



HAL
open science

Actes de la 16ème Conférence SAGEO

Paule-Annick Davoine, Didier Josselin, François Pinet

► **To cite this version:**

Paule-Annick Davoine, Didier Josselin, François Pinet. Actes de la 16ème Conférence SAGEO. Spatial Analysis and GEomatics, May 2021, La Rochelle, France. 2021, 978-2-910545-12-1. hal-03503666

HAL Id: hal-03503666

<https://hal.science/hal-03503666>

Submitted on 28 Dec 2021

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.

Public Domain

16th Spatial Analysis and Geomatics Conference



La Rochelle, 5-7 mai 2021

Actes de la conférence

Avant propos / Foreword

La conférence internationale francophone SAGEO, soutenue par le GDR MAGIS (<http://gdr-magis.imag.fr/>), est un événement annuel majeur dans le paysage de la **Géomatique, de l'Analyse Spatiale et des Sciences de l'Information Géographique**.

Ses objectifs sont de présenter des travaux de recherche récents et de réunir des chercheurs de disciplines diverses (informatique, géographie, archéologie, urbanisme, sciences de l'environnement, agronomie, humanités spatiales, etc.). Il s'agit également de promouvoir les échanges entre les différents acteurs du monde de la géomatique (chercheurs et enseignants-chercheurs, ingénieurs, gestionnaires, aménageurs, entrepreneurs...) qui œuvrent au développement et à la valorisation des recherches au sein et auprès des organismes publics, des entreprises et des collectivités territoriales.

Nous étions heureux à l'idée de vous accueillir à la Rochelle, la grande ville du littoral charentais à l'interface de deux mondes fertiles, la terre et la mer, comme peut l'être la géomatique. Il ne faut en effet rien moins que toute l'interdisciplinarité au cœur de la géomatique pour saisir la complexité d'un tel territoire, très attractif et en pleine croissance, mais également à haut risque en ces temps de changements et transition climatique. Le littoral urbain durable et intelligent que se propose de concevoir l'Université de la Rochelle doit composer avec ces différentes dimensions, géographiques, informatiques mais aussi écologiques, géophysiques, historiques, pour construire un avenir sécurisé et viable aux générations futures. Il s'agit de comprendre, analyser et anticiper la trajectoire d'un socio-éco-système littoral qui doit concilier les enjeux du développement durable, de la gestion du risque et de l'évolution vers les villes et les territoires intelligents à travers des contributions sur les problématiques littorales, urbaines ou environnementales. Ce sont les enjeux au cœur de cette édition 2021.

Les articles soumis ont été évalués à travers 106 rapports d'évaluation réalisés par le Comité de Programme de SAGEO'2021 que nous remercions.

Dans ce contexte, nous avons souhaité proposer divers types de communications qui mettent en relief la richesse et la complémentarité des travaux et des applications.

- **3 conférences invitées** traitant des différentes dimensions de la géomatique :

Michela BERTOLOTTO, *School of Computer Science de l'University College Dublin (Irlande)*

« Evaluation, gestion des risques et connaissances pour la gestion des inondations dans les zones urbaines »

Françoise GOURMELON, *directrice de recherche au CNRS en Géographie de l'Environnement, Plouzané (France)*

« Ouverture des données géographiques : nouveaux contextes, nouveaux enjeux de recherche »

Antoine DOUCET, professeur des universités au laboratoire L3i de La Rochelle Université (France)

« Extraction cross-lingue d'informations géographiques depuis des corpus numérisés, notamment historiques »

- **19 présentations** qui répondent pour moitié aux enjeux du *Littoral urbain durable et intelligent* à travers les trois premières sessions (analyse d'image pour l'observation et le suivi du littoral, transition environnementale et énergétique, et étude de l'environnement urbain),

et pour moitié à des questions fondamentales concernant la mobilité, la modélisation spatio-temporelle, et l'exploitation du Web de données.

- **25 posters**, mis en valeur à travers une galerie en ligne et la possibilité de voter pour le meilleur poster durant le temps de la conférence.

Par ailleurs, trois ateliers ont été sélectionnés afin d'animer les différentes communautés autour de la géomatique. Nous remercions les responsables des ateliers qui participent activement au foisonnement scientifique de ce domaine pluridisciplinaire.

Une session organisée avec l'AFIGEO montre que les rencontres et les échanges entre les laboratoires de recherche spécialisés en Géomatique et les autres acteurs de la géomatique (les entreprises, les collectivités territoriales) sont fertiles pour tous. Cette dynamique va se poursuivre, à travers notamment la participation du GdR CNRS MAGIS aux GeoDataDays qui se tiendront à Grenoble les 15 et 16 septembre 2021.

Enfin, nous remercions très chaleureusement les sponsors (GDR CNRS MAGIS, CNRS, l'Université de la Rochelle, la Communauté d'Agglomération de la Rochelle, l'IUT de la Rochelle), et Nathalie Renaudin (Blanchard) qui a piloté de multiples aspects pratiques de SAGEO'2021, Isabelle Héraudeau pour le suivi financier et l'ensemble des membres du comité d'organisation issus de deux unités de recherche (LIENSs et L3I) pour le travail réalisé avec entrain, enthousiasme et professionnalisme. Un grand merci également à l'accompagnement du comité scientifique, et à Nathalie Brachet pour la constitution des actes.

Pour finir, cette édition 2021 qui est entièrement virtualisée du fait de la crise sanitaire COVID 19 a largement mobilisé la communauté, inscrite en masse sur les différentes sessions proposées. Nous espérons avoir l'occasion ultérieurement d'accueillir SAGEO et les membres de la communauté en « vrai » dans cette belle ville de la Rochelle autour de fruits de mer et autres délicieux mets que nous offre ce beau littoral.

Le Comité d'organisation de SAGEO'2021

Comité scientifique / Program Committee

Chairs	Didier JOSSELIN	UMR ESPACE 7300 CNRS, Université d'Avignon
	Paule-Annick DAVOINE	Université Grenoble Alpes
	François PINET	INRAE
Members	Jean-Michel FOLLIN	Laboratoire GeF - Le Cnam
	Cyril RAY.....	Arts et Metiers Institute of Technology, École Navale, IRENav
	Eric DESJARDIN	URCA/CRéSTIC
	Cyril DE RUNZ.....	LIFAT, Université de Tours
	Anne RUAS.....	IFSTTAR
	Giovanni FUSCO	CNRS - UMR 7300 ESPACE
	Sébastien REY-COYREHOURCQ.....	UMR IDEES
	Marion LE TEXIER	Université de Rouen Normandie
	Mathieu ROCHE	Cirad, TETIS
	Gilles GESQUIERE.....	LIRIS
	Thérèse LIBOUREL	Université Montpellier - UMR Espace-Dev
	Frédéric FLOUVAT	University of New Caledonia
	Isabelle MOUGENOT.....	Université Montpellier EspaceDev
	Thomas LOUAIL.....	CNRS
	Pierre DIAS.....	Université Picardie Jules Verne
	Guillaume TOUYA	IGN
	Laurent ETIENNE	LABISEN
	Carmen BRANDO.....	École des hautes études en sciences sociales (EHESS)
	Didier LEIBOVICI	The University of Sheffield
	Jean-François GIRRES.....	UMR GRED - Université Paul Valéry Montpellier 3
	Jean-Pierre CHANET	INRAE
	Jérôme GENSEL	Laboratoire d'Informatique de Grenoble
	Ali HASSAN	Umanis
	André MIRALLES	Irstea, UMR TETIS
	Cécile DUCHENE	IGN-France, COGIT Laboratory
	Paule-Annick DAVOINE.....	Université Grenoble Alpes
	Alain BOUJU	La Rochelle University
	Thomas DEVOGELE	université François Rabelais (Tours)
	Eric GUILBERT	Laval University
	Sidonie CHRISTOPHE.....	Paris-Est University, IGN-ENSG, LaSTIG
	David SHEEREN.....	UMR DYNAFOR
	Christophe CLARAMUNT	Naval Academy Research Institute
	Florence LE BER	icube
	Lucile SAUTOT.....	AgroParisTech (Dép. SIAFEE, UMR TETIS)
	Roland BILLEN	University of Liege
	Gaëtan PALKA.....	UMR 6554 LETG Rennes
	Helene MATHIAN.....	CNRS- UMR 5600 EVS
	Marie-Hélène DE SÈDE-MARCEAU.....	laboratoire ThéMA
	Sandro BIMONTE	INRAE
	Géraldine DEL MONDO.....	LITIS
	Christian SALLABERRY	University of Pau et Pays de l'Adour
	Khaled REZEG.....	Mohamed Khider BISKRA UNIVERSITY
	Florence SEDES	I.R.I.T. Univ. P. Sabatier
	Michel MAINGUENAUD	INSA-Rouen
	Matthieu NOUCHER.....	CNRS UMR 5319 Passages
	Mireille BATTON-HUBERT	EMSE
	Aldo NAPOLI.....	MINES ParisTech - CRC
	Rodolphe DEVILLERS.....	Memorial University of Newfoundland
	Nathalie LONG	LIENSs , CNRS-Université La Rochelle
	Mauro GAIO.....	LMAP-UMR5142
	Eric KERGOSIEN.....	Laboratoire GERIICO
	Maguelonne TEISSEIRE.....	INRAE - UMR Tetis
	Ali FRIHIDA	École Nationale d'Ingénieurs de Tunis

Comité d'organisation / Steering committee

Chairs Alain BOUJU EA L3I, *La Rochelle*
Christine PLUMEJEAUD-PERREAU ... UMR 7266 LIENSs, *La Rochelle*

Members Mickaël COUSTATY
Cyril FAUCHER
Jamal MALKI
Damien MONDOU
Muzzamil LUQMAN
Cyrille SUIRE

Table des matières / Contents

Communications par Session

Analyse d'image pour l'observation et le suivi du littoral

Davis SHEEREN, DYNAFOR, Toulouse
[Chat] **Christine PLUMEJEAUD**, LIENSS, La Rochelle

Léa PALY, Véronique BEAUJOUAN, Nathalie CARCAUD and Lucien REY..... 13
[Construction des trajectoires paysagères de zones humides littorales – Apport de la géomatique à différentes échelles](#)

Soumia BENGOUFA, Simona NICULESCU, Mustapha Kamel MIHOUBI, Rabah BELKESSA and Katia ABBAD..... 25
[Cartographie de la vulnérabilité côtière par l'intégration des méthodes géo-spatiales : cas de Mostaganem \(ouest Algérien\)](#)

Approches géomatiques pour la transition

Françoise GOURMELON, LETG, Brest
[Chat] **Christine PLUMEJEAUD**, LIENSS, La Rochelle

Benjamin PILLOT, Nadeem AL-KURDI, Carmen GERVET and Laurent LINGUET 39
[Optical network ring analogy for effective solar PV site selection using GIS and optimization modelling](#)

Antoine MURY, Antoine COLLIN and Dorothée JAMES 51
[Prospective 2050 : étude de cas sur l'évolution du service écosystémique de protection côtière](#)
Mathilde LETARD, Antoine COLLIN, Dorothée JAMES and Antoine MURY..... 63
[Trajectoire socio-écologique d'une commune de la Baie du Mont-Saint-Michel entre 1947 et 2014](#)

Environnement urbain

María Jesús LOBO, LASTIG, Paris
Jacques GAUTIER, LASTIG, Paris
[Chat] **Alain BOUJU**, L3I, La Rochelle

Thomas LE JAN, Myriam SERVIÈRES, Thomas LEDUC and Vincent TOURRE 77
[Isovist computation of outdoor environment with semi-dense line SLAM and monocular camera](#)

Thomas LEDUC, Xenia STAVROPULOS-LAFFAILLE and Ignacio REQUENA-RUIZ 89
[Implementation of a solar model and shadow plotting in the context of a 2D GIS: challenges and applications for the cooling effect of tree-covered based greening solutions in urban public spaces](#)

Joan PEREZ, Giovanni FUSCO and Yukio SADAHIRO 101
[Spatial Evolution of Building Types: Methodological proposals](#)

Alessandro ARALDI, David EMSELLEM, Giovanni FUSCO, Andrea TETTAMANZI and Denis OVERAL 113
[Exploring building typologies through fast iterative Bayesian clustering](#)

Simulation, évaluation en géomatique

Jean-Michel FOLLIN, ESGT, Le Mans

[Chat] **Alain BOUJU**, L3I, La Rochelle

Alice JACQUIER and Carole ADAM..... 127

[Choix de mobilité : une approche SMA](#)

Robin CURA 139

[Pour une évaluation visuelle des modèles de simulation, au-delà de la face validation](#)

Marie HAVRET, Marion LE TEXIER and Sophie de RUFFRAY 151

[Localiser le foncier en 1828 : l'exemple du cadastre ancien de Rouen](#)

Extraction de contenu pour le GeoWeb

Boris MERISKAY, Rennes

[Chat] **Paul-Annick DAVOINE**, PACTE, Université de Grenoble-Alpes

Rodrique KAFANDO, Rémy DECOUPES, Lucile SAUTOT, Maguelonne TEISSEIRE and Christiane WEBER... 165

[Constitution de corpus thématique : Pour un meilleur suivi du territoire de la Métropole de Montpellier Méditerranée](#)

Marie GRADELER and Jean-Pierre CHERY..... 177

[Géoweb de narration et de modélisation pour les territoires littoraux : la plateforme NAMO Géoweb](#)

Jacques FIZE, Lucile SAUTOT, Martin LENTSCHAT, Ludovic JOURNAUX and Mohamed HILAL..... 189

[Grand débat national : une analyse automatique des contributions relatives aux transports](#)

Modélisation spatio-temporelle

Camille BERNARD, LIG, Grenoble

[Chat] **Paul-Annick DAVOINE**, PACTE, Université de Grenoble-Alpes

Kamaldeep Singh OBEROI and Géraldine DEL MONDO 203

[Graph-based Pattern Detection in Spatio-Temporal Phenomena](#)

Ba-Huy TRAN, Alain BOUJU, Christine PLUMEJEAUD-PERREAU and Vincent BRETAGNOLLE 215

[Une approche sémantique pour l'exploitation de données spatio-temporelles avec annotation Application à l'évolution d'assolement](#)

Diego VINASCO-ALVAREZ, John SAMUEL, Sylvie SERVIGNE and Gilles GESQUIÈRE..... 227

[Towards a semantic web representation from a 3D geospatial urban data model](#)

Posters

<i>Heythem ADJEROUD, Adrien LAMMOGLIA and Laurent CHAPELON.....</i>	241
<u>Etude des pratiques de déplacement dans une ville moyenne algérienne : le cas de Mila</u>	
<i>Raffaella BALZARINI, Estelle PLOYON, Mélanie MARCUZZI and Philippe BOURDEAU</i>	244
<u>Observations d'évolutions socio-écosystémiques en haute montagne : l'apport de la géomatique au dispositif Refuges Sentinelles</u>	
<i>Rabia Sarah CHERIGUENE, Anne PUISSANT and Habib MAHI.....</i>	247
<u>Apport des images ALSAT-2A pour la cartographie de l'occupation des sols - le cas de la ville d'Oran</u>	
<i>Antoine COLLIN, Tony DUROZIER, Gwenal HERVOUET and Stéphane COSTA</i>	250
<u>Détecter le processus de dépoldérisation par l'observation satellite à très haute résolution temporelle</u>	
<i>Matthieu DUBUIS, Rémi RATAJCZAK, Lény GRASSOT, Laure TOUGNE and Béatrice FERVERS</i>	253
<u>Détermination semi-automatique de l'occupation du sol grâce au logiciel GOURAMIC : évaluation des performances sur données historiques et opérateur-dépendance</u>	
<i>Amélie DUQUESNE, Christine PLUMEJEAUD-PERREAU and Jean-Michel CAROZZA</i>	256
<u>Application d'une approche semi- quantitative pour estimer les impacts cumulés des interventions anthropiques sur les systèmes fluviaux : l'exemple de la Charente entre Angoulême et Saintes (Ouest de la France)</u>	
<i>Jacques FIZE, Ludovic MONCLA and Bruno MARTINS.....</i>	259
<u>Résolution de toponymes par apprentissage profond à partir de cooccurrences et de relations spatiales</u>	
<i>Jean-Michel FOLLIN, Rovasoa VONIARINJANAHARY, Nathalie THOMMERET and Marie FOURNIER</i>	262
<u>FarmSIG : une application web pour l'analyse de l'évolution des espaces agricoles en lien avec les politiques publiques environnementales</u>	
<i>Dorothee JAMES, Antoine COLLIN and Mathilde LETARD</i>	265
<u>Étude spatio-temporelle d'un écosystème littoral : le pré-salé de Sainte-Anne en baie du Mont-Saint-Michel</u>	
<i>Yousra HAMROUNI, Olivia BERNARDOFF, Nicolas KARASIAK and David SHEEREN.....</i>	268
<u>Adaptation de domaine par transport optimal pour la cartographie à large échelle des peupleraies</u>	
<i>Jérémy KALSRON, Jean-Marie FAVREAU and Guillaume TOUYA</i>	271
<u>Le carrefour dont vous êtes le héros : description de carrefours pour les personnes déficientes visuelles</u>	
<i>Cécilia PIGNON-MUSSAUD, Ahmed LAATABI, Marion AMALRIC, Brice ANSELME, Elise BECK, Nicolas MARILLEAU and Nicolas BECU</i>	274
<u>Intégrer des données géospatiales dans un modèle multi-agents : cas de LittoSIM-GEN</u>	
<i>Louisa LE ROUZIC, Bruno GERVAIS, Jean-François BREILH and Axel CREACH</i>	277
<u>Mesure de la vulnérabilité des constructions résidentielles face au risque de submersion marine : proposition pour un outil opérationnel</u>	
<i>Arnaud MÉNARD, Vincent LAMBERT DE CURSAY, Cyril RAY, Charles GUENOIS, Maryam MASLEK ELAYAM and Maxime DRÉAU.....</i>	280
<u>Construction d'une méta-base de ports à l'échelle mondiale</u>	

<i>Jessica PIC, Etienne COSSART and Mathieu FRESSARD</i>	283
<u>Implications des infrastructures de drainage des versants dans la dynamique de transferts de sédiments et de produits phytosanitaires, le cas du Beaujolais viticole</u>	
<i>Marie RAMEL-DELOBEL, Thomas COUDON, Lény GRASSOT, Delphine PRAUD, Béatrice FERVERS and Pietro SALIZZONI</i>	286
<u>Exposition à la pollution atmosphérique et risque de cancer du sein : vers une prise en compte des mobilités quotidiennes dans les études épidémiologiques</u>	
<i>Samuel ROBERT, Brieuc CABIOCH, Eric CARROLL and Marie-Laure TRÉMÉLO</i>	289
<u>Vers une détermination automatique du nombre d'usagers sur une plage urbaine</u> <u>De l'observation au modèle à base d'intelligence artificielle pour la plage du Prophète à Marseille</u>	
<i>Nicolas ROBINET, Gaëtan JOLLY, Guy LEMPÉRIÈRE and Yves PETIT-BERGHEM</i>	292
<u>Apport de l'utilisation du drone et de transects écopaysagers pour l'échantillonnage spatial d'habitats : le cas des digues de la vallée de la Dives (Calvados)</u>	
<i>Loic SALMON, Nicolas MASLOV and Christophe CLARAMUNT</i>	295
<u>Processus de modélisation et optimisation pour la collecte de déchets plastiques en mer</u>	
<i>Magalie TÉCHER, Hassan AIT HADDOU and Rahim AGUEJDAD</i>	298
<u>Effets de la planification urbaine sur les températures de surface du sol (Montpellier, France)</u>	
<i>Nathalie THOMMERET, Delphine GRANCHER, Aurélien DOUSSERON, Daniel BRUNSTEIN and Clément VIRMOUX</i>	301
<u>L'état des piscines, construction d'un indicateur de la reprise du tourisme en situation post-catastrophe. L'île de Saint-Martin après le passage du cyclone Irma (sept 2017)</u>	
<i>Thi Huyen TRAN, Marion AMALRIC, Frédéric ROUSSEAU and Benoit GAUDOU</i>	304
<u>Sensibilisation des gestionnaires à la prévention des risques liés à l'eau : modélisation, simulation et participation dans le Delta du Mékong</u>	
<i>Quy Thy TRUONG, Anne RUAS and Serge LHOMME</i>	307
<u>Modélisation spatiale et positionnement de zones de dépôt pour la collecte de déchets post-ouragans</u>	
<i>Véronique VENTURINI and Caroline TAFANI</i>	310
<u>La planification littorale spatialisée de Balagne à l'épreuve de son opérationnalité : renouveler les méthodes d'identification des espaces vides</u>	
<i>Narimane ZAABAR, Simona NICULESCU, Mustapha Kamel MIHOUBI and Soumia BENGOUFA</i>	313
<u>Méthodes de Machine Learning pour l'analyse de l'occupation du sol des zones exposées au risque d'inondation : la ville côtière d'Ain Témouchent (Ouest Algérien)</u>	

Les articles et les posters ont été évalués par deux relecteurs du comité scientifique de SAGEO'2021.

Communications par Session

Analyse d'image pour l'observation et le suivi du littoral



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Construction des trajectoires paysagères de zones humides littorales – Apport de la géomatique à différentes échelles

Léa Paly¹, Véronique Beaujouan², Nathalie Carcaud³, Lucien Rey⁴

1. ESO – UMR 6590 CNRS, BAGAP, Institut Agro, Agrocampus-Ouest
2 rue André Le Nôtre, 491000, Angers, France
lea.paly@agrocampus-ouest.fr

2. UMR BAGAP, Institut Agro – Agrocampus-Ouest, centre d'Angers

3. ESO – UMR 6590 CNRS, Institut Agro, Agrocampus-Ouest, centre d'Angers

4. Master 2 GAED, mention Aménagement et Développement des Espaces Insulaires, Université des Antilles

RESUME. Les espaces littoraux sont l'objet de nombreux enjeux. Avec le changement climatique, les collectivités se doivent d'entreprendre des approches prospectives du développement territorial. Celles-ci doivent se nourrir de l'étude des territoires, de leurs trajectoires et des temporalités de changements. L'objectif de cette recherche est de construire les trajectoires paysagères du littoral Atlantique à l'aide d'analyses multivariées et de classifications. Les trajectoires à l'échelle régionale sont confrontées aux trajectoires à l'échelle locale, approfondies par l'analyse des changements d'occupation du sol. Cette comparaison a pour objectif de montrer la complémentarité d'analyse des trajectoires de zones humides littorales apportée par l'emboîtement des échelles.

ABSTRACT. Currently, coastal areas face many challenges. With climate change, local authorities have a duty to undertake forward-looking approaches of territorial development. This must be based on the study of territories, their trajectories, and temporalities of change. The purpose of this research is to construct landscape trajectories of the Atlantic coast using multivariate analysis and classification. Trajectories at the regional scale are compared with trajectories at local scale, deepened by analysis of changes in land use and land cover. The aim of this comparison is to show the complementary analysis provided by the nesting of scales for the study of coastal wetland trajectories.

MOTS-CLES : littoral, zone-humide, paysage, trajectoires, analyses multivariées, classification, transition, occupation du sol.

KEYWORDS: littoral, wetlands, landscape, trajectories, multivariate analysis, classification, transition, land-use / land-cover

1. Introduction

Le littoral, défini comme la zone d'interface entre l'espace continental et l'étendue maritime (Lévy et Lussault, 2013), est un espace flou et évolutif dans le temps court au gré des marées et dans le temps long notamment avec les variations du niveau de la mer (Lageat, 2019). Les espaces littoraux actuels résultent de la dernière transgression nommée transgression Flandrienne ayant eu lieu lors de l'Holocène (Verger, 2009).

Aujourd'hui il est communément accepté par la sphère scientifique de l'entrée du monde dans une nouvelle ère géologique, l'Anthropocène, dont le forçage anthropique confère sa particularité (Crutzen, 2002). La date de début fait encore débat chez les scientifiques (années 1950, première révolution industrielle, Néolithique, etc.) (Le Gall et al., 2017) et varie surtout localement (Garnier, 2019). C'est pourquoi les espaces littoraux, notamment caractérisés par un changement de nature et d'intensité des pressions anthropiques, se doivent d'être étudiés au prisme des projets de territoire dont ils ont fait l'objet. En effet, ces espaces anciennement surnommés « territoires du vide » (Corlay, 1995), ont longtemps inspiré crainte et rejet. Cette dévaluation, plus prononcée sur les zones humides littorales, a longtemps justifié des aménagements conséquents (Sajaloli, 1996). A l'issue d'une période hygiéniste, les sociétés se sont finalement rapprochées du rivage (Corbin, 1988). L'économie littorale s'est progressivement tournée vers le développement touristique et portuaire (Miossec, 2001).

Les interactions entre la Société et la Nature sont particulièrement bien illustrées sur les zones humides littorales. Ces espaces ont notamment subi depuis le Moyen-Age de nombreux aménagements (drainage, endiguement, poldérisation, etc.) à des fins d'usages anthropiques (agriculture, pisciculture, urbanisation...). Depuis les années 1960-1970, de nouveaux savoirs et regards ont conduit à la protection de ces zones humides (Poncet et al., 1979 ; Goeldner-Gianella et al., 2011). Les trajectoires paysagères de ces espaces sont donc rythmées entre faits de Nature et faits de Société.

Aujourd'hui le dérèglement climatique et l'élévation du niveau la mer, estimée à 0,4m – 0,6m d'ici 2100 (Horton et al., 2014), constituent un réel enjeu pour les zones humides littorales et leur paysage, dont les caractéristiques topographiques sont proches du niveau marin. Face à cet enjeu majeur, le développement territorial se doit d'avoir une vision prospective de ces espaces. De nombreux acteurs, dont le Conservatoire du littoral, se sont saisi de ce sujet avec par exemple le programme ADAPTO qui préconise une gestion douce du trait de côte et au sein duquel les restitutions de terres prises à la mer sont envisagées comme scénario d'évolution future¹. La connaissance de l'évolution passée de chaque espace est de ce fait essentielle pour le développement d'une approche prospective, d'autant plus que chaque territoire possède ses particularités.

¹ Baie de Lancieux - adapto, un projet LIFE : <https://www.lifeadapto.eu/baie-de-lancieux.html> - ©2020 Le Conservatoire du littoral

Trajectoires paysagères littorales et géomatique

Cette recherche cherche à construire les trajectoires paysagères² objectives de huit zones humides littorales de 1977 à 2000 à l'aide de l'outil cartographique et de données en libre accès. L'objectif est de pouvoir identifier les rythmes de changement de ces zones humides en partant de l'hypothèse que chacune de ces zones humides a eu des formes de rupture de sa trajectoire paysagère homogènes mais à des temporalités et des intensités différentes et que ces trajectoires homogènes à l'échelle régionale se distinguent à l'échelle locale. Une analyse sur les changements paysagers à l'échelle régionale viendra éclairer les grandes tendances d'évolution de ces espaces des années 1970 à aujourd'hui. Ces évolutions seront ensuite explorées à l'échelle locale par la détermination des trajectoires paysagères du site de Corsept (Loire-Atlantique).

2. Matériel et Méthode

2.1. Contexte de l'étude

Les zones humides analysées pour cette étude font partie d'un ensemble de 8 sites pilotes répartis sur la côte Atlantique (FIG 1.), étudiés dans le cadre d'un travail de doctorat. Ils ont pour point commun la présence d'un patrimoine naturel protégé actuellement qui est hérité d'aménagements anciens.

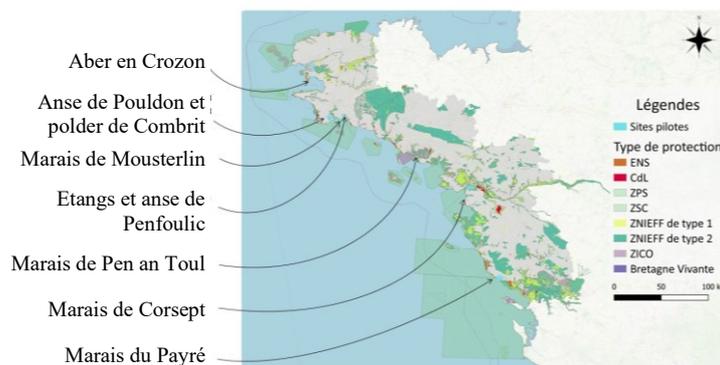


FIGURE 1. Situation géographique des huit sites pilotes étudiés

La construction des trajectoires paysagères des sites retenus repose sur la mobilisation de trois approches (FIG 2.). L'approche historique à l'aide d'archives et notamment de cartes et plans anciens apporte une compréhension de l'état passé

² La définition du paysage ici reprend celle de la Convention Européenne du Paysage. Celui-ci est défini comme « la partie du territoire telle que perçue par les populations et dont le caractère résulte de la combinaison de facteurs naturels et/ou humains et de leurs interrelations » (art.1a, convention européenne du paysage, 2000).

d'un site et, couplée à l'approche géoenvironnementale (décrite ci-dessous), permet de construire les trajectoires paysagères. Enfin l'approche émiqque qui se base sur le discours et les représentations des acteurs sociaux (De Sardan, 1998) utilise les entretiens pour aborder les représentations paysagères qu'ont les usagers des sites. Les rythmes de changements de ces paysages sont ainsi identifiés en partant de l'hypothèse que les moments de rupture sont différents notamment dans leur temporalité.

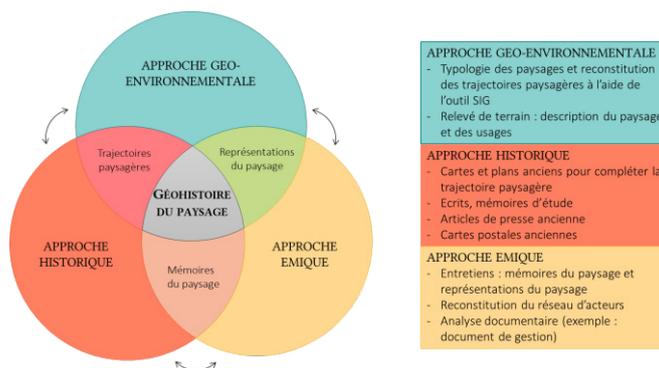


FIGURE 2. Schéma des trois approches utilisées

Cette étude présente les premiers résultats issus de l'approche géo-environnementale qui a pour objectif de mettre en lumière les grandes classes d'évolution à l'échelle régionale depuis les années 1970 et de les confronter aux spécificités locales actuelles des sites. Avec le développement numérique, les Systèmes d'Information Géographiques deviennent des outils incontournables pour l'étude du paysage (Brossard & Wieber, 1984). A partir de données d'occupation du sol et l'utilisation d'indicateurs paysagers, de nombreuses études font usage d'analyses statistiques multivariées aboutissant à des classifications de paysages et d'évolution (Guneroglu et al., 2015 ; Weng, 2007) ainsi qu'à la caractérisation de la succession de ces paysages le long de gradients (Modica et al., 2012; Vizzari et al., 2018; Vizzari & Sigura, 2015). Les matrices de transition ont aussi démontré leur intérêt vis-à-vis d'études spatio-temporelles pour décrire et analyser les changements surfaciques de la végétation (Delbosc et al., 2018; Dez et al., 2017) ou d'occupation du sol (Frondoni et al., 2011; Baud et al., 2015). Elles apportent en particulier une représentation quantitative et qualitative des trajectoires identifiées.

2.2. Analyse cartographique diachronique

Cette recherche s'appuie sur des classifications des types d'évolution du paysage dans le temps à l'échelle régionale et à l'échelle locale ; ainsi que sur l'utilisation de matrices de transition. Chaque analyse réalisée à l'échelle locale est étudiée par le

Trajectoires paysagères littorales et géomatique

site des marais de Corsept. La connaissance historique et locale de ce site permet d'avoir un recul critique sur la détermination des trajectoires paysagères.

2.2.1. Homogénéisation et traitement des bases de données spatiales

Cette étude mobilise trois couches d'occupation du sol en accès libre datant de 1977, 2000³ et 2013. Leur comparaison, en raison de nomenclatures thématiques différentes, a nécessité la création d'une nomenclature commune aux trois couches.

La couche d'occupation du sol IPLI de 1977 constitue le premier inventaire de l'occupation du sol du littoral sur l'ensemble du territoire national à l'échelle 1/25000^{ème}. Sa mise à jour, LITTOMOS, dans les années 2000 à 2006 reprend la même échelle dans le but de faciliter les analyses temporelles. Ces deux sources de données ont été élaborées par la photo-interprétation de photographies aériennes sur une bande d'environ 7km sur l'ensemble du littoral métropolitain. Enfin, l'occupation du sol à grande échelle produite en 2013 pour la Loire-Atlantique, la Vendée et le Morbihan au 1/5000^{ème} a été exploitée. Les descriptions de la couverture du sol et de l'usage du sol ont été couplées pour obtenir une couche d'occupation du sol (OCS) comparable aux précédentes sources pour ces trois départements. L'utilisation de la couche OSO créée par Copernicus, a permis de discriminer les prairies et cultures. Les couches IPLI et LITTOMOS présentent des disparités régionales en termes de nomenclature avec le regroupement pour IPLI des prairies et cultures sous la catégorie terres cultivées pour le Finistère (Sparfel, 2011). Ainsi la nomenclature de correspondance employée possède deux niveaux de détails. L'un, le plus large, sera adopté à l'échelle régionale, et le second, plus détaillé par suite d'une rectification des couches IPLI et LITTOMOS par photo-interprétation, sera utilisé à l'échelle locale. Cette nomenclature (TABLE 1.), gage de la correspondance entre les trois couches, comprend au maximum 9 postes décrivant les espaces urbanisés, les espaces naturels, les milieux humides, agricoles et forestiers.

TABLE 1. Nomenclature des différents postes d'occupation du sol aux trois dates

ANALYSE REGIONALE		ANALYSE LOCALE	
11	Zone urbanisées et espaces associés	11	Zone urbanisées et espaces associés
62	Milieux ouverts agricoles	20	Cultures
		23	Prairies
71	Milieux semi-ouverts	71	Milieux semi-ouverts
31	Milieux boisées	31	Milieux boisés
60	Milieux ouverts minéraux	61	Milieux ouverts
63	Milieux ouverts herbacées		
64	Milieux ouverts humides	40	Milieux humides
		42	Production maritime
50	Surface en eau	50	Surface en eau

³ La couche LITTOMOS a été réalisée en 2000 pour le Finistère et Morbihan, 2004 pour la Loire-Atlantique et 2006 pour la Vendée.

2.2.2. Calcul des indicateurs paysagers

La construction des trajectoires paysagères s'est établie à travers trois analyses opérant à différentes échelles spatiales. Pour chaque analyse, l'outil statistique R, avec les packages FactoMineR pour les ACPs et networkD3 pour les diagrammes de Sankey, a été utilisé ainsi que les logiciels de cartographie QGis et ArcGis. Chaque carte d'occupation du sol a été rastérisée à une résolution de 12,5m⁴ à l'échelle régionale puis analysée avec le logiciel Chloé (Boussard et Baudry, 2017) pour en extraire des indicateurs paysagers (TABLE 2.).

Des indicateurs de structure ont été calculés à l'échelle de mailles de 700m à l'échelle régionale et de 100m à l'échelle locale, tels que la pente moyenne et son écart-type, l'altitude moyenne et son écart-type, l'indice de position topographique (étendue), la distance minimale de la maille au trait de côte, la distance minimale de la maille au cours d'eau et la proportion de la maille depuis laquelle on voit le trait de côte.

Ensuite des indicateurs de composition (taille moyenne à l'échelle des taches et du paysage et leur écart-type, pourcentage des différents types d'OCS, nombre de taches en eau, indice de diversité de Shannon, pourcentage de couples de pixels en « milieux humides ») et de forme (longueur d'interface entre différents type d'OCS, indice d'agrégation des taches) ont été calculés avec le plugin Chloé sur des fenêtres d'analyse de 2500m à l'échelle régionale et de 1000m à l'échelle locale.

2.2.3. Analyse régionale : classification de l'évolution des paysages

La première analyse s'est appuyée sur la différence des valeurs des différents indicateurs par maille entre deux dates pour étudier l'évolution du paysage. La réalisation d'une analyse en composante principale (ACP) suivie d'une classification ascendante hiérarchique (CAH) a permis d'obtenir une classification des dynamiques paysagères observées à l'échelle régionale entre 1977 et 2000, 2000 et 2013. La classification utilisée, qualifiée de bottom-up, est choisie pour son opérationnalité. Une classification top-down, où les classes de paysages sont prédéterminées, aurait pu être utilisée mais une connaissance préalable des classes de paysages littoraux aurait été nécessaire ; de plus, la différence des résultats n'est pas conséquente (Van der Zanden et al., 2016).

2.2.4. Etude des changements d'occupation du sol à l'échelle des sites : cas de Corsept

Cette analyse vient alimenter les observations des changements de paysage par la détermination des changements d'occupation du sol (OCS) à l'échelle locale. Les OCS aux trois dates ont été découpées selon le périmètre du site de Corsept. Celui-ci a été délimité selon la géomorphologie du site et le zonage de l'espace reconnu pour son caractère naturel, attribué au périmètre des ZNIEFFs 1 et 2 (Zone Naturelle

⁴ Les cartes d'occupation du sol ont été rastérisées à 12.5m dans le but de garder le détail des éléments de faibles surfaces et d'optimiser les temps de calculs.

Trajectoires paysagères littorales et géomatique

d'Intérêt Faunistique et Floristique) de ce site. Après agrégation de ces trois couches, la nature des transitions entre deux dates et pour chacun des postes d'OCS est connue et exprimée en pourcentage absolu de surface passant d'un poste à un autre. La matrice de transition ainsi constituée est ensuite modélisée par un diagramme de Sankey. Ce même procédé a été répété sur un périmètre élargi par des zones tampon de 250m, 500m et 1000m afin de comparer les changements d'occupation du sol au voisinage proche de la zone humide

2.2.5. Analyse locale : évolution des paysages à l'échelle des sites : cas de Corsept

La dernière analyse consiste en l'examen de l'évolution du paysage à l'échelle des sites et est étudiée par le cas des marais de Corsept. Cette méthode permet de visualiser les trajectoires des sites sur le graphique des individus de l'ACP en isolant les points composant un site en 1977, 2000 et 2013. Cette étude a été entreprise afin de comprendre la trajectoire spécifique de ce site par analogie aux grandes classes d'évolution caractérisée à l'échelle régionale. Une ACP, dite historique, a été réalisée sur un tableau présentant en individus les mailles de la grille correspondant aux sites pour chacune des trois dates. L'ensemble de ces individus a été caractérisé par les mêmes indicateurs paysagers qu'à l'échelle régionale.

3. Premiers résultats

Ces résultats sont actuellement en cours de construction et d'interprétation. D'ici mai 2021, l'ensemble des analyses aura été effectuée et approfondie.

3.1. Typologie de l'évolution des paysages à l'échelle régionale

La première analyse réalisée sur une ancienne nomenclature entre 2000 et 2013 a identifié cinq grands types d'évolution des paysages à l'échelle régionale (FIG 3.).

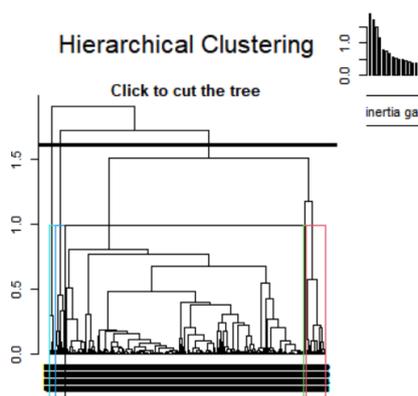


FIGURE 3. Dendrogramme issu de la CAH réalisée sur l'ACP concernant les évolutions entre 2000 et 2013, 5 classes sont retenues

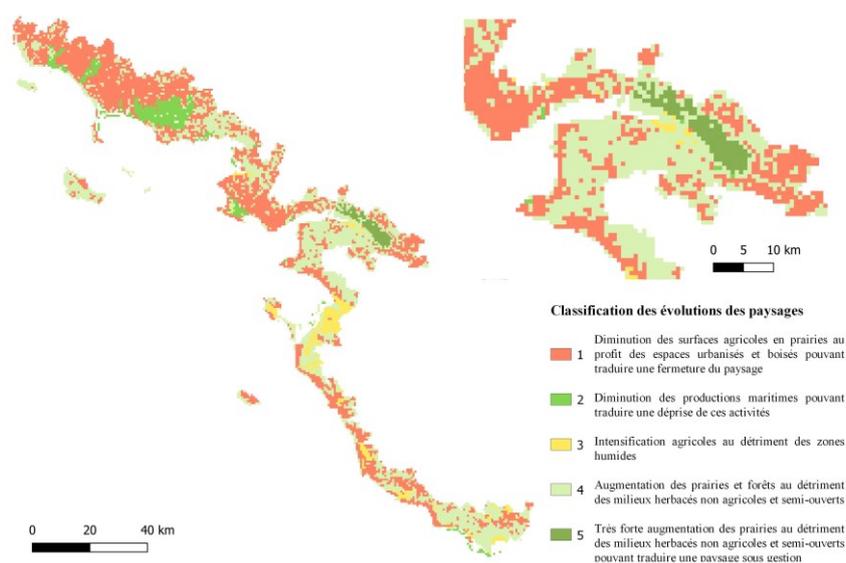


FIGURE 4. Typologie des évolutions des paysages entre 2000 et 2013 pour les quatre départements et le site de Corsept (Loire-Atlantique)

La classification fait ressortir une évolution propre à l'estuaire de la Loire avec la classe 5 entre 2000 et 2013 (FIG 4.). La première lecture des ACPs montre des processus généraux tels que l'intensification agricole et l'urbanisation mais aussi le verdissement des espaces qui deviennent plus « naturels » pouvant révéler une volonté politique de renaturation ou bien une déprise ou spécialisation agricole. Il est difficile à ce jour de plus s'avancer sur les résultats, l'interprétation étant encore en cours. Ces premières analyses, d'ici Mai, seront approfondies notamment dans le choix d'indicateurs paysagers spécifiques à l'évolution de ces espaces grâce à l'élaboration des matrices de transition à l'échelle régionale.

3.2. Matrices de transition de l'occupation du sol de Corsept

Les représentations des matrices de transition montrent que certains types d'occupation du sol ont plus varié que d'autres entre 1977 et 2013 (FIG 5, a. et b.). Les forêts ont augmenté entre 1977 et 2000 avant de diminuer en 2013 au profit des zones urbanisées. Les espaces naturels ont augmenté. Les zones en eau sont restées stables avec néanmoins une proportion plus importante en 1977. Les espaces urbanisés se sont légèrement accrus à chaque année (FIG 5. a.), et ont doublé entre 1977 et 2000 sur le périmètre élargi par la zone-tampon (FIG 5. b.) Les espaces de culture, en prairie et les zones humides représentent les plus grandes variations. Sur

Trajectoires paysagères littorales et géomatique

le périmètre de la zone humide, les prairies ont conséquemment augmenté entre 1977 et 2000 et de façon plus minime de 2000 à 2013 par transformation des zones humides principalement entre 1977 et 2000 et des cultures (1977 à 2013). Après les années 2000 la surface des zones humides reste stable. La même tendance d'évolution des prairies ressort sur le périmètre élargi par la zone tampon avec néanmoins un pourcentage de prairies à chaque année moins important, contrebalancé par une part de culture plus grande.

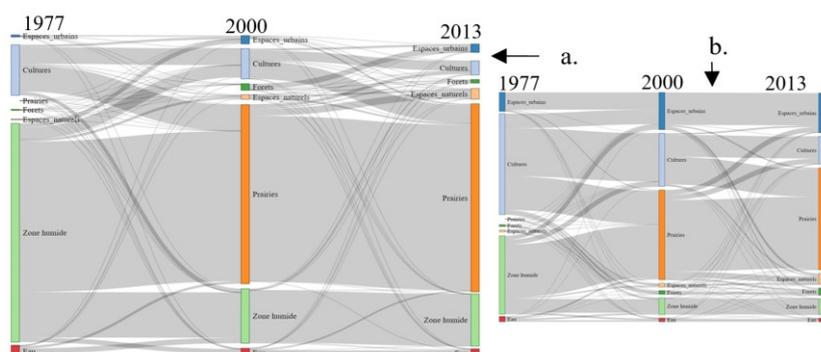


FIGURE 5. Transition entre les postes d'OCS entre 1977, 2000 et 2013 pour (a.) le périmètre du marais du Corsept, (b) étendu avec une zone tampon de 500m

3.3. Evolution des paysages à l'échelle du site de Corsept (Loire-Atlantique)

L'ACP historique apportera une lecture « paysagère » de ces évolutions entre 1977, 2000 et 2013 à l'aide des différents indicateurs décrits pour l'analyse régionale et adaptés aux spécificités locales du site. Celle-ci est en cours de réalisation.

4. Discussion

Les premiers résultats de ces analyses (en cours de réalisation et d'interprétation) soulèvent plusieurs questionnements au niveau de l'adéquation de la méthode. Une réduction des indicateurs utilisés est nécessaire afin de sélectionner les indicateurs discriminant les évolutions sur les espaces littoraux. L'usage de couches d'occupation du sol élaborées par différents acteurs et en libre accès apporte la simplicité d'accès mais demeure contraignant quant à leur mise en comparaison. Les méthodes de classification automatiques sont intéressantes, notamment sur de grands jeux de données, cependant leur utilisation doit être précautionneuse. La connaissance des variations des indicateurs utilisés est nécessaire pour mener cette analyse de façon optimale. De plus, l'interprétation doit prendre en compte l'imperfection des tables de correspondance, particulièrement pour les matrices de transition. La figure 5 (a.) a témoigné sur le site de Corsept d'un apport des zones

humides à la catégorie prairie de 1977 à 2000 avant de se stabiliser. Il est important de prendre en compte le fait que la majeure partie du site en 1977 était composée de prairies humides poste d'OCS qui a été intégré à la catégorie zone humide. Ainsi l'augmentation des prairies entre 1977 et 2000 est à relativiser et peut finalement caractériser l'assèchement des prairies humides présentes en 1977. Cette période correspond à la construction d'une digue et d'une douve collectrice des eaux des étiers en 1988-89 altérant alors les échanges entre la zone soumise aux marées et ces marais estuariens. Enfin, la connaissance du terrain est indispensable pour ce genre d'analyse. La transition des cultures vers les prairies peut traduire un phénomène de déprise ou de spécialisation agricole mais aussi un phénomène de patrimonialisation. En effet, la mise en place de mesures de gestion sur les zones humides implique souvent le maintien d'un paysage ouvert sous forme de prairies de fauche ou de pâturage. Ici, bien que le processus de patrimonialisation de cette zone humide et plus généralement de l'estuaire de la Loire a débuté avant les années 2000, les conséquences de cette patrimonialisation sur le site de Corsept sous forme de pratiques de gestion sont bien plus tardives. Ainsi cette évolution pourrait traduire plutôt un phénomène de déprise ou de spécialisation agricole pour ce site. On le comprend, les matrices de transition ont permis de révéler des erreurs de correspondance dans les nomenclatures des couches cartographiques. Malgré ces contraintes, l'utilisation de couches de données en libre accès permet la reproductibilité de ces analyses.

Les premiers résultats suggèrent une trajectoire de la zone humide de Corsept qui s'inscrit dans une trajectoire globale de changements des pratiques culturelles (augmentation des prairies au détriment des cultures, majoritairement) mais qui se distingue à l'échelle de l'espace protégé (peu voire pas d'urbanisation, part de prairies beaucoup plus importante, peu de culture), notamment en lien avec ses spécificités historiques (construction d'une digue, patrimonialisation récente). Les matrices de transition, couplées à une connaissance historique du terrain, offrent une analyse fine des trajectoires de ces espaces. Elles permettent à la fois une analyse quantitative en modélisant les pourcentages absolus des transitions d'OCS, qualitative par la connaissance des successions d'OCS mais aussi spatialisée avec l'étude de périmètres différents.

5. Conclusion

Au vu de ces premiers résultats, on comprend qu'il est important et nécessaire de prendre en compte l'histoire du paysage. Ce recul permet d'avoir un regard critique sur l'observation de phénomènes actuels. Ce regard est d'autant plus important sur l'espace littoral qui a été profondément modifié par l'Homme. Les sites pilotes identifiés peuvent à travers leur typologie présenter un paysage similaire. Or, en se penchant sur leur trajectoire paysagère, à travers l'évolution des paysages à l'échelle du site et son couplage à l'étude des transitions d'usage du sol, on s'aperçoit de la diversité dynamique de ces espaces. Ce travail montre aussi l'intérêt de varier les échelles d'analyse et la complémentarité des échelles régionales et locales pour comprendre les trajectoires paysagères des espaces littoraux. Hier, les grands projets

Trajectoires paysagères littorales et géomatique

de territoire ont souvent voulu appliquer un même projet de développement aux zones humides littorales comme l'ont montré les nombreux manuscrits techniques sur le dessèchement des marais du XIX^{ème} siècle (Dienne, 1891) et début XX^{ème} siècle (Lévy-Salvador, 1900). Aujourd'hui encore, les grands projets de restauration/renaturation des zones humides littorales peuvent conduire à banaliser ces paysages. La connaissance des héritages paysagers par la reconstitution des trajectoires paysagères permet de reconnaître les spécificités de ces territoires.

