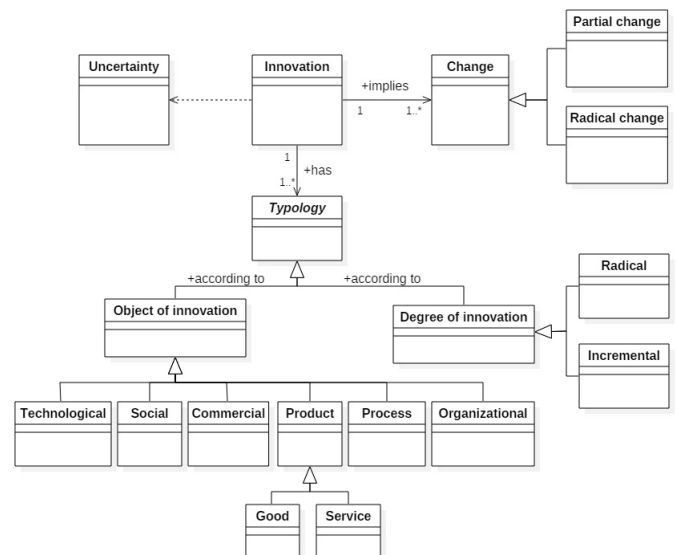


Figure 3 – Les concepts de modélisation selon la vision du cogniticien

affecté par des *risques* qui modifient la *valeur* à créer. Ces *risques* sont l'effet de l'incertitude mentionnée précédemment. Dans le tableau 4, nous définissons les concepts de modélisation de l'innovation vue de cet angle (voir figure 4).

4 APPLICATION EN TÉLÉMÉDECINE

Nous présentons succinctement une expérience d'innovation en télémédecine. Ce cas d'études est extrait d'une mission sous la responsabilité de l'ARS Occitanie sur laquelle nous sommes intervenus pour évaluer l'intensité des changements de pratiques professionnelles de cette innovation. Le projet était de déployer une solution de télémédecine dans plus de 20 EHPAD du département du Tarn pour réaliser des télé expertises (médecine de second recours), c'est-à-dire des demandes d'avis sur des patients auprès de spécialistes en gériatrie et psychiatrie d'établissements hospitaliers. Ces demandes étaient formulées par les médecins généralistes référents de patients dans les EHPAD. Le terminal de télémédecine est constitué par un service de visio-conférence pour les échanges synchrones entre les praticiens pendant la séance de consultation, et d'un outil collaboratif de gestion dématérialisée de documents pour la préparation de la séance (demande d'avis) et pour la production formelle de l'avis après la séance. L'intérêt de cette nouvelle forme de pratiques et la maturité atteinte par la technologie qui permet de la mettre en oeuvre est incontestable. Toutefois, au cours de cette mission, nous avons pu constater le manque de formalisation des processus métier autour de l'usage de la technologie et l'absence de référentiel

Concept	Définition
Changement	Passage d'un état initial à un état final.
Changement radical	La réalisation d'un concept totalement nouveau. (Boly 2004)
Changement partiel	L'amélioration d'un dispositif déjà existant. (Boly 2004)
Innovation radicale (ou de rupture)	Crée un nouveau marché, ou bien transforme en profondeur un ou plusieurs marché(s). (BPI 2015)
Innovation incrémentale	Améliore l'existant, contribue à la compétitivité et/ou la rentabilité de l'entreprise sans apporter de transformation significative. (BPI 2015)
Innovation de produit	Porte sur un bien ou un service. (BPI 2015)
Innovation de procédé	Porte sur un mode de fabrication, la manière dont l'entreprise organise son travail ou sa chaîne logistique. (BPI 2015)
Innovation commerciale	Porte sur la présentation, la distribution, la tarification, ou la promotion de l'offre. (BPI 2015)
Innovation sociale	Répond à des besoins sociaux, tant dans ses buts que ses modalités. (BPI 2015)
Innovation technologique	Crée ou intègre une ou plusieurs nouvelle(s) technologie(s). (BPI 2015)
Innovation organisationnelle	Mise en oeuvre d'une nouvelle méthode organisationnelle dans les pratiques, l'organisation du lieu de travail ou les relations extérieures de la firme. (Borgiel 2017)
Incertitude	c'est un manque conscient de connaissance d'un sujet, relative à un objet, non encore parfaitement défini, dans un contexte nécessitant une décision/action, Selon FonCSI.

Tableau 3 – Les concepts de modélisation selon la vision du cogniticien

sur le pilotage de l'innovation en elle-même. Nous avons donc décidé de réaliser ce travail a posteriori, fort des enseignements acquis sur le terrain pendant les six mois d'observation et enquête. Nous évoquons ici l'apport du modèle conceptuel introduit en section 3 pour décrire cette innovation et en faire une analyse. Cette analyse a fourni une base de connaissances sous la forme d'un modèle de processus opérationnel à un niveau plus détaillé afin d'avoir une base explicite pour réfléchir à l'évaluation de la pratique. L'expérience comprend une caractérisation des changements dans les pratiques professionnelles qui ont été associés à des activités définies du modèle de processus (Chaher, Belaud & Pingaud 2017). Les tableaux 5 à 8 montrent l'instanciation du méta-

Concept	Définition
Processus	Une structure holistique d'activités organisées dans le temps et dans l'espace, et exécutées par des acteurs compétents qui jouent des rôles bien définis pour réaliser un objectif. (Sienou 2009)
Idée	Élaboration originale de la pensée, permettant, en particulier, de répondre à une situation, d'être à l'origine d'une action, d'une oeuvre ou d'une invention originale.
Activité	La réalisation d'une tâche.
Projet	Ensemble unique de processus, constitué d'activités coordonnées et maîtrisées ayant des dates de début et de fin et entreprises pour atteindre les objectifs du projet qui requiert la fourniture de livrables conformes à des exigences spécifiques, selon la Norme ISO 21500.
Risque	Situation ou circonstance indésirable qui a à la fois une probabilité de survenir et une conséquence potentiellement négative sur un projet, selon la Norme ISO 17666.
Performance	Résultat obtenu dans un domaine précis.
Ressources	Elles sont mobilisées pour convertir des éléments d'entrée en éléments de sortie.

Tableau 4 – Les concepts de modélisation selon la vision opératoire

modèle pour réussir à caractériser de manière complète cette innovation. Sur le tableau 5, le rôle des praticiens de santé sur le parcours de soins apparaît. Cette vision met en évidence le besoin d'évaluer des changements de pratiques collaboratives du fait de l'usage du terminal de télémédecine. La capacité d'innovation tient dans la faculté d'un usage explicite des terminaux dans ce contexte collaboratif et dans l'exercice même des métiers. L'innovation est qualifiée de radicale dans le tableau 6. Ce sont des innovations de procédé, commerciale, sociale et technologique qui s'entremêlent. L'incertitude est due au temps d'apprentissage sur l'usage du terminal par les praticiens qui sont déjà en surcharge de travail sur les prises en charge. La fiabilité de l'outil doit être totale. Le changement radical est simplement concentré sur les caractères synchrone et à distance de la consultation entre médecins. La vision opératoire du tableau 7 fait apparaître la logique du projet et de sa mise en oeuvre sur le pilote de taille conséquente avec des indicateurs de performance et de risque associés. L'idée fait écho au changement radical énoncé dans la vision du cogniticien. Enfin, le tableau 8 mon-

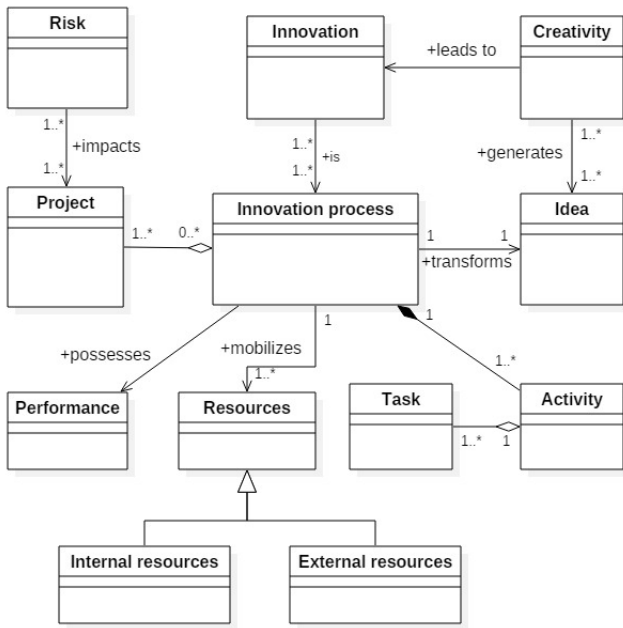


Figure 4 – Les concepts de modélisation selon la vision opératoire

tre que cette innovation dépasse le simple stade de l’invention pour satisfaire des exigences métiers dans un contexte qui appelle un besoin nouveau. La production de valeur explicite cela, avec une déclinaison sur le plan économique et sur le plan de l’usage, au bénéfice du système de santé.