L'outil cartographique et les méthodes de classification sont un bon moyen de déterminer des classes de trajectoires à des échelles spatiales différentes. Il serait néanmoins intéressant d'utiliser des données plus anciennes, notamment du début du XX^{ème} siècle avant le grand développement touristique des espaces côtiers. Cela permettrait de mieux cerner les moments de rupture liés à une pression anthropique trop importante et de les comparer aux autres sites. Pour les marais de Corse ce moment de rupture est récent et peu conséquent en termes paysagers alors que pour les marais du Payré (85), par exemple, ce moment était déjà initié au XII^{ème} siècle par la création des marais salants et à poissons. L'objectif de cette recherche est de poursuivre ces trajectoires du XVIII^{ème} au XXI^{ème} siècles grâce à la numérisation de cartes anciennes.

Bibliographie

- Boussard H, Baudry J. 2017 Chloé4.0: a software for landscape pattern analysis. See <https://www6.rennes.inra.fr/bagap/PRODUCTIONS/Logiciels>.
- Brossard, T., & Wieber, J.-C. (1984). Le paysage : Trois définitions, un mode d'analyse et de cartographie. *Espace géographique*, vol. 13, n°1, 5-12.
- Corbin A., 1988. *Le territoire du vide. L'occident et le désir de rivage*. Flammarion, Champs histoire, 2018, 399p
- Corlay, J.-P. (1995). Géographie sociale, géographie du littoral. *Noréis*, vol. 165, n°1, 247-265.
- Crutzen, P. J. (2002). Geology of mankind The Anthropocene. *Nature*, vol. 415, n°23, 1.
- Delbosq, P., Tanné, M., Bioret, F., Panaïotis, C., & Sawtschuk, J. (2018). Analyse spatio-temporelle du paysage végétal du massif du Haut-Vénacais (Corse). *Cybergeo*, n°846, 1-30
- De Sardan, J.-P. O. (1998). Émique. *L'Homme*, 38(147), 151-166. <https://doi.org/10.3406/hom.1998.370510>
- Dez, M. L., Sawtschuk, J., & Bioret, F. (2017). Les prairies de l'estuaire de la Loire : Étude de la dynamique de la végétation de 1982 à 2014. *Mappemonde*, n°119.
- Dienne, É. de (Cte). (1891). Histoire du dessèchement des lacs et marais en France avant 1789. *Bibliothèque Nationale de France*.
- Frondoni, R., Mollo, B., & Capotorti, G. (2011). A landscape analysis of land cover change in the Municipality of Rome (Italy): Spatio-temporal characteristics and ecological implications of land cover transitions from 1954 to 2001. *Landscape and Urban Planning*, vol. 100, 117-128.

- Garnier, E. (2019). Trajectoires d'une anthropisation dans la vallée du Doubs, XVIe-XXIe siècles. *Revue scientifique Bourgogne-Franche-Comté Nature*, vol. 30, 116-139.
- Goeldner-Gianella, L., Feiss-Jehel, C., & Decroix, G. (2011). Les oubliées du « désir du rivage » ? L'image des zones humides littorales dans la peinture et la société françaises depuis le XVIIIe siècle. *Cybergeo*.
- Guneroglu, N., Acar, C., Guneroglu, A., Dihkan, M., & Karsli, F. (2015). Coastal land degradation and character assessment of Southern Black Sea landscape. *Ocean & Coastal Management*, vol. 118, 282-289.
- Horton, B. P., Rahmstorf, S., Engelhart, S. E., & Kemp, A. C. (2014). Expert assessment of sea-level rise by AD 2100 and AD 2300. *Quaternary Science Reviews*, vol. 84, 1-6.
- Le Gall, J., Hamant, O., & Bouron, J.-B. (2017). Notion en débat : Anthropocène - Géoconfluences [Document]. *Géoconfluences*. <http://geoconfluences.ens-lyon.fr/informations-scientifiques/a-la-une/notion-a-la-une/anthropocene>
- Lévy-Salvador, P. (1900). Hydraulique agricole—LIVRE III - 4e à 8e PARTIE - Assainissements et dessèchements. Colmatages. Polders. Drainage. Utilisation agricole des eaux d'égout. Annexes. *Vve CH., EDITEUR*, 3(4-8), 574.
- Miossec, A. (2001). L'évolution de la géographie des océans et des littoraux face aux perspectives du développement durable au XXIe siècle. Quelles hypothèses envisager?//The geography of coasts and seas facing sustainable development: What could be the future? *Annales de Géographie*, vol. 110, n°621, 509-526.
- Modica, G., Vizzari, M., Pollino, M., Fichera, C. R., Zoccali, P., & Di Fazio, S. (2012). Spatio-temporal analysis of the urban-rural gradient structure: An application in a Mediterranean mountainous landscape (Serra San Bruno, Italy). *Earth System Dynamics Discussions*, vol. 3, n°2, 827-870.
- Sajaloli, B. (1996). Les zones humides: Une nouvelle vitrine pour l'environnement (Wetlands: a new showcase for environment). *Bulletin de l'Association de géographes français*, vol. 73, n°2, 132-144.
- Sparfel, L. (2011). *Etude des changements d'occupation des sols dans la zone côtière à partir de données hétérogènes: Application au pays de Brest*. [Sciences de l'Homme et Société / Géographie]. Université de Bretagne occidentale - Brest.
- Van der Zanden, E. H., Levers, C., Verburg, P. H., & Kuemmerle, T. (2016). Representing composition, spatial structure and management intensity of European agricultural landscapes: A new typology. *Landscape and Urban Planning*, vol. 150, 36-49.
- Verger F. (2009). Zones humides du littoral français. BELIN, Paris, 447p.
- Vizzari, M., Hilal, M., Sigura, M., Antognelli, S., & Joly, D. (2018). Urban-rural-natural gradient analysis with CORINE data: An application to the metropolitan France. *Landscape and Urban Planning*, vol. 171, 18-29.
- Vizzari, M., & Sigura, M. (2015). Landscape sequences along the urban-rural-natural gradient: A novel geospatial approach for identification and analysis. *Landscape and Urban Planning*, vol. 140, 42-55.
- Weng, Y.-C. (2007). Spatiotemporal changes of landscape pattern in response to urbanization. *Landscape and Urban Planning*, vol. 81, n°4, 341-353.

Cartographie de la vulnérabilité côtière par l'intégration des méthodes géo-spatiales : cas de Mostaganem (ouest Algérien)

Soumia BENGOUFA ^{1,2,3*}, **Simona NICULESCU** ², **Mustapha Kamel MIHOUBI** ², **Rabah BELKESSA** ¹, **Katia ABBAD** ¹

1. Ecole Nationale Supérieure des Sciences de la Mer et de l'Aménagement du Littoral, Dely Ibrahim, Alger, Algérie, bengoufasoumia@gmail.com, belkessa.r.littoral@gmail.com, abbadkatia@gmail.com,

2. UBO, Université de Bretagne Occidentale, Laboratoire LETG Brest UMR 6554 CNRS, Brest, France simona.niculescu@univ-brest.fr

3. Ecole National Supérieure de l'Hydraulique, Mobilisation et Valorisation des Ressources en Eau, NR 29; BP 31 BLIDA, Algérie, 09000 ; k.mihoubi@ensh.dz

RESUME. Dans le contexte de la croissance d'activités socio-économiques littorales et l'accentuation des risques côtiers, ce travail vise à évaluer la vulnérabilité d'un linéaire côtier de 32 km de la ville de Mostaganem (l'Ouest-Algérien). En utilisant des méthodes de télédétection et des Système d'Information Géographique (SIG), cinq variables physiques et socio-économiques, à savoir le taux d'érosion (trait de côte), le type du relief, la pente, la hauteur significative des vagues (Hs) et le type d'occupation du sol, sont traitées. Un modèle de SIG a été créé pour calculer un Indice de Vulnérabilité Côtière (IVC) en utilisant une méthode de pondération. Sur la base des valeurs de l'IVC, la côte est catégorisée en quatre classes de vulnérabilité, à savoir : faible (7,5 km), modérée (4.08 km), élevée (11 km), et très élevée (9,74 km). Cette étude pourrait contribuer à la cartographie de la vulnérabilité de la zone d'étude dans une optique de mitigation et prévention du risque de l'érosion côtière.

ABSTRACT. In the context of the increasing rate of coastal socio-economic activities and the exacerbating of the coastal hazards, this work aims to assess the coastal vulnerability of 32 km of the Mostaganem coastline (Algeria), using remote sensing methods and Geographic Information Systems (GIS). To achieve that, five variables, namely erosion rate (shoreline), relief type, slope, significant wave height (Hs) and land use type, are considered. A GIS model was created to calculate a Coastal Vulnerability Index (CVI) based on weighting method. According to the CVI values, the shoreline is classified into four vulnerability classes, LOW (42.5 km), MODERATE (23.46 km), HIGH (34.61 km), and VERY HIGH (7.5 km). Thus, this study is useful for risk mitigation and prevention of coastal erosion hazard.

MOTS-CLES : Indice de Vulnérabilité, Ouest-algérien, Machine Learning, SIG, Erosion

KEYWORDS: Vulnerability Index, West-Algeria, Machine Learning, GIS, Erosion,

1. Introduction

Les côtes sont des reliefs dynamiques, dont les changements dépendent des facteurs physiques qui les caractérisent (géologie, géomorphologie, etc.) et/ou du forçage de la mer (l'action des vagues, niveau de la mer, etc.). Compte tenu du fait que 21 % de la population mondiale y vit (Brooks et al., 2006), ces côtes sont sujettes à des concentrations importantes et croissantes d'activités et d'enjeux socio-économiques. En effet, le dynamisme naturel, les et les aléas météorologiques extrêmes rendent ces zones côtières particulièrement vulnérables. Ceci a motivé la recherche sur le concept de la vulnérabilité afin de la rendre mesurable. Le but étant de fournir des méthodes permettant aux acteurs engagés sur les questions littorales de traiter les causes et prévenir les conséquences pour une gouvernance efficace. Au vu de la complexité de cette zone, qui se trouve au carrefour de plusieurs facteurs, ces méthodes d'évaluations requièrent une approche systémique. Une des premières tentatives d'évaluations la vulnérabilité côtière par un calcul d'Indice de Vulnérabilité Côtière (IVC) était celle de Gornitz (1991). Suite à son étude, de nombreuses versions d'IVC ont été développées à différentes échelles.

Le sous-groupe de la gestion des zones côtières du Groupe Intergouvernemental d'Experts sur l'Évolution du Climat (GIEC) a encouragé la mise au point d'une Méthodologie Commune (MC) pour évaluer la vulnérabilité des différentes nations côtières. Cependant, les données nécessaires pour évaluer un ou plusieurs facteurs ne sont pas toujours disponibles, en particulier pour les pays en développement (Awange et al., 2018). De plus, la surveillance de l'environnement côtier est souvent une tâche fastidieuse en raison des techniques traditionnelles chronophages qui ne permettent pas de couvrir un système côtier en entier. La télédétection, les approches de la géomatique et des Systèmes d'Information Géographique (SIG) peuvent fournir des alternatives pour cette surveillance. Les images satellitaires fournissent une couverture synaptique et sont utiles pour les applications de surveillance détaillée (Mujabar & Chandrasekar, 2013). Ainsi, leurs propriétés spectrales peuvent être utilisées pour extraire plusieurs caractéristiques notamment celles du relief côtier (Boak & Turner, 2005) et de l'occupation du sol.

Dans cette étude, une version modifiée de l'IVC est proposée en traitant 5 variables propices aux côtes micro-tidales, en se basant sur l'apport de la télédétection et des SIG. L'objectif principal de cette étude est d'appliquer une méthode d'évaluation de la vulnérabilité côtière face au risque d'érosion, qui sera à la fois pratique et scientifiquement fondée sur des sources et de méthodes de traitement de données fiables. Le but est de dériver un outil reproductible d'aide à la planification, la prise de décisions et l'aménagement des villes côtières.

2. Zone d'étude

La zone d'étude est une des trois grandes Wilayas du Nord-ouest algérien. Située à 400 km de la capitale, Mostaganem est une zone touristique par excellence (Taïbi, 2016).

Cartographie de la vulnérabilité côtière

La demande croissante ainsi que le nombre insuffisant d'infrastructures ont conduit les autorités à augmenter le nombre de structures d'accueil. En effet, 16 zones d'extension touristique côtières sont prévues et sont en cours d'aménagement (Ghodbani et al., 2016). Compte tenu des enjeux sociaux, économiques et naturels, des constituent un sujet de préoccupation de plus en plus prégnant pour la Wilaya de Mostaganem. De ce fait, il convient de promouvoir une approche pragmatique afin d'évaluer son état de vulnérabilité actuel, pour un aménagement et une planification spatiale idoïne.

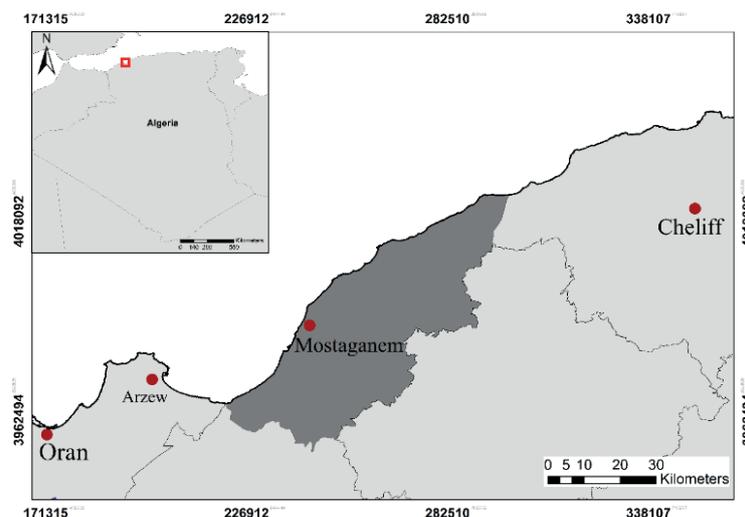


Figure 1. Situation géographique de la zone d'étude

3. Matériel et méthodes

Le présent travail formule un modèle SIG de l'IVC dans un contexte socio-éco-systémique, en prenant en considération les variables propices aux côtes micro-tidales, à l'instar des côtes du bassin ouest méditerranéen. Le modèle développé est composé de cinq variables géophysiques et socio-économiques : a) le taux d'érosion/accrétion (m), b) les formes de relief (géomorphologie), c) pente (%), d) hauteur moyenne significative des vagues (m), e) l'occupation du sol (fig. 2). Ce sont des variables de différentes dimensions. Cette hétérogénéité est éliminée par un système de pondération de 1 à 5 (fig. 2) selon les intervalles définis par McLaughlin & Cooper, (2010).

3.1. Traitement des données

Depuis une synthèse cartographique, jusqu'au recours à la modélisation passant par l'analyse des images satellitaires (fig. 2) divers types de sources de données ont été utilisées afin de traiter les variables retenues pour le modèle IVC :

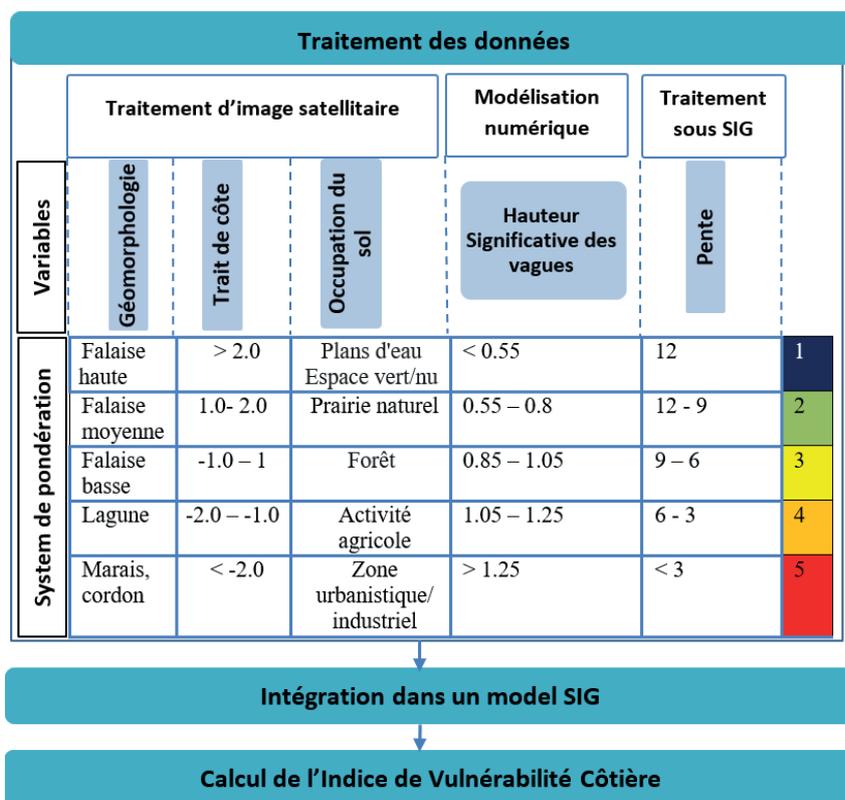


Figure 2. Méthodologie de travail et approche de traitement de donnée

a- La variable : le taux d'érosion/accrétion (Trait de côte)

Une analyse spatio-temporelle de l'évolution du trait de côte (TC) a été élaborée afin de calculer le taux d'érosion, en comparant l'ancienne position du TC issue des photos aériennes de 1992 avec sa position récente extraite automatiquement à partir des images de type Pléiades (0.5 m). L'acquisition de ces images a été programmée en juillet 2019 pour couvrir la zone d'étude le jour même de l'acquisition des données In-Situ par GPS utilisées afin de valider le modèle d'extraction automatique du TC. Ce modèle est basé sur une classification supervisée des images par une approche orientée objet. La méthode d'extraction automatique comprend trois étapes principales : (a) la segmentation de l'image (b) l'entraînement et la classification et (c) Extraction du trait de côte (fig. 3).

L'algorithme de segmentation de type Multirésolution a été appliqué en utilisant le logiciel ECognition. L'algorithme Random Forest (RF) a été utilisé pour la classification de l'image afin de détecter l'indicateur du TC retenu pour l'évaluation du taux érosion, qui est la limite supérieure du sable humide. C'est un algorithme d'apprentissage automatique (Machine Learning) non paramétrique (Breiman, 2001).

Cartographie de la vulnérabilité côtière

Après un processus d'optimisation par validation croisée, les valeurs du nombre d'arbres dans la forêt (Ntree) et le nombre de variables à chaque nœud (Mtry) ont été fixées à 500 et 60 respectivement. La classification a permis de détecter le TC qui a été extrait automatiquement sous SIG.

Une étape de lissage s'est avérée nécessaire afin de le comparer à celui extrait des photos aériennes (fig. 3). Les TC extraits ont été intégrés dans le modèle Digital Shoreline Analysis System (DSAS). Les valeurs de End Point Rate (EPR) ont été calculées sur des transects d'une équidistance de 20 m. Il s'agit d'un taux calculé en divisant la distance entre la position du TC initial (1992) et celle du plus récent (2019) par le temps écoulé (58 ans). L'EPR a l'avantage d'une exigence minimale pour les entrées (deux traits de côtes) (To et Thao, 2008).

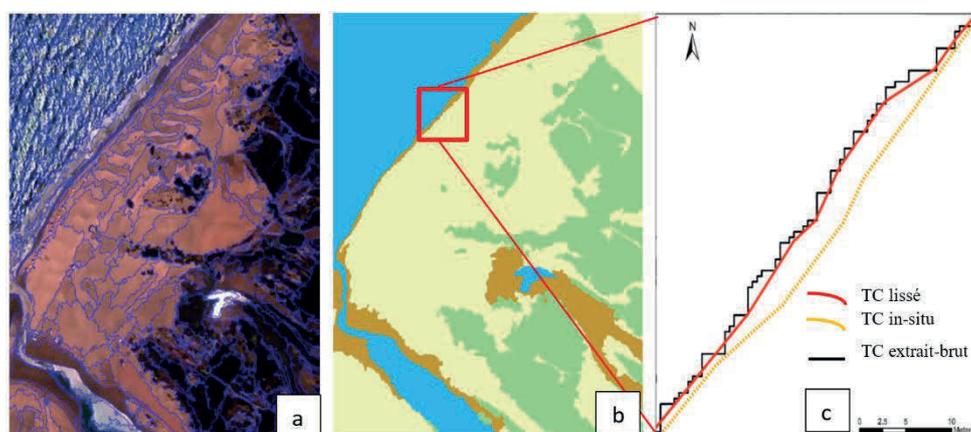


Figure 3. Extraction du trait de côte a) Segmentation, b) Classification, c) Extraction

b- La variable : relief côtier (géomorphologie)

Les données permettant d'analyser le degré de vulnérabilité liée à la géomorphologie de la côte étudiée sont tirées des Modèles Numériques de Terrain (MNT), obtenus via le portail en ligne de l'United States Geological Survey (USGS), ainsi que des images Pléiades. Les MNT ont été fusionnés avec les données issues des cartes topographiques et traitées sous SIG (ArcGis® 10.1). Le traitement des images Pléiades a permis de reconnaître les faciès géomorphologiques du littoral. A noter que les connaissances des traits géomorphologiques du site acquises lors des sorties terrain effectuées, ont facilité l'évaluation.

c- La variable : pente

La pente régionale de la zone côtière a été calculée à partir d'une grille d'élévations topographiques, élaborée par le traitement du MNT et la grille bathymétrique, dont les données numériques sont extraites de la carte marine établie par le Service Hydrographique des Forces Navales algériennes (SHFN). Ces données ont été fusionnées et traitées à l'aide du logiciel SIG (ArcGis® 10.1) afin d'établir une carte de pente.

d- La variable : la hauteur significative des vagues à la côte (Hs)

Basée sur des modèles mathématiques, la modélisation numérique décrit de manière suffisamment fiable les phénomènes physiques. En effet, un modèle numérique du MIKE21-DHI a été utilisé dans le cadre de cette étude afin de calculer les hauteurs significatives des vagues à la côte (Hs). Il s'agit dans un premier temps de recueillir l'ensemble des données nécessaires à la simulation (des données bathymétriques, météorologiques et le niveau de la mer). Les modèles en sorties ont montré que le degré d'exposition augmente lorsque les vagues proviennent du secteur NW (fig. 4), avec un temps de retour de 8s et une Hs au large de 2.5m qui sont les paramètres choisis pour l'évaluation de la sensibilité liée à cette variable.

e- La variable : occupation du sol

Le type d'occupation du sol a été déterminé via un traitement des images Pléiades par une approche de classification orientée objet. Une segmentation par l'algorithme Multiresolution a été effectuée avec le logiciel Ecognition. Une classification supervisée des objets résultants de la segmentation a été réalisée par l'algorithme RF avec une optimisation préalable des hyperparamètres (nombre d'arbres Ntree et profondeur maximale Mtry).

L'optimisation a été effectuée par la méthode de validation croisée afin de fixer les valeurs des paramètres de la segmentation Multirésolution et de l'algorithme RF. En effet, ces valeurs ont été choisies sur la base des modèles de classification qui ont obtenu la plus grande précision globale. Ainsi, les valeurs des paramètres de segmentation étaient : Sp=10, Sp= 0.5, Cp= 01 et les valeurs de RF étaient : Mtry=25 et Ntree=500. Une étape de validation a été entreprise pour calculer la matrice de confusion et les indices de pertinence (indice de précision globale et indice Kappa).

3.2. Calcul d'indice de vulnérabilité côtière (IVC)

Les résultats du traitement relatif à chaque variable ont été intégrés dans un modèle SIG pour une application d'un système de pondération de 5 degré de cotation (un code de 1 jusqu'à 5) (Fig.1) selon une grille de calcul par cellule de 20m*20m. A chaque variable est affectée un degré en fonction de l'ampleur potentielle de son exposition et sa contribution à des changements physiques de la côte. Les 5 degrés obtenus font l'objet de calculs en utilisant l'équation (1) développée par McLaughlin et al. (2001). L'éventail complet de l'IVC a été divisé en quatre, à parts égales, et le quart supérieur a été pris comme "côte à très haut risque".

$$CVI = \sqrt[2]{\frac{a \cdot b \cdot c \cdot d \cdot e}{6}} \quad (1)$$

a = Taux d'érosion ; b = géomorphologie ; c = pente ; d = hauteurs des vagues e = occupation du sol ;

4. Résultats et discussion

Les résultats du traitement des données relatifs à chaque variable ainsi que le degré de leur vulnérabilité sont présentés ci-dessous.

a. La variable trait de côte

Le modèle d'extraction automatique du TC a été validé par des données terrain. En observant la distance moyenne entre le TC extrait et celui du terrain, il est apparu que le TC produit par le modèle d'extraction présente une bonne précision (Distance max de 0.7m). La figure 4 illustre le taux d'évolution du TC, de la zone d'étude, estimé à partir des données historiques et récentes (1992 et 2019). Le taux d'évolution pendant la période étudiée révèle une érosion importante dans plusieurs parties de la côte de Mostaganem (plus de 2m/an), par conséquent, une valeur pondérée de 5 leur a été attribuée (fig.9). Cette forte érosion est notée au niveau des plages sableuses. Ceci pourrait être lié aux constructions massives sur les cordons dunaires et l'extraction du sable non réglementée (le pillage) (Senouci & Taibi, 2019). Les côtes marquant une érosion modérée à faible (moins de 2m/an), principalement les côtes rocheuses, ont une résistance considérable face aux forçages de la mer.

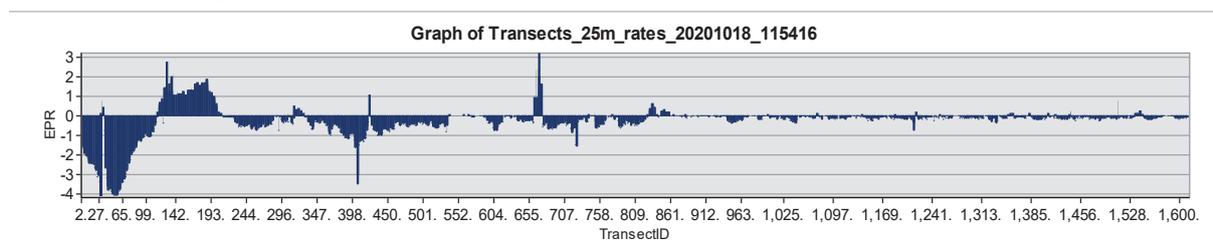


Figure 4. Taux net d'évolution du trait de côte par an EPR (1992-2019)

b. La variable : pente

La figure 5 montre que le gradient de la pente du tronçon côtier étudié est estimé entre 0,2 % et 9,3 %. Les zones ayant les gradients de pente inférieurs à 1,5%, à savoir les plages, sont plus susceptibles de subir un lessivage des sédiments à contre-courant (Jeyabal & Ramasamy, 2014). A l'opposé d'une côte en pente douce, où toute élévation du niveau de la mer inonderait de grandes étendues de terre, la conséquence de l'action de la mer sur une côte abrupte serait insignifiante (Rao et al., 2008). Par conséquent, à faible pente la côte devrait reculer plus vite que les côtes à pente raides, par conséquent, ces zones se voient attribuer les valeurs pondérées de 5.

c. La variable : relief côtier

Le littoral étudié présente une alternance de plages, falaises, baies rocheuses, dunes et cordons dunaires. En fonction du faciès rencontré, les risques d'érosion côtière sont plus ou moins importants. Selon McLaughlin et Cooper (2010) les reliefs côtiers qui sont à basse altitude (<10 m), à l'instar des plages sableuses, se voient attribuer la valeur 5 (très forte vulnérabilité) (fig.9), ces côtes ont un potentiel élevé aux inondations ou aux submersions marines. Les côtes à falaise et de rivages rocheux de hautes altitudes (>20 m) sont considérées moins vulnérables parce qu'elles sont plus résistantes au forçage de la mer et à la remontée du niveau marin. De ce fait, la valeur pondérée de 1 leur a été attribuée (fig.9).

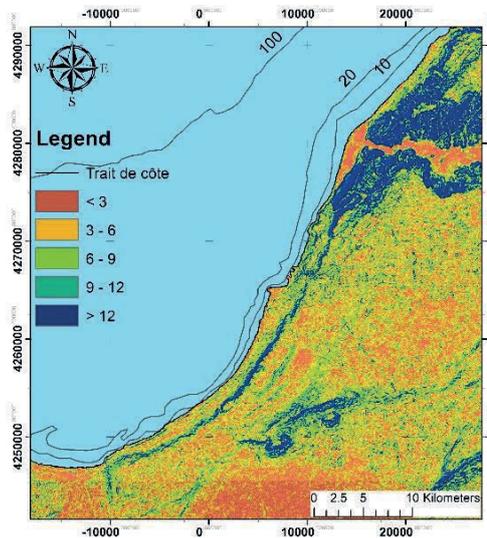


Figure 5. Carte des pentes (%)

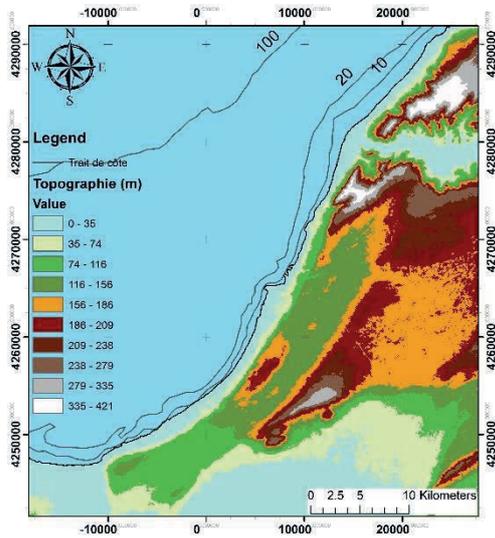


Figure 6. Carte topographique de la zone (dérivée du MNT)

d. La variable : hauteur des vagues

Les modèles en sorties montrent que la quasi-totalité du littoral de Mostaganem est exposée à l’action des vagues et des houles. Cette forte exposition est liée au fait que la morphologie côtière et sous-marine dissipe peu l’énergie marine. Les zones côtières du secteur ouest de la zone d’étude montrent une augmentation de la hauteur des vagues de plus de 1,4 m, libèrent plus l’énergie sur la côte, ce qui entraîne une érosion ou un affaissement important des plages. De ce fait, une valeur pondérée de 5 (très forte vulnérabilité) lui a été attribuée (fig.11). Les ondes de faible énergie produisent une hauteur d’onde plus faible (<1,0 m), ce qui entraîne des processus d’accrétion dus à l’accumulation des sédiments. Par conséquent, une valeur pondérée de 1 a été attribuée à ces zones de faible hauteur de vagues.

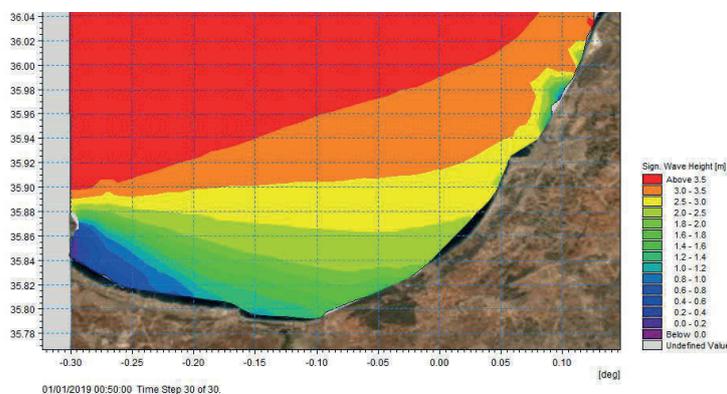


Figure 7. Hauteurs des vagues à la cote calculées par le model de Mike-21

Cartographie de la vulnérabilité côtière

e. La variable : occupation du sol

Les résultats de l'analyse de l'image Pléiade par une approche de classification orientée objet et l'algorithme RF sont représentés dans la figure 12, qui illustre la répartition spatiale du type d'occupation du sol du littoral de Mostaganem. Cette analyse a permis de détecter cinq classes d'occupation du sol. La classification par l'algorithme RF utilisant la segmentation Multirésolution a montré peu d'indications d'erreurs. Une évaluation de la précision a été effectuée afin de montrer l'efficacité de la classification basée sur les modèles optimisés. L'analyse de la matrice de confusion révèle que les indices de précision pour plusieurs classes étaient relativement cohérents avec un indice kappa de l'ordre de 93%.

L'analyse globale a révélé que la côte étudiée est fortement urbanisée au niveau de certains secteurs, principalement dans sa partie centre (ville de Mostaganem). L'activité agricole, quant à elle, est très présente, particulièrement dans le secteur ouest. En raison de l'impact de ces actions anthropiques et suite à la suppression du tissu végétal et aux empiètements, les cordons dunaires, qui sont considérés comme barrière naturelle protectrice, connaissent un taux de diminution de plus de 2,0 km² (Senouci & Taibi, 2019). Ces mêmes types d'occupation du sol, responsables de la destruction de l'écosystème dunaire, se trouvent sans protection et directement exposés à l'action de la mer. En effet, ils sont considérés à la fois comme enjeux socio-économiques à protéger et facteurs anthropiques accentuant la vulnérabilité côtière, la raison pour laquelle une valeur pondérée de 5 leur a été attribuée.

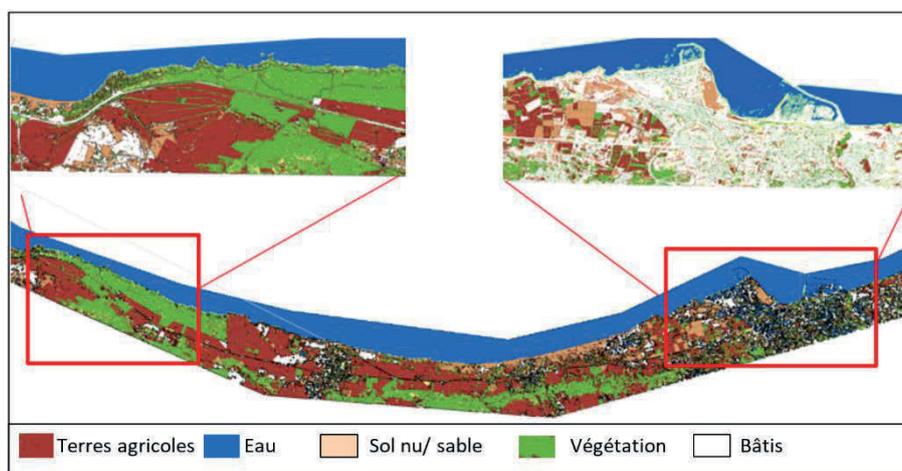


Figure 8. Carte d'occupation du sol de la frange côtière de Mostaganem

4.1. Indice de Vulnérabilité Côtière (IVC)

Le résultat des valeurs de l'IVC varient de de 3.09 à 19.36, avec une valeur moyenne de 10,41. Sur la base de cette fourchette de valeurs, les zones côtières sont catégorisées en quatre classes : vulnérabilité, faible, modérée, forte et très forte. La

carte de l'IVC (fig. 9.f) montre le degré de vulnérabilités côtières le long du littoral de Mostaganem. Le graphe (fig.10) montre les étendues en linéaire côtier (Km).

La côte étudiée s'étend sur un linéaire de 32 km, dont 30.13 % à 34.03 % présente un fort à très fort degré de vulnérabilité respectivement (IVC de 12.13 à 19.36). 12.62% de ce linéaire est caractérisée par une intensité modérée (IVC de 8.51 à 2.13). La carte de synthèse de l'IVC révèle que la forte intensité caractérise surtout le secteur central de Mostaganem, fortement urbanisé, ainsi que les reliefs de faible altitude tels que l'estuaire d'Oued Cheliff et les criques sableuses. Par ailleurs, 23.2% de ce linéaire présente un faible degré de vulnérabilité (IVC<8.51). Ces zones connaissent un phénomène d'accrétion dues aux affleurements rocheux au large des côtes (barrières naturelles). De même pour les côtes rocheuses et à falaise (élévation supérieure à 10,0 m) qui sont mieux protégées.

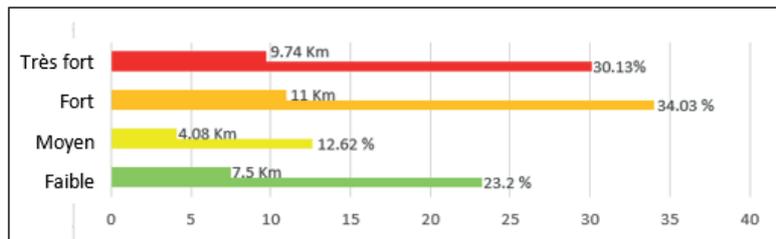


Figure 10. Présentation graphique du degré de vulnérabilité en termes de pourcentage du linéaire côtier

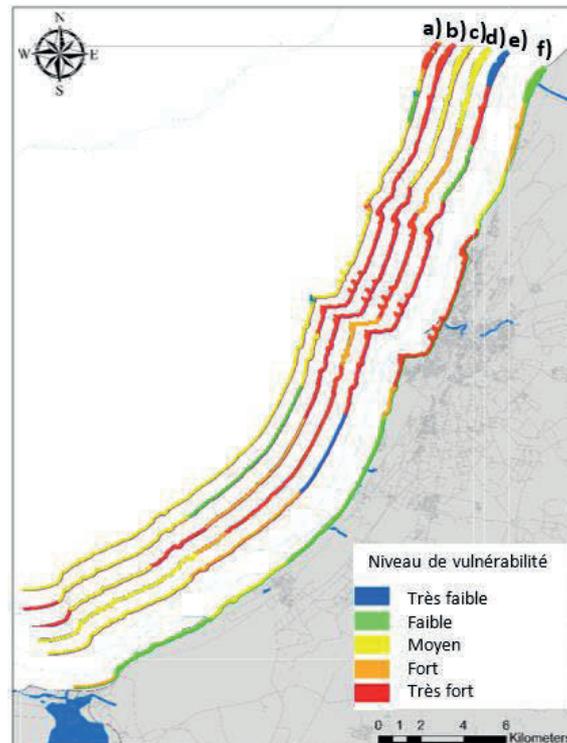


Figure 9. Carte la vulnérabilité côtière a) vulnérabilité par rapport à la variable trait de côte, b) vulnérabilité par rapport à pente, c) vulnérabilité par rapport à la variable relief côtier, d) vulnérabilité par rapport à la variable Hs, e) vulnérabilité par rapport à la variable type d'occupation du sol, f) Indice de Vulnérabilité Côtière de Mostaganem

5. Discussion

Plusieurs études (ex. Ayadi, et al., 2016 ; Kermani, et al., 2016 ; Bengoufa, et al., 2018) ont rapporté que le littoral algérien est soumis à une érosion considérable mettant en péril des enjeux socio-économiques. En effet, il est essentiel que les scientifiques contribuent à développer des méthodes d'évaluation des risques et d'assurer leur applicabilité et ainsi produire des données fiables et adaptées aux pays en développement, qui disposent peu de données (Awange, et al., 2018).

Les variables du modèle de l'IVC peuvent être choisies en fonction de la localisation et les caractéristiques des côtes (Koroglu et al. 2019). Dans le modèle ici développé 5 variables ont été traitées par des méthodes de la géomatique et de la télédétection. Ces méthodes intégrées permettent d'optimiser le temps et l'effort et offrent les solutions technologiques pour générer des bases de données sur le relief côtier et ses processus dynamiques (Kaliraj, S., & Chandrasekar, N. 2012). Ainsi, leur efficacité pour rendre mesurable la vulnérabilité côtière, a été prouvée grâce au système de cotation de 5 niveaux selon les intervalles d'intensité de chaque variable qui ont des unités hétérogènes, en homogénéisant leurs dimensions.

La performance de l'algorithme d'apprentissage automatique RF pour la détection du trait de côte, a été démontrée et validée par des mesures in-situ. En fait, un nombre limité d'études ont évalué la précision de ces algorithmes pour la détection du trait de côte. Ainsi, rares sont celles qui ont validé leurs résultats avec des données de référence (Dada, et al., 2016). En outre, pour analyser le niveau d'exposition des activités socio-économiques au risque d'érosion, le modèle d'IVC ici développé a intégré le type d'occupation du sol. Des études qui intègrent la composante socio-économique, ce qui était le cas de ce modèle d'IVC, permettront une planification urbanistique plus efficace et une définition des règles qui en découlent, y compris l'interdiction des constructions dans les zones exposées.

6. Conclusion

Dans une optique de spatialisation des risques, cette étude a permis de développer une méthode d'évaluation de la vulnérabilité face au risque d'érosion selon une approche géospatiale. D'après les résultats, plusieurs points chauds (forte vulnérabilité) ont été identifiés à différents endroits. Cette identification des zones permet de mieux tracer les plans d'aménagement côtier dans une optique de développement durable et une gestion d'un système de potentialités incontournables.

L'utilisation des méthodes intégrées de télédétection et de SIG a permis de générer une base de données relative au relief côtier, aux processus dynamiques et aux facteurs socio-économiques. La classification supervisée des images Pléiades a montré un grand potentiel dans le traitement des variables, notamment pour l'extraction du trait de côte et le type de l'occupation du sol. Ainsi, ces méthodes ont permis de rendre la vulnérabilité mesurable malgré l'hétérogénéité des variables qui entrent en jeu en les intégrant dans un seul modèle SIG. Ce modèle est une plateforme efficace de cartographie des zones potentiellement vulnérables aux dommages dus à l'érosion et il est utile aux scientifiques et décideurs pour une planification à long terme de la gestion des zones côtières.

Financement

Ce travail de recherche est soutenu par le Partenariat Hubert Curien PHC-Tassili.

Bibliographie

- Awange, J. L., Saleem, A., Konneh, S. S., Goncalves, R. M., Kiema, J. B. K., & Hu, K. X. (2018). Liberia's coastal erosion vulnerability and LULC change analysis : Post-civil war and Ebola epidemic. *Applied Geography*, 101, 56–67.
- Boak, E. H., & Turner, I. L. (2005). Shoreline definition and detection : A review. *Journal of coastal research*, 21(4 (214)), 688–703.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Brooks, N., Nicholls, R., & Hall, J. (2006). Sea-level rise : Coastal impacts and responses. *Norwich WBGU ISBN*, 3–936191.
- Chandrasekar, V., Keränen, R., Lim, S., & Moisseev, D. (2013). Recent advances in classification of observations from dual polarization weather radars. *Atmospheric Research*, 119, 97–111.
- Dada, OA., Guangxue, L., Qiao, L., Ding, D., Yanyan, M., Jishang X. 2016. Seasonal shoreline behaviours along the arcuate Niger Delta coast: Complex interaction between fluvial and marine processes. *Continental Shelf Research* 122: 51–67.
- Ghodbani, T., Kansab, O., & Kouti, A. (2016). Développement du tourisme balnéaire en Algérie face à la problématique de protection des espaces littoraux. Le cas des côtes mostaganemoises. *Études caribéennes*, 33-34.
- Gornitz, V. (1991). Global coastal hazards from future sea level rise. *Global and Planetary Change*, 3(4), 379-398.
- Jeyabal, G., & Ramasamy, S. (2014). Down core variation in sediment characteristics and trace element geochemistry of a core sample in Pichavaram Mangrove area, Tamil Nadu, Southeast Coast of India. *Enviro Geo Chimica Acta*, 1(3), 206–214.
- Kaliraj, S., & Chandrasekar, N. (2012). Geo-processing model on coastal vulnerability index to explore risk zone along the south west coast of Tamilnadu, India. *Int JEarth Sci Eng*, 5.
- Koroglu, A., Ranasinghe, R., Jiménez, J. A., & Dastgheib, A. (2019). Comparison of coastal vulnerability index. *Ocean & coastal management*, 178, 104799.
- McLaughlin, S., & Cooper, J. A. G. (2010). A multi-scale coastal vulnerability index : A tool for coastal managers? *Environmental Hazards*, 9(3), 233–248.
- Mujabar, P. S., & Chandrasekar, N. (2013). Shoreline change analysis along the coast between Kanyakumari and Tuticorin of India using remote sensing and GIS. *Arabian Journal of Geosciences*. 6 (3), 647-664
- Senouci, R., & Taibi, N.-E. (2019). IMPACT OF URBANIZATION ON COASTAL DUNE : CASE OF KHARROUBA, WEST ALGERIA. *Journal of Sedimentary Envi*, 4(1)
- Taibi, N.-E. (2016). Conflict Between Coastal Tourism Development and Sustainability : Case of Mostaganem, Western Algeria. *Eur Jrnl of Sust Development*, 5(4), 13–13.
- To, D. V., & Thao, P. T. P. (2008). A shoreline analysis using DSAS in Nam Dinh Coastal Area. *International Journal of Geoinformatics*, 4(1).

Communications par Session

Approches géomatiques pour la transition



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Optical network ring analogy for effective solar PV site selection using GIS and optimization modelling

Benjamin Pillot¹, Nadeem Al-Kurdi¹, Carmen Gervet¹,
Laurent Linguet¹

ESPACE-DEV, Univ Montpellier, IRD, Univ Guyane, Univ Réunion, Univ Antilles, Montpellier, France

ABSTRACT. Energy transition requires a holistic approach, involving land, resources, environmental and economical data and constraints. The core purpose of energy transition, is to migrate power systems towards renewable energy usage (solar, hydro, biomass, wind), in a technical and economical viable manner. In this article we address this challenge as a spatio-temporal analysis problem combined with decision support targeted for policy makers and investors. We show how the selection of potential PV sites, can efficiently be combined with the optimization of the connection to the grid in terms of operational costs. Our contribution lies in bringing forward a modeling analogy with the SONET problem, addressed in fiber network designs. Improved results compared to an existing GIS-optimization PV site placement approach are illustrated on a real case study and data from the French Guiana.

RÉSUMÉ. La transition énergétique nécessite une approche holistique, impliquant la gestion des sols, ressources potentielles, les données et contraintes environnementales et économiques. L'objectif central est de migrer les systèmes électriques vers l'utilisation des énergies renouvelables (solaire, hydroélectrique, biomasse, éolienne), de manière techniquement et économiquement viable. Dans cet article, nous abordons ce défi comme un problème d'analyse spatio-temporelle intégré à un modèle d'aide à la décision conçu pour les décideurs politiques et les investisseurs. Nous montrons comment la recherche de sites photovoltaïques peut être combinée efficacement avec l'optimisation du raccordement au réseau en termes de coûts. Notre contribution réside dans la mise en place d'une analogie de modélisation avec le problème SONET, défini dans les conceptions de réseaux de fibre optique. La qualité des résultats comparés à une approche existante de placement de parcs photovoltaïques est présentée sur un cas d'étude en Guyane française.

KEYWORDS: Energy planning, Spatial Decision support, Spatio-temporal data

MOTS-CLÉS : transition énergétique, aide à la décision spatiale, données spatio-temporelles

1 Introduction

Nowadays, the main goal of most countries' energy planning policy (Hache, Palle, 2019) in terms of *energy transition*, relies on the integration of renewable-based generation in power networks. However, increasing the share of renewable energy (RE) sources is still challenging, as a result of their inherent intermittent and geographical dispersion (Hache, Palle, 2019). Specific planning strategies (Gorsevski *et al.*, 2013; Ramirez Camargo, Stoeglehner, 2018) must be developed accordingly, in order to reach energy transition targets without threatening the existing infrastructures.

Recent works propose an integrated model framework, combining Geographic Information Systems and Robust Optimization, such as the GREECE-OPSPV (*Geographical REnewable Energy Candidate Extraction - Optimal Planning and Sizing of PV parks*) system (Al-Kurdi *et al.*, 2019; Pillot *et al.*, 2020). It identifies the optimal sites for solar photovoltaic (PV) generation at utility scale, according to specific energy planning targets. The approach connects both geographical information system (GIS) and robust linear optimization (RO) through the pipe depicted in Figure 1. The GIS component gathers large heterogeneous sets of spatiotemporal data, and allows for location and size of the best solar PV sites to be retrieved with respect to geographical constraints (restricted areas, land use, distance to grid, etc.), spatial dispersion of the resource, hourly global energy demand and generation, predefined planning scenarios, and the degree of risk adversity authorized by the decision maker. Through a set of spatio-temporal data layers and control parameters, the GIS module (GREECE) converts spatial constraints and parcels into *items* characterized by *de-spatialized* attributes and solar resource time series. Based on current electricity generation and demand time series as well as projection scenarios, the RO model (OPSPV) then looks for optimal site candidates (location, size and power), with respect to given temporal constraints (parcel size, hourly electricity demand, maximum penetration of intermittent RE power) and objectives (maximize energy generation and minimize total costs). The model eventually gives an estimate of the risk associated with solar PV investment at utility scale to the decision and policy maker, by means of a Pareto approach (cost vs. energy generation) and according to best and worst case scenarios.

In this integrated GIS and decision support model, when a solar PV site (location and size) is proposed, a new substation is considered for connection to the power network. This approach was justified as the model was designed for private investors in RE plants, thus each solar PV site required its own connection to the grid. However, in island networks (i.e. not interconnected), the network manager might have the authority to decide which site should be turned on and off, depending on the share of intermittent power (Notton

GIS and optimization modelling

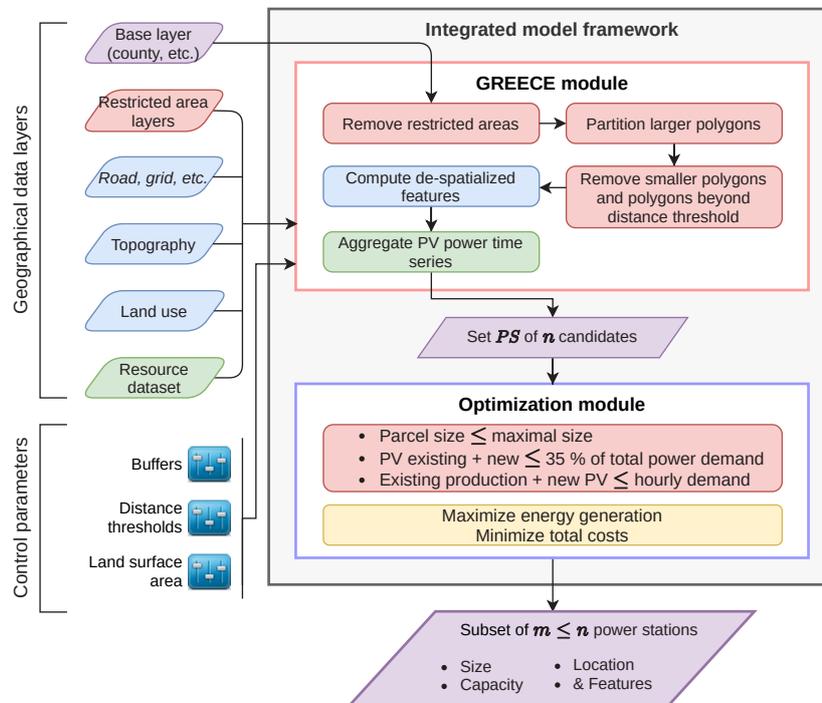


Figure 1. Schematic view of the original GREECE-OPSPV integrated model framework (Al-Kurdi et al., 2019)

et al., n.d.; Tapachès et al., 2019). Accordingly, some investors have moved on to diversifying their RE investments (hydro, biomass, etc.) or exploiting power plants with added storage capacity (Spaes, 2019; Tapachès et al., 2019). Another solution though is to rely on co-investment, whereby several sites belonging to multiple investors would be connected to the same substation. To our knowledge this approach hasn't been investigated in terms of simulation models. This would be a cost-effective solution as the maximum power per RE site is constrained by the technical features of the power network. In fact, for multiple small-scale solar PV facilities, the final cost would grow with the number of corresponding substations. In this paper we address this problem by deriving a model that would mitigate the cost by aggregating various power plants around one unique substation. This objective comes with new challenges on the GIS and optimization model.

The problem tackled is to maximize PV energy supply and minimizing costs through small-scale power facilities. Our new framework is based on an analogy with the Synchronous Optical NETWORK (SONET) (Pelleau et al., 2009; Goldschmidt et al., 2003), which is a popular network design in the field of fiber-optic technology. Typically, we first tailor the constraints of the SONET optimization problem to our own problem and then integrate them into the RO

model of the framework. Since this raises the computational complexity of the combinatorial optimization, the data range of those constraints is first pruned by enhancing the GREECE model with respect to the spatial constraints that apply to the potential sites. It ensures keeping the CPU time reasonably low without loss of solution quality. With the proposed architecture, the best solar PV sites are now gathered into *rings*, i.e. local aggregates of plants, connected to one sole substation. It results in lower energy unit cost (M€/GWh) as well as means for the decision makers and grid managers to manage their risk adversity, compared to the previous architecture.

To describe the resulting benefits, we compare both approaches, with and without considering constraints from the SONET problem, by applying the GREECE-OPSPV framework to the real case of French Guiana. The article is structured as follows: section 2 presents the revisited version of GREECE-OPSPV based on the SONET analogy; section 3 is the experimental section based on a real-world case study and compare both results from the former and the new version of GREECE-OPSPV; section 5 concludes the paper.

2 Applying the SONET design to RE sources in power systems

2.1 Planning small-scale PV sites for safer grid operation and improved risk management

Safer grid operation. When too much intermittent power is injected into the grid, it might be necessary for the network manager to disconnect some of the solar PV sites (Tapachès *et al.*, 2019; Notton *et al.*, n.d.). When there are only large power plants, the loss of one of them could affect grid stability. The use of smaller facilities therefore gives more flexibility when handling the energy supplied from intermittent sources, and eventually ensures safer grid operation once solar PV sites have been commissioned. This idea of targeting smaller solar PV sites is even more relevant in small not interconnected electricity networks, whose low *inertia* actually implies higher impact from intermittent RE sources on frequency variability and grid stability (Notton *et al.*, n.d.; Tapachès *et al.*, 2019). Typically, the GREECE-OPSPV model allows risk adversity to be assessed *upstream* using Pareto optimisation through best and worst case scenarios (Al-Kurdi *et al.*, 2019; Pillot *et al.*, 2020). In this paper, we consider smaller PV sites, and look for handling the risk *downstream*: whether the riskier solution is chosen or not, the grid manager must remain capable of eventually switching off a plant without jeopardizing the whole network.

Risk adversity management. When there is one unique investor deploying multiple sites, it is also more interesting to invest in various small-scale facilities rather than in few bigger ones. As a matter of fact, the grid manager will not care about the economical risk but rather about the technical risk: he will eventually shut down any solar PV site in case of too much intermittent power injected into the network (Notton *et al.*, n.d.; Tapachès *et al.*, 2019). This technical liability might not be related to technical constraints, as this is the

case for EDF in French Guiana which disconnects plants in the chronological order of their connection to the grid (CTG, 2017; EDF, 2019). As a result, while the investor may ensure for instance to still have 6 or 7 out of 8 small-scale PV plants running when some sites are eventually disconnected, he could lose half of the production if he only runs two large-scale facilities. This is an economical risk that some investors might not be willing to take.

Clusters of RE sources: towards the SONENT analogy. Previous works on the GREECE-OPSPV model, sought each solar PV power plant to be built along its own substation (Al-Kurdi *et al.*, 2019; Pillot *et al.*, 2020). As mentioned, this approximation is no longer valid when the number of sites dramatically increases, which is the case if we only allow small-scale plants to be eventually built. In order to keep the solutions economically relevant, we revisit this model so that several power plants might actually be connected to the same substation. We present the underlying methodology in the following section.

2.2 Adapting SONENT constraints to the RE planning problem

In the field of fiber-optic technology, one of the most popular network designs is the Synchronous Optical NETWORK (SONET) (Pelleau *et al.*, 2009; Goldschmidt *et al.*, 2003). In one possible topology, each customer is connected to exactly one *local ring* through add-drop multiplexers (ADM), and those local rings are all connected to one *federal ring* through a digital cross connector (DXC) (Pelleau *et al.*, 2009). This topology is depicted in Figure 2a. The cost

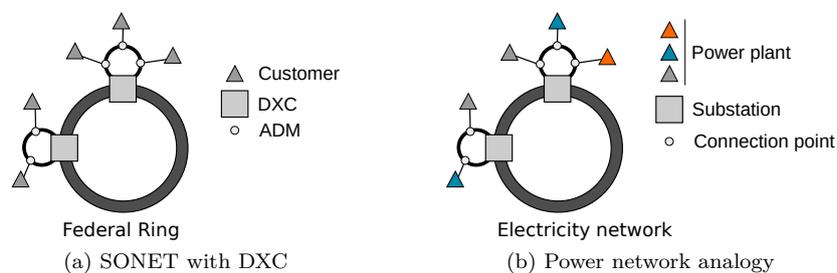


Figure 2. Fiber optical rings (Pelleau *et al.*, 2009) and power network analogy

of DXCs is much higher than that of ADMs, so the number of DXCs must be minimized, that is the number of rings. This is known as the *SONET Ring Assignment Problem (SRAP) with capacity constraints*. This can be formally depicted as a node-partitioning problem for a given graph G (Goldschmidt *et al.*, 2003). Nodes of G stand for the customers to be connected and the edge weights represent the traffic demand between sites. The analogy with power networks is depicted in Figure 2b and can be summarized as follows: we may think of the federal ring as the electricity network and of the DXCs as the different substations connected to it. Various power plants can then be connected to each substation/local ring with respect to the available hosting capacity. Those power plants can be of different type, such as dispatchable (biomass, hydroelectricity, geothermal, etc.) and non-dispatchable (solar, wind, etc.) RE sources

for example. Each ring may be seen as an aggregate of power plants that are spatially *close* enough to each other to be connected to the same substation.

We now show how constraints for this problem can be derived and adapted from the SRAP (Goldschmidt *et al.*, 2003), as well as developed to answer specific needs. In order to avoid a combinatorial explosion, the spatio-temporal GIS model is used to determine pairwise distance between all solar PV potential sites belonging to the set PS . For each ring k , we will therefore limit the available sites to the set B_k , that is the set of site indices for which sites PS_j have distance d_{kj} from site PS_k below a threshold D (including PS_k):

$$B_k = \{j \mid j \in N = \{1, \dots, n\}, d_{kj} \leq D\} \quad \forall k \in N = \{1, \dots, n\} \quad (1)$$

We then apply the constraints from the SONET problem to revisit the former optimisation problem (see Table 1 for the nomenclature):

$$\max_{<} b \quad \sum_k \sum_i \sum_h SA_{ik} \times Ppv_{h,i} \quad (2a)$$

$$\text{s.t.} \quad \sum_{i \in B_k} SA_{ik} \times Pnom \leq C \quad \forall k \in N = \{1, \dots, n\}, \quad (2b)$$

$$\sum_k x_{ik} \leq 1 \quad \forall i \in N = \{1, \dots, n\}, \quad (2c)$$

$$\sum_{i \in B_k} x_{ik} \leq M \times y_k \quad \forall k \in N = \{1, \dots, n\}, \quad (2d)$$

$$M \times \sum_{i \in B_k} x_{ik} \geq y_k \quad \forall k \in N = \{1, \dots, n\}, \quad (2e)$$

$$\sum_{i \in B_k} \sum_k SA_{ik} \times Ppv_{h,i} + Eint_h \leq 0.35 \times Dem_h \quad \forall h, \quad (2f)$$

$$\sum_{i \in B_k} \sum_k SA_{ik} \times Ppv_{h,i} + Eint_h + Ep_h \leq Dem_h \quad \forall h, \quad (2g)$$

$$SA_{ik} \leq Smax_i \times x_{ik}, \forall i \in B_k, \forall k \in N = \{1, \dots, n\}, \quad (2h)$$

$$x_{ik} \times Smin \leq SA_{ik}, \forall i \in B_k, \forall k \in N = \{1, \dots, n\}, \quad (2i)$$

$$x_{ik}, y_k \in \{0, 1\}.$$

Let $x_{ik} = 1$ if site PS_i is selected and belongs to ring k and $x_{ik} = 0$ otherwise. Let $y_k = 1$ if power plants are connected to ring k (i.e. ring is active) and $y_k = 0$ otherwise. Aggregate of PV power per ring is kept below a threshold C corresponding to the substation maximum hosting capacity (2b). Constraint (2c) ensures that a selected site only belongs to one ring. Constraints (2d) and (2e) guarantee that a ring with one or more selected PV sites is active. Constraint (2f) prevents the amount of intermittent energy from being over 35% of the total forecast energy demand (CTG, 2017; EDF, 2017).

GIS and optimization modelling

Satisfaction of the forecast energy demand is defined by constraint (2g), using existing resources augmented with new PV generation. Constraints (2h) and (2i) relate the size of the PV sites, lying between $Smin$ and $Smax_i$, to whether they are selected or not. This relationship is required to link both the energy production and the different costs. If a plant size is not null then the site is selected, and conversely if a site is not selected then its size is forced to be null.

Essentially the strategic energy planning over some given time horizon has two main objective functions: 1) to maximize the total hourly energy production over the year through new PV energy generation (2a), 2) to minimize the total costs related to PV installation, connection to the grid, etc. Since these functions are in different units they are not combined into a single weighted function that would not be meaningful, but instead solved by seeking the Pareto frontier, i.e. optimizing each function while constraining the other one. As a result, the optimization problem (2a)-(2i) is applied for every given constrained cost value in the range of the Pareto.

Minimize costs: Modeling non-linear functions. The total cost $Cost$ corresponds to the sum of the costs in every ring k . Each cost in ring k is the aggregate of the capital cost Cap_k of all the PV sites built in the ring, plus the connection cost to the grid $Ccon_k$, the substation cost $Csta_k$ and the operational & maintenance costs Cop_k :

$$Cost = \sum_k Cap_k + Cop_k + Ccon_k + Csta_k \quad (3)$$

Table 1. Nomenclature for the OPSPV problem formulation

$h \in H = \{1, \dots, 8760\}$	Hour (per year)
$i \in N = \{1, \dots, n\}$	Site index
$k \in N = \{1, \dots, n\}$	Ring index
n	Number of candidate sites derived from GREECE
x_{ik}, y_k	Boolean decision variables
B_k	Boundary set of nearest PV sites corresponding to ring k
C	Substation maximum hosting capacity (kW)
M	Big number to enforce Boolean inference
$Eint_h$	Current hourly production from intermittent energy sources (kWh)
Eph	Current hourly production from dispatchable energy sources (kWh)
Dem_h	Estimated global (forecasted) hourly power demand (kWh)
$Pnom$	Nominal power per unit area (kW/m ²)
SA_{ik}	Surface area of new selected PV site belonging to ring k (m ²)
$Smin, Smax_i$	Minimum and maximum area for each candidate parcel (m ²)
PS_i	Potential Site
$Ppv_{h,i}$	Estimated hourly production per PV unit (kWh/m ²)
dg_i	Minimal distance from the grid to PV site centroid (m)
d_{ij}	Distance between site PS_i and site PS_j
$Clan$	Transmission line unit cost (€/m)
$Ccap_k$	Capital cost of implementation of new PV power plants (€)
Cop_k	Annual operational cost per PV power plant (€)
$Ccon_k$	Connection costs for each ring, transmission lines (€)
$Csta_k$	Capital cost for new substation per ring (€)

Capital cost Cap_k and operational & maintenance costs Cop_k are defined as piecewise linear functions (Pillot *et al.*, 2020), and depend on the nominal power range of the PV sites implemented in ring k . Regarding connection costs $Ccon_k$, using the centroid of all sites as the connection bridge to the sites would in fact make the problem nonlinear. To keep linearity, we have thus considered in first approximation the maximum distance to the grid among PV sites implemented in ring k :

$$Ccon_k = Clan \times \max_{i \in B_k} (dg_i \times x_{ik}) \quad \forall k \quad (4)$$

where $Clan$ stands for the unit cost of transmission lines (€/km). Finally, following the same idea as for the plant capital cost, we also define the substation cost $Csta_k$ as a piecewise function. The piecewise model allows for the problem to remain linear and for economies of scale in building substations to be included into the analysis. The substation cost depends on the aggregated PV nominal power in ring k , which actually determines the final substation hosting capacity:

$$Csta_k = \begin{cases} a_4 \times Pnom \times (\sum_{i \in B_k} SA_{ik}) + y_4 & \text{if } 0_{MW} \leq (\sum_{i \in B_k} SA_{ik}) \times Pnom \leq 10_{MW} \\ a_5 \times Pnom \times (\sum_{i \in B_k} SA_{ik}) + y_5 & \text{if } 10_{MW} \leq (\sum_{i \in B_k} SA_{ik}) \times Pnom \leq 50_{MW} \\ a_6 \times Pnom \times (\sum_{i \in B_k} SA_{ik}) + y_6 & \text{if } 50_{MW} \leq (\sum_{i \in B_k} SA_{ik}) \times Pnom \end{cases} \quad (5)$$

3 Case study: *ring* vs. *site* approach in French Guiana

In this section, we compare the SONET-based *ring approach* with the *site approach* (Al-Kurdi *et al.*, 2019; Pillot *et al.*, 2020). We have applied both methods in French Guiana, according to characteristics and policy targets for the horizon 2030 in the region. In the context of the French *Energy Transition Act*, it is expected to triple RE source capacities by 2023 (CTG, 2017), first by increasing solar PV, then biomass.

3.1 Data and processing

The spatio-temporal GIS GREECE model first extracts the potential sites that will feed the OPSPV module depending on geographical constraints and land management. The model prunes the unrestricted territory in order to finally get suitable parcels with respect to minimum and maximum surface area. We refer the reader to (Pillot *et al.*, 2020) for further details about the methodology. It resulted in a set PS of 133 land parcels between 1.5 and 50 ha, depicted in Figure 3. They correspond to all the potential sites where solar PV plants could be built, such that all the constraints hold. We will rely on those plots to compare both ring and site approaches.

Beyond geographic location, each potential site generated by GREECE also comes along with specific features and resource time series (e.g. solar irradiation). Those features are extracted from heterogeneous data layers and remote sensing data. Essentially, GREECE allows for *de-spatialization* and discretiza-

GIS and optimization modelling

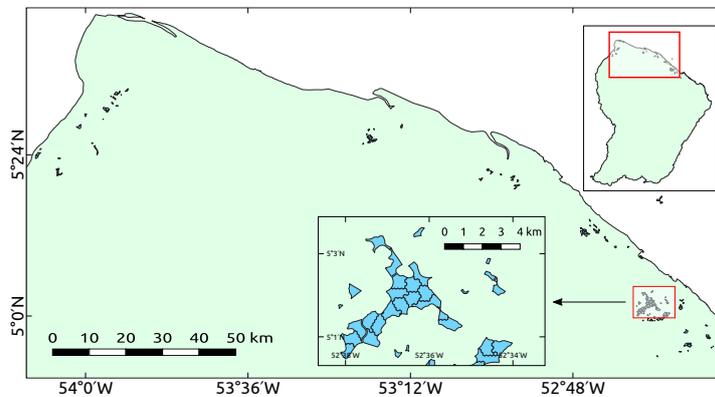


Figure 3. Set of potential sites derived by the GREECE model in Fr. Guiana.

tion of the potential sites in order to feed the combinatorial optimization model with *items* defined as tuples of digitalized attributes. In the present study, for simplicity purposes, items have been defined as tuples of 3 attributes: land surface area, distance to the grid, and time series of solar irradiation values. Our optimization model then converts those values into maximum nominal power, connection costs, and solar PV plant output power profiles.

Finally, hourly energy demand for the 2030 horizon has been projected according to 2016 records (EDF, 2019) and EDF estimations for worst case (5% annual growth) and best case (2% annual growth).

3.2 Results and analysis

In this paper, we compare two main aspects of relevance to the decision maker and network manager (power plant energy investors in French Guiana and EDF) between the GREECE-OPSPV model (*site approach*) and our proposed SONET-based model (*ring approach*). These aspects are related to what we previously stated as safer grid operations and better risk adversity management (see section 2.1). The comparison is based on the worst case scenario implemented in the site approach. In all cases, we only consider rings made of small-scale facilities whose nominal power is not greater than 5 MW. We first compare Pareto solutions for both approaches (Figure 4) in two fair distinct cases: (a) only small-scale facilities up to 5 MW can be built in the site approach, while no limitation is set on the ring hosting capacity (i.e maximum power that can be injected into the corresponding substation, sum of all PV sites' power attached to that ring); (b) solar PV sites up to 20 MW can be built in the site approach, and so the same applies to ring hosting capacity in the ring approach, which cannot exceed 20 MW as well. The 20 MW limit corresponds to the specific technical constraints of the French Guiana power network (CTG, 2017; EDF, 2017). This comparison investigates the relative contribution of each approach with respect to economical benefits, and risk management in terms of spatial distribution of the sites.

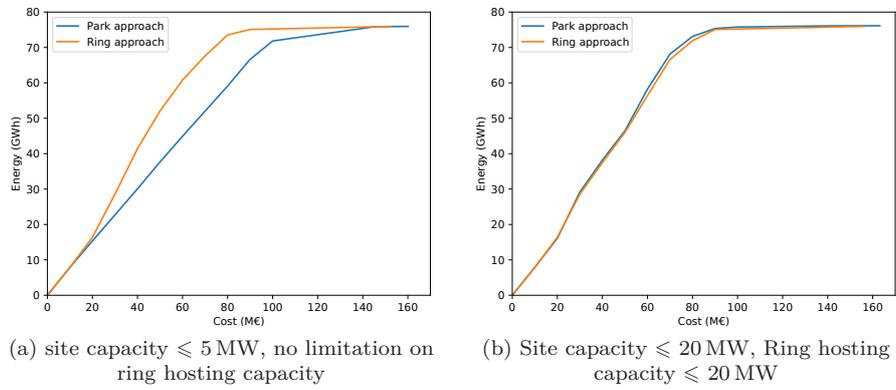


Figure 4. Pareto solutions for ring and site approach.

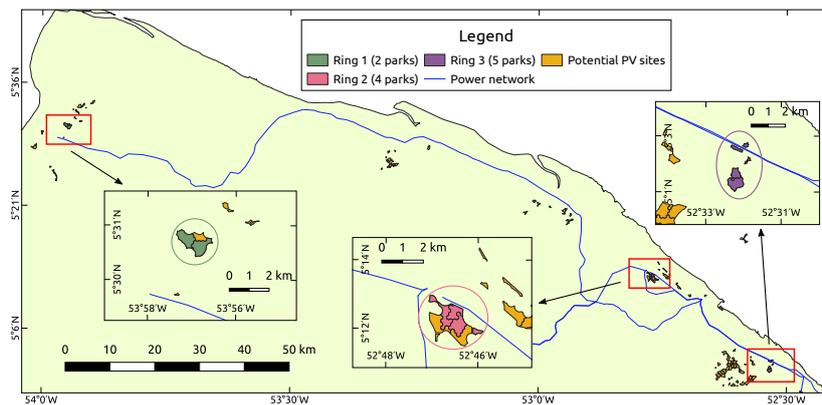
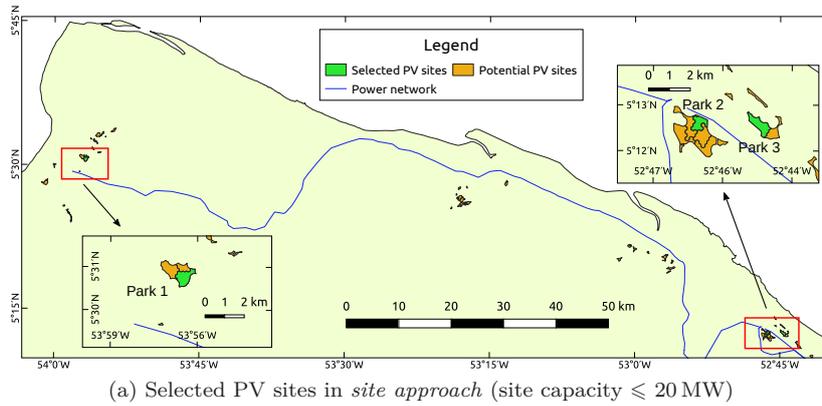


Figure 5. *site vs. ring approach spatial results for Pareto solution of Figure 4b (selected cost value = 80 M€).*

GIS and optimization modelling

When only small-scale solar PV sites can be integrated into a power network (a), the Pareto solutions depicted in Figure 4a show that the ring approach gives better results, with for instance 10 MW of extra installed power and 16 GWh of extra power generation compared to the site approach for the same overall cost (60 M€). When both approaches are subject to the same constraints (b), the resulting Pareto solutions remain similar in both cases regardless of the cost value (see Figure 4b). However, as stated in section 2.1, small-scale power stations allow for more manageable risk adversity once it is operating. Hence the ring approach gives the best energy planning strategy with respect to both the decision maker and the network manager needs.

To describe the resulting spatial difference between both models, we plot in Figure 5 the selected PV sites against the selected rings regarding the second case (Figure 4b) and with respect to the same Pareto solution (cost value = 80 M€). In both examples, the selected PV sites are scattered throughout the Northern shore of French Guiana, along the power network (i.e. cost is lower for sites close to the grid). The total installed power is similar (45.4 MW against 44.9 MW) but is allocated among 3 facilities in the site approach (Figure 5a), and among 3 rings and 11 small-scale facilities in the ring approach (Figure 5b). The characteristics of the PV sites in both examples are given in Table 2 and Table 3.

Table 2. Characteristics of the solar PV sites depicted in Figure 5a

site ID	Nominal power (MW)
site 1	6.87
site 2	20
site 3	18.52
Total	45.39

Table 3. Characteristics of the rings depicted in Figure 5b

Ring ID	Number of sites	Installed power (MW)	site capacities (MW)
Ring 1	2	5.56	{4.72, 0.84}
Ring 2	4	19.3	{5, 5, 4.74, 4.56}
Ring 3	5	19.99	{5, 5, 5, 2.8, 2.19}
Total	11	44.85	—

4 Conclusion

In this paper, we exploited the strengths of spatial decision support, applied to the field of energy transition. In particular, we studied the integration of spatial analysis on spatio-temporal data (solar radiation) as a pre-processing step to an optimization module. This module introduces a novel approach to planning the placement of solar PV sites, given the GIS computation of potential sites, such that the costs are further optimized compared to a more traditional approach. We showed and modelled an analogy with the SONET model, to aggregate

potential sites around one shared substation that feeds the power grid. The results show substantial gain in costs per KWh produced, as well as enhanced energy stability and risk management given a larger number of selected sites. Further and ongoing work include the extension of our GIS and optimization model, to assess biomass potentials and its energy power embedded into the grid.

References

- Al-Kurdi N., Pillot B., Gervet C., Linguet L. (2019). Towards robust scenarios of spatio-temporal renewable energy planning: A gis-ro approach. In *International conference on principles and practice of constraint programming*, pp. 729–747.
- CTG. (2017). *Programmation pluriannuelle de l'énergie (PPE) 2016-2018 et 2019-2023 de la Guyane*. Technical report. Collectivité Territoriale de Guyane.
- EDF. (2017). *Systèmes énergétiques insulaires Guyane - bilan prévisionnel de l'équilibre offre / demande d'électricité*. Technical report. Paris, France.
- EDF. (2019). *Open Data EDF Guyane*. Retrieved from <https://opendata-guyane.edf.fr/pages/home/> (Accessed: 2019-04-30)
- Goldschmidt O., Laugier A., Olinick E. V. (2003). SONET/SDH ring assignment with capacity constraints. *Discrete Applied Mathematics*, Vol. 129, No. 1, pp. 99–128.
- Gorsevski P. V., Cathcart S. C., Mirzaei G., Jamali M. M., Ye X., Gomezdelcampo E. (2013). A group-based spatial decision support system for wind farm site selection in Northwest Ohio. *Energy Policy*, Vol. 55, pp. 374–385.
- Hache E., Palle A. (2019). Renewable energy source integration into power networks, research trends and policy implications: A bibliometric and research actors survey analysis. *Energy Policy*, Vol. 124, No. September 2018, pp. 23–35.
- Notton G., Voyant C., Duchaud J. L. (n.d.). Difficulties of solar PV integration in island electrical networks - Case study in the French islands.
- Pelleau M., Van Hentenryck P., Truchet C. (2009). SONET network design problems. *Electronic Proceedings in EPTCS*, Vol. 5, pp. 81–95.
- Pillot B., Al-Kurdi N., Gervet C., Linguet L. (2020). An integrated GIS and robust optimization framework for solar PV plant planning scenarios at utility scale. *Applied Energy*.
- Ramirez Camargo L., Stoeglehner G. (2018). Spatiotemporal modelling for integrated spatial and energy planning. *Energy, Sustainability and Society*, Vol. 8, No. 1, pp. 29.
- Spaes J. (2019). *Inauguration de la centrale solaire avec stockage de Savane des Pères en Guyane*. <https://www.pv-magazine.fr/2019/10/17/inauguration-de-la-centrale-solaire-avec-stockage-de-savane-des-peres-en-guyane/>.
- Tapachès E., Perez R., Lauret P., Perez M., David M. (2019). Mitigation of the variability of a PV fleet via geographical dispersion and energy storage systems on the Reunion island non-interconnected grid. In *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, p. 2082-2086.

Prospective 2050 : étude de cas sur l'évolution du service écosystémique de protection côtière

Antoine Mury¹, Antoine Collin¹, Dorothée James¹

1. EPHE, PSL Université Paris, CNRS LETG,
15 Boulevard de la Mer, 35800 Dinard, France
antoine.mury@etu.ephe.psl.eu

RÉSUMÉ. Face aux changements environnementaux, induits par le changement climatique global, et notamment face la hausse du niveau moyen des mers, prévue et déjà observé à travers le monde par le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), certains systèmes éco-géomorphologiques font preuve de résilience et d'adaptation. C'est le cas notamment des marais littoraux, qui grâce à leur capacité à piéger les sédiments, parviennent à s'exhausser progressivement, en suivant le rythme de hausse du niveau marin. Cependant, il convient de s'interroger au sujet de l'aptitude de ces derniers à maintenir leurs niveaux de services écosystémiques face à ces changements environnementaux, et tout particulièrement leur capacité de protection contre le risque naturel de submersion marine. Ainsi, cette étude vise à modéliser de manière heuristique et prospective le service écosystémique de protection côtière assuré par les systèmes éco-géomorphologiques de la baie du Mont-Saint-Michel à l'horizon 2050, dont la topographie future a été simulée à partir des taux de sédimentation recensés dans la littérature.

ABSTRACT. Facing the environmental changes resulting from global climate change, and particularly the rise in mean sea levels, predicted and already observed worldwide by the Intergovernmental Panel on Climate Change (IPCC), some ecogeomorphological systems are showing resilience and adaptation. This is especially the case, in certain geographical contexts, of salt marshes, which, thanks to their capacity to trap sediments, manage to rise gradually, following the sea level rise rhythm. However, questions need to be asked about their ability to maintain their levels of ecosystem services in the face of these environmental changes, particularly their ability to protect against the natural risk of marine flooding. Thus, this study aims to model, in a heuristic and prospective manner, the ecosystem service of coastal protection provided by the ecogeomorphological systems of the Bay of Mont-Saint-Michel by 2050, whose future topography was simulated from the sedimentation rates identified in the literature.

MOTS-CLÉS : Submersion marine, service écosystémique, modélisation, adaptation, prospective.

KEYWORDS: Marine flooding, ecosystem service, modelling, adaptation, prospective

1. Introduction

Les marais littoraux sont des systèmes écosystémiques complexes, remplissant de nombreuses fonctions écosystémiques, notamment en matière de support pour la biodiversité (McOwen *et al.*, 2017), de fourniture de matière première, de séquestration du dioxyde de carbone atmosphérique (Mcowen *et al.*, 2017) ou encore de régulation face aux risques de submersion marine (Mury *et al.*, 2020) et d'érosion côtière. Or ces systèmes, et les fonctions qui y sont associées, se trouvent, de par leur situation géographique, à l'interface entre terres et mers, menacés par les changements environnementaux globaux, parmi lesquels la hausse du niveau moyen des mers (IPCC, 2014).

Néanmoins, les marais littoraux disposent d'un atout non-négligeable pour faire face à ces changements et ainsi s'adapter à ces modifications des conditions environnementales : leur capacité naturelle à s'exhausser en captant les sédiments (Stumpf, 1983 ; Allen, 2000 ; French, 2006 ; Détriché *et al.*, 2011 ; McKee *et al.*, 2012).

En effet, si le phénomène d'accrétion est un processus naturel au sein des baies, en raison de la dissipation de l'énergie qui s'y produit et de l'apport en matériel sédimentaire d'origine continentale par les fleuves côtiers, ce phénomène s'accroît en présence de surfaces de marais littoraux. Ceci s'explique par les caractéristiques éco-physiologiques des espèces végétales qui constituent les marais, et qui, par l'intermédiaire de leurs systèmes caulinaires et racinaires, présentent des capacités de piégeage et de fixation des sédiments en suspension lors des épisodes d'immersion qui les affectent, permettant ainsi l'élévation des sols (Stumpf, 1983).

Ce phénomène d'accrétion a notamment été quantifié en baie du Mont-Saint-Michel, par Détriché *et al.* (2011) pour les surfaces de marais littoraux, sur la période 1999-2005 et par Berger et Caline (1991), ainsi que par Larsonneur (1989), pour les surfaces sablo-vaseuses. Ces études réalisées en baie du Mont-Saint-Michel, font état de taux de sédimentation variables en fonction du gradient altitudinal et de la nature biogéomorphologique des sites d'observation. Ainsi, pour les surfaces de marais littoraux, trois taux ont été quantifiés, dépendant de la physiographie des observations : le bas schorre (3,45-38,11 mm/an), le moyen schorre (> 4,91 mm/an), et le haut schorre (>1,35 mm/an). Pour les surfaces sablo-vaseuses, deux taux moyens ont été déterminés par des méthodes de datation au ²¹⁰Pb (Berger et Caline, 1991), venant corroborer les analyses cartographiques réalisées par Larsonneur (1989), avec respectivement environ 4,5 mm/an et 15 mm/an, pour la slikke (vase) et la frange sablo-vaseuse du marais.

Ces taux de sédimentation, peuvent être comparés aux tendances régionales mesurées entre 1970 et 2016 en matière de hausse du niveau marin, qui font état d'une hausse annuelle comprise entre 2 et 3 mm, inférieure donc aux 3,3 mm/an prédits sur cette même période à l'échelle globale (IPCC, 2014). La comparaison de ces taux d'élévation, du niveau marin, d'une part, et des systèmes

Prospective 2050

écogéomorphologiques, d'autre part, mettant en évidence une vitesse d'accrétion du marais supérieure à la vitesse de hausse du niveau marin, permet de dresser l'hypothèse d'une résilience des systèmes écogéomorphologiques littoraux de la baie du Mont-Saint-Michel face à cette hausse du niveau marin. Se pose néanmoins la question de l'évolution des fonctions écosystémiques, notamment de protection côtière, dans ce futur contexte local.

Cette étude vise donc à modéliser de manière exploratoire et prospective, à l'horizon 2050, la capacité d'atténuation des vagues par les systèmes écogéomorphologiques côtiers de la baie du Mont-Saint-Michel, dont l'élévation future a été simulée à partir des taux de sédimentation précédemment évoqués. L'objectif étant *in fine* d'observer les effets de cette modification topographique sur l'évolution du service écosystémique de protection.

2. Matériels et méthodes

2.1. Site d'étude

Cette étude a été réalisée sur un site d'environ 3 km² dans la partie occidentale de la baie du Mont-Saint-Michel, présentant des enjeux humains (bâtiments, axes de communications) ainsi que des systèmes écogéomorphologiques (marais littoraux, cheniers coquilliers, estran sablo-vaseux) représentatifs du paysage de la baie et susceptibles de fournir un service écosystémique de protection contre la submersion marine (FIGURE 1).

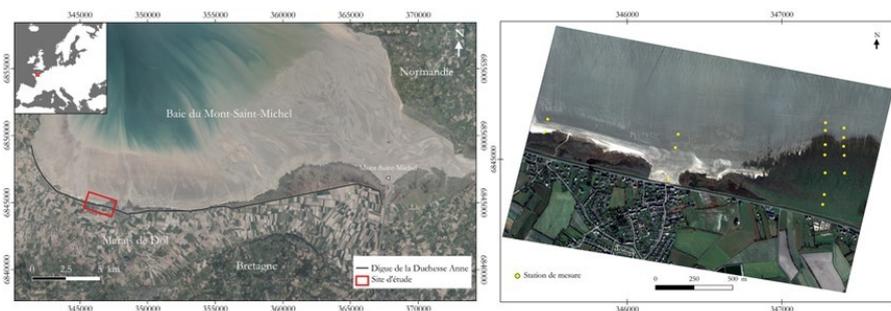


FIGURE 1. Situation géographique générale du site d'étude et localisation des stations de mesure des hauteurs significatives des vagues.

2.2. Mesure des vagues à très haute résolution et calcul des atténuations

2.2.1. Méthodologie d'acquisition des mesures de vagues

Les hauteurs significatives des vagues (H_{m0}) ont été acquises *in situ* à l'occasion d'une campagne de mesure de quatre cycles de marées consécutifs entre les 22 et 23

janvier 2019, à l'aide de 17 capteurs de pressions (NKE SP2T10) réparties sur 4 transects transversaux (FIGURE 1) et enregistrant à une fréquence de 2 Hz.

2.2.2. Méthodologie de calcul des atténuations

À partir des valeurs de Hm_0 acquises lors de la campagne de mesure *in situ*, des valeurs d'atténuation sont calculées en %/m pour chaque paire de stations de mesure. Ces valeurs d'atténuation constituent les réponses statistiques à la modélisation de l'atténuation des vagues, induite par les systèmes écotopographiques littoraux.

2.3. Données d'imageries spatioportées superspectrales WorldView-3 et aéroportées LiDAR

2.3.1. Données originales

Les 16 bandes spectrales de l'imagerie passive satellite WorldView-3 (WV-3) (acquise le 20 février 2019), ainsi que les produits dérivés de l'imagerie active LiDAR (*Light Detection And Ranging*, acquise entre le 2 mai et le 4 juillet 2018) sont utilisées comme prédicteurs de la modélisation de l'atténuation des Hm_0 , qui permet d'estimer la protection offerte par les systèmes écotopographiques littoraux contre le risque de submersion marine.

Les 16 bandes spectrales de l'imagerie satellite, d'une résolution spatiale originale de 1.20 m, ont été corrigées géométriquement et radiométriquement avant d'être affinées à une résolution spatiale de 0,30 m par l'intermédiaire d'une procédure de *pan-sharpening* (Gram-Schmidt) basée sur la finesse de la bande panchromatique (résolution spatiale 0,30 m × 0,30 m, cf. Mury et *al.*, 2020b). Des modèles numériques de surface (MNS) et d'élévation (MNE) sont réalisés à partir des données de l'imagerie LiDAR avec une résolution spatiale de 0,30 m × 0,30 m et une précision verticale inférieure à 0,2 m.

2.3.2. Données simulées

Afin d'observer les effets de l'évolution de la topographie du site sur le service d'atténuation des vagues, à l'horizon 2050, les valeurs des MNS et MNE dérivés de l'imagerie LiDAR ont été modifiées de manière à simuler l'élévation des systèmes écotopographiques à partir des taux annuels recensés au sein de la littérature. Pour chaque secteur du site, préalablement identifié par photo-interprétation, a été attribué un taux d'accrétion annuelle (FIGURE 2), lequel est ensuite multiplié par 32 (nombre d'années entre les deux dates) pour simuler l'élévation prévisionnelle entre 2018 et 2050.

Prospective 2050

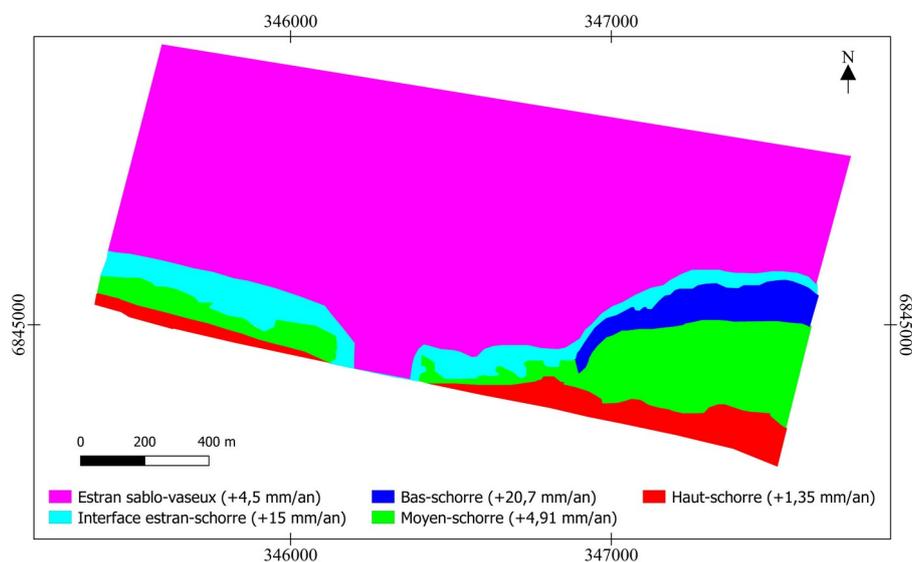


FIGURE 2. Valeurs d'accrétion sédimentaire par secteurs géographiques.

2.4. Modélisation spatiale de l'atténuation des Hm_0

De manière à évaluer le service écosystémique de protection contre la submersion marine, deux modélisations spatiales de l'atténuation des Hm_0 sont réalisées, à partir des valeurs d'atténuation calculées précédemment pour chaque paire de capteurs (cf. 2.2.) : utilisant (1) des prédicteurs spectraux et topographiques originaux, puis (2) utilisant les modèles numériques simulés (cf. 2.3.), par l'intermédiaire de régressions linéaires multiples, tel que :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Où, Y est la valeur d'atténuation, β_0 une constante, β_1 la valeur du coefficient qui est propre à X, la valeur du prédicteur.

2.4.1. Corrélation des prédicteurs avec les valeurs d'atténuation

Préalablement à la réalisation des modélisations spatiales, une sélection des cinq prédicteurs les plus corrélés avec les valeurs d'atténuation est réalisée afin de conserver un équilibre statistique, au sein de la modélisation, entre le nombre de prédicteurs et le nombre de réponses statistiques (les valeurs d'atténuation).

2.4.2. Conception des compartiments d'évaluation

Afin de pouvoir comparer aisément les capacités d'atténuation des Hm_0 quantifiées à travers les deux modélisations spatiales, les valeurs d'atténuation en %/

m sont moyennées au sein des compartiments de 30 m × 600 m orientés en fonction des houles dominantes affectant le site d'étude.

2.4.3. Taux global d'évolution du service de protection côtière

L'évolution du service d'atténuation des vagues, sous-jacente à l'évolution de la topographie, à l'horizon 2050, peut être étudiée par l'intermédiaire du taux global d'évolution calculé tel que :

$$Tx = \frac{(\text{Modélisation prospective} - \text{Modélisation originale})}{\text{Modélisation originale}} \quad (2)$$

Cette manipulation a pour objectif de mettre en évidence les compartiments dont les valeurs d'atténuation moyennes offertes par les éco- et géo-systèmes sont les plus affectées par les changements de topographie attendus d'ici 2050. Cela vise ainsi à observer la variabilité dans le temps du service d'atténuation des Hm_0 induit par les éco- et géo-systèmes du site d'étude.

2.4.4. Évolution de la capacité adaptative des enjeux

La capacité protectrice des éco- et géo-systèmes littoraux peut être abordée en tant que *proxy* de la capacité adaptative, qui constitue l'une des composantes de la vulnérabilité des enjeux (Mury *et al.*, 2020). La transposition de cette composante, depuis les modèles moyennés, aux éléments du cadre bâti (identifiés préalablement par un algorithme de classification supervisée (« maximum de vraisemblance »)), permet donc d'estimer un premier degré de vulnérabilité des enjeux face à l'aléa submersion marine. De manière à pouvoir être comparées de manière diachronique, ces valeurs de capacité adaptative des enjeux font l'objet d'une standardisation entre 0 et 1.

3. Résultats

3.1. Corrélation des prédicteurs avec les valeurs d'atténuation

Cinq prédicteurs des imageries ont été sélectionnés pour la modélisation spatiale de manière que le nombre de prédicteurs de la modélisation soit inférieur à la moitié du nombre de valeur d'atténuations calculées à partir des mesures *in situ*. Les cinq prédicteurs des imageries affichant les coefficients de corrélations les plus proches de 1 sont, en ordre croissant, SWIR1, PIR1, PIR2, MNE et MNS (TABLE 2). C'est donc la combinaison de ces prédicteurs, spectraux et topographiques, qui est utilisée pour la modélisation spatiale de l'atténuation des Hm_0 .

TABLE 1. Coefficients de corrélation entre les prédicteurs de la modélisation et les valeurs d'atténuation.

Prédicteurs	Coefficient de corrélation (Pearson)
-------------	--------------------------------------

Prospective 2050

PIR 1	0,66
PIR 2	0,69
SWIR 1	0,65
MNE	0,71
MNS	0,73

3.2. Modélisation spatiale de l'atténuation des Hm_0

3.2.1. Modélisation originale

La modélisation spatiale de l'atténuation des Hm_0 réalisée à partir des prédicteurs spectraux PIR1, PIR2, SWIR1 et des prédicteurs topographiques MNE et MNS originaux issus de l'imagerie LiDAR, affiche un coefficient de détermination R^2 de 0,70 et fait état de valeurs d'atténuation comprises entre -1,25 et 1,88%/m (FIGURE 3a).

Cette modélisation permet l'élaboration d'un modèle présentant des valeurs d'atténuation moyennées comprises entre 0,02 et 0,77%/m (FIGURE 3b).

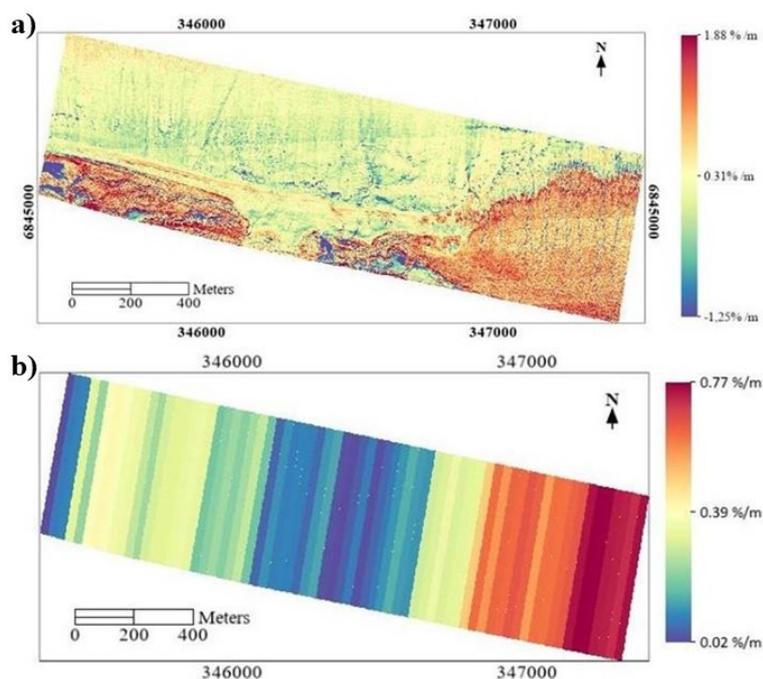


FIGURE 3. a) Modélisation spatiale originale de l'atténuation des Hm_0 et b) modèle moyenné associé.

3.2.2. Modélisation prospective

La modélisation spatiale prospective de l'atténuation des Hm_0 , réalisées à partir des prédicteurs spectraux PIR1, PIR2, SWIR1 et des MNS et MNE simulés, présente quant à elle des valeurs d'atténuation entre -1,35 et 1,79%/m (FIGURE 4a).

Le modèle « à compartiments » réalisé à partir de la modélisation prospective présente des valeurs moyennes d'atténuation des Hm_0 comprises entre -0,10 et 0,64%/m (FIGURE 4b).

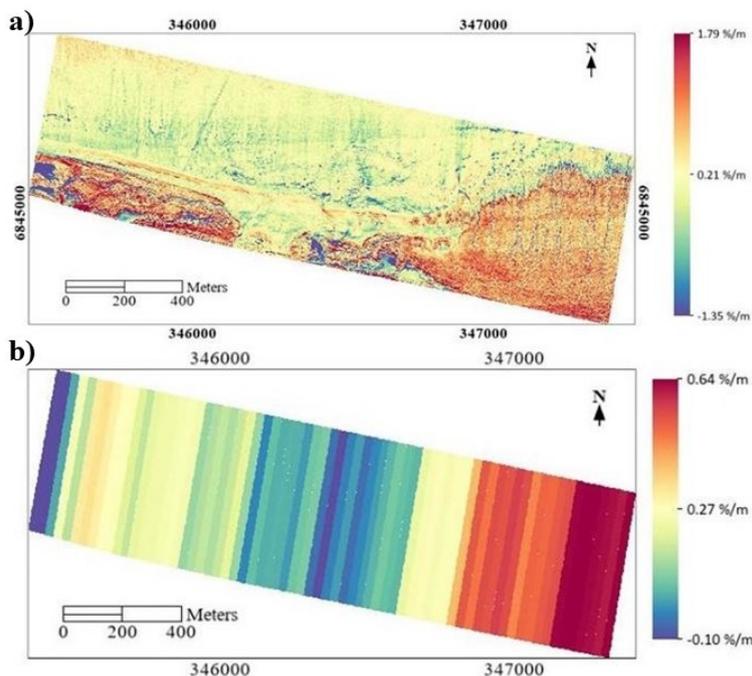


FIGURE 4. a) Modélisation spatiale prospective de l'atténuation des Hm_0 et b) modèle moyenné associé.