Ce sont 31 concepts qui sont ici instanciés pour caractériser cette expérience. La complémentarité des 4 points de vue facilite le passage à la pratique lorsqu’il faut faire cet exercice d’instanciation. Nous allons poursuivre le travail sur le cadre en mettant au point le modèle de processus d’innovation qui a été mis en oeuvre pour progresser dans le déploiement de la télémédecine à l’échelle de la région Occitanie.

5 CONCLUSIONS

Comme l’objectif de notre recherche est de fournir des formes de rationalisation de l’innovation. Nous avons entrepris une première fouille des définitions de l’innovation pour créer un méta-modèle ou modèle conceptuel.

Ce premier résultat permet d’ouvrir deux chantiers :
 - Un travail sur le cycle de vie par une représentation originale et réaliste du processus d’innovation. Les concepts évoqués ici y seront mobilisés comme des flux ou des ressources autour des activités,
 - Un travail d’extension du modèle conceptuel pour caractériser l’innovation ouverte. Nous pensons qu’il est possible de la représenter comme une extension de l’innovation. Nous travaillons sur un modèle avec une structure en deux sphères concentriques : le

Classe	Instance
Interrelation	Demandes d’avis
Acteurs	Praticiens de santé
Membres du projet	Équipes de soins, Agence Régionale de Santé
Chercheurs	Méthodes d’évaluation de changements de pratiques professionnelles
Connaissance tacite	Pratiques des soins
Connaissance explicite	Usages des terminaux de télémédecine
Information	Documents de cadrage du projet
Compétence	Soigner un patient âgé
Capacité	Adoption du terminal de télémédecine pour pratiquer

Tableau 5 – Instance du méta-modèle selon la vision du sociologue

noyau représente le modèle conceptuel de l’innovation présenté ici, et la première couche sphérique qui l’entoure représente le modèle conceptuel spécifique de l’innovation ouverte. La santé connectée est un secteur où l’innovation ouverte est une réalité du simple fait du grand nombre de parties prenantes à impliquer dans le processus d’innovation. Nous pouvons citer trois catégories : les citoyens/patients, les praticiens de la santé et du social, les fournisseurs de solutions organisationnelles et technologiques. Au delà de l’expérience vécue sur la télémédecine, nous avons donc des terrains expérimentaux sur lesquels nous pourrions éprouver le cadre dans ce contexte d’un fonctionnement en réseau. De manière tout à fait concrète, cette perspective s’alimente aussi d’une réflexion sur les lieux de médiation de l’innovation qui se multiplient (les Fab Labs et les Living Labs). Nous avons la possibilité de travailler sur des projets menés dans le cadre du living lab sur le parcours patient, le Connected Health Lab, à cette fin (Chaher et al. 2017). Ce travail permettra de mettre en évidence l’apport de ces nouvelles structures à une accélération de l’innovation ouverte.

Enfin, dans notre titre nous utilisons «vers une innovation dirigée par les modèles» pour bien préciser qu’il s’agit d’une première qui, à notre connaissance, n’a pas d’antécédent et ouvre de très nombreuses perspectives scientifiques et techniques.

References

Achi, A. & Salinesi, C. (2015). Proposition d’un modèle d’innovation par l’usage des si, *INFORSID*.
 Boly, V. (2004). *Ingénierie de l’Innovation-Organisation et méthodologies des entreprises innovantes*, Lavoisier.
 Borgiel, K. (2017). *Proposition d’une approche centrée usage pour l’accompagnement*

Classe	Instance
Changement	Évolution des pratiques
Changement radical	Échange synchrone en visio-conférence avec préparation préalable du dossier
Innovation radicale	Nouvelle forme de coordination entre professionnels sur un parcours patient
Innovation de procédé	Oui car la production de l'avis ne suit plus le même processus
Innovation de service	Une nouvelle offre de service de soins
Innovation commerciale	Oui, car il s'agit d'un service nouveau pour un EHPAD jouant sur son attractivité
Innovation sociale	Oui, car la collaboration entre praticiens se déroule autrement et un processus collaboratif relie plusieurs métiers
Innovation technologique	Oui car usage du terminal de la télémédecine
Incertitude	Niveau d'appropriation de l'outil par les médecins, temps à consacrer

Tableau 6 – Instance du méta-modèle selon la vision du cognitif

Classe	Instance
Processus	Enchaînement des activités en trois temps : Demande/Consultation/Avis
Idée	S'affranchir des problèmes de distance et réussir à se synchroniser pour réfléchir en liaison synchrone sur un dossier
Activité	Réalisation des actions sur les trois temps du processus par les acteurs compétents
Projet	Étude expérimentale sur un échantillon de 20 EHPAD et trois services hospitaliers avec plus de cent patients concernés
Risque	Non appropriation du terminal
Performance	Meilleure préparation des dossiers
Ressources	Équipes des médecins requérants et des requis

Tableau 7 – Instance du méta-modèle selon la vision opératoire

de l'innovation organisationnelle liée à l'introduction d'un outil numérique collaboratif dans un système d'information: application au contexte du «home care», PhD thesis, Université de Bordeaux.

BPI, F. (2015). *Le guide Innovation Nouvelle Généra-*

Classe	Instance
Invention	Couplage visio-conférence et outil de workflow
Utilisateur	Système de santé
Exigence	Gains en matière de prise en charge et de coûts
Valeur	Productivité des demandes d'avis
Valeur économique	Minimisation des déplacements en service spécialisé en établissement hospitalier
Valeur d'usage	Souplesse dans la mise en relation synchrone - favorise un échange entre équipes

Tableau 8 – Instance du méta-modèle selon la vision de l'économiste

tion.

Castro, E. (2010). *Contribution à l'ingénierie de l'innovation: proposition d'un cadre de modélisation pour un système d'innovation centré sur le produit*, PhD thesis, Université de Lorraine.

Chaher, Y., Belaud, J.-P. & Pingaud, H. (2017). *Managing open innovation in connected health through a living lab*, *International Conference on Engineering, Technology and Innovation (ICE)*, IEEE.

Cohen, R. (2016). *Concevoir et lancer un projet.*

Cortes Robles, G. (2006). *Management de l'innovation technologique et des connaissances*, PhD thesis, Institut National Polytechnique de Toulouse.

Estefan, J. A. et al. (2007). *Survey of model-based systems engineering (mbse) methodologies*, *In-cose MBSE Focus Group* **25**(8): 1–12.

Holbrook, M. B. (1994). *The nature of customer value: an axiology of services in the consumption experience*, *Service quality: New directions in theory and practice* **21**: 21–71.

ISO (2009). *Iso 31000, Risk Management*.

Lall, S. (1992). *Technological capabilities and industrialization*, *World development* **20**(2): 165–186.

Penide, T., Gourc, D., Pingaud, H. & Peillon, P. (2013). *Innovative process engineering: a generic model of the innovation process*, *International Journal of Computer Integrated Manufacturing* **26**(3): 183–200.

Rabelo, R. J. & Bernus, P. (2015). *A holistic model of building innovation ecosystems*, *IFAC-PapersOnLine* **48**(3): 2250–2257.