3.3. Évolution 2018-2050 du service de protection côtière

L'étude du taux global d'évolution entre les deux modélisations, montre des valeurs comprises entre -2,60 et 2,60% sur le site d'étude (FIGURE 5a). Lorsque ce taux est étudié par l'intermédiaire du modèle basé sur les valeurs moyennées, les valeurs d'évolution contenues dans les compartiments entre 2018 et 2050 sont comprises entre -1,12% et 0,78% (FIGURE 5b).

Prospective 2050

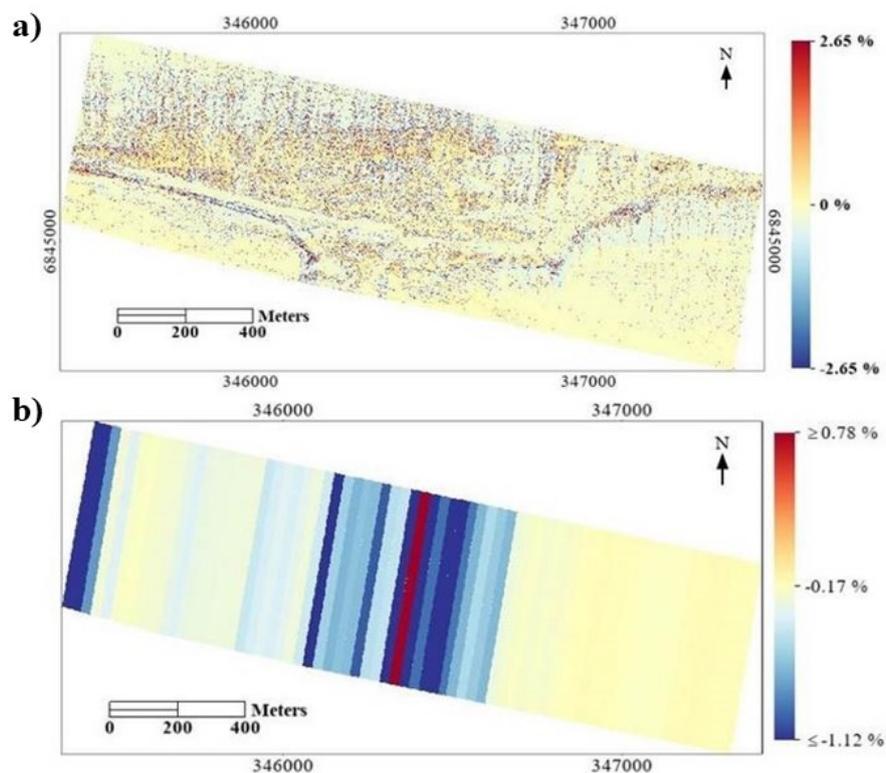


FIGURE 5. a) *Modélisation spatiale du taux global d'évolution* et b) *modèle moyenné associé*.

3.4. Évolution de la vulnérabilité des enjeux

L'application des différents modèles « à compartiments » dont les valeurs ont été standardisées, de 2018 et 2050, successivement, sur les enjeux situés en retrait du trait de côte, fait état de valeurs comprises entre 0,18 et 0,92 pour 2018 (FIGURE 6a), entre 0,18 et 1 pour 2050 (FIGURE 6b). Le taux global d'évolution entre les modélisations, appliqué aux enjeux, est également standardisé et présente des valeurs entre 0 et 1 (FIGURE 6c).

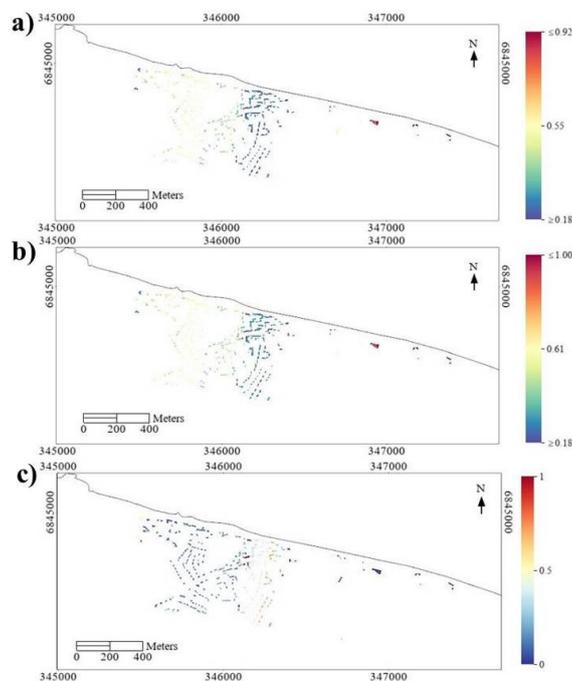


FIGURE 6. Application des modélisations sur les enjeux, à partir a) du modèle original, b) du modèle prospectif, et c) du modèle de taux global d'évolution.

4. Discussion/conclusion

4.1. Corrélation des prédicteurs avec les valeurs d'atténuations

La présence des bandes spectrales PIR1, PIR2 et SWIR1 de l'imagerie WV-3 et des MNE et MNS dérivés de l'imagerie LiDAR, parmi les prédicteurs présentant les meilleures corrélations avec les valeurs d'atténuation des Hm_0 , s'explique par leurs capacités respectives à fournir des informations sur les caractéristiques éco-physiologiques des communautés végétales de notre site d'étude, ainsi que sur les caractéristiques topographiques du site, qui sont les principaux éléments régissant la circulation des eaux. En effet, les bandes spectrales PIR1 et PIR2 sont particulièrement adaptées pour caractériser la structure anatomique des feuilles et sont utilisées, comme un indicateur du volume de la végétation (Collin *et al.*, 2018). La bande spectrale SWIR1, quant à elle, révèle l'absorption de la lumière par l'eau, et permet de déterminer l'état de santé et de développement de la végétation (Collin *et al.*, 2018). S'agissant des modèles numériques de surface et d'élévation dérivés de l'imagerie LiDAR, leurs fortes corrélations (0,73 et 0,71, respectivement) avec les

Prospective 2050

valeurs d'atténuation, s'expliquent par leur capacité à fournir des informations attenantes à la topographie du site, la rugosité de la surface et les pentes.

4.2. Modélisation spatiale de l'atténuation des Hm_0

Les modélisations spatiales de l'atténuation des Hm_0 , réalisées sur le site d'étude, mettent en évidence la capacité de protection qu'offrent les systèmes éco-géomorphologiques littoraux face aux vagues et à l'aléa submersion marine.

Ainsi, pour le modèle réalisé à partir des données originales, les valeurs d'atténuation des Hm_0 sont comprises entre -1,25%/m et 1,88%/m sur l'ensemble du site d'étude (FIGURE 3a). La modélisation prospective présente, quant à elle, des valeurs sensiblement similaires, entre -1,35%/m et 1,79% (FIGURE 4a), venant ainsi confirmer l'hypothèse d'un maintien de la capacité de protection offerte par les systèmes éco-géomorphologiques, à l'horizon 2050. Ces modélisations spatiales permettent la mise en lumière du rôle protecteur joué par les marais littoraux qui présentent les valeurs d'atténuations les plus importantes du site d'étude.

Ce rôle des systèmes éco-géomorphologiques, et spécialement du marais littoral, dans l'atténuation des vagues, est tout particulièrement observable à travers les modèles « à compartiments », avec des compartiments présentant des valeurs d'atténuation supérieures sur les secteurs présentant systèmes éco-géomorphologiques. Cette représentation des valeurs d'atténuation moyennées, permet ainsi, au-delà des valeurs, d'identifier spatialement les secteurs du site d'étude qui bénéficient d'une capacité de protection moindre de la part des systèmes éco-géomorphologiques. Ainsi, sur le site d'étude, il apparaît, à l'analyse des différents modèles, originaux comme prospectifs, que les secteurs offrant un service d'atténuation des vagues plus restreint se localisent sur les secteurs présentant peu ou pas de végétation et une faible topographie, et qui ne présentent donc pas de marais littoral développé.

4.3. Évolution 2018-2050 du service de protection côtière

L'étude du taux global d'évolution du service d'atténuation des Hm_0 , entre les modélisations originales et prospectives, notamment par l'intermédiaire du modèle « à compartiments », met en relief deux constats.

En premier lieu, ce taux global d'évolution montre la capacité d'adaptation des systèmes éco-géomorphologiques, qui par leur potentiel à s'exhausser naturellement, parviennent à maintenir un service écosystémique performant, bien qu'en léger recul, en matière d'atténuation des Hm_0 . Ceci est tout particulièrement tangible sur les secteurs de marais littoraux à l'Est et à l'Ouest de la zone d'étude. En second lieu, ce taux global d'évolution permet d'observer l'accroissement des inégalités, à l'horizon 2050, entre les secteurs bénéficiant d'un service de protection performant de la part des systèmes éco-géomorphologiques et ceux bénéficiant d'un service moindre, et dont les taux globaux d'évolution sont les plus négatifs.

Ainsi les enjeux identifiés comme étant ceux présentant une capacité adaptative moindre avec la modélisation originale (FIGURE 6a), le restent avec la modélisation

prospective 2050 (FIGURE 6b). Néanmoins les taux globaux d'évolution entre les deux modélisations, originale et prospective, font état d'un recul plus important du service de protection dont bénéficient les enjeux situés derrière les secteurs ne présentant pas de marais littoraux et de cheniers développés (FIGURE 6c).

Remerciements

Les auteurs remercient l'Agence spatiale européenne pour la fourniture des données d'imagerie satellite superspectrale WorldView-3, ainsi que le Service hydrographique et océanographique de la marine pour la fourniture des données d'imagerie LiDAR.

Références

- Allen, J.R.L. (2000). Morphodynamics of Holocene salt marshes: a review sketch from the Atlantic and Southern North Sea coasts of Europe. *Quaternary Science Reviews* 19, p. 1155–1231.
- Berger, G., Caline, B. (1991). Estimation of sedimentation rates in the Bay of Mont-Saint-Michel (France) by ²¹⁰Pb dating technique. A pilot study. *Océanologica acta, Proceedings of the International Colloquium on the environment of epicontinental seas, Lille, 20-22 March 1990, vol. sp. n°1, p. 137-142.*
- Détriché, S., Susperregui, A-S., Feunteun, E., Lefeuvre, J-C, Jigorel, A. (2011). Interrannual (1999-2005) morphodynamic evolution of macro-tidal salt marshes in Mont-Saint-Michel Bay (France). *Continental Shelf Research*, 31, p. 611-630.
- French, J. (2006). Tidal marsh sedimentation and resilience to environmental change: exploratory modelling of tidal, sea-level and sediment supply forcing in predominantly allochthonous systems. *Marine Geology* 235, p. 119–136.
- Collin, A. ; Lambert, N. ; Etienne, S. (2018). Satellite-based salt marsh elevation, vegetation height, and species composition mapping using the superspectral WorldView-3 imagery. *International Journal of Remote Sensing*, 39(17), p. 5619-5637.
- McKee, K.; Rogers, K.; Saintilan, N. (2012). Response of salt marsh and mangrove wetlands to changes in atmospheric CO₂, Climate, and sea level. *Global change and the function and distribution of wetlands*, p. 63-96.
- Mcowen C.J.; Weartherdon, L.V.; Van Bochove J-W.; Sullivan, E., Blyth S.; Zockler, C.; Stanwell-Smith, D.; Kingston, N.; Martin, C.S.; Spalding, M.; Fletcher, S. (2017). A global map of saltmarshes. *Biodiversity data journal* 5, e11764.
- Mury, A., Collin, A., Jeanson, M., James, D., Gloria, H., Pastol, Y., Etienne, S. (2020a), Mapping nature-based marine flooding risk using VHR wave, airborne LiDAR and satellite imagery: the case study of the Dol March (Bay of Mont-Saint-Michel, France), *Journal of Coastal Research*, Special Issue No. 95, p. 743-747.
- Mury, A., Collin, A., James, D. (2020b), Protection naturelle contre la submersion, apport de l'intelligence artificielle, *Comité français de cartographie, numéro spécial « Cartorisk »*, soumis.
- Stumpf, R. (1983). The process of sedimentation on the surface of a salt marsh. *Estuarine, Coastal and shelf science* 17, 5, p495-508.

Trajectoire socio-écologique d'une commune de la Baie du Mont-Saint-Michel entre 1947 et 2014

Mathilde Letard¹, Antoine Collin¹, Dorothée James¹, Antoine Mury¹

1. EPHE, PSL Université Paris, CNRS LETG
15 Boulevard de la Mer, 35800 Dinard, France
mathilde.letard@ephe.psl.eu;

RESUME. Les conséquences environnementales du changement climatique global nous renvoient à la nécessité de prendre en compte l'interconnexion des sociétés et des écosystèmes. Dans cette optique, cet article appréhende l'évolution d'une commune de la Baie du Mont-Saint-Michel comme celle d'un socio-écosystème, dont l'état est déterminé par les interactions entre ses composantes mais aussi par des forçages exogènes. Des séries temporelles de l'occupation du sol sont constituées à partir d'images aériennes anciennes et actuelles et permettent d'analyser la trajectoire socio-écologique de la commune. Les modèles de régressions linéaires attribuent en premier lieu l'anthropisation croissante de la zone à la croissance démographique, puis à des paramètres climatiques ($R^2 = 0.87$). Le développement de la population entraîne également la progradation du pré salé à proximité directe du bourg ($R^2 = 0.97$). Cependant, les prés salés de la zone semblent approcher d'un changement d'état : après une croissance parallèle à celle du bâti, ils connaissent un net recul qui semble lié à l'augmentation de la surface bâtie au-delà d'un certain seuil.

ABSTRACT. Climate change is bringing us back to considering the connection between societies and ecosystems. In this respect, this study aims at approaching the evolution of a small town of the Mont-Saint-Michel bay as the trajectory of a social-ecosystem influenced both by the interactions of its components and by external forcing. Airborne imagery is used to constitute land use time series in order to analyse the social-ecological trajectory of the town. Linear regression models show that population growth is the first explanation to increasing anthropization of the area, before climate factors ($R^2 = 0.87$). It is also the main factor of the progression of the salt marsh located right next to the city centre ($R^2 = 0.97$). However, salt marshes of the sector seem to come close to change of state: after having followed a growing trend parallel to that of the urban area, they are shrinking significantly, which seems to be linked to the augmentation of the urban area beyond a given threshold.

MOTS-CLES : socio-écosystème, séries temporelles, imagerie aérienne, modélisation, trajectoires

KEYWORDS: social-ecosystems, time series, airborne imagery, modelling, trajectories

1. Introduction

Les prés salés sont des écotones à végétation basse situés en zone intertidale. Ils fournissent des services importants aux sociétés côtières (Barbier *et al.* 2011), dont un support à la biodiversité locale (McOwen *et al.* 2017), le stockage de carbone atmosphérique (Johnson *et al.* 2014 ; McOwen *et al.* 2017) ou une protection naturelle des inondations et de l'érosion (Mury *et al.* 2020).

Régis par des processus complexes, leur suivi est nécessaire pour mieux les préserver des contraintes anthropiques et climatiques (Wiberg *et al.* 2019) auxquelles ils sont exposés. En effet, facteurs naturels et anthropiques se combinent pour déterminer l'évolution de la surface, de la biodiversité et de la morphologie des prés salés (Townend *et al.* 2010). Parmi eux, l'élévation, la végétation, la dynamique hydrosédimentaire (Townend *et al.* 2010 ; Fagherazzi *et al.* 2004 , 2012), l'eutrophisation (Deegan *et al.* 2012), les aménagements humains (Bouchard *et al.* 1995), l'invasion par divers usages interférant avec leur régime naturel (Allen 2000), ou encore les changements climatiques comme la montée du niveau marin (IPCC 2018).

Ces mécanismes variés contrôlant l'apport et la fixation sédimentaires essentiels à l'existence des prés salés (Townend *et al.* 2010 ; Fagherazzi *et al.* 2004 , 2012) expliquent la diversité des profils d'évolution de ces écosystèmes dans le monde. Si plusieurs études récentes font état d'un recul conséquent de leur surface globale (World Resources Institute 2005), les prés salés de certaines régions tendent au contraire à prograder (Détriché *et al.* 2011 ; Bouchard *et al.* 1995). C'est le cas en Baie du Mont Saint Michel, où la diversité des interactions qui conditionnent les marais littoraux est bien illustrée. Le fort marnage et le régime sédimentaire de la baie qui favorise la faible énergie des courants et le comblement sédimentaire leur permettent de progresser vers le large depuis plusieurs décennies. Cependant, la pression anthropique, l'exploitation des ressources naturelles et l'aménagement des côtes influencent également le développement des prés salés (Bouchard *et al.* 1995, Bromberg Gedan *et al.* 2009).

Le comportement des écotones côtiers mettant en jeu de nombreux paramètres, un raisonnement reconnaissant l'interconnexion de la société et de la biosphère et prenant en compte des forçages exogènes permettrait certainement une meilleure approche des changements qu'ils connaissent (Folke *et al.* 2011).

Cet article s'attache donc à évaluer l'évolution socio-écologique d'une commune côtière de la Baie du Mont Saint Michel *via* des séries temporelles d'évolution de l'occupation des sols constituées à partir d'images aériennes à haute résolution spatiale. Deux hypothèses sont émises et testées : (1) la démographie, la température, les précipitations, la vitesse des vents et le niveau marin conditionnent l'évolution du pré salé situé dans la zone et (2) expliquent l'anthropisation du site.

Trajectoire socio-écologique d'une commune entre 1947 et 2014

2. Matériel et méthodes

2.1. Site d'étude

La Baie du Mont Saint Michel est comprise entre la Péninsule du Cotentin et la Côte d'Émeraude, dans le Golfe Normand-Breton. Elle est caractérisée par un régime macro-tidal, son marnage faisant partie des plus élevés au monde (Archer 2013). La baie est délimitée au sud par le Marais de Dol, polder situé derrière la digue de la Duchesse Anne aux pieds de laquelle s'étendent d'importants prés salés recouvrant environ 40 km² (Valéry et Radureau 2014).

Le site d'étude choisi s'étend sur 6.5 km² autour de la commune du Vivier-sur-Mer, située dans la partie Ouest de la Baie du Mont Saint Michel, où les conditions hydrodynamiques sont plus calmes que dans la partie estuarienne de la baie. Délaissée par ses habitants dans la première partie du XX^{ème} siècle, la commune a connu un renouveau socio-économique avec l'essor de l'activité mytilicole et, plus largement, conchylicole, matérialisée par le développement du port « Le Vivier-sur-Mer/Cherruex » qui est aujourd'hui le plus important complexe conchylicole d'Europe. Le socio-écosystème comprend un village, des champs, des prés salés ainsi qu'une partie de la zone intertidale. Un fleuve, dont l'estuaire se trouve au niveau du port, assure un apport sédimentaire aux prés salés depuis les terres, en complément de l'apport de sédiments des courants de la baie.

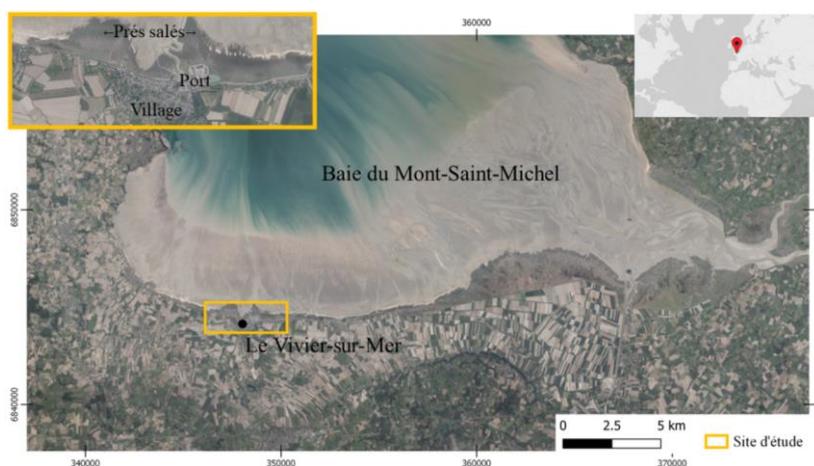


FIGURE 1. Localisation du site d'étude autour du Vivier-sur-Mer

2.2. Données

2.2.1. Images aériennes

Dix prises de vues aériennes de sources diverses mises à disposition par l'IGN sont utilisées pour cette étude. Leurs caractéristiques sont résumées dans la Table 1.

TABLE I. Prises de vues aériennes utilisées pour l'étude et caractéristiques

Date	Bande d'observation	Résolution spatiale
1947	Panchromatique	0.37 m
1952	Panchromatique	0.37 m
1961	Panchromatique	2.72 m
1966	Panchromatique	3.56 m
1978	Panchromatique	1.11 m
1986	Panchromatique	0.73 m
1989	Panchromatique	0.46 m
1996	Panchromatique	0.15 m
2002	Rouge + Vert + Bleu	0.5 m
2014	Rouge + Vert + Bleu + Infrarouge	0.5 m

2.2.2. Données climatologiques

Pour étudier l'influence du climat sur l'évolution paysagère, plusieurs variables extraites des simulations du climat entre 1951 et 2005 réalisées par Météo France, l'Institut Pierre Simon Laplace et le CERFACS sont exploitées. On dispose ainsi, pour chaque jour entre le 1^{er} Janvier 1951 et le 31 Décembre 2005, de températures (en K), de précipitations totales (en kg/m²/s), et de la vitesse des vents (en m/s).

2.2.3. Données démographiques et historiques

Pour étudier l'influence de la démographie sur l'évolution paysagère, les recensements de population effectués par l'INSEE tous les 5 à 9 ans entre 1968 et 2007 sur la commune du Vivier-sur-Mer ainsi que les taux de variation de la population – solde naturel et migratoire – entre chaque relevé sont utilisés. D'autres informations sur la démographie, ainsi que sur l'aménagement progressif du port sont extraites de travaux sur l'histoire du Vivier-sur-Mer.

2.3. Méthodes

2.3.1. Préparation des imageries aériennes

Dix orthomosaïques sont générées à partir des images de chaque campagne *via* le logiciel Metashape. Elles sont ensuite géoréférencées sur la base de l'Ortholittorale V2 (2013-2014) et découpées selon l'emprise de la zone d'étude (logiciel ENVI).

Pour assurer une continuité dans les données employées, une bande panchromatique composite est calculée en sommant les informations des bandes rouge, verte et bleue pour les dates de 2002 et 2014. Toutes les images sont ensuite

Trajectoire socio-écologique d'une commune entre 1947 et 2014

ré-échantillonnées à 1 mètre de résolution spatiale, moyenne raisonnable pour détecter les objets socio-écologiques à très haute résolution spatiale afin de suivre leur dynamique, avec la méthode du plus proche voisin (logiciel ENVI).

2.3.2. Cartographie de l'occupation du sol

Une cartographie de l'occupation du sol est ensuite produite à partir de chaque image. Pour cela, une classification par maximum de vraisemblance est effectuée sur une combinaison de 8 attributs de texture obtenus par l'application de filtres sur l'image panchromatique. En effet, la bande panchromatique seule ne suffit pas à distinguer les différentes occupations du sol efficacement. La variance, l'homogénéité, le contraste, la dissimilarité, l'entropie, le moment d'ordre 2, la corrélation, et la valeur moyenne sont donc calculés pour multiplier l'information disponible pour le classifieur. La détection obtenue est ensuite corrigée localement par photo-interprétation, pour éliminer la confusion persistant entre les classes dont la signature spectrale est très proche sur les images panchromatiques anciennes.

Les classes définies pour la classification correspondent aux types d'occupation du sol cartographiés : pré salé, sédiments vaseux, sédiments coquilliers, bâti, routes et chemins, arbustive et arborescente, champs (pelouses et cultures), sol nu, eau.

2.3.3. Création de séries temporelles d'occupation du sol

La surface de chacune des classes est calculée pour chaque date à partir de la carte correspondante, afin d'obtenir des séries temporelles de surface.

Pour certains types d'occupation, il est nécessaire de disposer d'une série de valeurs annuelles sur tout l'intervalle de 1947 à 2014. Des interpolations sont donc réalisées en ajustant une courbe de tendance issue de modèles statistiques aux observations. Pour le pré salé et le bâti, les interpolations sont faites avec un polynôme de degré 4. La somme des surfaces des champs et des arbres ainsi que la classe « routes/chemins » sont interpolées linéairement. A chaque fois, le modèle retenu pour l'interpolation – polynômial ou linéaire – est celui pour lequel la valeur du coefficient de Pearson obtenu était la plus élevée.

L'eau présente dans la scène ne représente que les lagunes et bassins portuaires créés progressivement au fil du développement du port. Les valeurs intermédiaires sont ainsi ajustées en fonction des connaissances de l'histoire de l'aménagement du port.

2.3.4. Création de séries temporelles climatologiques et démographiques

Les données climatiques sont moyennées annuellement pour obtenir une série d'une valeur par an pour chaque paramètre entre 1951 et 2005.

Les recensements de l'INSEE ne renseignent la population qu'à certaines dates. A partir des données d'évolution démographique, la population est estimée entre 1968 et 2014. Pour les dates antérieures, des données historiques sur la population du Vivier-sur-Mer en 1936, 1957, 1962, 1965 et 1967 sont utilisées et reliées par des valeurs interpolées linéairement.

2.3.5. Création de séries temporelles de statistiques paysagères

Outre la surface de chaque occupation du sol identifiée, l'indice de Shannon est employé pour évaluer la diversité paysagère. Cet indice est défini comme suit :

$$H' = - \sum_{i=1}^S p_i \log_2 p_i \quad (1)$$

Où H' est l'indice de diversité de Shannon, S le nombre de classes dans la scène et p_i la proportion de la classe i par rapport à l'ensemble des classes.

L'indice suivant quantifiant l'anthropisation du paysage sur la partie « terrestre » de la zone (excluant sédiments vaseux et coquilliers) est également utilisé :

$$A = \frac{\text{Surface de sol artificiel}}{\text{Surface de sol naturel}} \quad (2)$$

A est l'indice d'anthropisation. La surface de sol artificiel correspond à la somme des surfaces occupées par le bâti, les routes et chemins et l'eau (qui correspond aux bassins portuaires aménagés). La surface de sol naturel est quant à elle obtenue en additionnant les surfaces des champs, des arbres et arbustes, et du pré salé. Le sol nu n'est pas inclus car il est compliqué de caractériser sa présence comme un signe d'anthropisation ou non.

2.3.6. Modèles explicatifs à partir de régressions linéaires

Des régressions linéaires multiples sont réalisées pour tenter d'expliquer l'évolution de l'anthropisation du site d'étude et l'évolution de la surface du pré salé. Elles sont menées sur les données normalisées, afin de permettre la comparaison des coefficients de corrélation obtenus sans biais d'échelle.

3. Résultats

3.1. Cartes de l'occupation du sol pour chaque millésime

Les cartes obtenues par classification des 10 orthomosaïques sont présentées en Figure 2. 9 types d'objets sont distingués : les sédiments vaseux, les sédiments coquilliers, le pré salé, les routes et chemins, le bâti, le sol nu, l'eau, les champs (pelouses et cultures) et la végétation arborescente et arbustive.

Trajectoire socio-écologique d'une commune ente 1947 et 2014

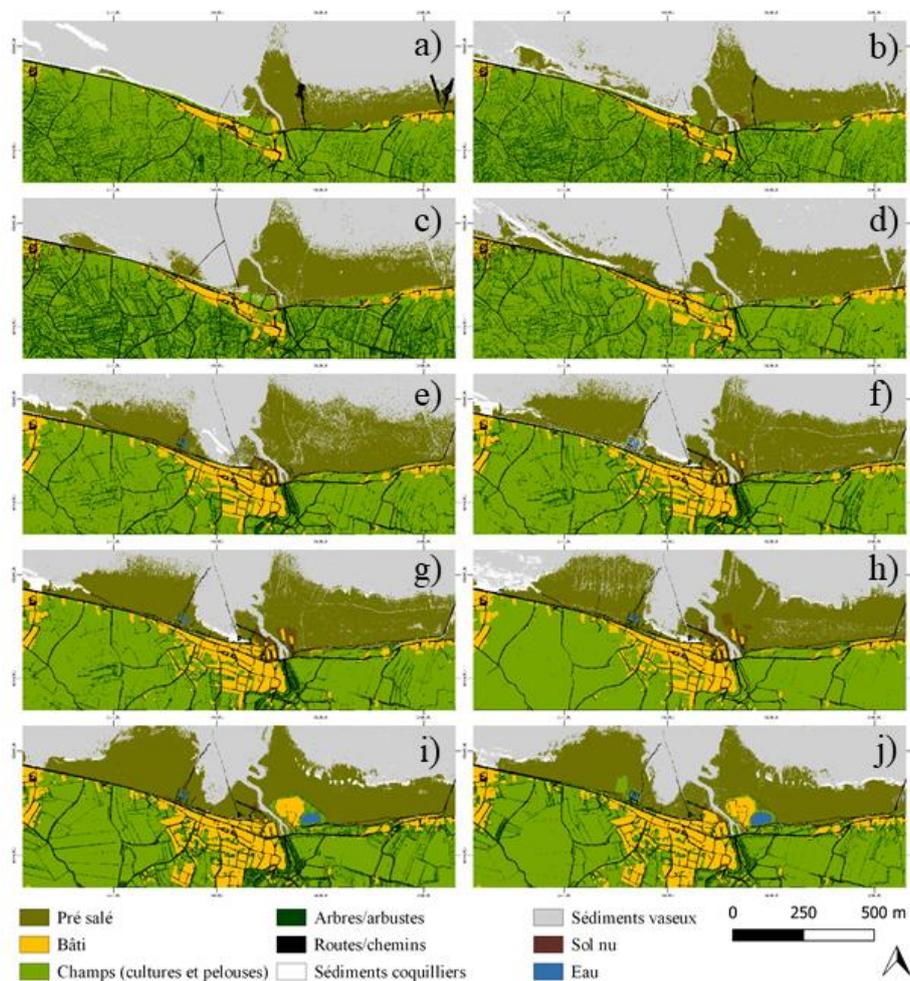


FIGURE 2. Cartes de l'occupation du sol autour du Vivier Sur Mer de 1947 à 2014 : a)1947, b)1952, c)1961, d)1966, e)1978, f)1986, g)1989, h)1996, i)2002, j)2014

3.2. Séries temporelles de l'évolution de la surface des différentes occupations du sol

Ces cartes permettent de dresser l'évolution de la surface des différentes catégories de sol présentée en Figure 3. Les valeurs s'étendent de 0 à 2 506 716 m².

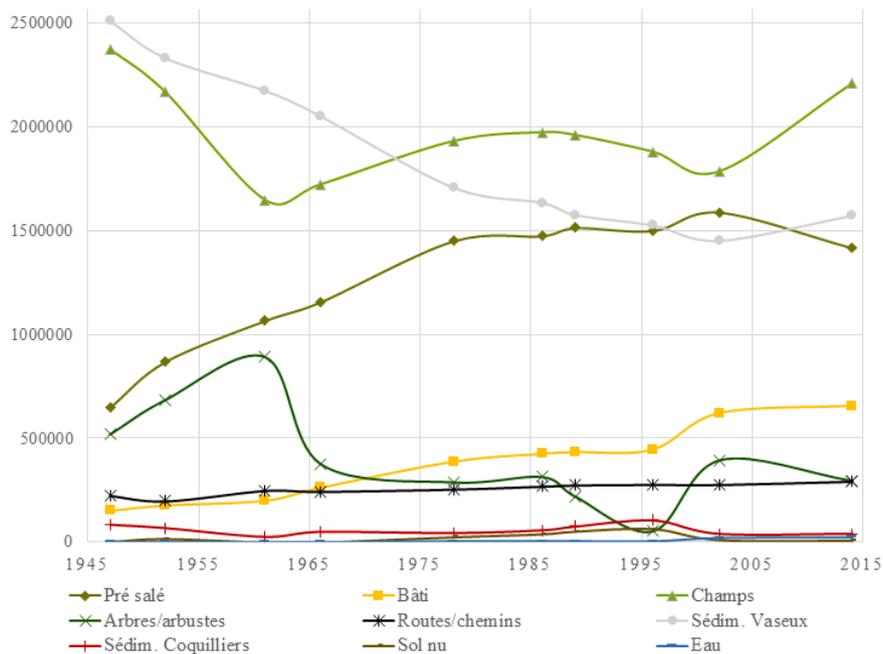


FIGURE 3. Évolution de la surface des classes d’occupation du sol en mètres carrés de 1947 à 2014

Les tendances d’évolution des classes principales sont présentées en Table 2.

TABLE 2. Tendances d’évolution des classes d’occupation du sol de 1947 à 2014

Sujet	Tendance	Coefficient de détermination
Pré salé (surface)	Positive (polynôme ordre 4)	0.98
Bâti (surface)	Positive (polynôme ordre 4)	0.97
Routes/chemins (surface)	Positive (linéaire)	0.83
Végétation (surface)	Négative (linéaire)	0.95
Diversité du paysage	Positive (polynôme ordre 2)	0.65
Anthropisation (indicateur)	Positive (polynôme ordre 4)	0.98

3.3. Modèles explicatifs de l’évolution socio-écologique

Les deux hypothèses testées s’avèrent pertinentes sur le plan statistique. L’hypothèse 1 est corroborée par le coefficient de corrélation de Pearson (évaluant

Trajectoire socio-écologique d'une commune ente 1947 et 2014

la pertinence du modèle) dont la valeur atteint 0.93 et par un coefficient de détermination R^2 (évaluant le pouvoir prédictif du modèle) de 0.87.

Pour l'hypothèse 2, le coefficient de corrélation de Pearson vaut 0.97 et le coefficient de détermination R^2 atteint 0.94. Celle-ci est donc également validée.

Les prédicteurs climatiques et démographiques n'ont pas tous le même impact sur l'anthropisation ou la surface du pré salé, comme l'illustre la Figure 4.

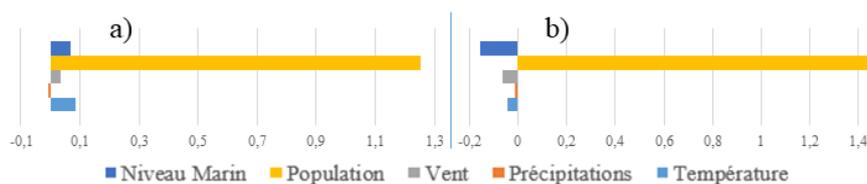


FIGURE 4. Influence des prédicteurs sur l'anthropisation (a) et sur la surface de pré salé (b) (les valeurs présentées sont celles des coefficients de régression linéaire)

Finalement, on trouve que l'anthropisation et la surface de pré salé peuvent être prédites avec les équations suivantes :

$$A = -0,10 + 1,26 \text{ Pop} + 0,08 \text{ T} + 0,07 \text{ Nm} + 0,03 \text{ V} - 0,006 \text{ P} \quad (3)$$

$$S_{\text{pré salé}} = 0,42 + 1,44 \text{ Pop} - 0,15 \text{ Nm} - 0,06 \text{ V} - 0,04 \text{ T} - 0,01 \text{ P} \quad (4)$$

Où A correspond à l'indice d'anthropisation à un instant donné, $S_{\text{pré salé}}$ à la surface de pré salé à un instant donné et Pop, T, Nm, V et P sont respectivement la population, la température, le niveau marin, la vitesse du vent et les précipitations à ce même instant.

4. Discussion

L'étude des séries temporelles de l'occupation du sol permet d'identifier les trajectoires d'évolution du socio-écosystème de la zone du Vivier-sur-Mer. Les cartes et les séries de surfaces obtenues montrent un changement radical du socio-écosystème. En deçà de la digue, on passe d'un bourg entouré d'un paysage bocager à une petite ville au bâti organisé (quartiers résidentiels à voirie curviligne, zone industrielle portuaire) entourée d'un *openfield*. Les séries temporelles corroborent cette observation : l'évolution de la surface des arbres et celle de la surface des champs sont symétriques. Lorsque les champs s'étendent, l'étendue des arbres et des haies diminue et inversement. L'augmentation significative de la surface bâtie traduit quant à elle le développement de l'activité portuaire et de la ville, et donc l'essor socio-économique à la suite de l'installation du port conchylicole.

Au-delà de la digue, le pré salé se développe : sa surface de 2014 représente plus du double de sa surface en 1947, témoignant d'un apport sous-jacent en sédiments important (Valéry et Radureau 2014).

Ces changements paysagers sont matérialisés par la croissance de la diversité du paysage, mais aussi et surtout par l'anthropisation, qui connaît une croissance polynomiale depuis 1947. Les résultats des régressions linéaires multiples montrent que l'anthropisation croissante et la progradation du pré salé sont la conséquence de forçages exercés par des paramètres exogènes sur le socio-écosystème. La pression anthropique – traduite par la croissance démographique – est le paramètre influençant le plus l'évolution du paysage. Ces résultats constituent non seulement un modèle prédictif rétrospectif mais aussi prédictif. Le coefficient de détermination élevé obtenu pour les deux modélisations laisse penser que l'évolution future du socio écosystème pourrait être modélisée à partir des projections climatiques du GIEC (IPCC 2018) et des projections de croissance démographique de l'INSEE ou des Nations Unies.

L'étude de l'interaction des composantes du socio-écosystème renseigne également sur les trajectoires d'évolution observables. Le graphique de la Figure 4 lie la surface du pré salé (en ordonnée) et celle du bâti (en abscisse). Les valeurs s'étendent de 151 551 pixels à 1 586 695 mètres carrés.

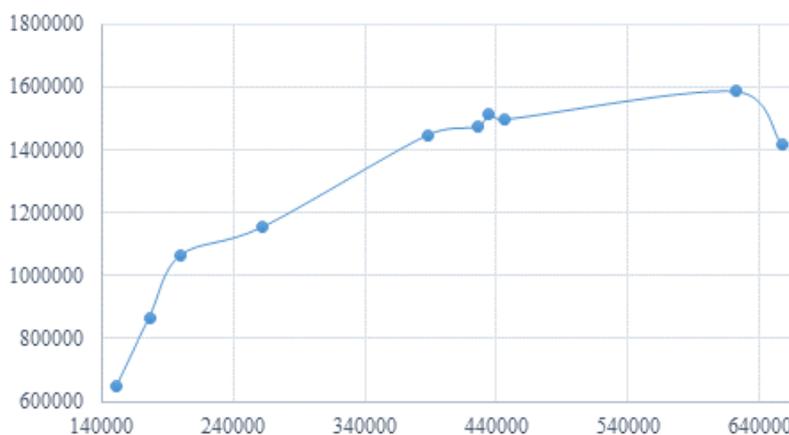


FIGURE 5. Surface des prés salés en fonction de la surface du bâti en mètres carrés de 1947 à 2014

L'étude de l'évolution de la surface du pré salé en fonction de celle du bâti – intrinsèquement liées comme le montrent les résultats des régressions – montre que le développement du bâti entraîne celui du pré salé, les courbes évoluant quasiment parallèlement entre 1960 et 2002. Ceci est cohérent, puisque l'artificialisation des sols change la dynamique d'écoulement des eaux, permettant un apport de sédiments au pré salé via le fleuve. Il semblerait toutefois qu'il existe un pallier d'augmentation du bâti au-delà duquel l'effet inverse se produit : depuis 2002, le

Trajectoire socio-écologique d'une commune entre 1947 et 2014

bâti continue d'augmenter mais le pré salé connaît un net recul. Ceci peut s'expliquer par l'invasion croissante du pré salé par les activités humaines : entre 2002 et 2014, les infrastructures portuaires s'agrandissent, grignotant toujours plus sur le pré salé, et la zone de pré salé défrichée pour des usages loisir à l'Ouest de la zone est plus étendue. Les facteurs naturels, matérialisés par une tempête en 2014, ont également un rôle-clé dans ce changement de trajectoire. Cependant, on peut supposer que cela laisse présager un changement d'état du pré salé, qui pourrait subir de manière plus marquée les forçages anthropiques (Folke *et al.* 2011).

Il pourrait être intéressant de pousser l'analyse de résilience du socio-écosystème plus loin, afin de mieux anticiper les changements à venir. Pour cela, une analyse de l'évolution des taux de croissance de chacune des classes pourrait être utile, et confirmerait peut-être l'arrivée d'une bifurcation dans la trajectoire d'évolution du pré salé (Scheffer 2010). Il serait également opportun d'exprimer statistiquement le phénomène de débocagisation observable par la simple observation des cartes. En effet, celui-ci a des conséquences évidentes sur le ruissellement des eaux et donc des sédiments terrestres vers le pré salé.

Finalement, l'étude montre que l'analyse des tendances d'évolution des différentes occupations du sol d'un socio-écosystème permet d'expliquer les dynamiques d'évolution des prés salés. Il est cependant important de noter que la méthode de classification retenue et le recours à une correction par photo-interprétation ont un impact sur la précision des valeurs de surfaces calculées. Si les données chiffrées pouvaient être précisées par une méthode de classification plus performante (ex., apprentissage profond, Letard *et al.* 2020), l'allure des séries temporelles obtenues permettrait toutefois bien d'envisager les trajectoires socio-écologiques rétrospectives et prospectives du site d'étude du Vivier-sur-Mer. Celui-ci semble se rapprocher d'une bifurcation dans son régime d'évolution : la pression anthropique, force motrice principale de la progradation du pré salé, pourrait avoir franchi un seuil et devenir un frein à celle-ci.

Bibliographie et références

- Allen J.R.L. (2000). Morphodynamics of Holocene salt marshes : a review sketch from the Atlantic and Southern North Sea coasts of Europe. *Quaternary Science Reviews*, vol. 18, n°12, p. 1155-1231.
- Archer A.W. (2013). World's highest tides: Hypertidal coastal systems in North America, South America and Europe. *Sediment. Geol.*, vol. 1, n°25, p. 284-285.
- Barbier E., Hacker S.D., Kennedy C., Koch E.W., Stier A.C. *et al.* (2011). The value of estuarine and coastal ecosystem services. *Ecological Monographs*, vol. 81, n°2, p. 169-193.
- Bouchard V., Digaire F., Lefeuvre, J-C., Guillon L-M. (1995). Progression des marais salés à l'ouest du Mont-Saint-Michel entre 1984 et 1994. *Mappemonde*, vol. 4, p.28-34.
- Deegan, L.A., Johnson D.S., Warren R.S., Peterson B.J., Fleeger J.W. *et al.* (2012). Coastal eutrophication as a driver of salt marsh loss. *Nature*, vol. 490, n°7420, p. 388-392.

- Détriché S., Susperregui A-S., Feunteun E., *et al.* (2011). Interannual (1995-2005) morphodynamic evolution of macro-tidal salt marshes in Mont-Saint-Michel Bay (France). *Continental Shelf Research*, vol. 3, n°6, p. 611-630.
- Fagherazzi S., Marani M., Blum L.K. (2004). *The ecogeomorphology of tidal marshes*. American Geophysical Union, Washington.
- Fagherazzi S., Kirwan M.L., Mudd S.M., Guntenspergen G.R., Termmerman S. *et al.* (2012). Numerical models of salt marsh evolution: ecological, geomorphic, and climatic factors. *Reviews of Geophysics*, vol. 50, n°1.
- Folke C., Jansson A., Rockström J. *et al.* (2011). Reconnecting to the Biosphere. *AMBIO*, vol. 40, n°719.
- Gedan K.B., Silliman B.R., Bertness M.D. (2009). Centuries of human-driven change in salt marsh ecosystems. *Annual Review of Marine Science*, vol.1, p.117-141.
- IPCC (2018). *Climate Change 2013: The physical science basis. Contribution of Working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge United Kingdom and New York, NY, USA, 1588 p.
- Johnson B.J., Lovelock C.E., Herr D. (2018). Climate regulation: salt marshes and blue carbon. *The Wetland Book*. Springer, Dordrecht, p. 1-12.
- Letard M., Collin A., Pastol Y. (2020). Improving mangrove and salt marsh mapping using superpixel-based deep classification of VHR superspectral satellite imagery and topobathymetric LiDAR data, *Sea Tech Week 2020, 12-15 Oct. 2020*, Brest.
- McOwen C.J., Weatherdon L.V., Van Bochove J-W., Sullivan E., Blyth S. *et al.* (2017). A global map of saltmarshes. *Biodiversity Data Journal*, vol. 5.
- Mury A., Collin A., Etienne S., Jeanson M. (2020). Wave attenuation service by intertidal coastal ecogeosystems in the Bay of Mont-Saint-Michel, France: Review and Meta-Analysis. *Estuaries and Coastal Zones in Times of Global Change*. Proceedings of ICEC-2018, p. 555-572.
- Scheffer M. (2010). Foreseeing tipping points. *Nature*, vol. 467, p. 411-412.
- Townend I., Fletcher C., Knappen M., Rossington K. (2011). A review of salt marsh dynamics. *Water and Environment Journal*, vol. 25, p. 477-488.
- Valéry A., Radureau, A. (2014). Evolution de la végétation des marais salés de la baie du Mont-Saint-Michel—Analyse cartographique 1984–2013. *Rapport Convention CREBS/Université de Rennes1/Conservatoire du Littoral/Inter-SAGE baie du Mont-Saint-Michel/Agence de l'Eau Seine-Normandie/Agence de l'eau Loire-Bretagne*. Rennes, France, 15p.
- Wiberg P.L., Fagherazzi S., Kirwan M.L. (2020). Improving predictions of salt marsh evolution through better integration of data and models. *Annual Review of Marine Science*, vol. 12, n°1, p. 389-413.
- Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington, DC.

Communications par Session

Environnement urbain



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Isovist computation of outdoor environment with semi-dense line SLAM and monocular camera

Thomas Le Jan¹, Myriam Servières¹, Thomas Leduc²,
Vincent Tourre¹

1. AAU, Ecole Centrale de Nantes, Nantes

thomas.le-jan@eleves.ec-nantes.fr,firstname.lastname@ec-nantes.fr

2. AAU, CNRS, ENSA Nantes, Nantes

thomas.leduc@crenau.archi.fr

ABSTRACT. Isovists are powerful tools for the morphological analysis of urban spaces, allowing us to account for the characteristics of the visible space from a given point of view. However, computing isovist field from digital model is limited by the quality of the model, and computing isovist field of the real environment with LiDaR sensors is limited by its cost and the setup of cumbersome equipment (eg. on a car). To allow the use of lightweight devices, we propose a method to compute isovists from a single monocular camera, catching a path in an urban environment. This method uses SLAM (Simultaneous Localization and Mapping) and ray casting algorithms. Our results are compared against digital model's isovists.

RÉSUMÉ. Les isovists sont des outils performants d'analyse morphologique des espaces urbains permettant de rendre compte des caractéristiques de l'espace visible à partir d'un point de vue donné. Cependant, le calcul des champs d'isovists à partir d'un modèle numérique est limité par la qualité de ce modèle, et le calcul des champs d'isovists d'environnement réel à l'aide de capteurs LiDaR reste limité par son coût et la mise en place d'un équipement encombrant (par exemple, sur un véhicule). Pour permettre l'utilisation d'équipement portable, nous proposons d'établir une méthode d'obtention d'isovists à partir d'une caméra monoculaire captant un parcours dans un environnement urbain. Cette méthode est basée sur les algorithmes de SLAM (Simultaneous Localization And Mapping) et de lancer de rayons. Les résultats obtenus par notre méthode sont comparés à ceux obtenus sur un ensemble d'isovists calculés sur un modèle numérique 3D simplifié de la ville.

KEYWORDS: Isovists, camera, space reconstruction

MOTS-CLÉS : Isovists, caméra, reconstruction de l'espace

1. Introduction

Isovists and isovist fields are conceptual tools used to carry out morphological analyses of architectural and urban spaces. They were developed and theorized in the 70s (Benedikt, 1979). An isovist determines the available space in a direct line of sight from a given position in an architectural or urban environment. As the visibility of the urban space is part of the urban design qualities identified by urban planners (Ewing, Handy, 2009), isovists are particularly relevant to evaluate the quality of an urban space. Depending on the dimension of the space, these isovists have a polygonal (2D) or polyhedral (3D) shape.

They're usually computed from a digital urban model extracted from a GIS, which is a simplified representation of the urban environment. Some methods can calculate isovists along a route caught with sensors, such as LiDaR (Schmid, Stülpnagel, 2018), but the use of these powerful sensors is not accessible to as many people as possible and this may be very inconvenient because of the cumbersome equipment that requires additional transport (e.g. on a car). Although there is more consumer LiDaR equipment available, here we focus on sensors that are embedded in everyone's smartphones. It is possible to reconstruct a map of the environment with lightweight sensors, for example monocular cameras, using SLAM algorithms (Simultaneous Localization And Mapping). SLAM addresses the computational issue of constructing or updating a map of an unknown environment while simultaneously keeping track of the observer's location. Time-of-flight camera are also available on several smartphones, but their actual depth performances and sensitivity to lighting conditions, among others drawbacks, prevent us to use them to reconstruct outdoor environments.

In this article, we propose a method to delineate an isovist field from a video caught on a simple monocular camera (video-based isovists), using SLAM algorithms. To check the consistency of our results, we will confront them to the isovists generated from a digital urban model (mockup-based isovists).

2. Previous Works

SLAM algorithms provide a fairly accurate estimation of the trajectory of a camera while reconstructing the environment in which the camera evolves. Usually, SLAM algorithms return a point cloud mapping the environment, so-called *reconstructed map*. Depending on the SLAM method, the point cloud can be sparse (e.g. ORB-SLAM2 (Mur-Artal, Tardos, 2017), PTAM (Klein, Murray, 2007)), dense (e.g. DTAM (Newcombe *et al.*, 2011)) or semi-dense (e.g. (Engel *et al.*, 2014)). The sparse methods detect feature points on keyframes and they are assigned 3D coordinates in the *reconstructed map*. The dense methods rather give depth to almost every pixel of a keyframe. The map is then reconstructed by overlapping the different depth maps. Finally, semi-dense methods fall midway between sparse and dense methods. They detect

Outdoor isovist computation with SLAM

areas of interest in keyframes and assign a depth to every pixel of these areas. Unfortunately, these methods each present some weaknesses. Indeed, a sparse point cloud may be fairly accurate but, due to its lack of density, it does not permit to reconstruct real obstacles (there are not enough feature points detected). Dense and semi-dense methods can return much denser point clouds but the returned maps will be very noisy.

Some recent methods (He *et al.*, 2018) can reconstruct a map of the environment with lines instead of points, which makes it easier to reconstruct accurate surfaces using Delaunay triangulation. The method implemented in this paper uses a semi-dense version of ORB-SLAM2 (Mur-Artal, Tardos, 2015) to generate a semi-dense point cloud and extract lines from it.

SLAM methods often include loop closure which improves tracking and mapping by recognizing places already visited by the user. Then the associated map points and camera poses are adjusted according to their previous values.

There are some methods that allow isovist computation from point clouds (Díaz-Vilariño *et al.*, 2018). These methods are based on space discretization. Once space is subdivided into voxels, we determine if the voxels are occupied or empty by the number of points they contain. Then by ray-casting, we determine if the voxels are visible or obstructed which gives us our isovist. There are limits to these kinds of methods. The first is that the returned isovists won't be a vector object, as it is computed in a discretized space. The second is that the method expects a dense and accurate point cloud to clearly model the environment, which is not the case of the previous common visual SLAM methods.

The other simple way to compute isovists as vector objects is to apply ray-casting on surfaces, instead of points. That is possible with semi-dense-line SLAM (He *et al.*, 2018) and its surface reconstruction method.

3. Proposal – Lightweight Real Isovist: LIRI

Our method's workflow is shown in Fig. 1. It is based on semi-dense-line SLAM algorithm (He *et al.*, 2018), which can reconstruct surfaces from semi-dense point cloud obtained by video processing (Fig. 2). We use a smartphone monocular camera but this method is standard enough to be applied to any type of monocular camera provided that we know its intrinsic parameters, which can be computed by camera calibration. So we have a 3D model of the environment made of triangles and an estimation of the sensor trajectory in the *reconstructed map*. The different poses returned by the algorithm are located in the coordinate system of the first detected keyframe. The global orientation of the *reconstructed map* then depends on the orientation of the camera lens during the detection of the first keyframe. In order to ease the comparison between video-based isovists and mockup-based isovists and to stabilize the

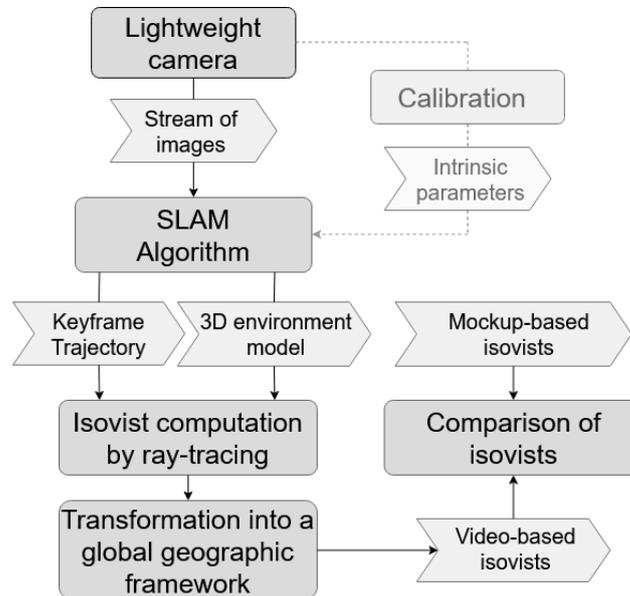


Figure 1. Workflow of our method.

skyline, we use an additional phone stabilizer, minimizing the camera orientation error. The phone stabilizer doesn't impact the standard and lightweight setup hypothesis. The resulting model has some outlier triangles that don't represent real assets so we apply a filter upon the area and the perimeter of the triangles to eliminate the outliers. This filter was set up experimentally.

We use a ray-tracing based method to compute isovists from the *reconstructed map* of the environment and estimated trajectory. We scan the environment using a panoptic approach in the horizontal plane of vision. To compute ray-tracing, we use the Möller-Trumbore intersection algorithm (Möller, Trumbore, 2005) that can quickly calculate the intersection point between a triangle and a ray in 3D space, without using plane equations. For each ray, we look at every triangle and determine the intersection point, if it exists, and we only keep the nearest intersection point from the ray origin. By connecting the intersection points obtained in the different directions we can reconstruct the isovist for a given position. Finally, if we aggregate the different isovists obtained by the positions along a trajectory, we have the isovists field related to that trajectory.

To compare our method to the topographic-based model, we have to transform our data in a geographical context. So we have to apply scaling, translation, and homothety (dilation) but also extract geographical coordinates of

Outdoor isovist computation with SLAM

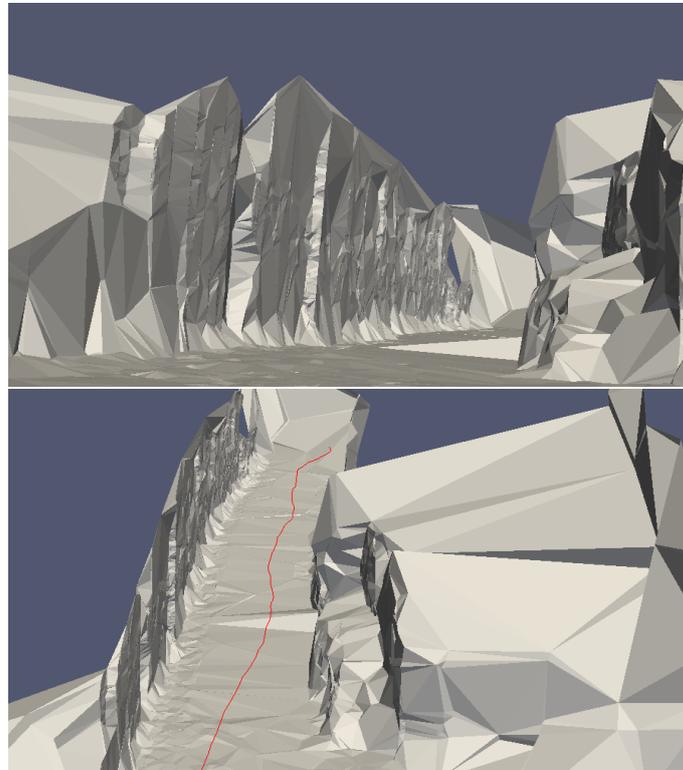


Figure 2. Surfaces reconstructed with semidense-line SLAM method. The red line represents the trajectory of the sensor.

one of the points of the trajectory. In our case we focus on the first point which corresponds to the first moment a keyframe is detected. The position of the camera associated with this first keyframe is predominant; we determine it empirically with the help of orthophotographs. We use a similar method to get the scale factor: we determine manually the geographical location of two points of the trajectory, using a GIS and orthophotographs, and we calculate the distance between them in the *reconstructed map* and in the topographic-based model. By applying this factor to our data, it is now possible to implement the estimation of the trajectory in the digital model to compute mockup-based isovists¹. We can also compare the isovists obtained by both models.

1. To calculate the mockup-based isovists, we use standard topographic data sets such as the ones provided by the IGN BD TOPO® database (June 2020 edition) and the t4gpd Python plugin (Leduc, Leduc, 2020).

4. Results

We present here the results on one dataset collected in the city of Nantes. We suppose that the user made a single one-way journey. We also implement an arbitrary limit distance for the ray-casting, based on the size of the *reconstructed map*. The computation has been conducted on a PC with 8GB RAM, 8x Intel Core i7 CPU 2.60 GHz, using Ubuntu 16.04 64 bits. With these characteristics, semi-dense-line SLAM ran for half an hour to fully map the environment from a one minute video. Then our method takes about two minutes to generate the 164 isovists related to the video.



Figure 3. Comparison of isovists field. Left: video-based isovists.
Right: mockup-based isovists.

A quick comparison (Fig. 3) of the video-based and mockup-based isovists shows the similarities and the differences between the two models. Thus, the video-based isovists from the first poses of the trajectory are very circular, because of the limited field of view of the camera that doesn't allow the SLAM method to reconstruct the environment behind the observer. As seen in Fig. 3 left, as the scene behind the observer is not reconstructed, there is no mask (i.e. no building facades) to stop the rays. This lack of mask casts the rays to the limits of the artificial horizon (which is here of very limited range) which explains the circularity of the resulting shape. The video-based and mockup-based isovists attached to the following poses seem pretty similar. Finally, the last poses show again some differences. This is caused by the fact that video-based isovists take into account walls delineating private plots while mockup-based isovists only consider buildings surfaces. The fences detected in the video are not present in the 3D model derived from standard datasets.

Outdoor isovist computation with SLAM

To go beyond visual comparison of the different isovists, we used four morphological indicators:

- Perimeter : P
- Area : A
- Ellipticity factor : $\epsilon = \pi \times L^2 / 4A$
- Convexity defect : $e_{conv} = A / A_{hull}$

With L , the largest distance between two vertices of the polygon and A_{hull} the area of the convex hull of the polygon.

The ellipticity factor reflects the compression of a circle to form an ellipsoid and the flattening of the isovist. The convexity defect reflects the convex aspects of an isovist, especially the presence of concavity.

All these four parameters give information upon the global form of an isovist. We then compute these parameters for one in ten of the isovists.

By observing the Fig. 4, 5 and 6 we can separate the isovists into three main groups whose limits are indicated by green dashed lines in the figures. The first group may contain the isovists corresponding to the first poses of the trajectory. In this group, the video-based and mockup-based isovists show some differences. Indeed the first video-based isovists of the trajectory are almost circular, because of the lack of information of the *reconstructed map* in the first poses of the trajectory. This lack of information is due to the camera's limited field of view. This period during which the SLAM algorithm has not reconstructed the environment with enough information yet could be considered as an initialization phase.

The second group contains the next isovists. Here, the video-based and mockup-based isovists show much more similarities than in the first group. This is because there is enough information in the video to fully map the environment and the digital model is similar to what we can see in the video.

Finally, the third group contains the isovists of the last poses of the trajectory. Here the video-based and mockup-based isovists show differences. This may be explained by the fact that the video shows the reality of the environment, with additional occlusions such as walls or fences delineating private plots, while the digital model only considers buildings as polygons. The environment around the latter part of the trajectory contains a lot of these additional occlusions.

It can also be seen from Fig. 7 that the convexity defect graph of video-based isovist is less constant and overall higher than the mockup-based isovist convexity defect graph. This indicates the presence of more occlusion leading to more concavity, which corresponds to the less smooth real environment.

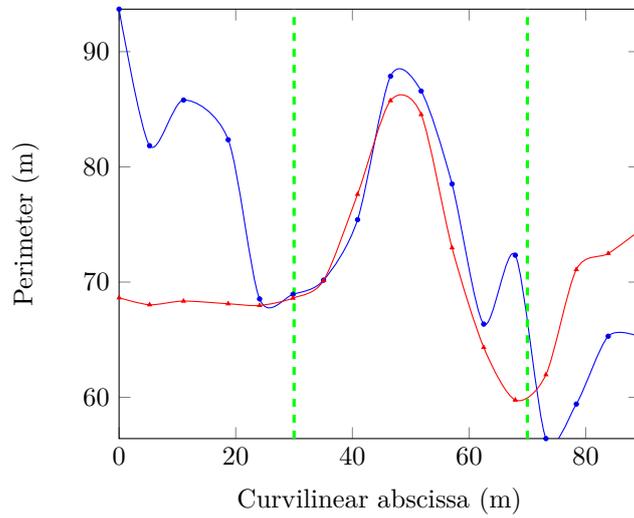


Figure 4. Perimeters (m) of video-based (blue) and mockup-based (red) isovists.

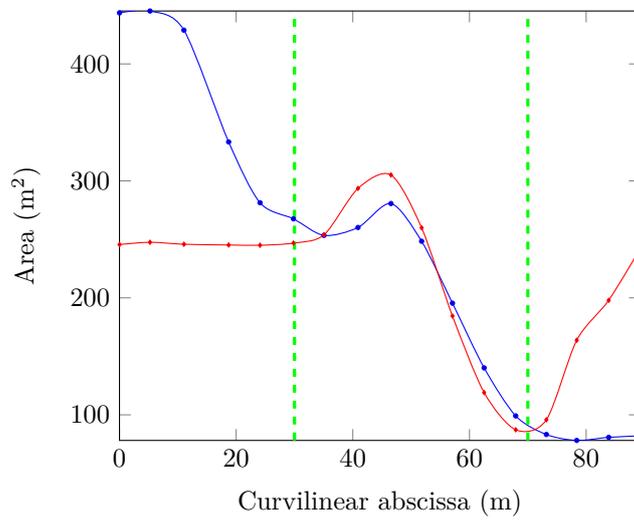


Figure 5. Areas (m²) of video-based (blue) and mockup-based (red) isovists.

Outdoor isovist computation with SLAM

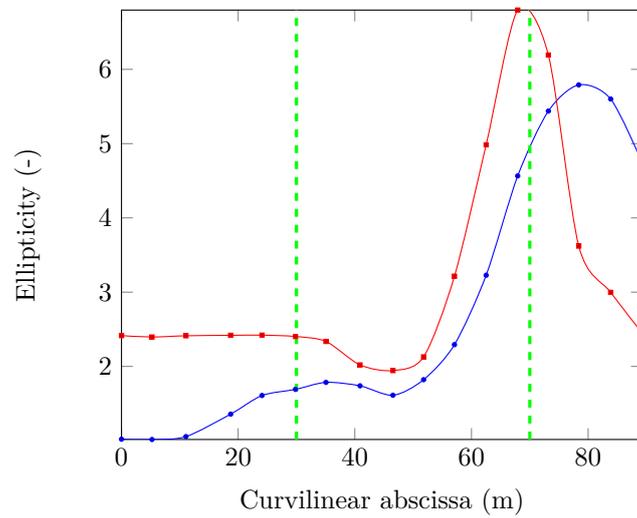


Figure 6. Ellipticity of video-based (blue) and mockup-based (red) isovists.

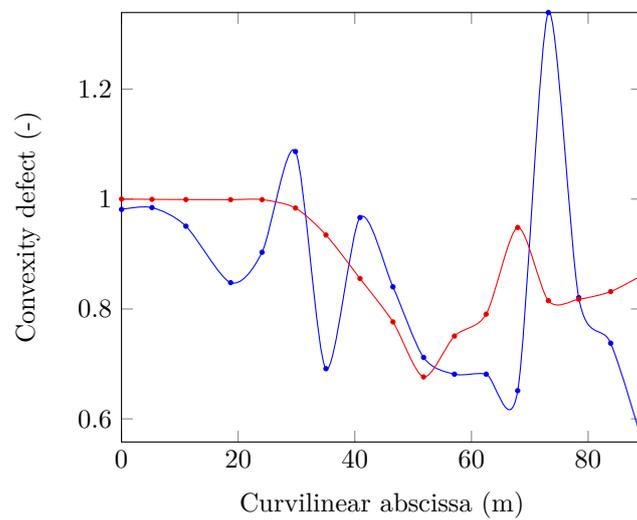


Figure 7. Convexity defect of video-based (blue) and mockup-based (red) isovists.

In this way, video-based isovists generated by our method more accurately reflects the reality of the environment, as long as there are enough data provided by the video.

5. Conclusion, Discussion and Perspectives

In this paper, we presented a process to generate isovists from a video recorded with a monocular camera, in an urban street. If the video shows enough information, the video-based isovists are pretty similar to the ones generated using standard topographic datasets, provided the viewing distance. There is even evidence to suggest that our method gives additional details upon the real environment, compared to the digital urban model, by taking into account some obstacles that are not houses or buildings.

Our method can be used to further improve the digital model by taking into account real assets obstructing the user's view represented by video-based isovists; thus tackling the well known updating problem in GIS. Moreover, the video-based reconstruction of the shape of the immediate surroundings could help several applications, eg. the display of geolocalised information in augmented reality.

Our method cannot run in real-time due to the important amount of calculation during the mapping of the environment in the semi-dense-line SLAM algorithm. This forced us to use short videos to test our method and thus to work on small environments, and this is why we limited the range of the isovists. This issue could be solved using a more powerful computer and longer videos of the environment. We could also use the loop closure of the SLAM algorithm to optimize the *reconstructed map* by recognizing places already visited by the user.

Another axis of improvement would be to automatize the acquisition of world coordinates and scale factor of the *reconstructed map* of the environment and the estimated trajectory. This may be done by getting the camera's meta-data giving GPS coordinates and orientation of the sensor. This would also allow an automatic correspondence of video-based and mockup-based isovists.

Our method is not dependant on any kind of device, and any monocular video – if we know the intrinsic parameters of the camera – can be taken as an entry. Considering that a limit of the video-based isovists is the limited field of view of a simple monocular camera, we may wonder about the possibility to extend this method to omnidirectional cameras. Indeed omnidirectional cameras can catch videos with a very large field of view and adapting this method to this kind of camera only demands to adapt the mathematical model of the SLAM algorithm from pinhole model to unified model (Caruso *et al.*, 2015). Using omnidirectional cameras implies processing more pixels which

Outdoor isovist computation with SLAM

may make the calculations heavier but we would have complete video-based isovists rather than half video-based isovists for the first poses of the trajectory.

Furthermore, it would be interesting to add semantic treatment to our video in order to optimize the surface reconstruction. It's an approach which is already used by MIT-SPARK (Rosinol *et al.*, 2020) but with stereo cameras. This approach may soon be extended to monocular cameras which could also be inserted in our method.

Another very recent method (Yang *et al.*, 2020) allows user to reconstruct in real-time a dense surface mesh of the environment with a mobile device equipped with an embedded monocular camera. This method has given interesting results in indoor environments. We could imagine applying it to outdoor environments to replace the semidense-line SLAM.

As future work, we could directly compare isovist field extracted from LiDaR point cloud with mock-up based isovist. Such a comparison could validate our method, taking the point cloud as a reference.

References

- Benedikt M. L. (1979). To take hold of space: isovists and isovist fields. *Environment and Planning B: Planning and Design*, Vol. 6, No. 1, pp. 47–65.
- Caruso D., Engel J., Cremers D. (2015, sep). Large-scale direct SLAM for omnidirectional cameras. In *2015 IEEE/rsj international conference on intelligent robots and systems (iros)*, pp. 141–148. IEEE. Retrieved from <http://ieeexplore.ieee.org/document/7353366/>
- Díaz-Vilariño L., González-deSantos L., Verbree E., Michailidou G., Zlatanova S. (2018, 9). From point clouds to 3D isovists in indoor environments. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-4, pp. 149-154. Retrieved from <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-4/149/2018/>
- Engel J., Schöps T., Cremers D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. In *European conference on computer vision (eccv)*, pp. 834–849. Retrieved from http://link.springer.com/10.1007/978-3-319-10605-2_54
- Ewing R., Handy S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design*, Vol. 14, pp. 65-84.
- He S., Qin X., Zhang Z., Jagersand M. (2018). Incremental 3D Line Segment Extraction from Semi-dense SLAM. *Proceedings - International Conference on Pattern Recognition*, Vol. 2018-August, pp. 1658–1663.
- Klein G., Murray D. (2007, nov). Parallel Tracking and Mapping for Small AR Workspaces. In *IEEE international symposium on mixed and augmented reality (ismar)*. IEEE. Retrieved from <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4538852>

- Leduc T., Leduc M. (2020). Minimum-area ellipse bounding an isovist: towards a 2D GIS-based efficient implementation. *International Journal of Geographical Information Science*, pp. 1–24.
- Möller T., Trumbore B. (2005). Fast, minimum storage ray/triangle intersection. *ACM SIGGRAPH 2005 Courses, SIGGRAPH 2005*, No. 1, pp. 1–7.
- Mur-Artal R., Tardos J. (2015, jul). Probabilistic Semi-Dense Mapping from Highly Accurate Feature-Based Monocular SLAM. In *Robotics: Science and systems xi*. Robotics: Science and Systems Foundation. Retrieved from <http://www.roboticsproceedings.org/rss11/p41.pdf>
- Mur-Artal R., Tardos J. D. (2017, 10). Orb-SLAM2: An open-source SLAM system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, Vol. 33, pp. 1255-1262. Retrieved from <http://IEEEExplore.IEEE.org/document/7946260/>
- Newcombe R., Lovegrove S., Davison A. (2011). DTAM: Dense tracking and mapping in real-time. In *Proc. of the intl. conf. on computer vision (iccv), barcelona, spain*, Vol. 1. Retrieved from http://www.doc.ic.ac.uk/~ajd/Publications/newcombe_et_al_iccv2011.pdf
- Rosinol A., Abate M., Chang Y., Carlone L. (2020, may). Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. In *2020 IEEE international conference on robotics and automation (icra)*, pp. 1689–1696. IEEE. Retrieved from <https://ieeexplore.ieee.org/document/9196885/>
- Schmid K., Stülpnagel R. von. (2018). Quantifying local spatial properties through lidar-based isovists for an evaluation of opinion-based vgi in a vr setup. In P. Kiefer, H. Huang, N. V. de Weghe, M. Raubal (Eds.), p. 173-178. ETH Zurich. Retrieved from <https://www.research-collection.ethz.ch/handle/20.500.11850/225613>
- Yang X., Zhou L., Jiang H., Tang Z., Wang Y., Bao H. *et al.* (2020). Mobile3Drecon: Real-time monocular 3D reconstruction on a mobile phone. *IEEE Transactions on Visualization and Computer Graphics*, pp. 3446-3456.

Implementation of a solar model and shadow plotting in the context of a 2D GIS: challenges and applications for the cooling effect of tree-covered based greening solutions in urban public spaces

Thomas Leduc¹, Xenia Stavropulos-Laffaille¹, Ignacio Requena-Ruiz¹

1. AAU-CRENAU, UMR 1563 CNRS/ECN/ENSAG/ENSAN

ENSA Nantes, 6 quai François Mitterrand, BP 16202, 44262 Nantes, France

thomas.leduc@crenau.archi.fr

RESUME. L'adaptation des espaces publics aux épisodes de chaleur intense est désormais un défi majeur pour la ville. Dans cette optique, cet article présente une contribution qui vise à délimiter et à traiter l'ombre au sol, qu'elle provienne des bâtiments ou du couvert arboré. Après une comparaison aux tracés des ombres obtenues via deux outils de référence, nous présentons un site urbain qui mélange des ombres d'origines différentes ainsi que divers indicateurs. L'objectif est ici d'aider les concepteurs urbains à trouver des solutions permettant de distinguer les contributions respectives des ombres, leur évolution annuelle et les éventuelles continuités d'ombres spatiales ou temporelles.

ABSTRACT. The adaptation of public spaces to episodes of intense heat is now a major challenge for the city. With this in mind, this article presents a contribution aimed at delineating and handling the ground shadow, whether it comes from buildings or from the tree canopy. After a comparison with shadows obtained via two reference tools, we present a real urban site that mixes shadows of different origins and, in addition, different indicators. The aim of these indicators is to assist urban designers with solutions that make it possible to distinguish the respective shadow contributions, their annual evolution and potential spatial or temporal continuities.

MOTS-CLES : carte des ombres, ombre urbaine, ombre du couvert arboré, microclimat urbain, conception basée sur l'ombre

KEYWORDS: shade map, urban shade, tree canopy shade, urban microclimate, shade-based design

1. Introduction

The urban heat island (UHI)¹ phenomenon (Sobstyl et al., 2018) is a bioclimatic response specific to compact high-rise urban districts, amplifies these overheating, exposes populations to high health risks, and significantly alters living conditions in the built environment (Davis and Gertler, 2015). Both climate and demographic pressures constraint decision-makers to organise urban lifestyles according to mitigation and adaptation strategies, which nowadays focus on how to cool down the urban space in short and mid-term.

(Aram et al., 2019) consider four groups of urban cooling strategies: i) vegetative cover, ii) stack night ventilation, iii) water bodies and iv) modification of surface albedos. (Ruefenacht and Acero, 2017) highlight the relevance of shading as a “key measure to improve Outdoor Thermal Comfort (OTC) and mitigate UHI because it leads to the reduction of air and surface temperature and can therefore result in cooling benefits”. In this regard, (Middel et al., 2016) enhance “the importance of solar access active management”. Considering outdoor air temperature difficult to control, (Aleksandrowicz et al., 2020) observe that sunlight exposure in streets is easier to regulate since it is directly linked to design decisions; therefore, they consider “outdoor shade as a primary comfort indicator”.

We argue that a tool developed in the context of a standard GIS (based on shade index, tree canopy cover, etc.) can help in quantifying the microclimatic qualities of urban spaces revealing, for instance, sunlit areas requiring shade intensification or, conversely, strategic shaded areas that deserve to be protected. The strictly geometric aspects of shading is essential in answering simple questions like: where are located the “shadow fences”² generated by the urban environment at a given date and time? What is the fraction of space, for a predefined study area - which is in shadow for a given date and time? What is the tree canopy contribution to the total shade cover? What is the evolution of this contribution over the summer period, the crown remaining unchanged? What areas are shaded for at least 3 hours on a hot summer day? Or even, to satisfy a constraint of temporal continuity, what areas are shaded for at least 3 consecutive hours on a hot summer day? Or even, do these areas guarantee spatial continuity in order to allow comfortable pedestrian mobility?

Simulation softwares such as Solene (Miguet and Groleau, 2002), Ladybug (Roudsari and Pak, 2013), UMEP-SOLWEIG (Lindberg et al., 2018) or ENVI-met³, already allow shadow studies. Some of them (Solene and Ladybug) run as plugins of computer-aided design programs that do not offer the functionalities of GIS (power of synthesis of 2D maps, capacity to process large geographic areas, etc.) and require

¹ As (Ruefenacht and Acero, 2017) summarizes: UHI “occurs because cities consume huge amounts of energy in electricity and fuel, have less vegetation to provide shade and cooling, and are built of materials that absorb and store energy from the sun”.

² We borrow this expression from (Knowles, 2003) to underline the fact that shadowing can also, in itself, generate constraints useful to the urban planner during a design phase (in reference to the solar envelope concept).

³ ENVI-met is a German trademark (No. 304 73 896), <https://www.envi-met.com/>

Solar model and shadow plotting in the context of a 2D GIS

a particular expertise in architectural design. Therefore, their implementation by the geographic information departments of municipalities is difficult. In the contrary, UMEP-SOLWEIG has been embedded into a GIS. Nevertheless, they require digital surface and terrain models. In addition, the raster-based processes do not allow the shadow fences to be handled as simply as standard vector entities.

This article aims to highlight the interest of a specific vectorial development in a standard GIS to answer the questions mentioned above. After a presentation of the solar model and the treatment of mineral and vegetal shadows that we have implemented, we will present the mechanism for calculating multiple shadow overlaps. In the next section, we will highlight the interest of these developments in one case study of the Coolscapes research program. Finally, we will discuss the contributions of our tool to a microclimate sensitive urban space design.

2. Method

Two main principles guide our methodological development: replicability and transportability on the one hand, and ease of implementation on the other. To guarantee these two principles, we use standard topographic data sets such as the ones provided by the IGN BD TOPO® database (June 2020 edition) and a widely distributed free and open source software (QGIS Development Team, 2020). It is noteworthy that the presented method of shadow simulation and mapping is potentially applicable at larger scales, e.g. district or city scale.

2.1. 3D processing in the context of 2D GIS

The sun path geometry and the delineation of the cast shadow generated by buildings and tree canopies are intrinsically three-dimensional urban issues. Our aim is to identify shaded areas that allow the city dweller to cool down; therefore, we simply delimit these areas on a horizontal plane at a pedestrian level (in this case, we choose the ground plane). This simplifying hypothesis allows us, at the end of a 3D processing presented later on, to produce a classical 2D cartography in the context of a standard GIS.

2.2. Implementation of the sun path model

Pysolar and PyEphem are Python implementations of ephemerides that allow estimating the trajectories and positions of the sun over time. To ensure consistency with the laboratory's simulation tools, we have implemented the ephemeris model of the Solene software in the context of our GIS simulation. A systematic comparison of the positions returned by Pysolar with those of our simulation shows that the respective solar heights differ by less than 1% on annual average.

2.3. Shadow cast from a collection of right prisms

The topographic data sets employed here provide a 3D metric and vector description (structured in objects) of the elements of the territory and its infrastructures, usable on scales from 1:5,000 to 1:50,000. More precisely, this

description consists in a set of polygons (sort of flat roof footprints) completed with an attribute equal to the height of the corresponding building. This twofold information allows us to generate a collection of right prisms each one representing a building in the urban model. The direction of the sun's beams being predetermined, it is then possible to i) calculate the ground shadow of each right prism, ii) and to proceed to a geometric union of all shadows.

2.4. Shadow cast from the tree canopies

Nantes Metropole's geographic information services provided us with a database of 75,156 trees in the conurbation. In addition to the point position of each trunk, this data includes various attributes such as the height of the tree, the circumference of the trunk, the species of the tree, etc. In the first version of tree modelling, we developed three different crown models: the first is spherical, the second cylindrical and the third conical. Based on the tree attributes and positions as well as the crown model and the sunbeam direction, we have then implemented a ground shadow model (the projection of a sphere being an ellipse for example). As in the case of the building shadows, at the end of the process we proceed to the geometric union of all tree shadows in order to deal with possible overlaps.

2.5. Management of shadow overlaps over several (consecutive) hours

To identify areas in the shade for several hours (possibly consecutively), we need to make multiple geometric intersections, the combinatory of which are potentially substantial (for example, being in the shade for 2 hours can be translated multiply: 10 am-11 am and 3 pm-4 pm, 1 pm-3 pm, etc.). To deal efficiently with this issue and bypass these numerous intersections, we proceed in three steps: i) geometric union (via the *unary union* operator) of the shadow contours, ii) *polygonization* of the network of polylines thus produced, iii) for each polygon resulting from the *polygonization*, calculation of the number of enclosing shadow zones.

3. Case study and experimental setup

3.1. Case study: Graslin district in Nantes (France)

The study site is located in the centre of Nantes (France), including parts of the Graslin district. This area consists of 13.6 ha, 293 building footprints representing a total floor area of about 6.4 ha, and 98 trees. Its altimetry varies from 6 to 30 meters and the average height of the buildings is about 15.7 meters (with a standard deviation set at 5.8 meters). The explained method is applied to this site with a special interest in the urban garden called "Cours Cambronne". This classic French garden, lined with buildings and trees, covers about 0.9 ha and presents various shaded areas (from both surrounding buildings and 64 trees) that modify comfort indexes and citizens practices (i.e., sitting places and trajectories). It is roughly 180 metres long and 50 metres wide. It also has the advantage of crossing shadows from the tree cover (rows of trees) and the building environment, all on a relatively flat ground.

Solar model and shadow plotting in the context of a 2D GIS

3.2. Experimental setup

In order to test our methodological proposal and respond to the research questions, we have defined several parameters as follows. Regarding the choice of representative dates and times for the annual sunshine simulation, we have chosen the 1st, 11th and 21st of each month. For each of these days, we have made a calculation with a hourly time step from 10 am to 4 pm. Regarding the choice of a representative summer day, we opted for July 21st. Regarding the tree modelling in this area, after empirical evaluation on site, we chose a cylindrical crown model with two distinct parameterizations. Eight trees in the Courtyard have a total height of 15 metres, a crown height of 10 metres and a crown radius equal to 5 metres. The rest of the trees are modelled with a total height of 9 metres, a crown height of 6 metres and a crown radius of 3 metres. It is noteworthy that although the trees in the area are deciduous, we do not consider the change in the crown due to seasonality.

4. Experimental verification

4.1. Comparison of shadows in the context of a CAAD software

The objective of this section is to use the solar model and the shadow engine implemented natively in the SketchUp software, to compare the shadows computed by the t4gpd plugin in the context of a GIS with those of a tool widely used by architectural practitioners. For that, we first import the building footprints and tree positions to build the corresponding 3D model using the SketchUp *pushpull* mechanism. We then import the shadow polygons produced by our simulations (through the SHP file format) and compare them visually to the shadows produced by the SketchUp shadow engine. Fig. 1 shows that these two shadow features match satisfactorily.



FIGURE 1. Comparison of the native SketchUp shadow features (in grey) with those from the t4gpd plugin (red polylines), on July 21 at 10 am UTC.

4.2. Comparison with the Urban Multi-scale Environmental Predictor (UMEP)

As the verification proposed in the previous section is purely visual, we suggest comparing our simulation results with those of the UMEP reference tool. Indeed, the latter embeds the shadow generator plugin capable of generating shadow plots from building, ground or even vegetation digital surface models, in the QGIS context. In order to use UMEP's "Daily Shadow Pattern" functionality, we first had to generate a digital terrain model, a digital model of the built-up surfaces, and a raster model of the tree canopy. We used a metric precision raster model such as the one proposed by the IGN's RGE Alti® to stick as closely as possible to the building layouts. Fig. 2 shows the relatively good correlation of the raster shadows produced by UMEP with the vector traces from our simulations.

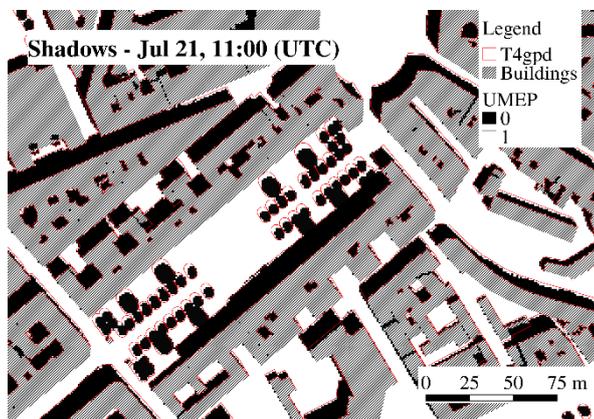


FIGURE 2. Comparison of the shadows from the UMEP Shadow generator (black pixels) with those from the t4gpd plugin (red polylines).

To go beyond this visual comparison, we measured the shadow-covered areas respectively produced by UMEP and our simulation tool at different times on 21st July. Table 1 shows variations ranging from 0.1% to 5.2%. Although this comparative analysis does not formally validate our simulation tool, it reinforces our ability to produce consistent results with the UMEP reference tool.

TABLE 1. Comparison of shadow-covered area measurements (in m²) assessed via the UMEP software vs. via the t4gpd plugin, on July 21st.

	10 am	11 am	12 am	1 pm	2 pm	3 pm
UMEP	35,498.6	29,062.9	27,838.8	27,677.2	28,900.2	36,852.8
T4gpd	33,694.2	29,093.6	28,036.8	27,533.8	28,560.2	36,161.6
Deviation	5.2 %	0.1 %	0.7 %	5.2 %	1.2 %	1.9 %

Solar model and shadow plotting in the context of a 2D GIS

5. Application to the French study site: Cours Cambronne (Nantes)

5.1. Shade ratios

5.1.1. Shade ratio at a given date and time

The first very simple application consists, in delineating the ground shadow due to the tree canopy and the built environment for a given territory, date and time. On 21 July at 12 am (UTC), the 293 buildings and 98 trees in the study site thus produce a ground shade of approximately 2.8 ha, i.e. approximately 20% of the total area. Furthermore, as shown in Fig. 3 (left), few streets oriented almost orthogonal to the sun's beams benefit from shade.

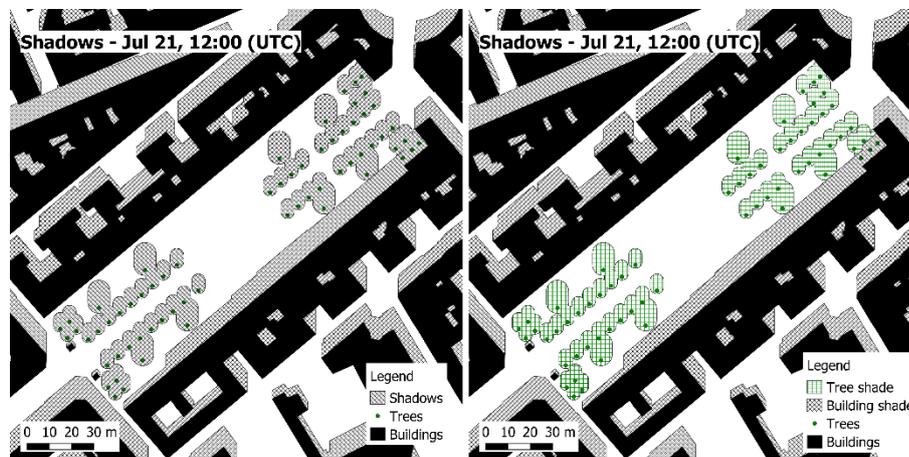


FIGURE 3. (left) The plugin allows evaluating both shadows of buildings and trees on a horizontal plane according to the date and time chosen. The result is a multi-polygon geometry (light grey area). (right) The plugin is able to distinguish, in the shadow cast on the ground, the contribution of trees (light green polygons) from that of buildings (light grey polygons).

5.1.2. Shade contribution of the tree canopy

In the context of urban greening policies, it is essential to distinguish the shadows generated by the built environment (i.e. the buildings) from those produced by the tree canopy. For the configuration shown in Fig. 3 (right), the shade produced by the tree canopy represents about 16% of the total shade on the ground. This contribution is naturally conditioned by the tree crown geometries as well as by their spatial distribution within the urban space (they may indeed be hidden by possible taller buildings).

5.1.3. Shade ratio evolution over the summer period

In this section, the yearly evolution of the shadow ratio for different times a day (between 10 am and 4 pm) is studied (Fig. 4). The considered period is chosen during

summertime, when the sun is at its highest path in the sky and, consequently, the solar energy input is the greatest. Considering the results, for the period between spring equinox and autumn equinox, it can be seen that, the shade provided by the tree canopy is significant. In fact, the tree shades cover on average one third of the courtyard. This contribution is almost systematically higher than the one of the surrounding buildings. At its peak (at 2 pm UTC), it is even around 6 times higher than the shade provided by the buildings.

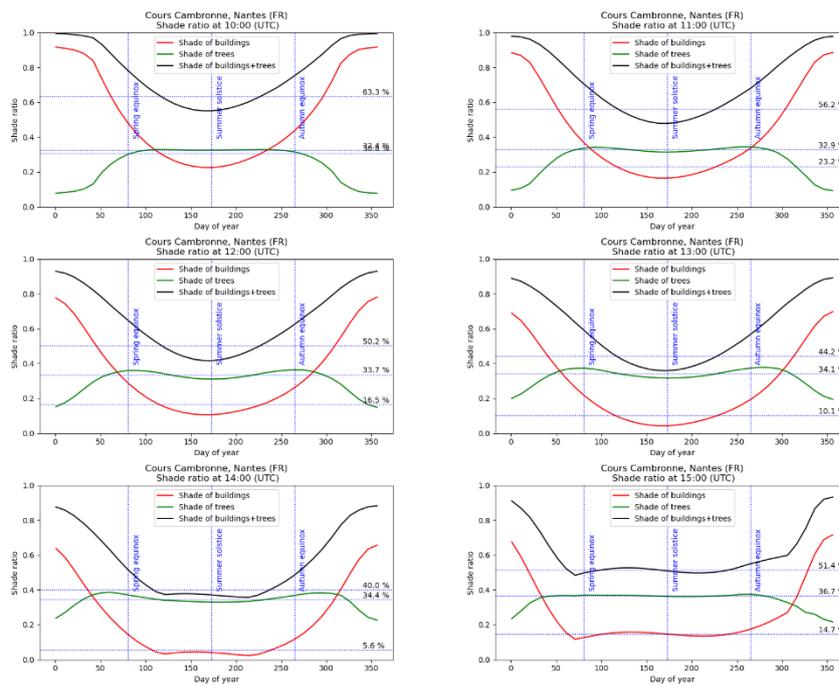


FIGURE 4. Evolution of shade ratios over the year for a few predetermined hours of daylight. The fall of foliage in autumn and winter is likely to modify the geometry of the crown and reduce the amount of shade.

5.2. Shadowing permanence

5.2.1. Delineation of shaded areas over several hours

The annual evolution of the ground shade presented in the previous section can also be studied for a given day. This allows distinguishing potential cool spots, during a heat wave event. Thus, the shade zones are cut out according to the number of hours of shade. With the special case of 21st July, Fig. 5 (left) shows the excessively fragmented nature of this cut within the shadow zone. It can be seen, however, that although the central strip situated halfway between the two rows of trees (longitudinal centre line of the courtyard) is finally fairly exposed, the tree roots are very well protected.

Solar model and shadow plotting in the context of a 2D GIS

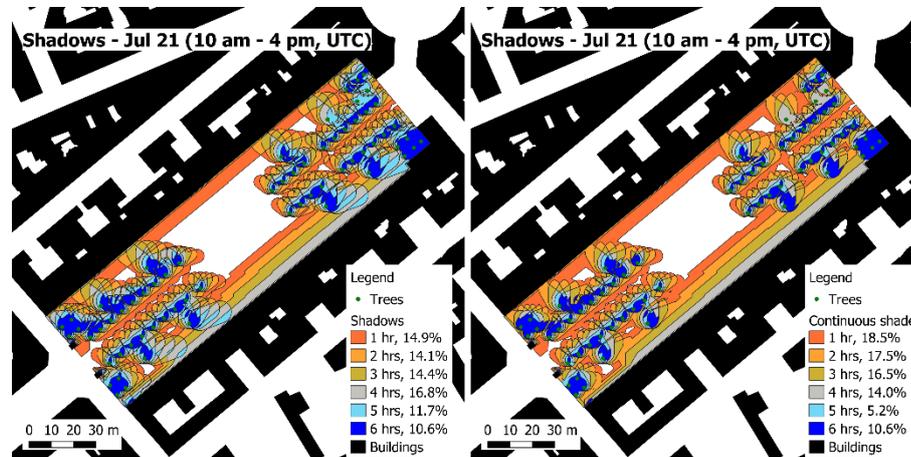


FIGURE 5. (left) Non-uniform distribution of shade in the courtyard during the hours when the sun is at its highest in summer. (right) Representation of areas in shade over several consecutive hours.

5.2.2. Delineation of shaded areas over several consecutive hours

The accumulation of shadow hours shown in Fig. 5 (left) does not include any notion of temporal continuity. However, it could be interesting to detect areas that are permanently in the shade (in order, for example, to design and deploy appropriate street furniture). Fig. 5 (right) shows a revised version of the indicator incorporating this dimension of continuity. The idea here is to represent areas as a function of their exposure to shade for several consecutive hours. As one can note, the overall fragmentation is much less, and patterns emerge that guarantee a certain spatial continuity.

6. Discussion

6.1. Pros: beyond descriptive analysis, shadow as a design object

While it is possible to vectorise the raster shadows obtained via UMEP, the underlying pixelisation is nevertheless problematic when seeking to use shadows for the design of urban spaces as given, for example, in the solar envelope concept (Knowles, 2003). From this point of view, our proposal is innovative in the specific context of geomatics as these vector shadows enable to delineate potential cool-paths and urban oases or cools spots in the urban fabric... all useful features for citizens during heat waves.

6.2. Contras: need for an energy-based approach

It should be noted, however, that the current method is not based on Digital Elevation Models (DEMs); thus, the interactions between the topography and the sun path model are not considered. This is a potential application limit especially for cities

with significant elevation changes. Furthermore, the simplified and strictly geometric approach is not sufficient to properly inform the complex notion of pedestrian thermal comfort in heterogeneous urban spaces. On the one hand, particular building geometries (e.g., walkways, porches, steep rooftops) are not considered. On the other hand, thermal comfort simulation tools such as ENVI-met allow studying the spatial distribution of thermal comfort indexes by integrating an energy-based approach. Supplementing our tool with a more representative geometric approach, based on 3D city models (ex. CityGML (Biljecki et al., 2018)) and a thermo-radiative (or even aerodynamic) model would make it possible to transform this prototype into a more relevant tool for city designers. Such an assessment will require considering the radiative models of the facades in a full 3D approach.

7. Conclusion

Our work contributes to an estimation, in the geomatics context, of tree-shaded zones to building ones, which is crucial for summer thermal comfort. It also makes it possible to go beyond the calculation and descriptive analysis of the shadows cast in the urban space and to use them as standard polygonal shapes. Continuing its development by integrating a comprehensive energy and a more realistic geometric model will make it operational for urban planners.

Acknowledgment

This study is supported by the French National Research Agency (ANR) under reference ANR-18-CE22-0003 (COOLSCAPES project).