- Rivière, A. & Mencarelli, R. (2012). Towards a theoretical clarification of perceived value in marketing, *Recherche et Applications en Marketing* **27**(3): 97–122.
- Schumpeter, J. A. (1934). *The schumptr: Theory economic development*, Harvard University Press.
- Sienou, A. (2009). *A Methodological framework for the integrated management of risks and business processes*, PhD thesis, Université de Toulouse, Centre de Recherche en Génie Industriel (Albi, France).
- Stewart, G. (1997). Supply-chain operations reference model (scor), *Logistics information management* **10**(2): 62–67.
- TOGAF (1998). The open group architecture framework, *Woburn, U.S.A* .
- Tomala, F., Senechal, O. & Tahon, C. (2001). Modèle de processus d'innovation, *3ème MOSIM, Troyes*.
- Von Hippel, E. (2005). Democratizing innovation: The evolving phenomenon of user innovation, *Journal für Betriebswirtschaft* **55**(1): 63–78.
- Xuereb, J.-M. (1991). Une redéfinition du processus d'innovation, *Revue française de gestion* **84**(1).
- Zaidat, A., Boucher, X. & Vincent, L. (2005). A framework for organization network engineering and integration, *Robotics and Computer-Integrated Manufacturing* **21**(3): 259–271.

VERS UNE METHODE DE GESTION INTEGREE RISQUES - PROCESSUS POUR LA SECURISATION DU CIRCUIT DU MEDICAMENT

Rafika THABET^(1,2), Amine BOUFAIED⁽¹⁾, Elyes LAMINE^(3,4), Ouajdi KORBA⁽¹⁾ et Hervé PINGAUD^(2,4)

⁽¹⁾ Laboratoire MARS, Université de Sousse, ISITCom Hammam Sousse, Tunisie

⁽²⁾ Laboratoire de Génie Chimique, CNRS, INP Toulouse, Université de Toulouse, France

⁽³⁾ Centre Génie Industriel, IMT Mines-Albi, Université de Toulouse, Albi, France

⁽⁴⁾ ISIS, INU Champollion, Université de Toulouse, Castres, France

rafika.thabet@gmail.com, boufaied.amine@gmail.com, elyes.lamine@mines-albi.fr,

ouajdi.korbaa@centraliens-lille.org, herve.pingaud@univ-jfc.fr

RESUME : *Devant la fréquence et l'ampleur des événements indésirables médicamenteux, les établissements de santé ont tout intérêt à mettre en place une démarche de gestion des risques dans le but d'assurer la sécurité du patient et, en particulier, de diminuer le risque de survenue des Erreurs médicamenteuses (EMs) liées au processus complexe de la Prise En Charge Médicamenteuse (PECM) du patient. Afin de les accompagner dans leurs démarches, nous proposons d'utiliser une méthode de gestion intégrée des risques et des processus. La littérature suggère diverses approches pour représenter l'information liée au risque dans les modèles de processus métiers, répertoriés sous le paradigme « Risk-aware Business Process Management ». La diversité de ces méthodes rend difficile le choix du langage et la méthode la plus adaptée au besoin d'une gestion efficace. Dans ce papier, nous proposons d'étudier le potentiel de ce paradigme pour la sécurisation du processus de la PECM contre le risque de survenu d'EMs.*

MOTS-CLES : *circuit du médicament, erreur médicamenteuse, gestion de processus, gestion de risque, ADOxx.*

1 INTRODUCTION

La Prise En Charge Médicamenteuse (PECM) est l'utilisation sécurisée, appropriée et efficace du médicament chez le patient pris en charge en établissement de santé (Arrêté du 16 avril 2011). Elle constitue un processus complexe et pluridisciplinaire, impliquant de nombreux professionnels et constitué de plusieurs étapes. Cette complexité entraîne un risque certain de survenue d'Erreurs Médicamenteuses (EMs) pouvant avoir des conséquences cliniques graves sur les patients. En effet, en 2015, la HAS estime que 40% des événements indésirables graves sont d'origine médicamenteuse. A ce titre, la sécurisation de ce processus est au cœur des préoccupations des tutelles et des établissements de santé (HAS 2013). En effet, la sécurisation de ce circuit passe par l'application de la réglementation, le respect des obligations professionnelles, la promotion de l'utilisation des nouvelles technologies et notamment par la mise en place d'une démarche de gestion des risques. Cette dernière a pour but d'assurer la sécurité du patient et des soins qui lui sont délivrés et, en particulier, de limiter le risque de survenue des EMs qui sont sources potentielles d'Événements Indésirables Médicamenteux (EIMs) évitables. Elle passe notamment par une phase d'état des lieux visant à identifier les risques potentiels, puis par une phase d'évaluation aboutissant à la définition puis à la mise en œuvre et au suivi d'un plan d'amélioration. Dans le domaine de la santé, plusieurs méthodes de gestion de ces risques existent (HAS 2013). Cependant, ces

méthodes ne prennent pas en considération le critère complexe de l'EM, sa forte liaison aux étapes du processus de la PECM, et les facteurs organisationnels et humains qui interviennent de manière importante dans la survenue de ces erreurs. A la lumière de ces constatations, il serait donc indispensable pour les établissements de santé de mettre en place une politique efficace de gestion qui prenne en considération tous les concepts liés aux risques de survenue d'EMs ainsi que leurs dépendances inhérentes au processus de la PECM.

Ainsi la gestion intégrée des processus métiers et des risques est au cœur de la démarche d'analyse dynamique que nous préconisons pour la sécurisation du processus de la PECM. En effet, cette gestion intégrée connue sous l'appellation « Risk-aware Business Process Management – R-BPM » est un paradigme récemment apparu (Sienou et al. 2010) (Suriadi et al. 2014). Il vise à intégrer les deux domaines traditionnellement séparés de la gestion des risques et de la gestion par les processus métiers. Cette intégration permet d'augmenter l'efficacité de l'identification, la détection, et la gestion des risques liés aux processus. C'est une approche qui facilite la détection, la prévention des risques, la planification des actions de protection et l'amélioration des décisions. Récemment, plusieurs approches R-BPM ont été proposées dans la littérature, notamment celle proposée par Sienou et al (Sienou et al. 2010), dénommée « Business Process-Risk management – Integrated Method (BPRIM) ». Dans le cadre de ce travail, nous proposons d'étudier le potentiel cette méthode pour gérer les risques de survenue d'EMs liées au processus de la PECM.

Ce papier est structuré comme suit : La section 2 présente un aperçu du processus de la PECM, l’ampleur et la fréquence de survenue des EMs dans ce processus et le besoin de sa sécurisation. La section 3 propose une étude comparative des approches existantes en R-BPM. Dans la section 4, nous présentons l’approche et la méthodologie adoptée. La section 5 donne un aperçu des premiers résultats obtenus suite à la mise en œuvre de notre modélisation. Le document se termine par une conclusion et des perspectives.

2 LA PRISE EN CHARGE MEDICAMENTEUSE

En 2006, selon la Société Française de Pharmacie Clinique (SFPC), le Circuit Du Médicament (CDM) était utilisé pour évoquer le processus de la PECM du patient hospitalisé. Ce processus comprenait les étapes de prescription, dispensation, d’administration et de suivi thérapeutique. Mais avec les nouvelles définitions proposées par l’Inspection Générale des Affaires Sociales (IGAS) et l’arrêté du 6 avril 2011, relatif au management de la qualité de la PECM et aux médicaments dans les établissements de santé, ce processus a évolué. Nous présentons dans la figure 1 l’évolution de ces nomenclatures.

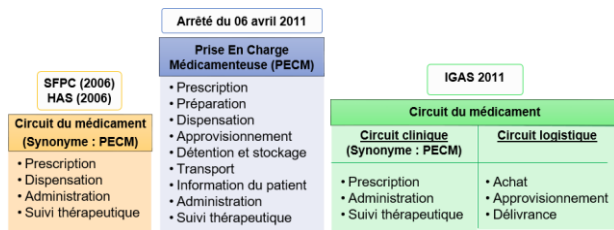


Figure 1 - Evolution des notions de CDM et la PECM

Selon (IGA 2011), la notion de CDM est vue comme un ensemble de deux circuits distincts et interconnectés :

- le premier, *clinique*, correspondant à la prise en charge thérapeutique à proprement parler ;
- le deuxième, *logistique*, concernant le médicament en tant que produit, de l’achat jusqu’à la délivrance dans l’unité de soins, rejoignant le circuit clinique au stade de l’administration du médicament au patient.