References

- Abdel-Aziz D M, Shboul A Al, Al-Kurdi N Y, 2015, “Effects of Tree Shading on Building’s Energy Consumption -The Case of Residential Buildings in a Mediterranean Climate” *American Journal of Environmental Engineering* **5**(5) 131–140
- Aleksandrowicz O, Zur S, Lebediger Y, Lerman Y, 2020, “Shade maps for prioritizing municipal microclimatic action in hot climates: Learning from Tel Aviv-Yafo” *Sustainable Cities and Society* **53**(September 2019) 101931, <https://doi.org/10.1016/j.scs.2019.101931>
- Aram F, Higuera García E, Solgi E, Mansournia S, 2019, “Urban green space cooling effect in cities” *Heliyon* **5**(4) e01339, <https://doi.org/10.1016/j.heliyon.2019.e01339>
- Biljecki F, Kumar K, Nagel C, 2018, “CityGML Application Domain Extension (ADE): overview of developments” *Open Geospatial Data, Software and Standards* **3**(1) 13, <https://opengeospatialdata.springeropen.com/articles/10.1186/s40965-018-0055-6>
- Coutts A M, White E C, Tapper N J, Beringer J, Livesley S J, 2016, “Temperature and human thermal comfort effects of street trees across three contrasting street canyon environments” *Theoretical and Applied Climatology* **124**(1–2) 55–68, <http://link.springer.com/10.1007/s00704-015-1409-y>

Solar model and shadow plotting in the context of a 2D GIS

- Davis L W, Gertler P J, 2015, “Contribution of air conditioning adoption to future energy use under global warming” *Proceedings of the National Academy of Sciences* **112**(19) 5962–5967, <http://www.pnas.org/lookup/doi/10.1073/pnas.1423558112>
- Demuzere M, Orru K, Heidrich O, Olazabal E, Geneletti D, Orru H, Bhawe A G, Mittal N, Feliu E, Faehnle M, 2014, “Mitigating and adapting to climate change: Multi-functional and multi-scale assessment of green urban infrastructure” *Journal of Environmental Management* **146** 107–115, <https://linkinghub.elsevier.com/retrieve/pii/S0301479714003740>
- Emmanuel R, Loconsole A, 2015, “Green infrastructure as an adaptation approach to tackling urban overheating in the Glasgow Clyde Valley Region, UK” *Landscape and Urban Planning* **138** 71–86, <https://linkinghub.elsevier.com/retrieve/pii/S0169204615000432>
- Ewing R, Handy S, Brownson R C, Clemente O, Winston E, 2006, “Identifying and Measuring Urban Design Qualities Related to Walkability” *Journal of Physical Activity and Health* **3**(Suppl 1) 223–240, http://forum.activelivingresearch.com/sites/default/files/JPAH_15_Ewing.pdf
- Feyisa G L, Dons K, Meilby H, 2014, “Efficiency of parks in mitigating urban heat island effect: An example from Addis Ababa” *Landscape and Urban Planning* **123** 87–95, <http://dx.doi.org/10.1016/j.landurbplan.2013.12.008>
- Gehl J, 2011 *Life Between Buildings: Using Public Space* (Island Press/Center for Resource Economics)
- Gillner S, Vogt J, Tharang A, Dettmann S, Roloff A, 2015, “Role of street trees in mitigating effects of heat and drought at highly sealed urban sites” *Landscape and Urban Planning* **143** 33–42, <http://dx.doi.org/10.1016/j.landurbplan.2015.06.005>
- Gunawardena K R, Wells M J, Kershaw T, 2017, “Utilising green and bluespace to mitigate urban heat island intensity” *Science of The Total Environment* **584–585** 1040–1055, <http://dx.doi.org/10.1016/j.scitotenv.2017.01.158>
- Kabisch N, van den Bosch M, Laforteza R, 2017, “The health benefits of nature-based solutions to urbanization challenges for children and the elderly – A systematic review” *Environmental Research* **159**(July) 362–373, <https://linkinghub.elsevier.com/retrieve/pii/S0013935117315396>
- Knowles R L, 2003, “The solar envelope: its meaning for energy and buildings” *Energy and Buildings* **35**(1) 15–25, <http://linkinghub.elsevier.com/retrieve/pii/S0378778802000762>
- Laforteza R, Chen J, van den Bosch C K, Randrup T B, 2018, “Nature-based solutions for resilient landscapes and cities” *Environmental Research* **165**(December 2017) 431–441, <https://linkinghub.elsevier.com/retrieve/pii/S0013935117317115>
- Lindberg F, Grimmond C S B, Gabey A, Huang B, Kent C W, Sun T, Theeuwes N E, Järvi L, Ward H C, Capel-Timms I, Chang Y, Jonsson P, Krave N, Liu D, Meyer D, Olofson K F G, Tan J, Wästberg D, Xue L, Zhang Z, 2018, “Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services” *Environmental Modelling & Software* **99** 70–87, <https://linkinghub.elsevier.com/retrieve/pii/S1364815217304140>
- Martilli A, Krayenhoff E S, Nazarian N, 2020, “Is the Urban Heat Island intensity relevant for heat mitigation studies?” *Urban Climate* **31**(January 2019) 100541, <https://linkinghub.elsevier.com/retrieve/pii/S2212095519300070>

- Middel A, Selover N, Hagen B, Chhetri N, 2016, “Impact of shade on outdoor thermal comfort—a seasonal field study in Tempe, Arizona” *International Journal of Biometeorology* **60**(12) 1849–1861, <http://dx.doi.org/10.1007/s00484-016-1172-5>
- Miguet F, Groleau D, 2002, “A daylight simulation tool for urban and architectural spaces—application to transmitted direct and diffuse light through glazing” *Building and environment* **37** 833–843, <http://www.sciencedirect.com/science/article/pii/S0360132302000495>
- Mullaney J, Lucke T, Trueman S J, 2015, “A review of benefits and challenges in growing street trees in paved urban environments” *Landscape and Urban Planning* **134** 157–166, <https://linkinghub.elsevier.com/retrieve/pii/S016920461400245X>
- QGIS Development Team, 2020, “QGIS Geographic Information System. Open Source Geospatial Foundation Project.”, <http://qgis.osgeo.org>
- Redon E, Lemonsu A, Masson V, 2020, “An urban trees parameterization for modeling microclimatic variables and thermal comfort conditions at street level with the Town Energy Balance model (TEB-SURFEX v8.0)” *Geoscientific Model Development* **13**(2) 385–399, <https://gmd.copernicus.org/articles/13/385/2020/>
- Roudsari M S, Pak M, 2013, “Ladybug: A parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design” *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association* 3128–3135
- Ruefenacht L A, Acero J A, 2017, “Strategies for Cooling Singapore: A catalogue of 80+ measures to mitigate urban heat island and improve outdoor thermal comfort”, <https://www.research-collection.ethz.ch/handle/20.500.11850/258216>
- Sobstyl J M, Emig T, Qomi M J A, Ulm F-J, Pellenq R J-M, 2018, “Role of City Texture in Urban Heat Islands at Nighttime” *Physical Review Letters* **120**(10) 108701, <https://link.aps.org/doi/10.1103/PhysRevLett.120.108701>
- Taha H, Akbari H, Rosenfeld A, Org E, 1991, “Heat Island and Oasis Effects of Vegetative Canopies: Micro-Meteorological Field-Measurements” *Theoretical and Applied Climatology* **44** 123–138, <https://escholarship.org/uc/item/5wt7072h>
- Whyte W H, 1980 *The Social Life of Small Urban Spaces* (Washington, D.C: Conservation Foundation)
- Yin L, 2017, “Street level urban design qualities for walkability: Combining 2D and 3D GIS measures” *Computers, Environment and Urban Systems* **64** 288–296, <http://dx.doi.org/10.1016/j.compenvurbsys.2017.04.001>

Spatial Evolution of Building Types: Methodological proposals

Perez Joan¹, Fusco Giovanni¹, Sadahiro Yukio²

1. Université Côte d'Azur, CNRS, ESPACE, Nice, France.

2. Department of Urban Engineering, University of Tokyo, Tokyo, Japan.

ABSTRACT. This paper presents different methods implemented with the aim of studying urban perforation and regeneration dynamics at the building level. The minimum requirement to perform these methods is a building footprint available for two different periods (time depth of at least 10 years) and heights (or floors) as attribute data. First, basic morphometric indicators are computed for each building: area, floor-area, number of neighbors, elongation, and convexity. Based on the availability of expert knowledge, different types of classifications and clustering are performed to obtain building types. A grid is superimposed on the test region of Osaka (Japan) and the number of building types per cell and for each period is computed, as well as the differences between each period. Mappings are then performed, showing that building types have specific locations and evolution trajectories. In some extreme cases, a specific building type can even gradually replace a type on a declining dynamic.

KEYWORDS: Spatial analysis, urban perforation, urban regeneration, building type, classification, clustering.

1. Introduction

Cities are in a perpetual flow of transformations with urban areas that constantly make, unmake and remake themselves. Different models are discussed in the literature, mostly focusing on the benefits and/or drawbacks of the increase (urban sprawl, compact cities, regeneration, etc.) or decrease (shrinking cities, perforation, etc.) of building densities. A factor that is often overlooked regarding the evolution of densities is the role of inherited urban forms. Occurrence of areas having difficulties in redefining themselves are often closely related in term of urban form. Of course, even if the form factor is often overlooked, this issue is nothing new. The numerous criticisms of modernist architecture (Jacobs 1961, Salingaros 2005), which brought large-sized and specialized buildings, unable to evolve with the needs of the cities, are a good illustration of the link between urban life and urban form. What has changed more recently however, are the increasing possibilities to explore this link through: (1) the existence of large-sized datasets related to urban form and available at different dates, such as the French BD TOPO ® or the Japanese Zenrin Maps ® and (2) an increase of computing power that allows running complex algorithms on the aforementioned datasets. Examples of such work in the literature include the classification of building footprints by a Random Forest Classifier (Hecht *et al.*, 2015), unsupervised Bayesian clustering for building types identification (Perez *et al.*, 2019a), etc. This paper seeks to emphasize on different needs and precautions that shall be taken while working on the link between urban form and space-time evolutions of urban areas. Suggestions and recommendations while using building footprints as the primary source of data are also provided throughout the paper.

Section 2 presents the data requirement, the test region and the computation of basic indicators of built-up form. Section 3 performs three different classifications, through a decision tree if reliable expert knowledge is available, a double k-means if only limited expert knowledge is available and finally, an unsupervised Self-Organizing Map if no expert knowledge is available. Section 4 creates a grid of a predetermined size over the test region in order to count the numerosity of each building type per cell and their evolution. Thematic maps are also displayed in this section. Section 5 concludes the paper.

2. Data Preparation

2.1. Minimum Data Requirement

The bare minimum dataset required to implement the different methods discussed in this paper is a GIS layer of building footprints with multiple years of consistency (at least two different periods). The gap in years between the layers is dependent upon the scope of the analysis, but also on specificities of the location under study. A gap of 10 years is enough in countries following a deconstruction/reconstruction model of buildings (some countries in Asia, Africa, etc.) while in countries where buildings are usually renovated or rehabilitated, such as in Europe, a minimum time depth of 20 to 30 years is needed to observe structural

Regeneration and Perforation of Building Types

changes in the distribution of building types. Each building shall be digitized as a single unit. Corrections should thus be made to GIS layers which aggregate contiguous buildings (like row houses) into single units. In addition, one attribute is required: building height, and another one is valuable, but not mandatory: building specialization. Specialization attribute allows filtering non-residential buildings to solely focus on residential patterns and dynamics of regeneration and perforation. Yet, due to mixed-use buildings, the distinction between residential and non-residential buildings is hardly ever straightforward. Thus, according to the source of data, attention should be paid to how specialization is encoded.

2.2. Test Region and Data Presentation

The test region is an area of 15 by 18 km in Japan containing central Osaka and its surroundings (**Fig.1 a.**). The peculiarities of Japan are that houses and small collective residential complexes are easily demolished to be reconstructed or to make way to new urban projects (Shelton, 2012) and that urban areas are experiencing urban shrinkage phenomena (Buhnik, 2010). For example, single-family houses have a lifespan of only 30 years (MLIT, 2007), which is in sharp contrast with European countries where houses are renovated and passed down from generation to generation. A short lifespan for buildings is an undeniable advantage for the study of the evolution of building types. Indeed, even now, high-quality datasets of building footprints are uneasy to access and, if they are, historic data of equivalent quality are usually not. Regarding urban shrinkage, demolitions without reconstructions is another interesting phenomenon to monitor over time.

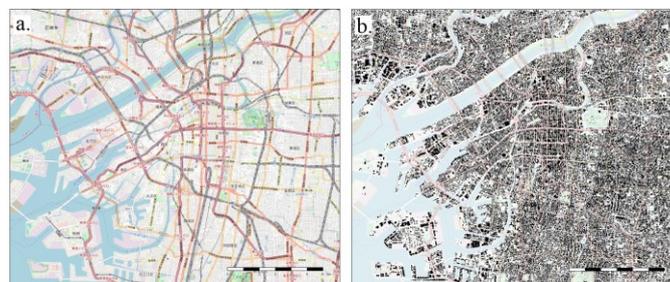


FIGURE 1. a. Osaka, OpenStreetMap. b. Buildings in Osaka in 2013-14 (Zmap TOWN II).

ZENRIN Residential Maps (Zmap TOWN II¹), which are digital maps focusing on the building footprints throughout Japan, are extracted and compiled into a GeoPackage file for the extent of the test region in 2003/04 and in 2013/14 (**Fig 1.**

¹ ZENRIN is a private map information company that holds the top share in the Japanese market for local residential and car navigation maps.
<https://www.zenrin.co.jp/product/category/gis/basemap/zmaptown/index.html>

b.). The original GeoPackage file only contains the two aforementioned building layers, respectively named BU0304 and BU1314. BU0304 is made of 760.067 inputs and BU1314 of 739.536, thus showing a decrease of 2.7% of the raw number of buildings in 10 years. ZENRIN Residential Maps possess several attribute data from which two are of interest in this research: the height of the buildings, expressed in number of floors, and building specialization. Building specialization is an encoded attribute, with codes as follows: 1200 for official and religious buildings (schools, administrative buildings, temple, etc.), 1363 for collective housing, 1364 for single-family homes, 1365 for private offices and mixed-use buildings, etc. In this paper, we focus on residential buildings, both collective and individual housings (1363 and 1364), which concern 469.519 buildings in 2003/04 and 428.875 in 2013/14. Each layer within the GeoPackage file originally contains 3 attribute data: an identifier, the number of floors and specialization.

2.3. Indicator Computation

The first step consists in computing a basic series of morphometric indicators for each building made of: the building footprint surface (area), the total amount of usable floor area, elongation, convexity and the number of adjoining neighbors. Elongation, convexity and the number of adjoining neighbors are detailed in Perez *et al.*, (2019a). They respectively provide a measure of how buildings are elongated compared to the most compact equivalent shape (a circle), how buildings have intricated or squared shapes, and finally, if buildings are free-standing detached structures or possess adjoining neighbors. The final set of variables is made of the newly calculated morphometric indicators plus number of floors and specialization. **Supplementary materials algorithm 1** shows how indicators are easily computed within R using the “sf” and “lwgeom” packages. This algorithm is applied to both BU0304 and BU1314. For the number of adjoining neighbors, we operate a small buffer (**algorithm 1**, # neighbors 1/2) with the aim of correcting buildings that shall be considered as adjoining, but are not due to low geolocation accuracy. Even if such occurrences are negligible within the ZENRIN Residential Maps, this precautionary step ought to be taken before computing the number of adjoining neighbors. Lastly, *range01* is a user defined function that allows normalizing elongation and convexity.

3. Residential Buildings: Classification of Types

Once a basic set of indicators has been calculated for each input layer, the next step is to perform a classification of building types that can be applied to different time points. Three options arise. First, reliable expert knowledge is available, and as such, the number and the characteristics of the relevant building types are known in advance. Second, limited expert knowledge is available, which leads to partial information known in advance, such as the number of types sought, but not their characteristics. Lastly, no expert knowledge is available, thus leading to a situation in which both the number and the content of the building types are unknown.

Regeneration and Perforation of Building Types

3.1. Classification with Expert Knowledge

When expert knowledge is available, a simple and efficient method of classification is found in a manually built decision tree. Successive tests on attributes split the original dataset into different class labels. When a solution is found and deemed robust by the expert, the conditional control statements can be used as benchmarks for other datasets possessing similar attributes. **Supplementary materials algorithm 2** is a user defined function named *DTs* which allows classifying each row of a given dataset using the aforementioned morphometric indicators. Each test follows a *if condition* then *outcome* structure. Several conditions per test are possible, and the outcome is always a label attribution. Using the *apply* function, *DTs* can be used on any dataset possessing the relevant indicators, in our case BU0304 and BU1314.

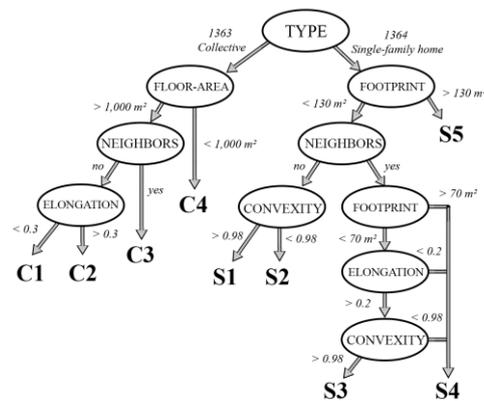


FIGURE 2. Visualization of the tests implemented in the *DTs* function

Fig.2 is a flowchart visualization of the successive tests implemented in *DTs*. They are 9 different outcomes, in which 4 are for collective housings (C1 to C4) and 5 are for single-family homes (S1 to S5). Each label attribution is based on expert knowledge. For example, C1 are high rise residential buildings following a “tower in the park” model, C2 are elongated residential complexes representative, amongst other structures, of Danchi housing (団地) which are complexes of apartment buildings built following western models after the Second World War, S3 are small and elongated row houses, typical of traditional wooden Nagaya (長屋), etc. The relevance of these building types in Japanese urban areas is derived from literature (e.g. Shelton, 2012; Bonnin *et al.*, 2014; Perez *et al.*, 2019b).

3.2. Classification with limited Expert Knowledge

If expert knowledge is limited, but still available in some extent, it is possible to make several decisions prior to a series of unsupervised cluster analyses. For

example, we could know the number of required clusters, as well as the fact that some inputs must not be grouped together. This could for instance be the case for collective and single-family homes, for specialized and residential buildings, or for building possessing large and small footprints. In such cases, standard unsupervised clustering techniques that minimize intra-cluster distances and maximize inter-cluster distances can be separately used for each subset, such as K-Nearest Neighbors, hierarchical clustering, *k*-means, classification trees, random forests, neural networks, etc. To illustrate this point, one of the simplest technique, *k*-means clustering, which aims at partitioning inputs into clusters such that the sum of squares from inputs to the assigned cluster centers is minimized, is applied as follows: **Supplementary materials algorithm 3** is a R protocol performing two different *k*-means, one for collective (attribute 1663) and one for single-family homes (attribute 1664). Two random subsets are draw and merged from BU0304 and BU1314 for collective buildings only (sample size of 20% of the inputs), and, once again, from BU0304 and BU1314 for single-family homes only (20% again). Drawing and merging random samples made of both periods altogether allows avoiding giving more weight to one period over the other. A clustering analysis is then performed for each subset (collective housing and single-family homes), with a number of clusters per model chosen *a priori*. In **algorithm 3**, we select the same number of clusters as within the DTs model (section 3.1: 4 clusters for collective housing and 5 clusters for single-family homes). Data are scaled and, in order to avoid redundancy, floor and area are removed in favor of floor-area. The two clustering analyses are thus performed on convexity, elongation, number of direct neighbors and floor-area. Once the two models are trained, for both periods, and for both collective and single-family home inputs, the whole datasets are mapped to the nearest clusters. This mapping is made using the *k*-nearest neighbor searching algorithm of the “FNN” package.

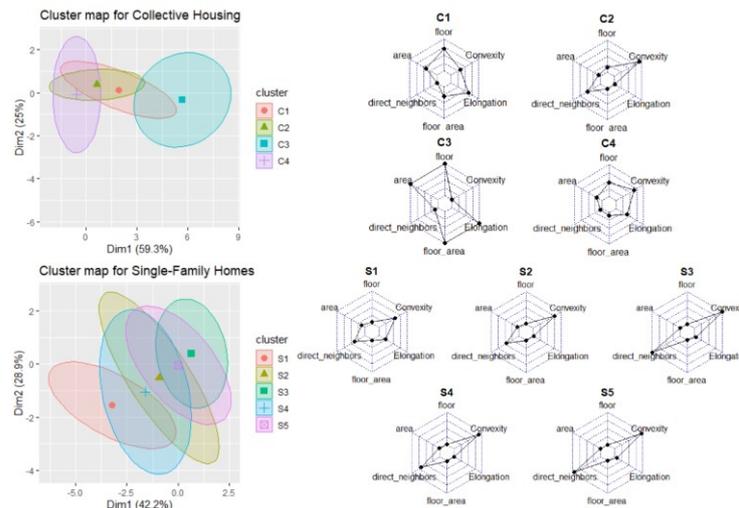


FIGURE 3. Cluster Map and Profiles for individual Clusters for Collective Housing (top) and Single-Family Homes (bottom)

Regeneration and Perforation of Building Types

The left side of **Fig.3** is a possible representation of the two clustering within two-dimensional spaces made by extracting the first two principal components (package “factoextra”) along which the variation in the data is maximal. Quality metrics in these clustering analyses are the share of the total variance in the data set that is explained by the clustering, given by dividing the between sum of square (BSS) by the total sum of square (TSS). These values are 94.13% for 4-class collective housing clustering and 93.82% for the 5-class, respectively. The right side of **Fig.3** shows radar charts for each variable. It is possible to add variables in the radar charts that have not been used during the clustering processes, such as area and floor in **Fig.3**. These plots are made with the package “fmsb”. The charts point at highly differentiated profiles, especially for collective housing, with profiles that are partially matching the DTs model. C3 are for example regrouping both high rise residential buildings and Danchi housings (団地), C4 appears to be the equivalent of DTs C4: small-sized complexes, etc. Cluster concordances with the DTs model are less clear for single-family houses, except for S2: articulated detached houses (equivalent to DTs S2) or S1: large villas (DTs S5). S3 and S5 are row houses with close characteristics, also close in the cluster map. In fact, even if literature or partial expert knowledge say that we shall seek for 5 different profiles of single-family homes, it appears that 5 is not a robust number of clusters for the k-means algorithm. In this respect it is possible to either (1) add/remove variables, such as using floor and area, instead of floor-area, or add newly computed morphometric indicators, (2) re-run the procedure (k-means is sensitive to the location of starting centroids) or try another algorithm, such as one of the algorithm discussed at the beginning of this section (3) re-run the algorithm with a different number of sought clusters (4) manually merge the clusters that have close characteristics.

3.3. Classification with no Expert Knowledge

Finally, in the case of a total lack of expert knowledge, any widely known clustering algorithms can be used, such as k-means, neural networks, or Bayesian networks. The main issues with unsupervised clustering are the assessment of the quality of the partitions, as well as the number of “interesting” partitions (Haldiki *et al.*, 2001). Since there is no way to *a priori* determine the most suitable number of clusters, researchers tend to use heuristics, such as looking at a cutoff point regarding the improvement of the explained variance (or log-likelihood in the case of Bayesian clustering) for each additional partition (elbow method). An unsupervised Bayesian clustering on BU1314, based on the indicators detailed in section 2.3 (plus specialization) has been performed in Perez *et al.*, (2019a). This clustering used an expectation-maximization algorithm (Dempster *et al.*, 1977) to perform one thousand clustering analyses and a MDL score (combining log-likelihood and a penalization function for the growing number of clusters) to define the optimal number of partitions. This application yields seven interesting clusters, from detached small houses to isolated high-rise buildings. In what follows, we will propose an alternative approach to unsupervised clustering based on Self-Organizing

Maps (SOM) (Kohonen, 1982). Bayesian and SOM clustering have already been compared in unsupervised approaches (Fusco and Perez, 2019).

Supplementary materials algorithm 4 is a R protocol performing SOM clustering on a sample made of BU0304 and BU1314 (sample size of 20% of the inputs). This time, we do not distinguish collective from single-family homes buildings. Once again, the sample is made with inputs draw from both periods altogether. The same indicators than in section 3.2 are used. SOM use a neighborhood function for each input to find the Best Matching unit (competitive training). Units, also called nodes, are distributed on a two-dimensional space (map). Once the best matching unit is found, a radius parameter allows updating the neighboring nodes, thus giving topological properties to the map that can be investigated through their proximity. There are two ways to identify clusters within a SOM. First, the number of nodes within the two-dimensional space can be set to a small value. Each node is then considered as a cluster. The second approach, which is the one used in **algorithm 4**, is to parameterize many nodes for the map (225, i.e. 15 by 15 in our application). The distances between couple of nodes form a distance matrix (U-Matrix). The advantage of this approach is that it allows visualizing each variable distribution across the map. Depicted by colors, these graphical representations are called heatmaps. Since similar values are aggregated in the same areas, heatmaps provide relevant information that can be used prior to the segmentation of the U-Matrix. The segmentation can then be performed using a simple hierarchical clustering, as in **algorithm 4**, in which the number of divisions is set manually.

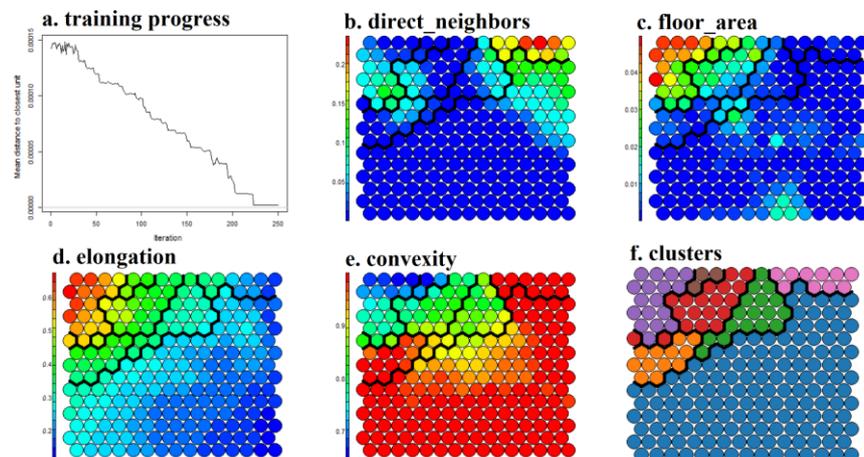


FIGURE 4. Self-Organizing Map outputs for a sample of BU0304 and BU1314 a. Training Progress b. Heatmap for direct neighbors c. Heatmap for floor-area d. Heatmap for Elongation e. Heatmap for Convexity f. Clustering results.

Fig.4 displays the outputs of a SOM clustering performed on a sample made of BU0304 and BU1314 using the “kohonen” package. The first thing to control is that the training curve flattens, as in **Fig.4 a**. The training is not over if it doesn't, the

Regeneration and Perforation of Building Types

*r*len parameter in **algorithm 4** must thus be set to a bigger value. **Fig.4 b. to e.** are the different heatmaps, showing that similar values are aggregating in the same areas. **f.** is the cluster map, where each color is a different building type. In this example, we set the number of clusters to seven. The black lines are the cluster boundaries, which have been retroactively mapped to the heatmaps. As in k-means, the ratio between BSS and TSS can be calculated for different numbers of clusters. For the 7-class solution, it reaches 84.01%. The “blue” cluster is made of different kinds of small sized single-family houses, the “pink” is for elongated row houses such as wooden Nagaya (長屋), the “purple” is for Danchi housing (団地), the “red” is representative of small collective residential complexes, etc. Once the SOM is trained with the sample, BU0304 and BU1314 data are mapped into the model using the *map.kohonen* function.

4. Spatial Analysis of Perforation and Regeneration

4.1. Automated Grid Creation and Evolution Count

In the previous section, we performed three different classifications that have been applied to the different periods associated with our datasets: 2003/04 and in 2013/14. To summarize, **algorithm 2, 3** and **4** are individually able to provide a new variable: a class label, that can be compared between two periods. This is the only input required in this section: a class label associated to at least two spatial datasets describing the same study area at different time points. **Algorithm 5** creates a grid with predetermined cell size, 250 meters, that are superimposed on the extent of the test region. Cells that are not intersecting any buildings are filtered out of the grid. Then, using a *tapply* structure, a recursive intersect is performed for each remaining cell to count the buildings grouped by labels. Once the *tapply* function has been ran for both datasets, a final line of code groups the results of both periods in a single dataframe. This whole process takes only a few seconds with an Intel® Core™ i7-8700 CPU @ 3.20GHz, 16gb RAM, while, on a traditional GIS software, for each and every class label, an intersect with a grid of this extent takes several minutes to run. For each cell, it is then possible to calculate the difference per building type, in order to map the temporal evolutions of each type.

4.2. Mapping

To map the count and evolution of building types, we use the “tmap” package, which allows building thematic maps following a layer-based structure. **Supplementary materials algorithm 6** shows a layered structure leading to the map of the count of the number of building for the C3 profile of the DTs model (**Fig.5 a.**). Before setting the maps, we first import an OpenStreetMap background. C3 values are then discretized into several categories (natural breaks), sequentially added as new layers to the map. **Algorithm 6** is interesting on both counts. First, for each category, cells are dissolved into a single multipart polygon before being

plotted. It allows plotting only the geometries thus substantially reducing computation time. Second, for each category, polygons are smoothed in order to make localized aggregates stand out.

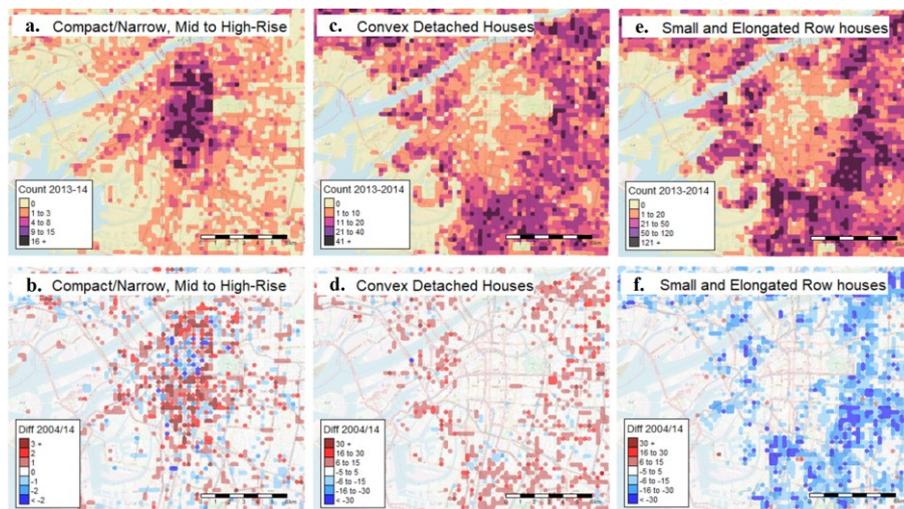


FIGURE 5. 2013-14 Count and 2003-4/2013-14 Evolution of building types for selected Profiles of the DTs model a. Count for C3 b. Evolution for C3 c. Count for S1 d. Evolution for S1 e. Count for S3 f. Evolution for S3

Fig.5 shows the count and evolution of three building types from the DTs model. Mappings within the above figure can be realized for any building type and for any of the three models discussed in section 3 using the structure of **algorithm 6**. C3 are mid to high rise compact/narrow buildings, mostly located in the central part of Osaka. The evolution map shows different hot and cold spots within this central area, thus pointing at a building type which is self-regenerating. S1 are convex detached houses mostly located outside of the center of Osaka. The evolution map shows that this kind of single-family homes is increasing, which is consistent with the literature pointing at a spread of small-sized prefabricated homes in Japan (Buntrock, 2017). S3 are small and elongated row houses, typical of traditional wooden Nagaya (長屋). This type is also located outside central Osaka, but also highly concentrated in specific neighborhoods. The evolution map points at a gradual disappearance of this traditional building type, even perhaps at a replacement if we link this disappearance with the increase of modern prefabricated homes.

5. Conclusion and Discussion

This paper put forward several recommendations and suggestions with the aim of studying the spatial location and temporal evolution of building types in large metropolitan areas. The first and most important condition is that high quality GIS

Regeneration and Perforation of Building Types

layers of building footprints with heights (or floors) as attribute data are available for at least two different time points. Another attribute is valuable, but not mandatory: building specialization. Attention should be paid to the time depth between the layers, as it must be sufficient to analyze structural changes in the spatial organization of building types. The temporality behind these structural changes is also dependent on cultural specificities regarding planning, building norms and urban development traditions (deconstruction, dismantlement, renovation, etc.). If the prerequisites are fulfilled, then it is possible to compute a basic set of indicators for each building, such as the one presented in section 2.3. Based on the kind of available expert knowledge, different types of classifications have been performed, from manually built decision trees to fully unsupervised clustering. Regardless of the method, quality metrics show that significantly different clusters can be found. It is possible to check the validity of each classification by drawing random samples and go on the field, or use a street-based urban imagery (such as Google Street View), then calculate the usual sensitivity and specificity metrics. The automated superimposition of a grid then allows counting occurrences of each type, as well as quantify the difference per cell between the different time points. For each category, mappings are performed, with cells that are discretized in the same category dissolved into a single multipart polygon, thus highlighting localized aggregates and, in the case of temporal evolution, hot and cold spots of perforation and regeneration of building types.

This paper focused on methodological suggestions, proposals and recommendations for studying building evolutions. Regarding the clustering applications, the methodological propositions could be improved, especially by focusing on maintaining output coherence over different settings or initializations. A matrix of similarity indexes can for example be calculated among the clustering results to evaluate the robustness of the protocols to pseudorandom number generation (seed), and cross-validation methods, such a k-fold or Jackknife resampling, can be used to evaluate the robustness of the outputs to different random samples. Yet, interesting preliminary results already stand out regarding specific locations and evolution trajectories of certain building types in Osaka, Japan. It shows that the perpetual mutation that characterize urban areas is linked, at least to some extent and to different degrees, to urban form. This research opens interesting perspectives, such as the study of the disappearance of some building types, or the gradual replacement of some types by other. Several methods, from standard correlation matrices to geographically weighted regressions, could allow exploring these dynamics in conjunction with other factors, such as population evolution, socio-demographic characteristics, planning policies, etc. These results must be analyzed together with literature and fieldwork related to the metropolitan area of Osaka.

6. References

Bonnin, P., Nishida, M. and Inaga, S. (2014) *Vocabulaire de la spatialité japonaise*. CNRS Editions, 560 p.

Buhnik, S. (2010) From Shrinking Cities to *Toshi no Shukushō*: Identifying Patterns of Urban Shrinkage in the Osaka Metropolitan Area. *Berkeley Planning Journal*, 23(1), pp. 132-155.

Buntrock, D. (2017) “Prefabricated housing in Japan”, in Smith, E. and Quale, J. (Eds.), *Offsite Architecture Constructing the Future*, Routledge: New York, NY, USA, pp. 190–213.

Dempster, A., Laird, N. and Rubin, D (1977) "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society. Series B*, 39(1), pp. 1-38.

Fusco G. and Perez J. (2019) Bayesian Network Clustering and Self-Organizing Maps under the Test of Indian Districts. A comparison, *Cybergeo: European Journal of Geography*, paper 887.

Hecht, R., Meinel, G. and Buchroithner, M. (2015) Automatic identification of building types based on topographic databases – a comparison of different data sources. *International Journal of Cartography*, 1(1), pp. 18-31.

Haldiki, M., Batistakis, Y. and Vazirgiannis, M. (2001) "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, Vol.17, pp. 107-145.

Jacobs J. (1961) *The Death and Life of Great American Cities*, Random House, New York, 458 p.

Kohonen, T. (1982) “Self-organized formation of topologically correct feature maps”, *Biological Cybernetics*, volume 43, pp. 59-69.

MLIT. (2007) “White Paper on Land, Infrastructure, Transport and Tourism in Japan”, *Ministry of Land, Infrastructure, Transport and Tourism Publication*, Ministry of Land: Tokyo, Japan, 68 p.

Perez, J., Fusco, G., Araldi, A. and Fuse, T. (2019a) “Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka”, *Asia-Pacific Journal of Regional Science*, 4, pp. 193-217.

Perez, J., Araldi, A., Fusco, G. and Fuse, T. (2019b) “The Character of Urban Japan: Overview of Osaka-Kobe’s Cityscapes”, *Urban Science*, 3(4), 105.

Salingeros, N. (2005) *Principles of Urban Structure*, Techne Press, Delft, 254 p.

Shelton, B. (2012) *Learning from the Japanese City: West Meets East in Urban Design*. Taylor & Francis, 224 p.

7. Supplementary Materials

<https://zenodo.org/record/4551462>

Acknowledgment: This study is supported by Joint Research Program no. 774 at CSIS, UTokyo (Zmap TOWN II 2013/14 Shapefile Osaka prefecture).

Exploring building typologies through fast iterative Bayesian clustering

Alessandro Araldi¹, David Emsellem², Giovanni Fusco¹, Andrea Tettamanzi³, Denis Overal²

1. Université Côte d'Azur, CNRS, ESPACE,
98, Bd Herriot, BP 3209, 06200 Nice, France.
[\[alessandro.araldi, giovanni.fusco\]@univ-cotedazur.fr](mailto:[alessandro.araldi, giovanni.fusco]@univ-cotedazur.fr)
2. KCITYLABS - Kinaxia Group,
80, route des Lucioles, 06560 Valbonne, France.
[\[david.emsellem, denis.overal\]@kcitylabs.fr](mailto:[david.emsellem, denis.overal]@kcitylabs.fr)
3. Université Côte d'Azur, I3S
2000, route des Lucioles, 06900 Sophia Antipolis, France.
andrea.tettamanzi@univ-cotedazur.fr

RESUME. L'identification et la description des typologies de bâtiments jouent un rôle fondamental dans la compréhension de la forme de l'espace bâti. Un nombre croissant de travaux développe et implémente de nouveaux protocoles sophistiqués de géomatique pour l'identification des typologies de bâtiments et leur organisation. Cet article partage le même objectif. Une procédure innovante, basée sur l'analyse quantitative des données est présentée ici avec, comme objectif, l'identification et la description non supervisée des types de bâtiments et de leur organisation. Après une procédure de pré-traitement spécifiquement adaptée à notre donnée source, nous développons un protocole de clustering non supervisé combinant un algorithme novateur d'inférence Bayésienne Naïve avec des approches ascendantes hiérarchiques; le tout, reposant sur cinq caractéristiques morphométriques intrinsèques de chaque bâtiment. Ce protocole permet d'identifier des groupes de bâtiments partageant des caractéristiques morphologiques similaires spécifiques ainsi que leur structure globale à différents niveaux d'agrégation. La méthodologie proposée est implémentée et évaluée dans l'espace d'étude du Département des Alpes-Maritimes, France.

ABSTRACT. The identification and description of building typologies play a fundamental role in the understanding of the overall built-up form. A growing body of works is developing and implementing sophisticated, computer-aided protocols for the identification of building typologies and their organisation. This paper shares the same goal. An innovative data-driven procedure for the unsupervised identification and description of building types and organisation is here presented. After a specific pre-processing procedure, we develop an unsupervised clustering combining a new algorithm of Naïve Bayes inference and hierarchical ascending approaches relying on five morphometric features of buildings. This protocol allows us to identify groups of buildings sharing specific similar morphological characteristics and their overall structure at different aggregation levels. The proposed methodology is implemented and evaluated in the case study of the Alpes-Maritimes Department, France.

MOTS-CLES : Bâtiments, Typo-morphologie, Clustering Bayésien Naïve, CAH

KEYWORDS: Building, Typo-morphology, Bayesian Naïve Clustering, HCA

1. Introduction

An interdisciplinary and growing body of research investigates the relationship occurring between the form and the functioning of cities (Carmona, 2019). In the typo-morphological tradition, the properties of the urban form are defined by specific spatial combinations of its constitutive elements: streets, plots and buildings (Moudon, 1997). While streets and plots show a higher inertia over time, buildings change at a faster pace depending on the specific historical, socio-economic and technological context. Moreover, while some buildings (and building types) are easily torn down and replaced, others can endure for several decades or centuries showing higher resilience to urban transformation. Both geography and urban typomorphology developed multiple approaches and protocols for the identification and analysis of building types based on the comparative analysis of buildings. As defined by Sheer (2017, p.171) “*a building type is an abstraction, a pattern, where we observe formal similarities between one building and another even though the buildings may have different architectural expressions. [] buildings might share many common formal characteristics, but are very different in color, materials, style, and expressiveness.*”

Building types can result from a given time period, in a specific regional context with certain easily recognizable stylistic patterns (i.e. the English terraced house; the American bungalow house; the Parisian Haussmann apartment-building). They could also be more overarching, grouping together similar buildings produced in different time periods, but sharing consistent common features. Different information at different description levels can be used for the identification and description of building types. The combination of specific features can thus result in a comprehensive building typology, i.e. an organization of building types having a given logic and inner coherence. Three levels of details can be distinguished: i) aesthetical and stylistic features (such as façades materials and composition, colours, etc.), ii) the internal organisation of the building (including some structural considerations) iii) the overall external hull of a building (shape, footprint area, height etc.). Nonetheless, formal definitions of building types, categories and structures describing the whole building stock of a large urban region are extremely rare (Orford and Readcliffe 2007). Similarly, there is no agreement about the definition of which combination of formal characteristics is required for the identification of building types and their differentiation, since different sets of features underlie the definition of each building type. When the goal of the study is the identification of building types over a large urban region, databases encompass few features, often limited to the simple geometrical form. The study of the external building hulls allows typifying what might be called the *skeletal form* of the building.

Two main approaches are currently used for the identification of building types (Hecht et al. 2015). *Knowledge-based approaches* have been proposed based on the study of small but exhaustive datasets including building features at each of the three detail levels. These approaches are based on expert knowledge of the relevant building types and features in a given study area but are hardly adaptable to other regions, or in the presence of a poorer description of buildings. Computer-aided protocols have been

developed since the 70s, especially by the Centre for Land Use and Built Form Studies (LUBFS) at Cambridge University (Steadman 2016). Thanks to more recent data analysis developments, together with the increasing computing power availability, systematic and quantitative protocols have been developed: *Data-based approaches* have been proposed especially from digital cartography, with the goal of cartographic generalisation or for the identification of urban structures, building detection and building pattern recognition. While the study of building typologies with expert-based protocols and based on highly detailed and historical datasets finds its origin in the urban typo-morphology tradition since the 1950s, the identification of building types from footprint-based data is a more recent field of study (Hecht 2015, 2016).

Two subgroups might be further specified in data-based approaches. Supervised protocols (similarly to knowledge-based analysis) require prior knowledge of the target groups we want to identify within the dataset: features are attributed to each group based on similarity rules. Examples of classificatory approaches can be found in Orford and Radcliffe 2007, Hech et al 2015, Hartmann 2016, etc. Unsupervised approaches encompass clustering protocols where the identification of groups is based on algorithms looking for internal similarity among features and without prior knowledge of the target groups or user intervention. Clustering protocols automatically determine natural partitions (clusters) from the input arising from the specific data structure without imposing a predefined identification of the classes. Examples can be found in the works of Schirmer and Axhausen 2015, Perez et al. 2019, etc. In fact, expert-based knowledge is never completely absent. In supervised protocols, it is required at the beginning of the analysis to define the target groups, their numerosity and their overall organisation. In clustering approaches, it is needed for the interpretation of outcomes. The former allows the analyst to better identify specific predefined building typologies, while the latter allows a more exploratory analysis where natural groups emerge from the data structure and are later interpreted and related to the specific characteristics of the study region. When focusing on clustering approaches, group identification is also influenced by the underlying algorithmic rules. As discussed in Fusco and Perez (2019) most of the traditional approaches (such as K-means) impose the sphericity of clusters (i.e. intra-cluster homogeneity) on all the descriptive variables, which could not always be coherent with the complexity of the context under study. Bayesian Clustering (BC) allows us to overcome these limitations. Still, as for most of the clustering approaches, BC, even when using Naïve clustering models, imposes other kinds of constraints and can be particularly time-consuming when exploring possible solutions in parameter space. This paper presents a specific protocol aiming to identify building typologies of a large urban region and their overall organisation combining unsupervised, incremental Naïve Bayesian and Hierarchical clustering approaches. The protocol is implemented on a few selected features describing the morphological characteristics of building hulls. The paper is organised as follows. Section 2 describes the proposed clustering protocol. In Section 3, the protocol is implemented to the study area of the Alpes-Maritimes Department, France, and its results are described and interpreted. Section 4 closes the paper with a discussion on limitations of the protocol, improvements and future perspectives.

2. Method

2.1. Data preprocessing

The protocol presented in this paper is implemented on French Department of Alpes-Maritimes (Section 3.1). In France, the BD TOPO® by the National Geographic Institute (IGN) is an exhaustive dataset of metrical precision providing the information on building footprint and height, retrieved from satellite or aerial imagery. Since April 2019, a new version of this dataset has been released (BD TOPO®, V3.0): its main novelty consists in the combination of the original data with information from the national computerized cadastral plan (MAJIC). Several new features are made available such as the number of dwellings, the age of construction and the number of stories. Nonetheless, this dataset has some limitations, among which three are of concern for our research, requiring specific pre-processing protocols.

The first issue is related to the building footprint definition. Part of the building dataset, enriched by the cadastral plan information, have a detailed definition of their footprint: while one (or more) polygon(s) corresponds to the main built-up body(ies), a number of extensions (such as terraces, loggias, porches, etc.) are separately modelled as adjoining polygons. These extensions are identified by a specific feature attribute, namely “light structure”. The same attribute is also associated with independent structures such as greenhouses, garages, small and large industrial sheds. Another share of buildings is defined from satellite/aerial based methods: in this case, the delineation of the footprint is defined with imagery detection algorithms where the footprint corresponds to the external demarcation of the overall built-up structure, therefore including both the main building body and all extensions. Thus, a harmonization protocol has been specifically developed in order to re-aggregate buildings made by several constituent parts. The second issue is related to the definition of the functional specialization attribute. The feature “Building Nature” provides the distinction between several specialized structures (i.e. religious, industrial, agricultural, etc.). When the ‘*overall architecture or aspect of a building does not reveal with exactitude its function*’ (ibid.) the building is classified as undifferentiated. After a manual assessment of this field on a subspace encompassing about 15 thousand buildings, we observed that the specialization attribute is always correctly assigned, although limited to 5.1% of the actual functional specialized buildings (overall accuracy 54.2%). Thus, this feature is enriched through another IGN BD TOPO layer, namely ‘Activity Zones’, where specialized buildings are retrieved and collected from other national authoritative sources. A set of specific rules and filters are defined and implemented to associate the specialized function to the original building dataset: on the same 15-thousand features subset, the resulting enriched definition of the field shows 86.8% accuracy score in identifying specialized buildings. This allows us to filter them and implement our clustering protocols only on ordinary residential or mixed function buildings. Indeed, as discussed in the typomorphological literature, specialized buildings (industrial, commercial, public, etc.) have often specific and extreme morphological properties values (Caniggia et Maffei 2001) introducing important outliers in our dataset, biasing the final outcomes.

Once the subset of non-specialized buildings is redefined, morphometric descriptors can be implemented. Three indicators are directly computed from building footprint: Surface (S), Elongation (E) and Convexity (C). One indicator, Topological Contiguity (TC), is defined as the number of neighbours within a continuous built-up unit of adjoining neighbours. Finally, the building Height (H) is provided by the original dataset. These five indicators represent a set of minimal description of the building form obtained by a simplified 3D dataset (LOD0+ of the CityGML data model in Biljecki et al., 2016). Nonetheless, building Height (as for several features of the latest BD Topo V0.3) can be unknown. As we will see in the next section, the clustering protocol developed in this work is specifically conceived to deal with partial missing information. As previously introduced, BC algorithm is based on a probabilistic framework. This approach requires qualitative or categorical data. A discretization of the five morphometric descriptors is therefore necessary: for our case study the discretisation was obtained through a mix of univariate data analysis and expert knowledge. This last pre-processing step produces a first reduction of the overall complexity of the original data: in our specific case study, for example, it allows us to pass from about 300 thousand building units to 2.2 thousand building tuples, each one corresponding to a specific combination of our features modalities.

2.2. Iterative Naive Bayesian Inference Agglomerative Clustering

Bayesian inference is a powerful probabilistic option for quantitative and qualitative multivariate data clustering using simple model architectures as the Naïve Bayesian Classifier, where the cluster variable is conceived as the common parent of all the other variables, and conditional independence among them is assumed knowing the cluster variable (Duda and Hart 1973). EM algorithm is normally used to identify an optimal clustering solution in terms of log-likelihood, for a given number of clusters (Dempster et al. 1977). Exploring solutions with varying number of clusters can be done with a random walk in solution space, using a clustering score combining log-likelihood and a penalization for model complexity, i.e. number of clusters (like in Fusco and Perez 2019). McCaffrey (2013) offers an interesting alternative to the EM algorithm for BC: Iterative Naive Bayesian Inference Agglomerative Clustering (INBIAC). So far, the implementation of the INBIAC algorithm can be only found in Carneiro et al. (2015) for credit card fraud detection. INBIAC is a much faster algorithm than EM as it replaces recursive batch inference of cluster assignments for all records to an iterative assignment of individual records which are randomly extracted from the database and assigned to the highest likelihood cluster at that given moment of the clustering procedure. The higher speed of the INBIAC algorithm can be used to perform a higher number of clustering solutions.

Just like EM, INBIAC results are sensitive to the clustering initialization. In EM, initialization implies randomly assigning all records to clusters. In INBIAC, a k-cluster solution needs the use of k records as initial cluster seeds. McCaffrey (2013) proposes a preliminary phase of seed initialization, randomly choosing k seeds and finally keeping the set of seeds with maximum Hamming distance. In our algorithm we improved McCaffrey's protocol in several respects. Firstly, to better represent the

ordinal structure of our data, we used Manhattan distance instead of Hamming distance, after normalizing for the cardinality of the ordinal values. Secondly, the usual Laplacian smoothing in the initialization of conditional probability tables, is reinforced by a further smoothing on the ordinal values which are contiguous to those of the seeds. After seed selection, all remaining records are assigned one by one to the cluster having maximum log-likelihood, and the cluster conditional probability tables will be updated after each record assignment. Iterative random removal of individual records from their current clusters to be re-evaluated and reassigned to a better cluster can thus begin. Given the Naïve model architecture and the resulting additive formula of model log-likelihood, the local optimization of log-likelihood in record assignment to clusters produces a global log-likelihood improvement for the whole model. More precisely, iterative record clustering is done in two phases. In the first phase, cluster priors are considered equal for all k clusters. In a second phase, cluster priors are determined from the current cluster probability distributions, allowing a more precise calculus of posterior probabilities in Bayes' formula.