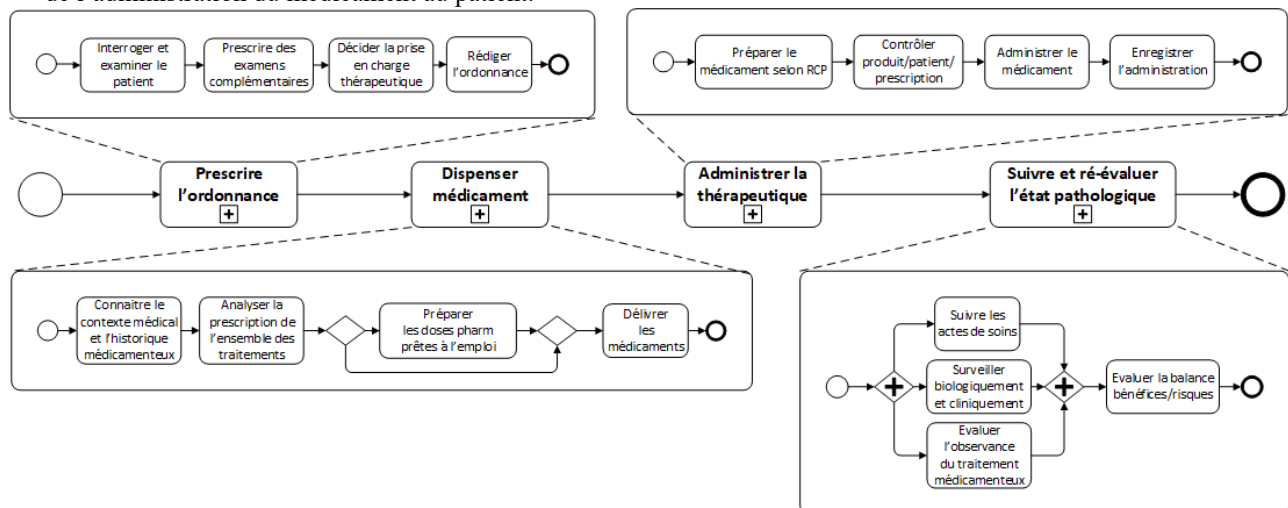


Figure 2 - Modèle du processus de la PECM avec BPMN

Dans le cadre de ce travail, nous nous intéressons essentiellement au circuit clinique du CDM : le processus de la PECM. Ce dernier est l’un des processus de soins les plus complexes notamment par :

- La diversité des pathologies, leur degré d’urgence, leur gravité couplée à la diversité des thérapeutiques et aux terrains des patients ;
- L’intervention de différents acteurs de santé : le médecin prescripteur, le pharmacien dispensateur, les aides-soignantes, l’infirmier(ère) qui administre le médicament et enfin le patient qui doit être informé sur son traitement ;
- L’enchaînement de nombreuses étapes du processus qui se passent en des lieux géographiques différents ;
- La circulation de l’information qui doit rester accessible.

Nous présentons, dans la figure 2, une description des étapes du processus de la PECM en utilisant la notation BPMN. Les sous-processus et les activités ont été inspirés du macro-processus présenté dans le rapport de la HAS en 2013 (HAS 2013).

La complexité de ce processus expose au risque de survenue d’EMs. Ces erreurs peuvent être à l’origine d’EIMs avec des conséquences cliniques potentiellement graves pour le patient. D’ailleurs, il est établi que le médicament est une des causes principales dans la survenue d’événements indésirables graves chez les patients hospitalisés (HAS 2013). Ainsi une bonne maîtrise du processus de la PECM passe notamment par une bonne connaissance des EMs et des EIMs (Morin, Souchon, et Saint-Lorant 2015). Pour ce faire, nous allons présenter, dans la suite, les concepts liés à l’EM.

2.1 Caractérisation de l’erreur médicamenteuse

Plusieurs définitions existent, en littérature, décrivant la notion d’EM. Nous proposons dans le tableau 1 une liste non exhaustive des définitions qui ont retenu notre attention.

Références	Définitions de l'EM
ANSM ¹	« omission ou la réalisation non intentionnelle d'un acte survenu au cours du processus de soins impliquant un médicament, qui peut être à l'origine d'un risque ou d'un événement indésirable pour le patient. »
SFPC	« écart par rapport à ce qui aurait dû être fait au cours de la prise en charge thérapeutique médicamenteuse du patient »
AAQTE ²	« événement iatrogène médicamenteux évitable, résultant d'un dysfonctionnement non intentionnel dans l'organisation de la prise en charge thérapeutique médicamenteuse du patient. »
OMS ³	« événement évitable au cours duquel une dose de médicament reçue par le patient diffère de ce qui avait été prescrit ou ne correspond pas à la politique et aux procédures de l'hôpital »

Tableau 1 - Définitions de l'erreur médicamenteuse

En 1995, le NCC MERP (National Coordinating Council for Medication Error Reporting and Prevention) a créé un référentiel de classification des EMs. Ce référentiel a été par la suite traduit par le Réseau d'Epidémiologie de l'Erreur Médicamenteuse – REEM et présenté par la SFPC, en 2006, dans son « Dictionnaire français de l'erreur médicamenteuse » (Schmitt et al. 2006). Ce dernier caractérise l'EM par :

- **Etape de réalisation :** Il s'agit de l'étape de la PECM durant laquelle l'erreur peut survenir. En effet, plusieurs études nationales et internationales confirment le fait que les EMs surviennent dans toutes les étapes du CDM. Nous présentons dans la figure 3 un résumé de ces études.

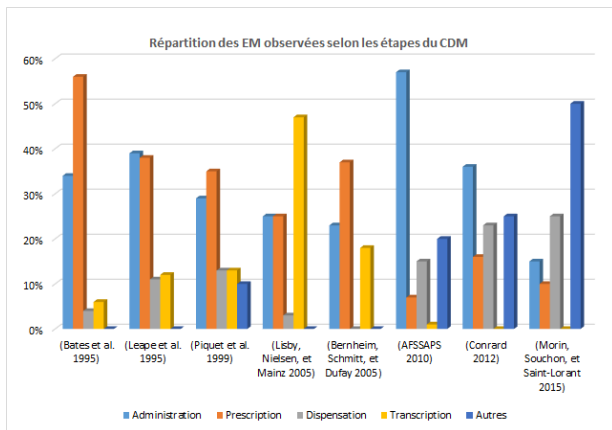


Figure 3 - Répartition des EM observées selon les étapes du CDM

- **Types d'erreur :** selon les données de la littérature on distingue 13 types d'EM (HAS 2013).
- **Nature de l'erreur :** Il s'agit d'une caractéristique qui renseigne le degré de réalisation de l'EM. Elle peut être de nature avérée, c'est-à-dire erreur qui s'est effectivement produite et est parvenue jusqu'au patient sans avoir été interceptée, ou à l'opposé, de nature potentielle, à savoir détectée et interceptée par un professionnel de santé, un patient

¹ Agence Nationale de Sécurité du Médicament
² Association pour l'Assurance Qualité en Thérapeutique et l'Evaluation
³ Organisation Mondiale de la Santé

ou son entourage avant l'administration du médicament au patient.

- **Gravité des conséquences cliniques :** Pour décrire la gravité des conséquences cliniques sur le patient, la SFPC a défini neuf catégories de A à I, la catégorie I correspondant au niveau maximal de gravité : le décès du patient. La catégorisation du niveau de gravité dépend de 3 paramètres :
 - le degré de réalisation (erreur latente, potentielle ou avérée) ;
 - les effets indésirables chez le patient (absence d'effet indésirable, dommages temporaires, dommages irréversibles, pronostic vital engagé voire décès) ;
 - les moyens supplémentaires mis en œuvre pour optimiser la prise en charge du patient suite à l'erreur (surveillance accrue, allongement de la durée de l'hospitalisation, nécessité de thérapeutiques supplémentaires).
- **Causes de l'erreur :** Les causes possibles d'une EM sont multifactorielles. On retrouve dans la littérature plusieurs études, notamment l'étude Multicentrique pour l'Évaluation de la ReVue des Erreurs et leur Iatrogénie Liées aux médicaments (MERVEIL) (Conrard 2012) et l'étude de (Morin, Souchon, et Saint-Lorant 2015), systèmes de déclaration des EMs, tel que le réseau REEM, et des revues de littératures, tel que le rapport de la HAS en 2013 (HAS 2013), qui visent à tracer un bilan de l'existant concernant les principales erreurs médicamenteuses et leurs causes.

En se basant sur les définitions et les caractéristiques de l'EM déjà présentées, nous proposons, dans la figure 4, un méta-modèle décrivant l'EM dans son contexte. L'analyse des études nationales et internationales sur les EMs montre l'aspect complexe et non maîtrisé de ces erreurs. En effet, l'apparition de ces erreurs met en question la robustesse des méthodes de gestion utilisées actuellement pour sécuriser le processus de la PECM dans les établissements de santé.

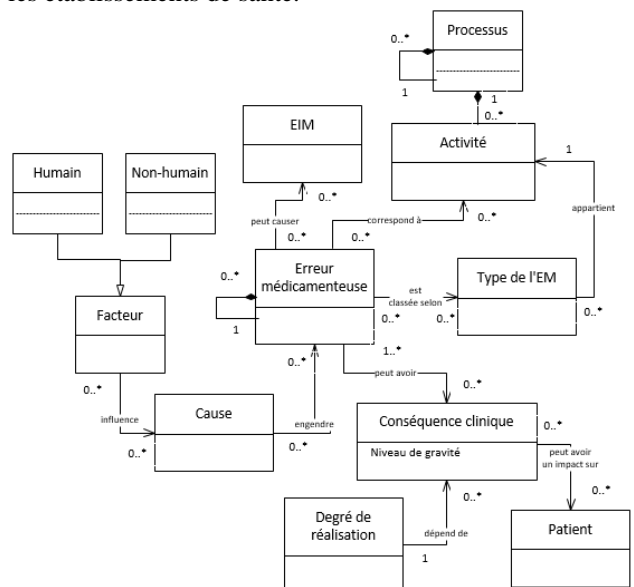


Figure 4 - Méta-modèle simplifié de l'EM

2.2 Besoin de la sécurisation de la PECM

Depuis la publication du rapport “*To err is human*” (Donaldson, Corrigan, et Kohn 1999), en 1999, la prise de conscience dans le domaine de la sécurité des soins s’est accélérée. Aux États-Unis, le nombre de patients confrontés chaque année à une EM est estimé à 1,5 million et le nombre de patients qui en décèdent est estimé à 7 000. En France, ces données internationales ont été confirmées lors des Enquêtes Nationales sur les Événements Indésirables associés aux Soins (ENEIS) qui ont été réalisées en 2004 et 2009 à la demande de la direction générale de la santé et de la Direction Générale de l’Organisation des Soins (DGOS), afin de connaître l’incidence des événements indésirables graves (EIG) dans les établissements de santé, leur part d’évitabilité et d’en analyser les causes immédiates. En combinant les résultats des deux enquêtes, environ 30 % des EIG seraient liés à un EIM, dont environ 50 % seraient évitables donc assimilés à des EMs (Michel et al. 2011). Il n’est donc pas étonnant que l’arrêté du 06 avril 2011 cible, plus particulièrement dans son article 9, la mise en place par la direction de l’établissement d’une organisation en charge de l’analyse des événements indésirables, EMs ou dysfonctionnements liés à la PECM, et la planification des actions nécessaires pour en améliorer la sécurité (c’est-à-dire la mise en œuvre d’une démarche a priori et a posteriori de gestion des risques médicamenteux) (Arrêté du 16 avril 2011).

Dans le domaine de la santé, plusieurs méthodes de gestion des risques existent et leur utilisation est recommandée par la HAS (HAS 2013). Cependant, ces méthodes ne prennent pas en considération certains critères de l’EM et du processus de la PECM du patient. En effet, à notre connaissance, il n’existe aucune méthode de gestion utilisée actuellement qui prend en considération :

- toutes les caractéristiques de l’EM ;
- la forte relation qui existe entre l’EM et son (ou ses) étape(s) de survenue dans le processus de la PECM
- et les facteurs organisationnels et humains qui interviennent de manière importante dans la survenue des EMs.

Dans l’optique de garantir la sécurité du processus de la PECM du patient, les établissements de santé sont alors incités à mettre en œuvre une politique efficace de détection, prévention et maîtrise des EMs liés au processus complexe de la PECM. Pour ce faire, nous préconisons l’exploration de nouvelles pistes de recherches, notamment la gestion intégrée des risques et des processus.

3 GESTION INTEGREE DES RISQUES ET DES PROCESSUS METIERS

Au cours de ces dernières années, un intérêt majeur de recherche est porté à l’intégration des deux domaines, traditionnellement séparés, de la gestion des risques et la gestion par les processus métiers dans un concept connu sous le sigle R-BPM. L’importance de cette intégration a

été confirmée dans la communauté des chercheurs (Zur Muehlen et Ho 2005), dans les lignes directrices de l’industrie (COSO 2004), et dans beaucoup d’études (Sienou et al. 2010)(Becker et al. 2010) (Suriadi et al. 2014).

Dans cette partie, nous proposons une étude bibliographique sur les approches R-BPM existantes, entre 2010 et 2017. Nous proposons, ainsi, plusieurs méthodes pour les classer. La première méthode de classification consiste à comparer les approches selon le niveau d’intégration du concept risque dans le cycle de vie du BPM. Le résultat de cette comparaison est illustré dans le tableau 2.

En effet, les approches R-BPM peuvent être divisées en deux groupes, à savoir :

- **Approches R-BPM au niveau conceptuel** : regroupe les approches s’intéressant à la gestion des risques pendant la phase conceptuelle (Design-time) des processus métiers ;
- **Approches R-BPM au niveau opérationnel** : regroupe les approches qui se focalisent plutôt sur la gestion des risques pendant et après l’exécution des processus métiers (Run-time).

		Approches
Cycle de vie du BPM	Niveau conceptuel	ANC1 : (Jakoubi et al. 2010), (Tjoa et al. 2011) ANC2 : (Sienou et al. 2010) ANC3 : (Cope et al. 2014) ANC4 : (Weiss et Winkelmann 2011) ANC5 : (Rotaru et al. 2011) ANC6 : (Betz, Hickl, et Oberweis 2011), ANC7 : (Strecker, Heise, et Frank 2011), ANC8 : (Panayiotou et al. 2010) ANC9 : (Bai et al. 2012), ANC10 : (Fenz 2011), (Fenz et al. 2013) ANC11 : (Pittl, Fill, et Honegger 2017) ANC12 : (Shabnam et al. 2014) ANC13 : (Lhannaoui, Kabbaj, et Bakkoury 2014) ANC14 : (Shah et al. 2017)
	Niveau opérationnel	ANO1 : (Conforti et al. 2016) ANO2 : (Jans et al. 2011) ANO3 : (Wickboldt et al. 2011) ANO4 : (Haggag, Khedr, et Montasser 2015) ANO5 : (Kim et al. 2017) ANO6 : (Pika et al. 2016) ANO7 : (Rogge-Solti et Weske 2015) ANO8 : (Metzger et Bohn 2017)