A further improvement of McCaffrey's original INBIAC algorithm has been the treatment of missing values. Under the Missing at Random assumption, likelihoods and posterior probabilities of cluster assignments are calculated only on the observed values, but missing value imputation is later performed based on the most probable values within the assigned cluster. Imputed values are iteratively erased and re-imputed within the INBIAC procedure, and the final log-likelihood of the clustering solution includes the contribution of imputed values. The clustering iterations within INBIAC stop when no record can be reassigned to a different cluster. Finally, buildings are weighted by their footprint surface in the algorithm, giving the same importance to each m^2 of built-up surface (the clustering solution would otherwise be biased by the overrepresentation of small buildings). The implementation of the protocol for several numbers of clusters k , ranging in a user-defined interval, allows to explore different clustering solutions. The optimal clustering solution(s) can be then selected through a simple graphic method: similarly to the well-known *elbow method*, the log-likelihood loss scores are plotted across different numbers of clusters k (Fig.1.a-b). The presence of strong variations might reveal structural changes in the data clustering corresponding to suboptimal solutions at a given number of classes k .

2.3. HCA

The implementation of the INBIAC protocol, allows us to identify one (or more) optimal solution(s) based on the optimization of specific scores and descriptive parameters. Nonetheless, independently from the specific set of parameters used for the model evaluation, the selected clustering(s) solution(s) would always provide a specific partition of the original dataset defined for a given number of groups k . It is thus interesting to study the variation of building clustering across k . Buildings constantly grouped in the same clusters could reveal stronger structural patterns within the data. We thus implement an agglomerative Hierarchical Clustering Analysis (HCA) using as input the INBIAC best outcomes for each k clustering solution in the interval explored. The rationale underlying this methodological choice is the

following: the subset of n best clustering solutions can be used to partition our original dataset (or similarly, the 2.2 thousand tuples) in smaller subgroups of elements (kernels) always being clustered together independently of the number of clusters k . These kernels represent the finest partition for which the highest inter-level consensus is observed: no cluster at any level further divides the element in finer groups. Within our specific context, these kernels correspond to a highly detailed meta-cluster solution of specific building sub-types; few kernels gather most of the buildings (more precisely of the built-up surface, given our weighting scheme), and vice versa a large number of kernels encompass less built-up surface with less recurrent shapes. HCA will be here implemented with Gower's dissimilarity metric for cluster distance and Ward-linkage agglomerative principle among clusters. Implementing an HCA allows us to produce hierarchically nested groupings based on the similarities within this elementary kernel partition. Kernels are arranged in a hierarchical manner. Thus, the combination of the INBIAC and HCA protocol combines the advantages of the two protocols. On the one side, we keep the ability of probabilistic Bayesian inference to select non-spherical clusters defined with a maximum log-likelihood approach on subgroups of features. On the other, an overall hierarchical structure allows the analyst to observe the overall data clustering organization similarly to knowledge- and ontological-based classifications. Moreover, the main advantage compared to regular HCA applied on raw data, is that the outcome variability produced by the high sensitivity to the initial clustering settings is strongly reduced. This approach shares the same underlying hypothesis of consensus clustering protocols (Monti et al., 2003), where several cluster solutions are combined in order to achieve a more robust solution. Nonetheless while consensus clustering looks for similarities within a larger number of clustering solutions for a given number of clusters k , in our case we use a more 'controlled' subset of n suboptimal solutions at different levels k . Moreover, instead of implementing the same clustering protocol at two different stages of the analysis, we combine a non-hierarchical and a hierarchical protocol in the first and second stage of the analysis, respectively. Finally, our protocol provides a profile for each cluster as probability distributions of its members over the values of each clustering variable. These profiles are later used for cluster interpretation (section 3.2).

3. Application

3.1 Study Area

The protocol presented in this work, is implemented on the building stock of the Alpes-Maritimes Department in southern France. This study area is made up of a large coastal conurbation and of its alpine hinterland. The coastal conurbation of the French Riviera stretches over 60 km from the French-Italian border to the Esterel mountains. Its western section includes the cities of Cannes, Grasse and Antibes, counting 74.2, 51 and 73.8 thousand inhabitants, respectively. Nice, with its 343 thousand inhabitants represents the largest municipality of the French Riviera and the administrative centre of the Department. The enclave of Monaco and the border city of Menton have respectively 38 and 28 thousand inhabitants. Spread around these main centres, 295 thousand people find their home in smaller cities, villages and hamlets. The alpine

hinterland represents over one third of the overall departmental surface but has a population of only 15 thousand inhabitants (1% of the departmental population). The building stock of the Alpes-Maritimes shows a hyper-concentrated urban fabric along the coast, a strong sprawl in the western sector, and an urban development in the eastern sector strongly influenced by its higher topographic constraints. With the exclusion of Grasse, the other five main urban centres are located along the coastline, with other 11 municipalities, gathering 42% of the overall departmental buildings (51% building surface). When including the whole of the coastal conurbation 33% of the departmental surface encompasses almost 97% of the building stock.

3.2 Implementation and results

From the original 500 thousand polygons, the data pre-processing protocol allows us to identify about 300 thousand buildings: 30% are specialized and the remaining 70% have a residential or mixed use. Our clustering protocol will be applied here only to this larger stock of ordinary buildings. The segmentation of the five features (S, H, C, E, TC) further reduces the variability of our dataset resulting in 2.2 thousand observed combinations. Then, the INBIAC protocol is implemented exploring 100 solutions with different initial seeds, for each number of clusters k between 5 and 15 producing a total of 1.1 thousand models (Fig.1.a). The quality parameters of these solutions vary with k : loglikelihood loss is maximal for the 5-cluster solution and minimal for the 15-cluster solution. Fig.1.a shows the scatter plot of the log-likelihood loss vs the category utility score (Gluck and Corter, 1985) for all the models implemented: important gains are observed for the 6-, 8- and 11-cluster solutions, further confirmed by the elbow method applied on the log-likelihood loss scores (Fig.1.b).

According to our approach, we will nevertheless use the information derived from all the 11 best clustering solutions. Their combination further defines 280 kernels, 15 of which encompass 57% of the overall built-up surface. The implementation of the agglomerative HCA on these kernels, allows us to identify and describe the overall organization of the building types in the study area (Fig.1.c-e). The dendrogram in Fig.1.c shows the succession of cluster agglomerations along a distance axis, starting from the 280 kernels (below) and arriving to the complete amalgamation of clusters (top). The length of the segment on the distance axis during which a given k -cluster solution is present is indicative of its importance in structuring the building typology in the study area. The first four solutions showing the longest segments on the distance axis of the dendrogram correspond to 4, 10, 21 and 8 clusters (Fig.1.d). The ten-cluster solution is described as follows.

A.1 corresponds to compact small detached single-family houses and duplexes while A.2 to compact small detached two and three stories houses; they account for 23.5 and 12.6 % of the overall built-up surface of the study area, respectively. These groups correspond to two of the most common building types of the French Riviera, known as the *niçois house*. These building types often do not exceed 150 m² and three stories; they represent a specific production of the early XX century and in the second after war. While A.1 one corresponds to a single dwelling house, A.2 might also include few dwellings initially intended for different members of the same family group. In

the same building type, we also find some more recent small real estate development projects, often used as second homes and for touristic purposes.

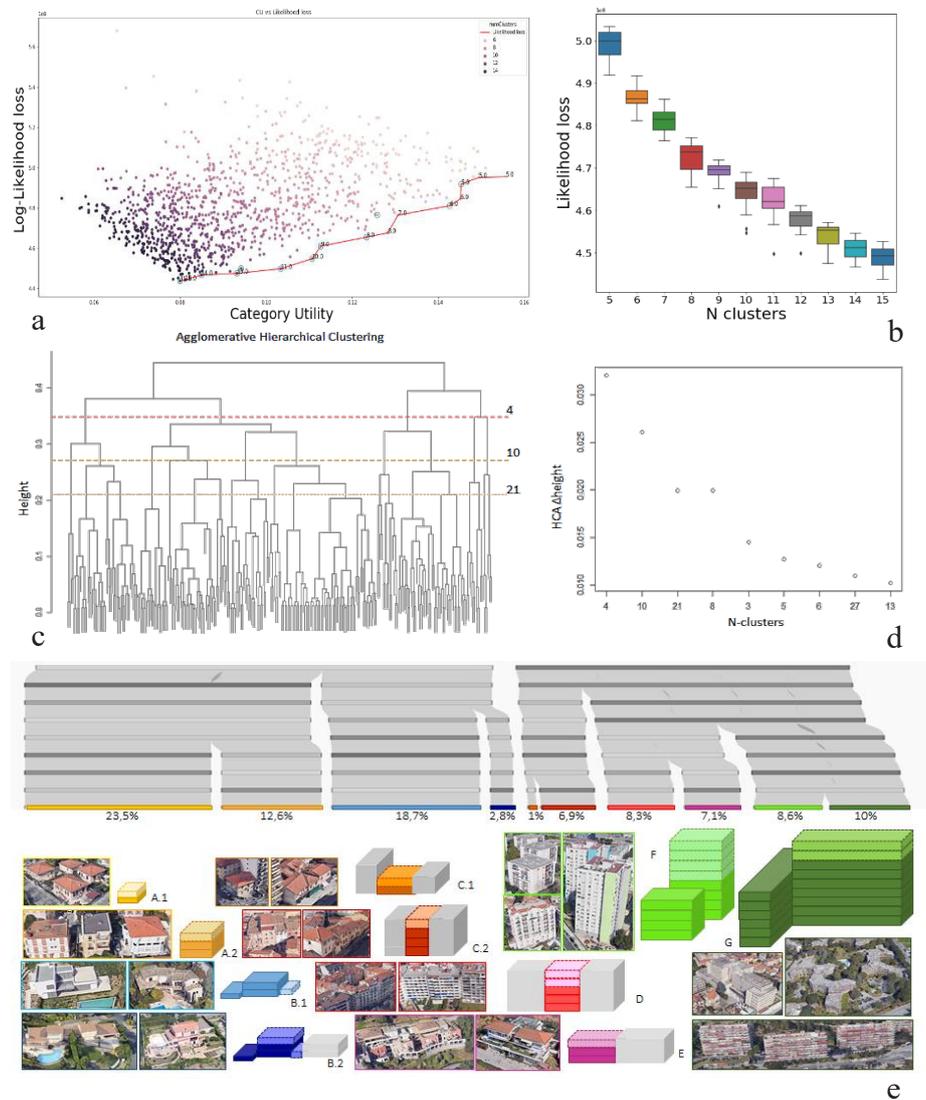


FIGURE 1. Outcomes of the building classification: a) Scatter plot of the 1000 models scores obtained with the INBIAC clustering: log-likelihood loss vs. category utility (CU); b) Box-plots of the top 20 INBIAC models scoring the highest loglikelihood-loss scores for each number of cluster k in the interval $[5;15]$; c) HCA dendrogram implemented on the outcomes of the 11 best scoring INBIAC models; d) identification of the partitions corresponding to the highest dendrogram depths; e) interpretation of the 10-cluster HCA solution: the dendrogram partitions between 2 and 10 are weighted by the building surface assigned to each group.

B.1 Identifies articulated homes and villas while B.2 corresponds to very articulated large villas (18.7 and 2.8 % of the overall built-up surface, respectively). Both groups combine large villas and (semi) detached houses with a footprint surface mainly between 150 and 300 m². On the one side, we find large individual compact homes and villas produced before the '50s which have been enriched by later functional additions/extensions (garages, porches, etc.). On the other side, more recent large articulated villas resulting from the suburban sprawl of end of the XX century and the beginning of 2000. What mainly differentiate the B.2 from B.1, is a more complex shape together with higher building heights (2-3 stories). High standing villas often occupy geographic locations with specific peculiar positions such as the top of hills, capes etc.; large gated communities are mainly made of these two building types.

C1. and C.2 are small-sized (<300 m²) mid-rise adjoining buildings (1 and 6.9% of the total building surface, respectively). Most of the medieval centres of cities and villages are made up of these groups, as well as more informal settlements along connective axes. C1 differs from C2 for its lower height (1-2 stories) and its larger surface; moreover, C1 can often be found in more recent urban fabric as a residual of a previous settlement, surrounded (and often substituted) by mid-rise buildings D.

D. Adjoining mid-size and mid-rise apartment buildings represents the most common building type characterizing city centres outside their historical core (8.3% of the total building surface). Despite their different aesthetical properties (ranging from belle époque, art deco or more modern and contemporary stylistic features), these buildings share similar footprint size (between 300 and 600 m²), elevated contiguity and height (above 4 stories, but for long time limited at 6 stories by cityscape rules).

E. Two building types can be found within cluster E (7.1% of the building surface). On the one side we find mid-sized and mid-rise detached and semi-detached large apartment buildings. On the other we find some huge (historical) villas and manors, often occupying the most priced locations of the French Riviera (such as capes) and immersed in residential areas made by large villas of type B.

F. Free standing mid-to-high rise towers (8.6% of the total building surface). Towers are apartment buildings whose simple compact footprint is relatively small in comparison to their vertical development. More precisely, cluster F encompasses two different building types: traditional small mid-rise towers (3 to 6 storeys and building footprint between 300 and 600 m²) and more modern, larger high-rise towers (more than 6 storey-tall but rarely skyscrapers in our study area).

G. Freestanding very large and high-rise apartment buildings, with articulated shapes, (10% of the total building surface). Buildings of G type have a footprint surface above 1200 m²; they have very elongated and/or composite shaped buildings; they can be freestanding or belonging to large and complex urban residential development projects. Most of the housing production is located in the close peripheries; their production only started from the late 60s and it encompasses both high standing apartment buildings and large projects developed by local and national housing programs. Some of these exceptional buildings can also be found in more central and

compact contexts (i.e. demolition and reconstruction of a large articulated buildings built over an entire block).

4. Conclusion and perspective

In this paper we presented an innovative protocol for the identification and description of building types and their overall organisation combining Naïve Bayesian and Hierarchical clustering protocols. The outcomes of this systematic and quantitative analysis allow the data-driven derivation of a system of typologies of buildings, hierarchically organised. The protocol is described and implemented for the real-world contemporary case study of the Alpes-Maritimes Department. Ten building types have been identified and described.

Several research perspectives can be outlined. From the methodological point of view, sensitivity analyses should be implemented to assess the robustness of the three main steps of the protocol presented in this paper. The first step considers the role of the variable discretisation: the same protocol should be evaluated both with a general binning method and with segmentation approaches based on the specific statistical distribution of variables observed in the (sub)region under analysis. The second and third phases correspond to the two clustering protocols: both INBIAC and HCA can be assessed considering different distance measures and evaluating their validity under different parametric conditions. Independently of the specific parametric choices, a comparative analysis should also be carried out with more traditional approaches (i.e. k-means, DBSCAN) in order to further identify relative strengths and weaknesses of the protocol here proposed, both in the specific context of building type identification but also within other thematic fields.

From the thematic-related point of view four major directions for further developments can be outlined. Firstly, this same protocol can be tested with an incremental number of descriptors of the building envelope aiming at testing and identifying the role played by individual morphometrics into the building typologies. This work might contribute to the debate on the definition of a reliable and universally accepted set of characters and variables for the identification of building envelope typologies. Moreover, internal layout and details of style, facade, roof coverage might also be included: the implementation of the same clustering approach with different levels and granularity of information can shed a new light on the relative role played by skeletal, internal and stylistic features in the identification and definition of building typologies. Secondly the protocol proposed in this work should be tested on different study areas and at larger scales, to allow a better appreciation of the methodological robustness, its computational and geographical scalability; more important, the reproducibility will permit systematic, quantitative and fine grained comparative analysis on a large number of case studies (both with national-wide datasets and with volunteered geographic large dataset such as OpenStreetMap at the international level). Thirdly, the analysis of the spatial organisation of buildings types and their relative cooccurrences represent a key factor in the definition of streetscapes, urban fabrics and morphological regions, allowing to investigate the multi-scalar

nature of cities based on the finest level of the building unit. The work here described represents indeed only the first step of an undergoing larger research aiming at understanding the building types and the urban fabrics of French contemporary cities.

Acknowledgment: This research was funded by the IDEX UCA JEDI, within the AAP Partenariat 2019 (action 6.5).

References

- Caniggia G, Maffei GL (1979). *Lettura dell'edilizia di base*. Firenze: Alinea.
- Carmona M, (2019) Place value: place quality and its impact on health, social, economic and environmental outcomes, *Journal of Urban Design*, 24:1, 1-48.
- Carneiro EM, et al. (2015) Cluster analysis and artificial neural networks: A case study in credit card fraud detection. 2015 *12th Int. Conference on Information Technology*. IEEE, 2015.
- Dempster AP, Laird NM, Rubin DB (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*. 39 (1): 1–38
- Duda RO, Hart PE (1973) *Pattern classification and scene analysis*. NY: John Wiley & Sons.
- Fusco G, Perez J (2019) Bayesian Clustering and SOM Neural Networks under the Test of Indian Districts. A comparison. *Cybergeog*, n°887, .<https://journals.openedition.org/cybergeog/31909>
- Gluck MA, Corter JE (1985), Information, uncertainty, and the utility of categories, *Program of the Seventh Annual Conference of the Cognitive Science Society*, pp. 283–287
- Hartmann A, Meinel G, Hecht R, et Behnisch M (2016). A workflow for automatic quantification of structure and dynamic of the German building stock using official spatial data. *ISPRS International Journal of Geo-Information*, 5(8), 142.
- Hecht R., Meinel G. & Buchroithner M. (2015). Automatic identification of building types based on topographic databases – a comparison of different data sources, *International Journal of Cartography*, 1:1, 18-31.
- IGN (2020) BD TOPO®, Version 3.0. Descriptif de contenu. https://geoservices.ign.fr/ressources_documentaires/Espace_documentaire/BASES_VECTORIELLES/BDTOPO/DC_BDTOPO_3-0.pdf
- McCaffrey J. (2013) *Data Clustering Using Naive Bayes Inference*. <http://msdn.microsoft.com/en-us/magazine/jj991980.aspx>.
- Monti S, Tamayo P, Mesirov J et Golub T, (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1-2), pp.91-118.
- Moudon, A. V. (1997). Urban morphology as an emerging interdisciplinary field. *Urban morphology*, 1(1), 3-10.
- Orford, S., Radcliffe, J. (2007). Modelling UK residential dwelling types using OS Mastermap data: A comparison to the 2001 census. *Computers, Env. & Urban Systems*, 31(2), 206–227.
- Perez J., Fusco, G., Araldi, A., & Fuse, T. (2020). Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka. *Asia-Pacific Journal of Regional Science*, 4(1), 193-217.
- Scheer, B. C. (2018). Urban Morphology as a Research Method. *Planning Research and Knowledge*. Abingdon, UK: Routledge, 167-181.
- Steadman P (2016). Research in architecture and urban studies at Cambridge in the 1960s and 1970s: what really happened. *The Journal of Architecture* 21, no. 2 (2016): 291-306

Communications par Session

Simulation, évaluation en géomatique



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Choix de mobilité : une approche SMA

Alice Jacquier¹, Carole Adam¹

Univ. Grenoble-Alpes, Laboratoire d'Informatique de Grenoble (LIG)
prenom.nom@univ-grenoble-alpes.fr

RÉSUMÉ. La mobilité a beaucoup changé récemment, et va continuer à évoluer pour faire face à de nouveaux défis, en particulier environnementaux. Le projet SwITCh vise à créer un outil pour aider les planificateurs urbains à concevoir des villes intelligentes et durables. Cet article y contribue en proposant un modèle agent des décisions de la population concernant leur modalité de transport pour leur trajet quotidien domicile-travail. Nous avons implémenté ce modèle avec la plateforme de simulation GAMA, et utilisé ce simulateur pour tester 2 hypothèses. Nos premiers résultats montrent que l'amélioration des infrastructures cyclables conduit à une augmentation du nombre de cyclistes, mais que par contre une simple augmentation du prix de l'essence ne décourage pas les automobilistes. Des travaux futurs étudieront donc d'autres stratégies pour augmenter la part des mobilités douces dans les trajets quotidiens.

ABSTRACT. Means of mobility have changed a lot recently and will continue to evolve, in order to meet new environmental challenges. The SwITCh project aims to create a tool to help urban planners designing smart and sustainable cities. This paper contributes a first step by providing an agent-based model of the population's decisions regarding transport modality for their daily home-work commute. We implemented our model with the GAMA simulation platform, and used this simulator to test 2 hypotheses. Our first results show that better cycling infrastructures result in an increase in the number of cyclists; and that a simple petrol price rise is not enough to discourage car users. Future work will thus investigate better strategies to improve the part of sustainable mobility in daily commutes.

MOTS-CLÉS : Simulation multi-agents, modèle de décision, mobilité urbaine

KEYWORDS: Agent-based simulation, decision model, urban mobility

1. Introduction

Nous faisons face à une crise environnementale, et la question de la durabilité de nos pratiques quotidiennes se place de plus en plus au centre de nos préoccupations. Durant ces dernières décennies, la distance moyenne quotidienne

parcourue par les français a augmenté considérablement (5 km en moyenne dans les années 50, contre 45 km en moyenne de nos jours (Viard, 2011)), faisant de la mobilité une problématique importante dans cette recherche de durabilité. Nous voulons donc proposer un outil pédagogique à destination du grand public pour leur permettre de comprendre les raisons et l'impact de leurs choix de mobilité.

La simulation sociale à base d'agents est très utile pour étudier des phénomènes sociaux et proposer de tels outils. Elle permet en particulier de répéter et contrôler exactement les paramètres de scénarios définis, afin d'en explorer les implications, et de tester des stratégies. Dans ce contexte, nous présentons donc un modèle à base d'agents de la population d'une ville (ici Castanet-Tolosan), et de ses décisions concernant le mode de transport domicile-travail. Ce modèle attribue à chaque agent, représentant un habitant, une modalité préférée, en fonction des caractéristiques de l'environnement et de ses attributs sociaux et psychologiques. Il permet ensuite de simuler différents scénarios en faisant varier les paramètres environnementaux comme le prix de l'essence, les infrastructures de la ville, ou la météo du jour. Dans cet article nous testons dans notre simulateur deux interventions pour essayer de faire baisser la part de la voiture. A terme, ce premier modèle servira de base à un jeu sérieux dans lequel les citoyens pourront explorer diverses interventions.

La Section 2 introduit un état de l'art pluridisciplinaire des travaux sur le choix modal, en sociologie et en informatique. La Section 3 décrit notre modèle conceptuel basé sur ces travaux. La Section 4 décrit son implémentation avec GAMA ainsi que les premières expérimentations et leurs résultats. Enfin la Section 5 conclut cet article et discute des perspectives d'amélioration.

2. Choix du mode de transport : un état de l'art pluridisciplinaire

Le choix modal a été étudié en sociologie sous plusieurs axes : définition de profils d'utilisateurs (Section 2.1), ou analyse des critères de décision (Section 2.2). Divers modèles informatiques en ont aussi été proposés (Section 2.3).

2.1. Sociologie : profils d'utilisateurs

L'étude Mobil'Air (Chalabaev *et al.*, 2018) présente 4 profils d'utilisateurs de la voiture par opposition aux autres modes de transport : automobilistes ouverts à tout mode ; automobilistes pro mode individuel (préférant un mode de transport assurant l'indépendance) ; automobilistes à contre coeur (contraints d'utiliser la voiture) ; automobilistes convaincus (aimant conduire et prenant la voiture pour tous leurs trajets). L'avantage de ce projet est de fournir les pourcentages de la population dans chaque catégorie. Par contre, son point négatif est de ne considérer que les automobilistes. Ces profils illustrent le principe de contrainte de mode de déplacement selon son motif (*e.g.* emmener un enfant à

Approche SMA de la mobilité

l'école nécessite d'avoir un siège passager sur son mode de transport) et la force de l'habitude (certains usagers ayant tendance à reproduire la même routine).

Rocci (2007) propose 6 autres profils d'usagers : les automobilistes passionnés en adhésion à l'automobile ; les automobilistes passionnés en opposition à un autre mode ; les multimodaux rationnels, en adhérence à la voiture mais qui utilisent les transports en commun par moment ; les multimodaux en opposition à la voiture, mais contraints de la prendre par moment ; les alternatifs passionnés en adhérence à leur mode ; et les alternatifs en opposition à leur mode. Cette classification aborde de manière moins tranchée la dualité entre voiture et autres modes, et est aussi plus détaillée, allant au-delà des seuls automobilistes. On y retrouve cependant des similitudes, notamment cette idée d'usage contraint de la voiture. Rocci détaille d'ailleurs une autre raison à cet usage contraint : ce qu'elle appelle le "capital mobilité", ou la capacité à utiliser un mode de transport (posséder un vélo ou une voiture, savoir lire une carte de métro, avoir le permis...). Cela recouvre aussi la notion d'accessibilité (tous les transports en commun ne sont pas -facilement- accessibles). Enfin elle montre que chaque personne a une vision personnelle des différents modes de transport : les automobilistes convaincus, par exemple, vont avoir tendance à sous-estimer le prix de la voiture, et surestimer le prix des transports. De plus, les transports en commun ont aussi une image associée au stress, agression et dépendance (Kaufmann *et al.*, 2010). Par contre, Rocci ne fournit aucune statistique de répartition de ces profils dans la population fournie.

Nous pouvons trouver dans la littérature d'autres études qui ont créé des profils, par exemple en utilisant des modèles de mélange (Briand *et al.*, 2016). Cette étude ne concerne que les transports en commun, et souligne différents profils concernant leur utilisation. Ces modèles sont intéressants sur la méthodologie, mais pour notre travail nous nous baserons sur les profils cités ci-dessus, moins complexes et qui explicitent mieux les raisons des choix de chaque profil.

2.2. Sociologie: facteurs de décision

Il existe une infinité de facteurs que nous pourrions prendre en compte dans notre modèle. Cependant, nous désirons suivre le paradigme KISS ("Keep it simple, stupid", (Axelrod, 1997 ; Adam, Gaudou, 2016)) et concevoir un modèle simple, afin de mieux en comprendre les résultats. Comme il est destiné à servir de base pour un jeu sérieux, nous voulons aborder de manière simple les mécanismes de prise de décision. Nous présentons donc ici une sélection non-exhaustive des principaux facteurs de décision.

2.2.1. L'habitude

Les habitudes sont au coeur des décisions de mobilité. Elles peuvent se changer dans les moments décisifs des cycles de vie des personnes (Brette *et al.*, 2014). Les changements de contexte, c'est-à-dire quand les autres habitudes

sont bouleversées (changement d'emploi du temps, déménagement, naissance d'un enfant, emménagement avec un conjoint, etc.), sont des moments propices pour briser une ancienne habitude ou en intégrer une nouvelle, en ciblant les interventions sur ces populations (Rothman *et al.*, 2015). Enfin, les habitudes peuvent modifier les perceptions (temps, coût, etc). Ainsi un usager habitué de la voiture va surestimer le temps de trajet en transport public et minimiser le temps de trajet en voiture (Rocci, 2007) (Betsch *et al.*, 2001).

Il nous semble donc essentiel de prendre en compte l'habitude dans notre modélisation du comportement humain de choix modal. Par ailleurs, l'instrumentalisation des changements de cycle de vie est une piste intéressante pour les scénarios d'interventions à tester dans notre simulateur.

2.2.2. *Le contexte de l'environnement*

L'environnement impose des contraintes dans la décision du mode de transport (Rocci, 2007). La voiture est utilisée pour des longs trajets, et pour combler une carence en transports en commun (pour les activités nocturnes). La voiture est aussi utilisée pour diminuer la charge mentale car elle permet un trajet direct. En effet, la réduction de la charge mentale est pour certains un argument pour ne pas prendre la voiture, mais seulement si le trajet de transport en communs est facile ou direct (i.e un seul métro, tram, bus à prendre, pas de changement à faire). Si le trajet est trop compliqué, alors la préférence va à la voiture. Enfin, ce choix dépend aussi du temps total de trajet, incluant l'accès aux arrêts de transport en commun, et donc de la répartition du réseau sur l'agglomération. Un voyageur qui n'a pas d'arrêt à distance raisonnable de chez lui et/ou de sa destination va aussi préférer la voiture.

Ces informations montrent que l'environnement joue un rôle dans la décision du mode de transport. Il nous faudra donc modéliser assez précisément ces aspects de l'environnement : position des arrêts, fréquence des bus, etc.

2.3. *Informatique: modèles de décision du mode de transport*

Il existe déjà des simulateurs de trafic (par exemple le simulateur SUMO, (Behrisch *et al.*, 2011)). Différents modèles de décision de mode de transport ont aussi été proposés selon différentes approches. D'un côté, le champ de la recherche opérationnelle sur les MCDA (multi-criteria decision aid) propose des méthodes comme Prométhée (Behzadian *et al.*, 2010) permettant de trouver une solution optimale.

Gandit (Gandit, 2007) propose un modèle de décision de mode de transport (uniquement entre voiture et transports en commun) avec une approche par apprentissage automatique. Il réalise une régression logistique sur la base de trois types de critères : socio-démographiques, structurels, et psychologiques. Sa classification des critères pour chaque individu est intéressante, et est basée sur 3 études menées dans le cadre de sa thèse (une pour chaque type de

Approche SMA de la mobilité

critère). Mais cette approche est très complexe et ne laisse pas comprendre ce qu'il se passe, car elle prédit le résultat du processus de décision mais n'explique pas ce processus lui-même. Cette approche ne convient donc pas pour concevoir notre jeu sérieux, qui doit faire comprendre au public le cheminement de pensée des agents simulés pour avoir un impact sur leur propre raisonnement. D'autres projets ont d'ailleurs préféré l'approche SMA pour ces mêmes objectifs : comprendre le processus de formation du résultat, et non pas trouver le résultat optimal. C'est le cas de Wester (Wester, 2015) qui cherche à comprendre la construction du réseau de transports en commun de Brazzaville. Nous choisissons donc nous aussi une approche SMA.

2.4. Notre proposition

Notre but est de créer un jeu sérieux pédagogique à destination du grand public, afin d'expliquer et explorer des scénarios d'aménagement urbain. De plus, le jeu pourrait permettre de mettre en lumière des comportements indésirables (comme la prise de la voiture, par exemple). L'approche par simulation multi-agents est donc une bonne option. En effet, la modélisation à l'échelle microscopique et individuelle nous offre une vision différente et réaliste du processus du choix modal. De ce fait, les conséquences des actions menées dans le jeu sont compréhensibles, et explicables à un public non expert. Cependant nous n'avons trouvé aucun modèle à base d'agents décrivant le processus de décision d'un mode de transport, à l'échelle individuelle. Nous décrivons donc dans la section suivante notre propre modèle de choix modal individuel, basé sur les facteurs de choix identifiés dans l'état de l'art ci-dessus.

3. Modèle de décision du mode de transport

Cette section décrit notre modèle conceptuel de la décision modale des usagers : le modèle de l'environnement urbain (Section 3.1) ; et le modèle des agents humains, chacun représentant un usager (Section 3.2).

3.1. L'environnement, une variable à ne pas négliger

Éléments géométriques : nous commençons par définir l'environnement dans lequel les usagers évoluent, qui contient les éléments suivants :

- Road (route), qui composent les arêtes d'un graphe représentant le réseau de routes, et qui a un attribut "has_bike_lane" qui informe de la présence d'une piste cyclable ou non.
- Building : id pour identifier les bâtiments. type le type du bâtiment. sub_area, le lieu du bâtiment. nb_households, le nombre de ménage dans ce bâtiment. size, la taille du bâtiment.

– Crossroad : **type**, le type du carrefour. **crossing**, qui donne des informations sur ce carrefour.

Nous utilisons des graphes pour représenter les réseaux de transports (un graphe par type de transport) ainsi que le réseau routier. Les bâtiments sont représentés par des polygones, selon les informations obtenues sur Open Street Map. Nous n'avons pas considéré le relief des trajets, mais pour l'instant uniquement leur distance. Cependant, nous n'excluons pas la possibilité d'intégrer ultérieurement un attribut "dénivelé" qui permettrait de prendre en compte le relief, grâce à l'intégration d'un modèle numérique de terrain. Dans l'implémentation du simulateur, ces éléments pourront être importés depuis Open Street Map pour n'importe quelle ville. Dans nos résultats (cf Section 4) nous avons choisi par exemple la ville de Castanet-Tolosan.

Éléments contextuels : Dans l'environnement, nous prenons aussi en compte des éléments contextuels pertinents dans le choix modal (météo, informations sur le réseau de transport public, prix de l'essence, infrastructures pour les cyclistes). Cependant, nous avons dû simplifier dans un premier temps beaucoup de ces éléments. Ainsi notre modèle des transports en commun comporte uniquement des informations de confort (affluence) et temporelles (fréquence) mais pas encore d'informations spatiales (distance à l'arrêt). Pour la météo, nous faisons simplement un tirage aléatoire entre 3 types de temps (ensoleillé, nuageux, pluvieux), pondéré par les informations relatives à la ville modélisée (temps d'ensoleillement, informations pluviométriques). Cela est suffisant pour expérimenter des scénarios réalistes (moins de cyclistes par temps pluvieux), et pourra être complexifié dans le futur si nécessaire. Tous ces attributs jouent un rôle dans le choix du mode de transport, et sont pris en compte dans le modèle de décision décrit dans la section 3.2.

3.2. *Agents usagers*

Dans notre modèle, chaque usager est représenté par un agent, doté d'un certain nombre d'attributs : un âge, un genre, une catégorie socio-professionnelle, un lieu de travail, un lieu de résidence, un niveau de sportivité et trois listes d'agents représentant respectivement sa famille, ses amis, et ses collègues. Chaque agent a également une liste contenant les notes associées à chaque mode de transport. Le calcul de ces notes est expliqué dans le paragraphe suivant 3.2. Ils ont de plus une liste de critères auxquels ils apportent plus ou moins d'importance, voir 3.2. L'habitude est représentée par un coefficient qui augmente la note du mode de transport préféré courant. Nous avons aussi un attribut qui permet de déterminer si l'agent possède un vélo, ou encore une voiture. Certains de ces attributs ne sont pas encore utilisés dans le modèle, mais sont déjà implémentés. Les listes représentant la famille, les amis, et les collègues par exemple serviront dans le futur à modéliser la pression sociale, ou encore la diffusion de pratiques parmi les pairs. La catégorie socio-professionnelle n'est

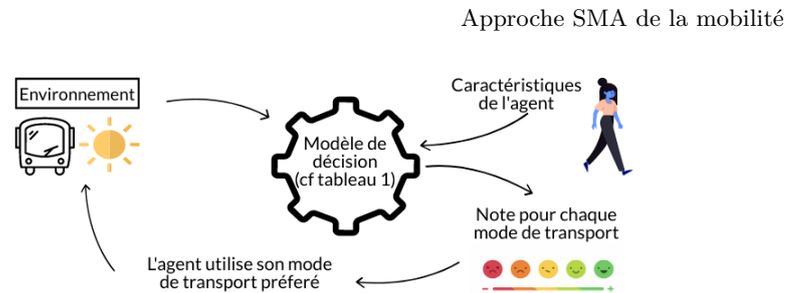


FIGURE 1. Schéma récapitulatif du processus de prise de décision

pas non plus utilisée ; elle le sera ultérieurement pour déterminer la possession ou non d'une voiture.

Calcul des priorités des modalités : Chaque agent calcule ensuite sa note de préférence pour chaque mode de transport, en fonction des attributs ci-dessus. En se basant sur les études de (Rocci, 2007) et (Kaufmann *et al.*, 2010), nous avons pu **définir 6 critères : prix, temps, confort, sécurité, simplicité, écologie**. Chaque agent accorde une importance plus ou moins grande à chacun de ces critères. En fonction de l'environnement de l'agent et de l'importance donnée à chaque critère, nous calculons alors une note individuelle de priorité p_i de chaque mode de transport i , pour chaque agent, avec la formule suivante :

$$p_i = \frac{\sum_{c \in \text{critères}} (val_{i,c} \times pref_c) + \text{habit_coeff} \times p_i}{\text{nombre de critères} + 1}$$

La valeur $val_{i,c}$, comprise entre 0 et 1, correspond à une évaluation par l'agent de la valeur du critère c pour le mode de transport i , selon l'état de l'environnement et les caractéristiques de l'agent. Par exemple le vélo pourrait obtenir une note de 1 sur le critère écologie, mais sa note de confort dépendra de la distance à parcourir et du niveau de sportivité de l'agent, et sa note de sécurité dépendra des infrastructures cyclables de la ville. La note de la voiture sur le critère de prix dépendra entre autres du prix de l'essence, alors la note de simplicité dépendra de la possession ou non d'un véhicule, etc. Ce calcul est résumé au Tableau 1. Ce calcul est par ailleurs pondéré par un coefficient d'habitude coeff_habit , expliqué dans le paragraphe suivant.

Prise en compte de l'habitude : Lorsqu'un agent décide quel moyen de transport il préfère utiliser, le coefficient d'habitude va pondérer de manière plus ou moins forte la valeur de la priorité. Cela permet de modéliser le fait qu'un choix "routinier" n'est plus réévalué quotidiennement, au contraire il faut un changement majeur pour remettre en question un choix. Ainsi, si cela fait plusieurs années que l'agent utilise la voiture, il va de moins en moins remettre en question son choix ou analyser les autres moyens de transports autour de lui, et il lui sera plus difficile de réaliser quand un autre mode de transport correspond mieux à ses critères. Ce coefficient d'habitude est calculé

Tableau 1. Facteurs entrant en compte dans le calcul de la note de chaque mode de transport sur chaque critère. Si la valeur est fixe, elle est indiquée directement dans la case.

	Vélo	Bus	Voiture	Marche
Confort	Distance sportivité météo	Affluence	1	Distance sportivité météo
Prix	0	Prix abonnement	Prix essence distance	0
Ecologie	1	0.8	0	1
Simplicité	Distance	#correspondances (pour l'instant ≈ 0.7)	1	Distance
Sécurité	Ratio routes avec pistes cyclables	Genre, heure du jour	Niveau de trafic	Genre heure du jour
Temps	Distance vitesse moyenne vélo (pondérée par sportivité)	Distance Vitesse moyenne bus (pondéré par trafic)	Distance Vitesse moyenne voiture (pondéré par trafic)	Distance, vitesse moyenne marche

selon l'ancienneté de la valeur d'une priorité (ancienneté notée n), de manière logarithmique : $\text{coeff_habit} = \log n$. La croissance de ce coefficient est donc de plus en plus faible, à mesure que n augmente à chaque pas de temps. Dès que l'utilisateur change de mode de transport préféré, n est réinitialisé à 0.

4. Implémentation et résultats

4.1. Implémentation et méthodologie

La plateforme GAMA. GAMA (Taillandier *et al.*, 2019) est une plateforme de simulation multi-agent open source, permettant d'implémenter des modèles agents de haut-niveau, avec un langage simple, accessible aux non-informaticiens. GAMA fournit différentes architectures d'agents (automate, réflexes, architecture cognitive, émotions...). Elle permet aussi d'inclure des données de divers types, y compris des données issues de Systèmes d'Information Géographique. Le code de notre simulation sera partagé sur OpenABM¹ ultérieurement, une fois stable.

Création de l'environnement. Nous avons récupéré les données relatives à la ville de Castanet-Tolosan via Open Street Map (graphe des routes, informations relatives aux bâtiments, etc) afin de générer la carte utilisée dans notre modèle de déplacement. Nous fixons le ratio de pistes cyclables à 0.5, et une météo ensoleillée. Nous fixons également : le prix de l'essence à 1,5€ au litre ; le prix mensuel de l'abonnement de transports en commun à 30€ ; la capacité des bus à 50 personnes, et leur fréquence de passage toutes les 7 minutes.

Création de la population. Nous avons lancé la simulation avec une population synthétique de 10 000 habitants. Il existe déjà des travaux sur la

1. <https://www.comses.net/codebases/>

Approche SMA de la mobilité

génération de populations synthétiques réalistes (par exemple le projet Genstar (Chapuis *et al.*, 2019)), qui seront réutilisés ultérieurement pour améliorer le réalisme de notre simulateur. Pour l'instant, dans un souci de simplification, nous considérons qu'un agent sur deux possède une voiture, un agent sur deux possède un vélo, et l'initialisation des autres attributs (lieu de résidence et de travail, priorité de chacun des critères) se fait aléatoirement. Avec ces informations, l'agent choisit un mode de transport (voiture, vélo, bus ou marche) pour son trajet domicile-travail.

Méthodologie. Nous mesurons en sortie la distribution (statique) des agents par mode de transport à une date donnée, sur 2 scénarios : 1) en variant le ratio de pistes cyclables ; et 2) en variant le prix de l'essence (tous les autres paramètres fixés). Les autres paramètres ne changent pas (météo ensoleillée, initialisation des agents avec la même graine). Chaque situation a été simulée 10 fois et nous avons fait la moyenne des distributions de mode de transport résultantes, pour lisser la part d'aléatoire.

Scénario 1 : part de cyclistes selon les infrastructures. Nous voulons tester l'hypothèse que le taux d'infrastructures cyclables influence le nombre d'agents choisissant le vélo : plus il y a de pistes cyclables, plus il y a de cyclistes, et vice-versa. Pour cela, nous avons simulé deux situations : 1) aucune infrastructure pour le vélo, et 2) toutes les routes équipées d'infrastructures cyclables. La figure 2 illustre la distribution moyenne des usagers sur les différents modes de transport dans ces 2 situations. Ces résultats valident notre hypothèse : la quantité d'infrastructures influe bien sur le taux de cyclistes.

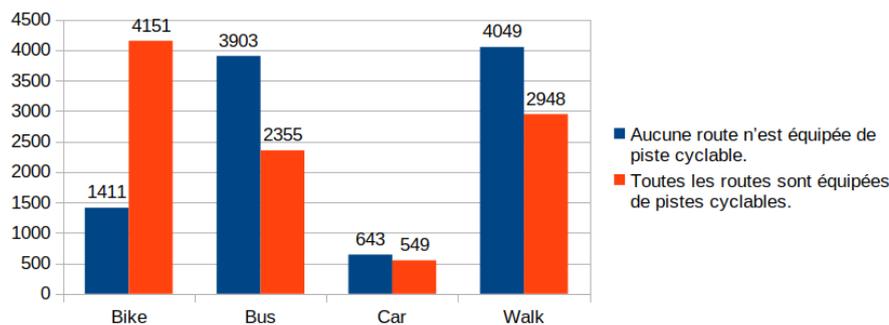


FIGURE 2. Distribution selon ratio de pistes cyclables

Scénario 2 : part d'automobilistes selon le coût de l'essence. La deuxième hypothèse que nous émettons est que le prix de l'essence influence le nombre d'automobilistes : plus l'essence est chère, moins il y a d'automobilistes. Nous testons pour cela 2 situations : 1) prix de l'essence à 0.2 € par litre ; et 2) prix de l'essence augmenté à 5 euros par Litre. La figure 3 montre les résultats de ces deux simulations. Nous constatons un résultat assez mitigé : le taux d'automobilistes ne change pas, peu importe le prix de l'essence. Ce résultat semble incorrect, et pourtant il reste crédible. La voiture coûte cher,

et les personnes ayant pour critère important le prix ne choisiraient pas la voiture dans un premier temps ; donc lorsque le prix augmente, cela a peu d'incidence. Cela correspond aussi aux observations dans le monde réel : le prix de l'essence augmente régulièrement depuis des années, pourtant la part de la voiture ne diminue pas ; c'est d'ailleurs encore plus le cas actuellement avec la crise sanitaire qui décourage le recours aux transports en commun. Il faut donc chercher d'autres moyens pour faire augmenter la part des mobilités douces en milieu urbain.

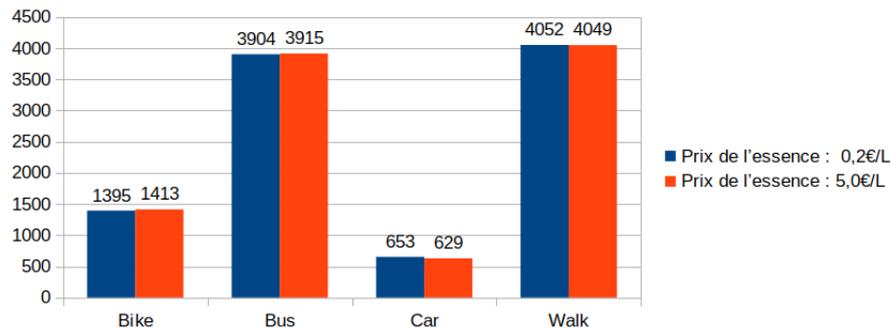


FIGURE 3. Distribution selon prix de l'essence.

4.2. Discussion

La principale limitation de ce modèle est sa validation difficile. Les résultats sont crédibles mais il est difficile de vérifier le raisonnement réel des habitants. Par ailleurs une fois le modèle plus stable, il sera intéressant d'utiliser OpenMole pour en faire une analyse de sensibilité.

Ce modèle est aussi très simplifié, du point de vue de l'environnement comme des agents. Dans le futur, nous envisageons donc de donner plus de réalisme à l'environnement (météo basée sur des relevés pluviométriques, comme fait dans Vigiflood (Adam, Andonoff, 2019) ; pollution de l'air, comme dans HoanKiemAir (Duc *et al.*, 2020) ; prise en compte du dénivelé des trajets grâce à un Modèle Numérique de Terrain...) et de tester d'autres scénarios (ajout d'un mode de transport, gratuité des transports publics, piétonnisation du centre...).

Des travaux sont aussi en cours pour générer et intégrer une population synthétique réaliste, incluant par exemple la provenance sociale. Pour l'instant nous avons supposé que les habitants vivent et travaillent dans la même ville, ce qui augmente artificiellement la part de la marche dans les déplacements. Les décisions des agents sont aussi basées uniquement sur des critères rationnels pour l'instant, mais nous allons maintenant étudier d'autres facteurs plus subjectifs : influence de l'entourage ou de leaders d'opinion (par analogie avec la diffusion d'innovation, largement documentée dans la littérature (Vodopivec

Approche SMA de la mobilité

et al., 2021)); émotions (Bourgais *et al.*, 2018); biais cognitifs (Arnaud *et al.*, 2017); ou encore habitude, dont le rôle est central dans l'inertie des pratiques modales (Brette *et al.*, 2014).

5. Conclusion

Nous avons présenté dans cet article un modèle du choix du mode de déplacement domicile-travail. Notre modèle prend en entrée des caractéristiques liées à l'environnement (météo, réseau de transports en commun, infrastructures cyclables, prix de l'essence), et d'autres liées aux usagers (genre, âge, préférences, lieu de résidence et de travail, capital mobilité). En sortie, le modèle donne pour chaque agent un mode de transport préféré. Nos résultats ont montré que l'amélioration des infrastructures cyclables augmentait la part des cyclistes (mais plutôt aux dépens du bus), et que l'augmentation du prix de l'essence ne décourageait pas les automobilistes. Malgré certaines limitations, ces résultats sont crédibles. Nous prévoyons dans le futur d'améliorer le réalisme, que ce soit de l'environnement, ou bien du processus cognitif des agents. A terme, ce modèle sera utilisé dans un jeu sérieux afin d'impliquer et de sensibiliser les citoyens de manière ludique à l'impact de leurs choix individuels sur la collectivité. Ils pourront en particulier participer à l'élaboration de nouveaux scénarios. En effet la transition écologique ne peut se faire sans la participation active de tous.

Remerciements. Ce travail résulte d'un stage de M2R et se poursuit en thèse. Il est financé dans le cadre du projet SwITCh, qui a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-19-CE22-0003.

Bibliographie

- Adam C., Andonoff E. (2019). Vigiflood: a serious game for understanding the challenges of risk communication. In *International conference on information systems for crisis response and management*. Valencia, Espagne, ISCRAM association.
- Adam C., Gaudou B. (2016). Bdi agents in social simulations: a survey. *Knowledge Engineering Review*, vol. 31, n° 3, p. 207–238.
- Arnaud M., Adam C., Dugdale J. (2017, May). The role of cognitive biases in reactions to bushfires. In *Is cram*. Albi, ISCRAM.
- Axelrod R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration* (vol. 3). Princeton university press.
- Behrisch M., Bieker L., Erdmann J., Krajzewicz D. (2011). Sumo—simulation of urban mobility: an overview. In *Third international conference on advances in system simulation (simul)*. Barcelona, Spain, IARIA.
- Behzadian M., Kazemzadeh R. B., Albadvi A., Aghdasi M. (2010). Promethee: A comprehensive literature review on methodologies and applications. *European journal of Operational research*, vol. 200, n° 1, p. 198–215.