Tableau 2 - Niveau d’intégration du concept risque dans le cycle de vie du BPM

La deuxième méthode de classification consiste à comparer les approches selon la maturité de la caractérisation du risque ou le degré de formalisation du risque. Le résultat de cette comparaison est illustré dans le tableau 3. Cette comparaison a été inspirée de concepts génériques des méthodes de modélisation présentés dans (Domenik Bork et Fill 2014). Dans l’ensemble, ces approches se concentrent principalement sur la définition de la syntaxe concrète des construits proposés du risque. Par exemple, l’approche proposée par (Jakoubi et al. 2010), introduit un ensemble de notations graphiques pour représenter les éléments du risque (tels que les menaces, les ressources,

les contre-mesures et les actions de récupération) pouvant être associés à des activités de processus métier. D'autre part, l'approche proposée par (Sienou et al. 2010) introduit de nouvelles notations graphiques pour représenter les éléments du risque (facteur risque, événements risques, situation risque, valeur, impact, etc.) en étendant le langage EPC (Event-driven Process Chain). Cependant, peu d'approches ont tenté de formaliser la syntaxe abstraite des construits proposés du risque. Parmi ces approches, on retrouve les travaux de (Sienou et al. 2010), (Cope et al. 2014), (Betz, Hickl, et Oberweis 2011), et Strecker et al. (Strecker, Heise, et Frank 2011) qui utilisent le langage UML pour définir la syntaxe abstraite de leurs construits, et l'approche proposée par Weiß et Winkelmann (Weiss et Winkelmann 2011) qui utilise plutôt les diagrammes Entity Relationship (ER). La majorité de ces approches ne sont guidées par aucune norme de risque existante, à l'exception du travail de Weiß et Winkelmann (Weiss et Winkelmann 2011), et celui de Sienou et al. (Sienou et al. 2010). La troisième classification consiste à comparer ces approches selon les fondements méthodiques utilisés pour concevoir une approche R-BPM. Le résultat de cette comparaison est illustré dans le tableau 4. Parmi toutes ces approches, la méthode proposée par (Sienou et al. 2010), dénommée Business Process-Risk

management – Integrated Method (BPRIM), a retenu notre attention. Nous détaillerons cette méthode dans la section qui suit.

Approches R-BPM	Caractérisation des composants du risque	Norme du risque	Formalisation du risque			Analyse du risque	
			Syntaxe abstraite	Syntaxe concrète	Sémantique		
							ANC1
ANC2	+	+	+	+	-	-	-
ANC3	±	-	+	+	-	±	-
ANC4	+	+	+	+	-	-	-
ANC5	+	-	+	+	±	-	-
ANC6	±	-	+	+	-	-	-
ANC7	-	-	+	+	-	-	-
ANC8	-	-	-	+	-	±	-
ANC9	-	-	+	-	-	+	-
ANC10	±	±	NA	NA	NA	+	-
ANO1	-	-	+	+	-	+	-
ANO2	-	-	NA	NA	NA	-	-
ANO3	+	+	+	-	-	±	-
ANC11	±	+	+	+	+	+	-
ANC12	-	-	+	-	-	+	-
ANO8	-	-	+	-	-	+	-
ANO7	-	-	+	-	-	+	-
ANO4	-	-	+	+	-	+	-
ANC13	±	-	-	+	-	±	-
ANO5	±	-	+	-	-	+	-
ANO6	±	-	+	+	-	+	-
ANC14	+	±	+	-	-	+	-

Tableau 3 - Maturité de la description du risque

Approches R-BPM	Méta-modèle risque-processus	Langage de modélisation risque-processus	Règles de modélisation	Cycle de vie synchronisé	Méthodologie d'application	Implémentation
ANC1	-	-	-	-	-	±
ANC2	+	+	+	+	+	-
ANC3	-	±	±	-	+	-
ANC4	+	±	±	-	±	-
ANC5	±	±	±	-	-	-
ANC6	-	±	±	±	+	+
ANC7	+	±	±	-	±	-
ANC8	-	-	-	-	±	-
ANC9	-	-	-	-	±	+
ANC10	-	-	-	-	-	±
ANO1	±	+	±	+	±	+
ANO2	±	-	-	-	±	+
ANO3	-	-	-	±	±	+
ANC11	±	±	±	±	+	+
ANC12	-	-	-	-	-	-
ANO8	±	-	-	-	-	-
ANO7	±	-	-	-	±	+
ANO4	-	-	-	±	±	+
ANC13	-	±	±	-	-	-
ANO5	±	±	+	-	+	±
ANO6	-	±	-	-	-	+
ANC14	-	-	-	-	±	-

Tableau 4 - Finesse des fondements méthodiques

Légende : - Ne supporte pas le concept,
 + Supporte largement le concept
 et +/- Supporte partiellement le concept

4 APPROCHE ADOPTÉE

D'un point de vue général, la méthode BPRIM (Sienou 2009) est la seule, parmi toutes les approches R-BPM présentées, qui propose un cadre méthodologique conceptuel complet composé notamment de :

- Un cycle de vie synchronisé de l'ingénierie des processus et de l'ingénierie des risques. Il comporte essentiellement quatre phases à savoir le cadrage, l'appréciation, le traitement et le suivi ;
- Un méta modèle unifiant risque et processus parfaitement défini, basé sur la norme ISO 19440 et compatible avec la norme ISO 31000 ;
- Un langage de modélisation commun des processus métiers et des risques, basé sur l'extension du langage EPC (Scheer 1999) avec la proposition d'un ensemble de modèles et de diagrammes (Modèle EPC, Modèle EPC étendu aux risques, Diagramme de taxonomie des risques, Diagramme d'analyse des risques, Diagramme des relations risques, Diagramme du contexte du risque) permettant la cons-

truction d'une démarche conceptuelle de processus métier dirigée par les risques.

Nous avons choisi d'éprouver ce cadre conceptuel sur notre domaine d'application, celui de la gestion des EMs liés au processus de la PECM. En effet, cette méthode est la seule qui permet, à la fois, de :

- modéliser le processus de la PECM dans un modèle appelé : EPC ;
- modéliser le processus de la PECM étendu aux EMs dans un modèle appelé : EPC étendu aux risques ;
- typer les risques d'EMs dans un modèle appelé : Diagramme de taxonomie des risques ;
- modéliser les relations entre les EMs dans un modèle appelé : Diagramme de relation des risques ;
- décrire l'EM dans tout son contexte, à savoir cause (facteur risque et évènement risque) et conséquence (situation risque, valeur, impact, etc.), et analyser

l'EM dans un modèle appelé : Diagramme d'analyse du risque ;

Tous ces points permettent de décrire parfaitement toutes les caractéristiques d'une EM, d'analyser ces erreurs, de modéliser le processus de la PECM étendu aux EMs et d'analyser l'impact de ces erreurs sur les activités et les ressources du processus de la PECM du patient.

Cependant, cette méthode présente quelques limites :

- Il s'agit d'une méthode conceptuelle riche, mais elle n'intègre pas de méthodes et d'algorithmes permettant d'analyser les modèles construits ;
- La validité de la formalisation des construits proposés n'a pas été vérifiée ;
- Il n'existe pas d'outil qui supporte la méthode.

Ainsi pour consolider cette méthode, nous proposons de l'étendre, en adoptant la démarche suivante (figure 5).

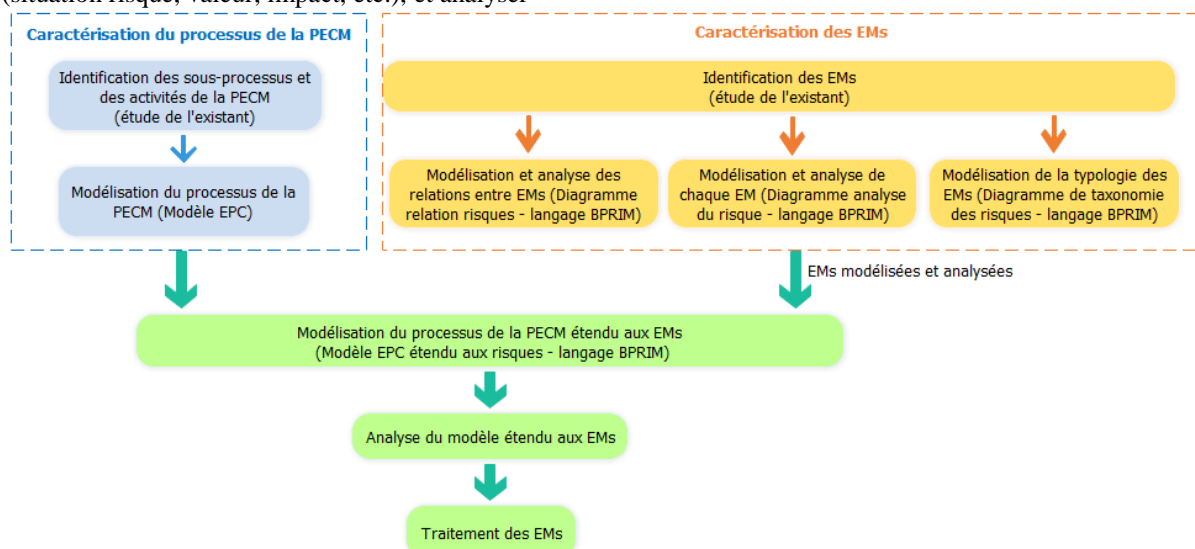


Figure 5 – Démarche adoptée

5 PREMIERS RÉSULTATS

Pour mettre en pratique la démarche adoptée, et l'appliquer à notre domaine d'études, nous avons choisi d'utiliser le logiciel ADOxx (Efendioğlu et Woitsch 2016), une plateforme de méta-modélisation. Bien que la communauté ADOxx.org se concentre sur la réalisation technique d'outils de modélisation, elle est également souvent impliquée dans les spécifications conceptuelles des approches basées sur des modèles dans le contexte de projets de recherche européens.

ADoBPRIM, l'outil qui naît de la démarche proposée, correspond à l'extension des concepts proposés dans la méthode BPRIM, sur la plateforme ADOxx. La mise en œuvre de cette démarche nous a permis de vérifier la validité des construits et des modèles proposés dans BPRIM et de l'étendre pour construire toute une méthode de modélisation. Une description détaillée de notre outil de modélisation (Thabet et al. 2018) est disponible via un projet au sein du « Open Models Laboratory », une communauté mondiale de modélisateurs et de développeurs de méthodes de modélisation (Dominik Bork et

Miron 2017). Les premiers résultats obtenus par ADoBPRIM sont présentés dans la figure 6. Cette dernière illustre quelques instanciations des modèles construits appliqués au processus de la PECM étendu aux risques de survenue d'EMs.

CONCLUSIONS

L'idée principale de cet article est la mise en œuvre d'une politique efficace de gestion des EMs liées au processus de la PECM en tirant profil du concept R-BPM. Nous avons présenté, dans ce travail, une démarche de modélisation commune des risques et des processus métiers basée la méthode BPRIM en utilisant la plateforme de méta-modélisation ADOxx.

La mise en œuvre et la vérification de la bonne validité de notre méthode de modélisation par son application dans un cas d'étude réel, nous a motivé pour poursuivre l'amélioration de cette méthode pour intégrer des mécanismes et des algorithmes d'analyse des risques de survenue d'EMs liés au processus de la PECM du patient dans les établissements de santé, et qui fera l'objet des prochaines publications.

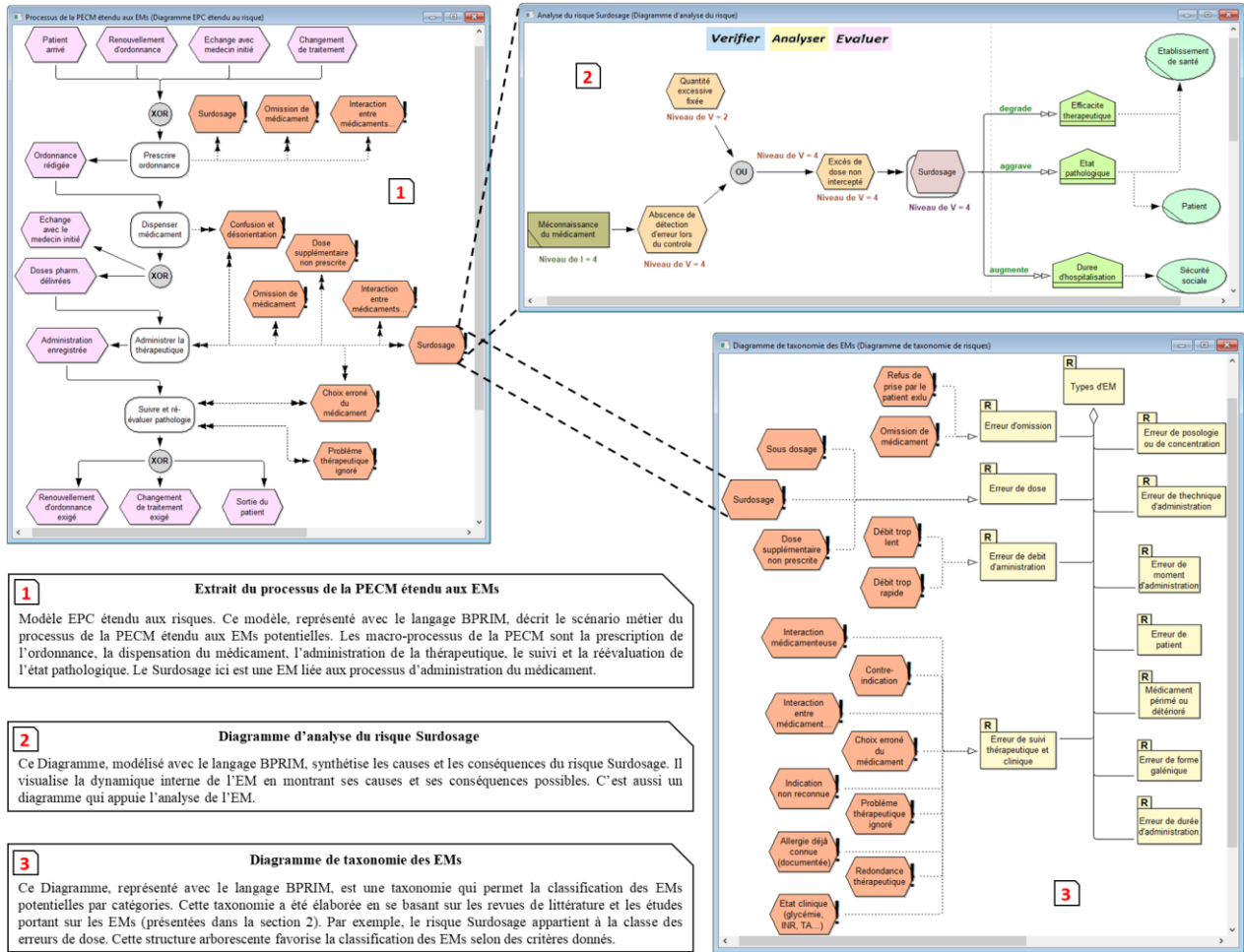


Figure 6 - Quelques modèles instanciés avec notre outil de modélisation ADoBPRIM

REFERENCES

Arrêté du 16 avril. 2011. « Arrêté du 6 avril 2011 relatif au management de la qualité de la prise en charge médicamenteuse et aux médicaments dans les établissements de santé [JORF du 16 avril 2011]. »

Bai, Xue, Ramayya Krishnan, Rema Padman, et Harry Jiannan Wang. 2012. « On risk management with information flows in business processes ». *Information Systems Research* 24 (3): 731–749.

Becker, Jörg, Dominic Breuker, Burkhard Weis, et Axel Winkelmann. 2010. « Exploring the status quo of business process modelling languages in the banking sector—an empirical insight into the usage of methods in banks ». In *Proceedings of the Twenty-first Australasian Conference on Information Systems (ACIS'10)*.

Betz, Stefanie, Susan Hickl, et Andreas Oberweis. 2011. « Risk-aware business process modeling and simulation using XML nets ». In *Commerce and enterprise computing (cec), 2011 IEEE 13th conference on*, 349–356. IEEE.

Bork, Domenik, et Hans-Georg Fill. 2014. « Formal aspects of enterprise modeling methods: a comparison framework ». In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 3400–3409. IEEE.

Bork, Dominik, et Elena-Teodora Miron. 2017. « OMiLAB-An Open Innovation Community for Modeling Method Engineering ».

Conforti, Raffaele, Sven Fink, Jonas Manderscheid, et Maximilian Röglinger. 2016. « PRISM—A predictive risk monitoring approach for business processes ». In *International Conference on Business Process Management*, 383–400. Springer.

Conrard, E. 2012. « La REMED: la Revue des Erreurs liées aux Médicaments Et aux Dispositifs associés. Les résultats de l'étude MERVEIL conduite pour évaluer la REMED en tant que méthode d'évaluation des pratiques professionnelles ».

Cope, Eric W., Lea A. Deleris, Dominik Etzweiler, Jana Koehler, Jochen M. Kuester, et Bonnie K. Ray. 2014. System and method for creating and expressing risk-extended business process models. 8862491, issued octobre 2014.

Donaldson, Molla S., Janet M. Corrigan, et Linda T. Kohn. 1999. *To err is human: building a safer health system*. Vol. 6. National Academies Press.

Efendioglu, Nesat, et Robert Woitsch. 2016. « Modelling Method Design: An Adoxx Realisation ». In *Enterprise Distributed Object Computing Workshop (EDOCW), 2016 IEEE 20th International*, 1–8. IEEE.

Fenz, Stefan. 2011. « From the resource to the business process risk level ». In *Proceedings of the South African Information Security Multi-Conference: Port*

- Elizabeth, South Africa, 17-18 May 2010, 100. Lulu.com.
- Haggag, Mohamed H., Ayman E. Khedr, et Hend S. Montasser. 2015. « A risk-aware business process management reference model and its application in an Egyptian university ». *International Journal of Computer Science and Engineering Survey* 6 (2): 11.
- HAS, Haute Autorité de Santé. 2013. « Outils de sécurisation et d'auto-évaluation de l'administration des médicaments ». *Saint-Denis La Plaine: HAS*.
- Jakoubi, Stefan, Simon Tjoa, Sigrun Goluch, et Gerhard Kitzler. 2010. « A formal approach towards risk-aware service level analysis and planning ». In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*, 180–187. IEEE.
- Jans, Mieke, Jan Martijn Van Der Werf, Nadine Lybaert, et Koen Vanhoof. 2011. « A business process mining application for internal transaction fraud mitigation ». *Expert Systems with Applications* 38 (10): 13351–13359.
- Kim, Jinwoo, Jinyoun Lee, Jooseok Lee, et Injun Choi. 2017. « An Integrated Process-Related Risk Management Approach to Proactive Threat and Opportunity Handling: A Framework and Rule Language ». *Knowledge and Process Management* 24 (1): 23–37.
- Lhannaoui, Hanane, Mohammed Issam Kabbaj, et Zohra Bakkoury. 2014. « Analyzing risks in business process models using a deviational technique ». In *Software Engineering and Applications (ICSOFT-EA), 2014 9th International Conference on*, 189–194. IEEE.
- Metzger, Andreas, et Philipp Bohn. 2017. « Risk-Based Proactive Process Adaptation ». In *International Conference on Service-Oriented Computing*, 351–366. Springer.
- Michel, P., M. Lathelize, R. Bru-Sonnet, S. Domecq, M. Kret, et J. L. Quenon. 2011. « Enquête Nationale sur les Événements Indésirables graves liés aux Soins 2009 (ENEIS2): description des résultats 2009 ». *Rapport final à la DREES (Ministère du travail, de l'emploi et de la Santé)–Février*, 3–3.
- Morin, P., J. Souchon, et G. Saint-Lorant. 2015. « Complémentarité de trois dispositifs de déclaration des événements indésirables médicamenteux ». *Le Pharmacien Hospitalier et Clinicien* 50 (4): 386–394.
- Panayiotou, N., Stylianos Oikonomitsios, Christina Athanasiadou, et S. Gayialis. 2010. « Risk assessment in virtual enterprise networks: A process-driven internal audit approach ». *Managing Risk in Virtual Enterprise Networks: Implementing Supply Chain Principles, Hershey, PA: IGI Global*, 290–312.
- Pika, Anastasiia, Wil MP van der Aalst, Moe Thandar Wynn, Colin J. Fidge, et Arthur HM ter Hofstede. 2016. « Evaluating and predicting overall process risk using event logs ». *Information Sciences* 352: 98–120.
- Pittl, Benedikt, Hans-Georg Fill, et Gerald Honegger. 2017. « Enabling Risk-Aware Enterprise Modeling using Semantic Annotations and Visual Rules ».
- Rogge-Solti, Andreas, et Mathias Weske. 2015. « Prediction of business process durations using non-Markovian stochastic Petri nets ». *Information Systems* 54: 1–14.
- Scheer, August-Wilhelm. 1999. « ARIS-business process framework ». *Berlin: Springer*.
- Schmitt, Etienne, Daniel Antier, Chantal Bernheim, Edith Dufay, Marie Caroline Husson, et Edgar Tissot. 2006. « Dictionnaire français de l'erreur médicamenteuse ». *Société Française de Pharmacie Clinique*.
- Shabnam, Luba, Farzana Haque, Moshir Bhuiyan, et Aneesh Krishna. 2014. « Risk measure propagation through organisational network ». In *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*, 217–222. IEEE.
- Shah, Liaqat Ali, Alain Etienne, Ali Siadat, et François Vernadat. 2017. « Process-oriented risk assessment methodology for manufacturing process evaluation ». *International Journal of Production Research* 55 (15): 4516–4529.
- Sienou, Amadou. 2009. « Proposition d'un cadre méthodologique pour filemanagement intégré des risques et des processus d'entreprise ». *A Methodological framework for the integrated management of risks and business processes*, INP Toulouse.
- Sienou, Amadou, Elyes Lamine, Hervé Pingaud, et Achim P. Karduck. 2010. « Risk driven process engineering in digital ecosystems: Modelling risk ». In *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*, 647–650. IEEE.
- Strecker, Stefan, David Heise, et Ulrich Frank. 2011. « RiskM: A multi-perspective modeling method for IT risk assessment ». *Information Systems Frontiers* 13 (4): 595–611.
- Suriadi, Suriadi, Burkhard Weis, Axel Winkelmann, Arthur HM ter Hofstede, Michael Adams, Raffaele Conforti, Colin Fidge, Marcello La Rosa, Chun Ouyang, et Michael Rosemann. 2014. « Current research in risk-aware business process management: overview, comparison, and gap analysis ». *Communications of the Association for Information Systems* 34 (1): 933–984.
- Thabet, Rafika, Elyes Lamine, Amine Boufaied, Ouajdi Korbaa, et Hervé Pingaud. 2018. « Business Process Risk management – Integrated Method ». <http://austria.omilab.org/psm/content/BPRIM>.
- Tjoa, Simon, Stefan Jakoubi, Gernot Goluch, Gerhard Kitzler, Sigrun Goluch, et Gerald Quirchmayr. 2011. « A formal approach enabling risk-aware business process modeling and simulation ». *IEEE Transactions on Services Computing* 4 (2): 153–166.
- Weiss, Burkhard, et Axel Winkelmann. 2011. « Developing a process-oriented notation for modeling operational risks—a conceptual metamodel approach to operational risk management in knowledge intensive business processes within the financial industry ». In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, 1–10. IEEE.
- Wickboldt, Juliano Araujo, Luís Armando Bianchin, Roben Castagna Lunardi, Lisandro Zambenedetti Granville, Luciano Paschoal Gaspar, et Claudio Bartolini. 2011. « A framework for risk assessment based on analysis of historical information of workflow execution in IT systems ». *Computer Networks* 55 (13): 2954–2975.
- Zur Muehlen, Michael, et Danny Ting-Yi Ho. 2005. « Risk management in the BPM lifecycle ». In *International Conference on Business Process Management*, 454–466. Springer.