- Betsch T., Haberstroh S., Glöckner A., Haar T., Fiedler K. (2001). The effects of routine strength on adaptation and information search in recurrent decision making. *Organizational behavior and human decision processes*, vol. 84, n° 1.
- Bourgais M., Taillandier P., Vercouter L., Adam C. (2018, March). Emotion modeling in social simulation: a survey. *JASSS*, vol. 21, n° 2.
- Brette O., Buhler T., Lazaric N., Marechal K. (2014). Reconsidering the nature and effects of habits in urban transportation behavior. *Journal of Institutional Economics*, vol. 10, n° 3, p. 399–426.
- Briand A. S., Come E., El Mahrsi M. K., Oukhellou L. (2016). Classification à base de modèle de mélange pour l'identification de profils temporels types d'usagers de transport public. In *Fr. int. conf. on data science*. Maroc, AAFD & SFC.
- Chalabaev A., Chardonnel S., Bouscasse H., Duché S., Isoard-Gatheur S., Mathy S. *et al.* (2018). Les apports de la littérature et des analyses préalables pour définir une intervention visant un report modal durable. In *MSFS*. Tours, France, MSFS.
- Chapuis K., Thiriot S., Taillandier P., Gaudou B., Amblard F. (2019). Gen*: an integrated tool for realistic agent population synthesis. In *SSC*. Germany, ESSA.
- Duc P. M., Chapuis K., Drogoul A., Gaudou B., Grignard A., Marilleau N. *et al.* (2020). Hoankiemair: simulating impacts of urban management practices on traffic and air pollution using a tangible agent-based model. In *Int. conf. on computing and communication technologies (riuf)*, p. 1–7. IEEE.
- Gandit M. (2007). *Déterminants psychosociaux du changement de comportement dans le choix du mode de transport : le cas de l'intermodalité*. Thèse de doctorat non publiée, Univ. Grenoble. (Thèse de doctorat)
- Kaufmann V., Tabaka K., Guidez J.-M., Louvet N. (2010). *Et si les français n'avaient plus seulement une voiture dans la tête?*. Rapport technique. France, C. d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques.
- Rocci A. (2007). *De l'automobilité à la multimodalité? analyse sociologique des freins et leviers au changement de comportements vers une réduction de l'usage de la voiture*. Thèse de doctorat non publiée, Sociologie, Univ. R. Descartes - Paris V.
- Rothman A. J., Gollwitzer P. M., Grant A. M., Neal D. T., Sheeran P., Wood W. (2015). Hale and hearty policies: How psychological science can create and maintain healthy habits. *Perspectives on Psychological Science*, vol. 10, n° 6, p. 701–5.
- Taillandier P., Gaudou B., Grignard A., Huynh Q.-N., Marilleau N., Caillou P. *et al.* (2019). Building, composing and experimenting complex spatial models with the gama platform. *GeoInformatica*, vol. 23, n° 2, p. 299–322.
- Viard J. (2011). *Éloge de la mobilité: essai sur le capital temps libre et la valeur travail*. France, Editions de l'Aube.
- Vodopivec N., Adam C., Chanteau J.-P. (2021). Modeling opinion leader's role in the diffusion of innovation. *arXiv preprint arXiv:2101.11260*.
- Wester L. (2015). Modélisation multi-agents de transports collectifs artisanaux : structures émergentes et stratégies individuelles. In *SAGEO*. Tunisia.

Pour une évaluation visuelle des modèles de simulation, au-delà de la *face validation*

Robin Cura

UMR Géographie-cités, CNRS
robin.cura@parisgeo.cnrs.fr

RÉSUMÉ. Nous proposons une approche méthodologique dédiée à l'évaluation de modèles de simulation stochastiques descriptifs caractérisés par une faible possibilité de recours à des données empiriques. En l'absence de données structurées de contrôle – ou quand celles-ci sont trop peu nombreuses, incertaines ou lacunaires –, l'évaluation peut être réalisée d'après les connaissances expertes, de manière qualitative. Pour que ce processus soit mené de manière rigoureuse, robuste et reproductible, il ne doit pas être uniquement basé sur des intuitions ou premières impressions. L'évaluation peut prendre la forme d'une comparaison systématique entre les sorties du modèle et les hypothèses thématiques sous-jacentes à la modélisation, tirées des connaissances expertes. Cette comparaison est menée à l'aide d'une méthode d'« évaluation visuelle » que nous définissons et appliquons ici. L'approche méthodologique proposée doit ainsi permettre et faciliter l'évaluation de modèles de simulations par le biais de connaissances expertes explicitées, en s'appuyant sur l'utilisation d'un environnement d'exploration interactive des données de sortie de simulation.

ABSTRACT. In this talk, we propose a new methodological framework, the “visual evaluation”, that enables or facilitates the evaluation of complex and stochastic simulation models that mostly rely on expert knowledge rather than empirical data for their hypotheses. This method has been developed for the modelling of space-time processes and relies on the thorough use of qualitative visual comparison between the model's outputs and the underlying empirical knowledge. Visual evaluation is based on a remediation of face validation for massive simulation datasets, and thus requires an interactive exploratory data analysis platform dedicated to simulation outputs.

MOTS-CLÉS : Simulation multi-agents, méthodes d'évaluation, évaluation visuelle, analyse interactive de données, modélisation en interdisciplinarité

KEYWORDS: Agent-based modelling, evaluation methods, visual evaluation, face validation, Exploratory Data Analysis, interdisciplinary modelling

1 Introduction

Pour comprendre et modéliser une structure spatiale modelée par l'héritage de ses états antérieurs, il est nécessaire d'en comprendre l'émergence, souvent bien antérieure. Comme l'écrit Durand-Dastès (1999, p. 23), l'explication d'un « objet spatial » contemporain n'est en effet possible qu'en montrant « à la fois comment il est né et comment il a persisté ». La modélisation de processus spatiaux génériques en géographie implique donc de comprendre les origines et raisons de leur apparition, et donc de faire appel à des connaissances en dehors du champ strict de la modélisation en géographie. Si, dans cette discipline, la modélisation de dynamiques spatiales à l'aide de modèles de simulation multi-agents est désormais bien ancrée, la « culture modélisatrice » est moins présente dans les champs disciplinaires historiques et archéologiques. La conception d'un modèle de simulation spatial requiert alors d'une part la capacité d'intégrer les connaissances expertes nécessaires via un processus de co-construction (Cura, 2020a, pp. 26–29, 340–346)¹ et d'autre part, de manière inhérente à une démarche de modélisation, la capacité de soumettre le modèle conçu à une évaluation rigoureuse.

1.1 Contexte et cas d'étude

La proposition méthodologique ici développée se veut générique, mais a été développée dans le cadre d'un projet interdisciplinaire de modélisation de transformations des structures spatiales sur le temps long, l'ANR TRANSMONDYN. Au sein de ce projet, un groupe composé de géographes-modélisateurs, d'archéologues et d'historiens a conçu et développé un modèle de simulation, SIMFEODAL, cherchant à reproduire les dynamiques de polarisation, de hiérarchisation et de fixation de l'habitat rural de l'Europe du Nord-Ouest au tournant du 1er millénaire de notre ère, en modélisant les mécanismes de migrations et de concentration de foyers paysans. Le modèle et sa démarche de construction ont fait l'œuvre de publications spécifiques (Cura *et al.*, 2017 ; Tannier *et al.*, 2020), et nous en donnerons donc ici un descriptif très succinct. SIMFEODAL est un modèle de simulation à base d'agents stochastique, à visée exploratoire, qui suit l'approche « KIDS » formalisée par Edmonds et Moss (2005). Il se classe ainsi dans la catégorie des modèles descriptifs, avec de nombreux types d'agents (foyers paysans, seigneurs, églises paroissiales, pôles d'attraction, etc.) dont les interactions multiscalaires font émerger les dynamiques étudiées à l'aide de mécanismes complexes et nombreux sur une longue durée simulée de 400 ans.

1.2 Objectif

Nous proposons ici une méthode d'évaluation dédiée à des modèles de simulation à base d'agents co-construits selon un paradigme descriptif. La mé-

1. Cet article s'appuie sur une thèse de doctorat, et plus spécifiquement sur le matériel conceptuel, méthodologique et technique qui y est développé dans les chapitres 3 et 4 (Cura, 2020a, pp. 107–180, 181–262).

Pour une évaluation visuelle

thode proposée est particulièrement pensée pour l'évaluation de modèles qui s'appuient avant tout sur des connaissances à dire d'experts plus que sur des données structurées. Cette méthode d'évaluation, que nous nommons « évaluation visuelle », constitue une extension d'une étape classique de l'évaluation de modèle : la *face validation*, c'est-à-dire l'évaluation de la plausibilité des sorties d'un modèle. Cette méthode d'évaluation est fondée sur une analyse visuelle et subjective de la plausibilité du modèle, en demandant à l'évaluateur de se conforter à des intuitions sur le comportement attendu. L'utilisation de cette méthode est souvent considérée comme une mesure préalable relevant du bon sens plus que de l'évaluation strictement dite. L'évaluation visuelle étend le processus de *face validation* en systématisant l'analyse des sorties du modèle de manière qualitative. Il s'agit ainsi de définir en amont une grille d'analyse exhaustive, couvrant tous les aspects du modèle. C'est au prisme d'une comparaison systématique entre cette grille de critères attendus et des « indicateurs de sortie de simulation » (indicateurs visuels, quantitatifs ou qualitatifs) que le modèle sera évalué. Pour permettre à un public d'experts diversifié de mener cette évaluation visuelle directement, sans faire appel à un expert modélisateur ou à un analyste de données, nous proposons un outil interactif d'analyse visuelle et de comparaison des simulations effectuées par le biais de leurs données en sortie.

2 Approche conceptuelle : pour une évaluation visuelle

2.1 Évaluer un modèle

Depuis les travaux précurseurs en simulation informatique (Naylor, Finger, 1967 ; Hermann, 1967), la plupart des chercheurs mettent en avant qu'un modèle de simulation non évalué n'a ni utilité – pour Naylor et Finger (1967) notamment –, ni validité. Sans caricaturer ces écrits, on peut noter que tous cantonnent les modèles non évalués à des « jeux » ou encore, pour les plus modérés, à des outils pédagogiques. Nous ambitionnons de rendre possible l'évaluation de modèles pour lesquelles les méthodes plus classiques sont soit inadaptées, soit insatisfaisantes.

2.1.1 Différentes approches d'évaluation

Le terme d'évaluation en lui-même porte une certaine polysémie et peut être désambiguïsé, en reprenant les termes de Amblard *et al.* (2006), en « validation interne » et « validation externe ». Dans le premier cas, il s'agit de s'assurer de la bonne conception du modèle – « *building the model right* » (Balci, 1994, p. 123) – et dans le second, de montrer que le modèle est adapté à ce qu'il cherche à représenter – « *building the right model* » (Balci, 1994, p. 121). Ces deux composantes permettent de définir une procédure générale d'évaluation qui peut être résumée entre des cycles d'allers-retours et d'itérations entre les hypothèses du modèle, la validation interne de l'implémentation qui en est faite, et la validation externe des sorties produites (Klügl, 2008). Pour chacune de ces composantes, de nombreuses méthodes existent, sont généralement préconisées dans l'une ou l'autre des étapes de l'évaluation, et peuvent être classées en

différentes catégories, selon qu'elles utilisent une approche formelle, statique, dynamique ou encore informelle (Balci, 1998).

2.1.2 Quelle évaluation pour SIMFEODAL ?

Toutes les méthodes ne sont pas pour autant applicables à tous les modèles. Par exemple, les méthodes formelles sont souvent peu adaptées à l'évaluation de modèle de simulation à base d'agents. On emploie en effet généralement ces modèles parce qu'ils encouragent une approche anthropomorphique, plus aisément compréhensible et requérant moins de connaissances mathématiques que d'autres approches, mais aussi car il est extrêmement difficile d'exprimer des systèmes dotés de multiples interactions, qui plus est multi-scalaires, sous forme de réseaux d'équations. Les méthodes statiques ne sont aussi que partiellement applicables, et renvoient davantage à la validation interne des modèles, via une vérification logique des interactions entre les mécanismes des modèles sur un plan algorithmique. Pour la validation externe, le plus classique est de faire appel aux méthodes « dynamiques » au sein d'une étape communément nommée validation statistique. Il s'agit de confronter quantitativement les données produites par le modèle aux données empiriques qu'elles cherchent à reproduire, c'est-à-dire à s'assurer de la qualité de l'ajustement des données simulées. Ces méthodes dynamiques s'appuient donc intrinsèquement sur la présence de données empiriques directement comparables avec les sorties du modèle, chose commune dans la modélisation de phénomènes contemporains ou récents.

Dans le cas de SIMFEODAL, et plus généralement des modèles de processus anciens et inscrits dans la longue durée, les méthodes dynamiques sont difficilement mobilisables. Les données empiriques disponibles sont en effet des traces, incertaines et peu précises de par leur ancienneté, et extrêmement incomplètes (absentes et lacunaires) de par la durée de 400 ans modélisée – d'après la typologie de De Runz (2008). SIMFEODAL est fondé avant tout sur des hypothèses thématiques et des connaissances expertes, et les méthodes dynamiques ne sont alors pas exploitables. Dans cette situation, les théoriciens de l'évaluation recommandent, à défaut, de tout de même faire appel aux méthodes informelles (Petty, 2010), telle que la « *face validation* » qui est la plus couramment mobilisée.

2.2 De la *face validation* à l'évaluation visuelle

La *face validity* a été définie dès les débuts de la simulation comme une « impression initiale du réalisme d'un [modèle] » (Hermann, 1967, p. 221). Klügl (2008) en donne une définition plus explicite² et dresse une typologie des temps

2. « Face validity can be seen as the result of face validation. Under this paradigm I want to subsume all methods that rely on natural human intelligence. Examples are structured walk-through, expert assessments of descriptions, animations or results. Thus, face validity shows that processes and outcomes are reasonable and plausible within the frame of theoretic basis and implicit knowledge of system experts or stake-holder », Klügl (2008, pp. 39–41).

Pour une évaluation visuelle

et modes d'intervention de la *face validation*, c'est-à-dire du processus d'évaluation permettant de juger de la *face validity* d'un modèle. Cette méthode a l'avantage de pouvoir être mobilisée dans toutes les étapes du cycle d'évaluation (Balci, 1998) et de systématiquement mettre en avant les connaissances expertes des modélisateurs et thématiciens sur l'objet modélisé.

2.2.1 Limites de la *face validation* et propositions de résolution.

C'est toutefois une méthode assez critiquée par les théoriciens de l'évaluation (Hermann, 1967 ; Balci, 1994), qui considèrent que la *face validation* ne peut qu'être une étape préalable à des méthodes d'évaluation plus quantitatives et formelles. Parmi ces auteurs, Hermann (1967, p. 222) en pointe plusieurs limites – à partir desquelles nous construisons notre proposition d'évaluation visuelle – que l'on peut résumer ainsi.

Un manque de connaissance experte. Pour Hermann, la *face validation* a le défaut d'être trop accessible et de ne pouvoir être menée par n'importe qui, y compris par des gens qui n'auraient qu'une expérience limitée des phénomènes modélisés. Cette remarque est généralisable, et s'applique d'ailleurs quelle que soit la méthode d'évaluation : une expertise ne vaut que par la qualité de l'expert impliqué. Hermann considère ainsi que c'est au modélisateur uniquement de mener cette phase d'évaluation. Il se positionne en cela dans une logique de séparation forte entre les acteurs de l'élaboration d'un modèle : les modélisateurs conçoivent et évaluent le modèle seuls, et les thématiciens n'en sont qu'utilisateurs. L'évaluation visuelle est conçue en opposition à ce principe, et ne peut être pensée sans une collaboration étroite entre les porteurs de connaissances thématiques expertes sur le phénomène modélisé et les porteurs d'une connaissance méthodologique experte sur les manières de le modéliser.

Invraisemblance de certains comportements. Hermann met ensuite en avant le fait que dans un modèle, tous les mécanismes n'ont pas vocation à être vraisemblables et qu'une recherche de plausibilité de tels mécanismes fausserait l'analyse du modèle. Cela revient à ignorer l'importance du dialogue entre modélisateur et évaluateur. Si le modélisateur connaît les « substitutions » opérées dans le modèle, il pourra se garder de chercher à juger de leur vraisemblance. À l'inverse, un expert thématicien pourrait être étonné par certains comportements microscopiques si et seulement si les correspondances entre éléments du modèle et éléments du système modélisé n'étaient pas explicitées. Plus généralement, cette limite est présente s'il n'y a pas d'explicitation des « substitutions » implémentées, c'est-à-dire de la correspondance entre le domaine conceptuel et le domaine du modèle implémenté : le niveau attendu de plausibilité individuelle de chaque comportement doit être décidé pendant la construction du modèle. L'explicitation des choix de modélisation, et l'assurance que leurs enjeux et effets sont bien compris de tous les membres d'un projet de modélisation, sont donc des préalables indispensables à une évaluation visuelle.

Besoin d'explicitation préalable des objectifs. Hermann note enfin que la *face validation* ne peut constituer une méthode d'évaluation adaptée si l'on ne spécifie pas, en amont, les critères qu'elle doit s'attacher à examiner. Le risque est alors d'introduire encore plus de subjectivité dans cette analyse, et en particulier de briser la capacité de reproductibilité ou de justification d'une évaluation. Quand un modèle résulte d'un travail collaboratif les évaluations d'un même résultat peuvent en effet varier, et doivent alors être explicitées autant que possible, ne serait-ce que pour se doter d'un outil de discussion et de débat commun. Il apparaît alors primordial de définir en amont de la modélisation les critères d'évaluation. Cela ne limite aucunement la nécessaire subjectivité et complémentarité des évaluateurs experts, mais permet au contraire d'inscrire leurs discours dans un référentiel de comparabilité. Dès lors, en matière d'évaluation visuelle, la mise au point d'une grille d'analyse systématique nous semble indispensable. Celle-ci doit expliciter, en amont des analyses, l'ensemble des éléments qui seront analysés dans le modèle, les « substitutions » dans le domaine empirique auxquelles ils ont trait, et dès lors, les comportements attendus pour chacun.

2.2.2 Définition et pré-requis de l'évaluation visuelle

C'est à partir de ces remarques que l'on peut proposer une définition de l'évaluation visuelle, qui consiste alors à évaluer la plausibilité du comportement d'un modèle à partir des données qu'il produit. Cette évaluation s'appuie sur une comparaison systématique menée entre 1) les sorties du modèle et 2) les connaissances expertes sur les processus modélisés. Les critères de cette comparaison sont définis en amont de l'évaluation, et doivent expliciter pour chaque sortie analysée la ou les formes qu'elle devrait prendre. Ces éléments, les critères d'évaluation, ne peuvent être formulés par n'importe qui : si le modélisateur autant qu'un expert externe peuvent les spécifier, il convient de s'assurer de l'expertise – thématique et de la connaissance des processus tels que modélisés – de l'évaluateur. L'ensemble des critères d'évaluation forment alors une grille d'analyse qualitative, et la visualisation des différentes sorties du modèle permet de la compléter. Il devient alors possible d'apprécier globalement l'écart entre le modèle et les processus empiriques qu'il représente, sans chercher pour autant à quantifier ni à mesurer cet écart.

2.2.3 Composition de l'évaluation visuelle

Pour établir la correspondance entre le domaine empirique et celui du modèle, nous proposons une dichotomie simple en systématisant l'usage d'un vocabulaire souvent employé de manière plurielle. Nous caractérisons ces éléments en deux grands ensembles : les indices empiriques et les indicateurs de sortie.

Afin d'évaluer la capacité du modèle à reproduire un phénomène observé, il est nécessaire de disposer dans le domaine empirique de « points de repère », qui doivent relever de l'ensemble des échelles spatiales et temporelles que l'on choisit de mettre en scène. Chacun doit pouvoir être estimé, c'est-à-dire être en capacité d'être reproduits et comparables avec d'autres mesures. Nous les qua-

Pour une évaluation visuelle

lifions d'« **indices empiriques** » et les regroupons en deux catégories basées sur la précision avec laquelle ils peuvent être décrits : les « ordres de grandeur », mono-dimensionnels et d'une certitude variable, et les « formes stylisées », qui reposent sur l'agrégation ou la composition de plusieurs ordres de grandeur.

Dans le modèle de simulation, il est nécessaire de définir des « **indicateurs de sortie** », c'est-à-dire des variables informatiques produites par le modèle que l'on pourra ensuite comparer aux « indices empiriques » définis. Comme pour les indices empiriques, les indicateurs de sortie de simulation peuvent être distingués en des formes numériques simples (des scalaires), et en des plus complexes, les indicateurs multi-dimensionnels, qui peuvent être mis en correspondance avec les « formes stylisées » du domaine empirique.

On peut illustrer cette typologie et sa **mise en correspondance** avec un exemple tiré de SIMFEODAL . Dans le domaine empirique, on sait que la part de l'habitat rural dispersé a diminué au cours de la période étudiée, depuis environ 95% en 800 jusqu'à environ 20% en 1200. Ces deux valeurs sont des indices empiriques, de type « ordre de grandeur ». Dans le modèle de simulation, un indicateur de sortie comparable est calculé et correspond au taux de foyers paysans isolés (fig. 2, C). Le taux de foyers paysans isolés en début de simulation est un indicateur de sortie scalaire, de même que celui en fin de simulation. Dans le domaine empirique, la part de l'habitat dispersé a diminué régulièrement au cours du temps (forme stylisée), et on évaluera donc la capacité du modèle à reproduire un processus de concentration semblable. La courbe d'évolution du taux de foyers paysans isolés au cours de la simulation – constituée d'une série d'indicateurs de sortie « scalaires » – constitue alors, dans le domaine du modèle, un indicateur de sortie multi-dimensionnel que l'on pourra analyser : si la courbe part de 95% en début de simulation pour atteindre 20% en fin de simulation, avec une diminution régulière, alors l'indicateur de sortie multi-dimensionnel sera en adéquation avec l'indice empirique de type forme stylisée « évolution de la part de l'habitat rural dispersé ». En définissant en amont de la modélisation un ensemble d'indices empiriques, d'indicateurs de sortie correspondants ainsi que des propriétés attendues des ces derniers, on constitue alors une « grille d'analyse » qui permettra de mener l'évaluation visuelle du modèle.

3 Mise en œuvre pratique de la démarche d'évaluation visuelle

Dans le cas de SIMFEODAL, nous avons identifié une cinquantaine d'indices empiriques utiles à mobiliser, c'est-à-dire à comparer systématiquement avec les indicateurs de sortie de simulation correspondant – ces indices sont présentés, ainsi que leur forme attendue, dans Cura (2020a, pp. 142–150). Cette masse s'explique notamment par le fait que, dans ce modèle descriptif, près d'une dizaine de types d'agents coexistent et interagissent. Il est dès lors important d'observer les comportements (au moins agrégés) de ces différents types d'agents. On peut citer comme exemple le suivi temporel du taux de foyers paysans dispersés, la répartition spatiale des agrégats de population (villes, vil-

lages, etc.), l'évolution du maillage des paroisses ou encore le suivi du nombre de châteaux créés. Ces différents éléments permettent d'appréhender globalement la multi-dimensionnalité de la transformation de la société entre 800 et 1200, et sont nécessaires à la bonne compréhension de sa composante spatiale qui constitue le cœur d'intérêt de SIMFEODAL. La multiplication des indicateurs de sortie, nécessaire à une bonne appréhension de la diversité des comportements d'un modèle, pose alors, en premier lieu, une difficulté concrète ne serait-ce que pour les visualiser.

3.1 *Visualiser de nombreux indicateurs de sortie de simulation*

La masse des indicateurs de sortie (voir 2.2.3) rend, en effet, peu évidente la consultation « en direct » de l'exécution du modèle. De plus, en raison de la stochasticité du modèle, il est nécessaire d'effectuer plusieurs répliques de chaque simulation afin d'en visualiser la tendance agrégée. Le nombre d'indicateurs ainsi générés est important, et nous avons choisi, pour en garantir une consultation aisée, d'organiser ces indicateurs dans un outil interactif dédié à la mise en œuvre de l'évaluation visuelle. La visualisation doit donc se faire a posteriori, et implique dès lors de mener une collecte importante des données générées.

3.1.1 *Une collecte de toutes les données de sortie de simulation*

Cette collecte doit être aussi large que possible, afin d'être en mesure de fournir une base de calcul à l'ensemble des indicateurs de sortie de simulation convenus. Au-delà même des indicateurs de sortie définis, et afin de garantir une reproductibilité des analyses, il est nécessaire de veiller à généraliser la collecte des données produites par le modèle, c'est-à-dire d'enregistrer aussi les données qui ne devraient pas être étudiées. En effet, pour faire face à de potentielles futures modifications du modèle ou assurer une rétro-compatibilité complète, il est obligatoire de disposer des données fines dont l'agrégation permet le calcul des indicateurs et leur visualisation. De plus, au fur et à mesure de l'évolution du modèle, le nombre d'exécution du modèle dont il faut pouvoir conserver les données augmente. Par conséquent, la masse de données produites augmente d'autant, et devient ainsi rapidement trop importante pour être gérée de manière manuelle. Dans le cas de SIMFEODAL par exemple, la masse totale de données collectées sur le modèle au cours des nombreuses étapes de son évolution – plusieurs milliers de simulations – représente des dizaines de millions d'enregistrements en base de données.

3.1.2 *Gestion et organisation d'une importante masse de données*

Pour assurer l'universalité et la simplicité de l'interrogation de ces données massives, nous avons choisi de les organiser au sein d'une base de données relationnelle. La masse de données imposait des contraintes fortes en matière de performance des requêtes, notamment en raison de notre choix de calculer les indicateurs « à la volée ». L'interrogation et l'agrégation de millions d'enregistrements dans un temps acceptable pour une visualisation interactive nous a

Pour une évaluation visuelle

poussé à réaliser un *benchmark* de systèmes de gestion de bases de données (SGBD) relationnels performants (Cura, 2020a, fig. 4.10, p. 226). En raison de la généricité d'interrogation (SQL) et de l'extrême rapidité de traitement des requêtes qu'il permet, nous avons choisi de mobiliser un SGBD « analytique » libre, orienté colonne et haute-performance, OmniSciDB (Root, Mostak, 2016).

3.1.3 Mise en place d'un pipeline de visualisation

Afin de garantir une certaine reproductibilité à nos analyses, nous avons opté pour une chaîne de traitement basée sur des solutions libres et programmables plutôt que sur des outils graphiques, afin que le partage des codes-sources (et des données) permette de reproduire les visualisations de manière stricte. La production graphique des indicateurs de sortie de simulation s'inscrit ainsi dans un « pipeline de visualisation » (fig. 1) formalisé avec le langage libre R, qui orchestre l'interrogation des données, leur filtrage, la production des visualisations (carto)graphiques et tabulaires, ainsi que leur organisation et présentation au sein d'une plateforme interactive dédié, SIMEDB (fig. 2).

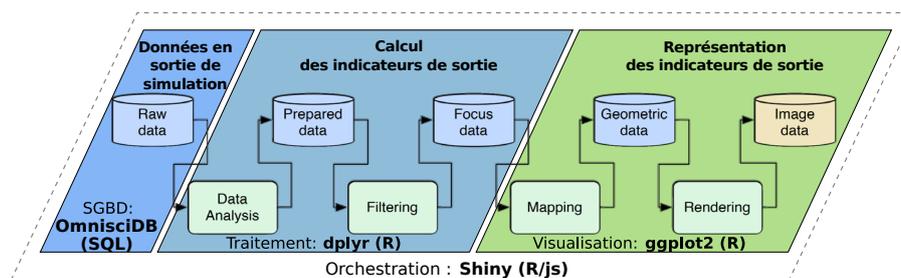


FIGURE 1. Le « pipeline » de visualisation – d'après Keim et al. (2010) – et son implémentation dans SIMEDB.

3.2 Comparer visuellement des simulations

La co-construction d'un modèle de simulation s'appuie sur une itération fréquente entre évolutions du modèle conceptuel, implémentation et évaluation des sorties. Chaque version du modèle entraîne en effet des modifications dans les valeurs de paramètres ou dans le détail des mécanismes, et il est donc important de pouvoir comparer les effets de ces changements. Pour appliquer une démarche d'évaluation visuelle, il est alors nécessaire de procéder régulièrement à une comparaison visuelle entre les simulations.

3.2.1 Visualiser les différences entre simulations

Dans la définition de l'évaluation visuelle (2.2.2), nous explicitons le besoin de recourir à une comparaison entre les indicateurs de sortie du modèle et des formes attendues d'indices empiriques exprimées dans une grille d'analyse.

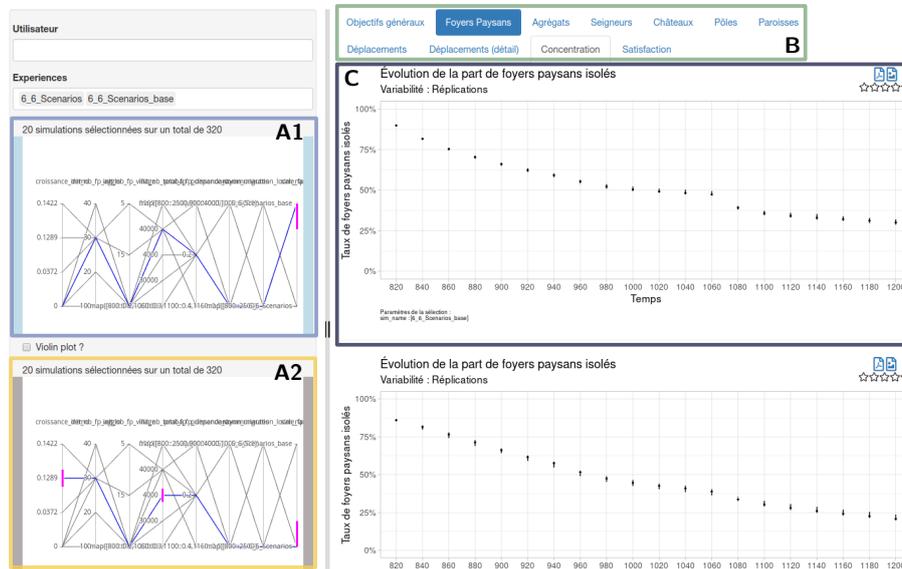


FIGURE 2. Interface de la plateforme SIMEDB.

À mesure que la construction du modèle évolue (calibration notamment), les écarts entre ces valeurs diminuent, jusqu'à atteindre des valeurs trop faibles pour être jugées thématiquement significatives. Il devient alors nécessaire de mener une évaluation plus fine, en comparant directement les indicateurs de sortie des simulations les unes avec les autres (évaluation relative) plutôt qu'avec le référentiel absolu que constituent les indices empiriques. SIMEDB est donc orientée vers une approche comparative plus que normative : si une partie des indices empiriques sont figurés dans certaines visualisations, l'accent est majoritairement placé sur la visualisation des différences entre simulations.

3.2.2 De la visualisation à la comparaison interactive

Au cours de son « cycle de vie », le modèle SIMFEODAL est passé par sept versions majeures et près d'une cinquantaine de sous-versions, sans compter les essais de variations de jeux de paramètres au sein même de ces sous-versions. L'identification des configurations du modèle (versions concurrentes, valeurs des jeux de paramètres mobilisés) devenait alors un problème en elle-même. Nous avons alors choisi de renforcer l'interactivité de SimEDB, en permettant notamment une sélection des simulations à comparer par un outil graphique. Dans l'interface de SimEDB, le choix des simulations dont on souhaite visualiser les indicateurs de sortie se fait ainsi via une sélection graphique des valeurs de paramètres à isoler, via un *brushing* sur des graphiques interactifs en coordonnées parallèles (fig. 2, A1 et A2). En répétant ces opérations de « brushing »

Pour une évaluation visuelle

dans l'interface , on peut ainsi comparer rapidement les indicateurs de sortie des simulations (accessibles via le double-menu d'onglets B). L'outil étant interactif, nous préférons le présenter sous forme de vidéo plutôt que d'en réaliser une description textuelle forcément moins adaptée. Une présentation de SIMEDB est ainsi disponible dans un dépôt dédié (Cura, 2020b).

4 Conclusion

Nous proposons une méthode d'évaluation de modèles de simulation stochastiques, descriptifs, et fondés sur la connaissance experte plus que sur des données empiriques. Cette méthode en elle-même a été conçue, raffinée et mise en œuvre collectivement dans le cadre de la co-construction d'un modèle de simulation ancré dans un passé lointain et sur le temps long. Au vu des évolutions du modèle SIMFEODAL que l'évaluation visuelle a accompagné, nous pouvons dire que le résultat de cette démarche est concluant. Le modèle produit des résultats comparables aux attentes que nous en avons et donc conforme aux connaissances empiriques sur lesquelles il s'appuie. L'exécution systématique de ce processus d'évaluation visuelle a facilité la réalisation d'allers-retours nombreux et réguliers entre le modèle conceptuel, son implémentation générale, le détail technique des mécanismes et les valeurs de paramètres testées. Les dernières versions du modèle, les mieux calibrées, nous permettent ainsi de valider la plausibilité des hypothèses thématiques que le modèle cherchait à expérimenter. La robustesse de SIMFEODAL a enfin été testée par l'application d'une analyse de sensibilité, elle aussi menée de manière visuelle (Cura, 2020a, pp. 290–322). Il nous semble donc que l'évaluation visuelle, enrichissant l'étape préalable et peu formalisée qu'est la *face validation*, est une méthode tout à fait valable pour l'évaluation de modèles de simulations qui se prêteraient mal aux méthodes plus formelles. Plus généralement, nous gageons que la construction et l'évaluation de tous les types de modèles de simulation gagneraient à mobiliser plus fréquemment des méthodes visuelles.

Bibliographie

- Amblard F., Rouchier J., Bommel P. (2006). Evaluation et validation de modèles multi-agents. *Modélisation et simulation multi-agents. Applications pour les Sciences de l'Homme et de la Société*, Hermès, p. 103–140.
- Balci O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of operations research*, vol. 53, n° 1, p. 121–173.
- Balci O. (1998). Verification, validation, and accreditation. In *Proceedings of the 1998 WSC*, vol. 1, p. 41–48.
- Cura R. (2020a). *Modéliser des systèmes de peuplement en interdisciplinarité : co-construction et exploration visuelle d'un modèle de simulation*. Thèse de doctorat non publiée, Université Panthéon-Sorbonne. (<https://tel.archives-ouvertes.fr/tel-02944405>)
- Cura R. (2020b). *Présentation de SimEDB*. DOI: 10.5281/zenodo.4554338.

- Cura R., Tannier C., Leturcq S., Zadora-Rio E., Lorans E., Rodier X. (2017). Transition 8 : 800-1100. Fixation, polarisation et hiérarchisation de l'habitat rural en Europe du Nord-Ouest. In L. Sanders (Ed.), *Peupler la terre - De la préhistoire à l'ère des métropoles*, p. 301–332. Presses universitaires François Rabelais.
- De Runz C. (2008). *Imperfection, temps et espace : modélisation, analyse et visualisation dans un SIG archéologique*. Thèse de doctorat non publiée, Université de Reims - Champagne Ardenne. (<https://tel.archives-ouvertes.fr/tel-00560668>)
- Durand-Dastès F. (1999). Jamais deux fois... Ou : de quelques précautions à prendre avec le temps. *Travaux de l'Institut de Géographie de Reims*, vol. 26, n° 101, p. 5–23.
- Edmonds B., Moss S. (2005). From KISS to KIDS – An 'Anti-simplistic' Modelling Approach. In P. Davidsson, B. Logan, K. Takadama (Eds.), *International workshop on multi-agent systems and agent-based simulation*, p. 130–144. Springer Berlin Heidelberg.
- Hermann C. (1967). Validation Problems in Games and Simulations with Special Reference to Models of International Politics. *Behavioral Science*, vol. 12, n° 3.
- Keim D., Kohlhammer J., Ellis G., Mansmann F. (2010). *Mastering the Information Age Solving Problems with Visual Analytics*. Eurographics Association.
- Klügl F. (2008). A Validation Methodology for Agent-based Simulations. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, p. 39–43. New York, NY, USA, ACM.
- Naylor T. H., Finger J. M. (1967). Verification of Computer Simulation Models. *Management Science*, vol. 14, n° 2, p. B92-B106.
- Petty M. D. (2010). Verification, validation, and accreditation. *Modeling and simulation fundamentals: Theoretical underpinnings and practical domains*, p. 325–372.
- Root C., Mostak T. (2016). MapD: A GPU-powered Big Data Analytics and Visualization Platform. In *ACM SIGGRAPH 2016 Talks*, p. 73:1–73:2. New York, NY, USA, ACM.
- Tannier C., Cura R., Leturcq S., Zadora-Rio E. (2020). An agent-based model for exploring the combined effects of social and demographic changes on the concentration and hierarchy of rural settlement patterns in North-Western Europe during the Middle Ages (800–1200 CE). *Journal of Anthropological Archaeology*, vol. 59, p. 101204.

Localiser le foncier en 1828 : l'exemple du cadastre ancien de Rouen

Marie Havret¹, Marion Le Texier², Sophie de Ruffray³

1. UMR 6266 IDEES, Université de Rouen Normandie

marie.havret@univ-rouen.fr

2. UMR 6266 IDEES, Université de Rouen Normandie

marion.le-texier@univ-rouen.fr

3. UMR 6266 IDEES, Université de Rouen Normandie

sophie.deruffray@univ-rouen.fr

RÉSUMÉ. *Le développement des outils informatiques a permis de porter un nouveau regard sur de nombreuses sources historiques, tant dans le recueil des données que dans l'usage de nouvelles approches pour les analyser. C'est notamment le cas du cadastre ancien qui permet de collecter des données foncières fiables et homogènes, à l'échelle intra-urbaine pour l'ensemble du territoire français dès le début du XIXe siècle. Cet article compare deux méthodes de localisation automatique des adresses permettant de répondre à l'absence de plan de référence. Il se base sur l'étude du cadastre ancien de la ville de Rouen, réalisé à partir de 1827.*

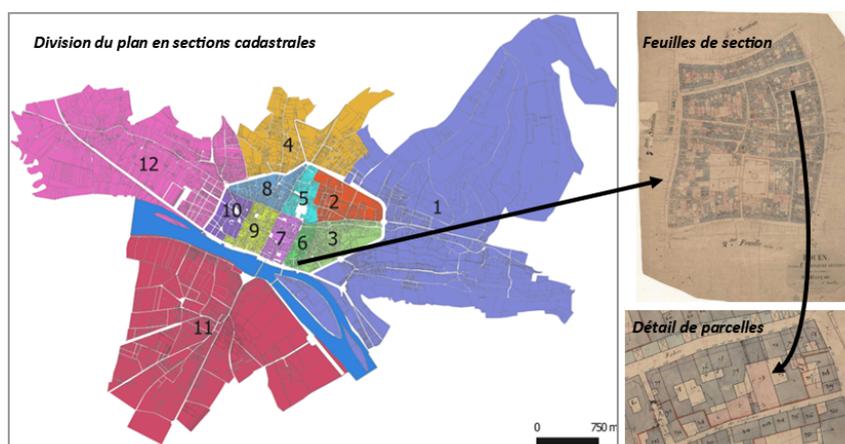
ABSTRACT. *The use of historical sources is renewed by the latest developments in computer tools, both in the collection of data and in the approaches used to analyse them. This is notably the case of the Napoleonic cadastre, a major historical source to gather reliable and homogenous intra-urban land data (dwelling) for the whole of France since the beginning of the 19th century. This article describes two methods to locate cadastre data in the absence of a reference plan. It is based on a study of the Napoleonic cadastre of the city of Rouen (France), carried out from 1827.*

MOTS-CLÉS : *base de données géohistorique, cadastre, localisation intra-urbaine, reproductibilité, géocodage*

KEYWORDS: *geohistorical database, cadastre, intra-urban location, reproducibility, geocoding*

1. Introduction

Elaboré à partir du début du XIXe siècle en France, le cadastre est une source de données extrêmement riche pour les chercheurs, puisqu'il permet de recueillir des informations fiables, exhaustives et homogènes sur les propriétés foncières pour l'ensemble du territoire national, de manière comparable dans le temps. Il est composé de deux documents distincts : le plan cadastral (FIG. 1), qui cartographie la géométrie des parcelles, et la matrice cadastrale (FIG. 2), qui recueille les informations attributaires : propriétaire (nom, profession, ...), propriété (adresse, superficie, ouvertures ...) et annexes (jardin, ...).



Source: Archives départementales de la Seine-Maritime, 1827

FIGURE 1 – Plan cadastral de Rouen

Les matrices cadastrales sont mises à jour à mesure des changements de propriétaires, des modifications du bâti et des réévaluations périodiques. Cependant, le plan parcellaire réalisé lors de l'introduction du cadastre (entre 1808 et 1848 selon les territoires infra-nationaux) a rarement été mis à jour avant 1970 et ne permet donc pas de suivre les modifications de propriétés (agrégation ou morcellement du parcellaire ; création, disparition ou modification du bâti). La seule information permettant de localiser les parcelles recensées dans le cadastre au cours du temps est l'adresse, renseignée par un numéro et un nom de rue dans les matrices.

La littérature rend compte d'un ensemble d'approches localisant des objets historiques à partir d'adresses (TAB. 1), dont la mise en oeuvre rencontre des pré-requis et des difficultés techniques plus ou moins importants et surmontables selon le contexte de recherche. Dans cet article, nous évaluons la possibilité de localiser des parcelles à partir de leurs adresses sur un plan de rues dans un temps limité à partir de l'exemple du cadastral historique rouennais de 1828.

Localiser le foncier en 1828

NOMS, PRÉNOMS, PROFESSIONS ET DEMEURES des propriétaires.	N.° du PLAN.	CANTONS, TRIAGES ou LIEUX DITS.	NATURE des PROPRIÉTÉS.	CONTENANCES en mesures métriques.			CLASSEMENT.	REVENU. francs. c.	NOMBRE D'OUVERTURES IMPOSABLES.			
				arp.	p.	mit.			Portes coulées et de magasin.	Fenêtres ou cloisons à deux ouvertures.	Maisons du quatrième étage.	Maisons n'ayant qu'une porte.
Juquin Guillaume ou Théophile	1016	Rue Bihorel	Jardin	22	52		quatrième	2	52			
cf. com.	1015	Rue Bihorel	Maison			165		165		22		
			Jardin			61	premier		61			
M. de la Fontaine Christophe ou Jacques	1016		Jardin	05	51		quatrième	3	41			

Source: Archives départementales de la Seine-Maritime, 1828

Monsieur Juquin, tailleur de profession et résidant au 11 Rue de l'Hôpital à Rouen, détient une propriété située au 19 Bis Rue Bihorel. Elle est composée de deux parcelles, la maison et le jardin (respectivement n°1015 et n°1014)

- **Première ligne** : Le jardin, d'une superficie de 252 m² (colonne "Contenance"), est évalué à 2,52 francs (colonne "Revenu"). La quatrième classe (colonne "Classement") correspond aux jardins d'agrément.

- **Deuxième ligne** : La maison est évaluée à 165 francs et dispose de 22 portes et fenêtres (colonne "Nombre d'ouvertures imposables"). Sa superficie au sol est connue grâce à la ligne suivante.

- **Troisième ligne** : La superficie au sol de la maison est de 61 m², correspondant à un revenu de 0,61 francs.

FIGURE 2 – Extrait d'une matrice cadastrale (4e section, Rouen, 1828)

2. Le cadastre rouennais de 1828

La ville de Rouen est cadastrée au cours des années 1827 et 1828. Le plan est d'abord levé, puis les matrices cadastrales sont rédigées. Un registre est tenu par section cadastrale. Seul le plan cadastral de 1827 est disponible à Rouen, mais les matrices sont mises à jour, et ce jusqu'en 1971.

L'ensemble des documents cadastraux de la ville sont disponibles aux Archives départementales de Seine-Maritime (AD76). Les données des matrices cadastrales de 1828 ont été saisies manuellement dans un tableur, représentant 13 911 parcelles bâties. Le plan cadastral de Rouen a été scanné par les AD76¹ puis vectorisé dans le cadre du projet de recherche *Morphologie urbaine*, porté par l'Université de Rouen et la Région Haute-Normandie (2012-2015)². Lors de

1. Les plans sont consultables à l'adresse suivante : <http://www.archivesdepartementales76.net/rechercher/archives-en-ligne/plans-du-cadastre/>

2. http://www.inventaire.culture.gouv.fr/pdf/2015_Haute_Normandie.pdf

Méthode		Prérequis	Avantages	Inconvénients	Exemple
Vectorisation des adresses à partir des plans anciens	Manuelle	Plan ancien géoréférencé	Simplicité technique Qualité des informations	Besoin humain et temps conséquent (selon le volume d'informations à traiter)	Orford et al, 2005
	Edition collaborative	Plan ancien géoréférencé Plateforme d'édition collaborative	Qualité des informations (si protocole de vérification) Possibilité de recueillir un volume important de données	Besoin humain et temps conséquent (selon le volume d'informations à traiter) Plateforme d'édition collaborative nécessaire Protocole de vérification	Projet Building Inspector (NYPL)
	Semi automatique	Plan ancien géoréférencé	Automatisation du processus	Complexité technique Performance limitée pour la reconnaissance du texte sur des documents anciens	Yu et al, 2016
Géocodage des adresses anciennes	A partir d'une base de données actuelles	Base de données d'adresses actuelles	Simplicité technique Automatisation d'une partie du processus	Vulnérabilité à la qualité des adresses et à la transformation de la ville (modification des rues, des numéros, des noms de rues)	Walford, 2018
	A partir d'une base de données actuelles et d'un répertoire historique	Base de données d'adresses actuelles & répertoire d'adresses anciennes	Automatisation du processus	Vulnérabilité à la qualité des adresses et à la transformation de la ville Temps conséquent (création du répertoire historique)	Lan et Longley, 2019 (Projet GB1900)
	A partir d'un répertoire historique et par édition collaborative	Répertoire d'objets géohistoriques Plateforme d'édition collaborative	Qualité des informations (si protocole de vérification) Possibilité de recueillir un volume important de données	Temps conséquent (création du répertoire historique) et besoin humain Plateforme d'édition collaborative Protocole de vérification	Cura et al, 2018
Par positionnement proportionnel ou équidistant		Vectorisation des rues	Simplicité technique Automatisation du processus	Vulnérabilité à la morphologie urbaine Nécessité de définir le premier et dernier numéro de chaque rue étudiée	Varet-Vitu et al, 2020

Les deux méthodes présentées dans l'article s'inspirent des méthodes apparaissant en orange

TABLE 1 – Méthodes de localisation des adresses anciennes

la vectorisation du plan, les membres du projet n'ont pas renseigné l'identifiant unique des parcelles dans la table attributaire, élément qui aurait permis de passer automatiquement du plan à la base de données.

Nous avons constitué un échantillon de contrôle afin de comparer les positions réelles et les positions estimées par les méthodes de localisation automatique. Les numéros de parcelle de cet échantillon ont été saisis dans la table attributaire du *shapefile* reprenant le plan cadastral. Trois feuilles de section situées dans des zones variées en termes de morphologie urbaine et d'ancienneté du bâti ont été sélectionnées, pour un total de 733 parcelles (FIG. 1):

- Une feuille de la section 6 : située dans l'hypercentre, cette section est caractérisée par un bâti dense et ancien.
- Une feuille de la section 4 : elle correspond à l'extension urbaine du début du XIXe siècle, à l'extérieur des boulevards intra-urbains. Elle est majoritairement constituée de maisons individuelles disposant de jardins privés.
- Une feuille de la section 11 : située sur la rive gauche de la ville, cette feuille parcellaire est caractérisée par une faible urbanisation, concentrée le long des rues principales.

Localiser le foncier en 1828

Enfin, afin de localiser les numéros et rues des propriétés bâties, nous avons vectorisé et renseigné les 674 rues figurant sur le plan cadastral dans un SIG.

3. Méthodes de localisation automatique des adresses

3.1. *Positionnement par géocodage à partir d'une base de données actuelles*

La méthode de positionnement par géocodage utilise les adresses actuelles pour définir la localisation des adresses anciennes (FIG. 3). Elle postule donc une relative stabilité entre les adresses anciennes et actuelles, en dépit des modifications de la structure urbaine.

Afin d'attribuer des coordonnées géographiques aux parcelles bâties à partir des adresses recensées dans les matrices, nous avons utilisé la librairie R "banR", créée par Joël Gombin et Paul-Antoine Chevalier, qui permet d'interroger directement l'API de la Base Adresse Nationale (BAN), mise à jour de façon hebdomadaire, et de récupérer les coordonnées géographiques de l'adresse demandée (Giraud *et al.*, 2019). Les parcelles sans numéro sont positionnées aléatoirement dans la rue. Aucune localisation n'est attribuée aux parcelles dont les rues ne figurent pas dans la BAN.

3.2. *Positionnement proportionnel*

La méthode de positionnement proportionnel déduit la position de chaque parcelle le long de la rue à partir de son numéro et nom de rue, du nombre de parcelles bâties et de la longueur de la rue (FIG. 4). Cette méthode positionne les numéros de façon régulière le long de la rue, en faisant correspondre le début de la rue au sens géométrique (*startpoint*) au premier numéro de la rue³. Cette méthode postule donc une régularité de la distribution des habitations le long des rues.

Cette méthode de positionnement proportionnel ne fonctionne donc que si le premier et le dernier numéros de rue ainsi que le sens de la numérotation sont connus. Nous avons localisé le *startpoint* suivant la logique édictée pour Paris en 1805 et généralisée à l'ensemble des communes françaises par l'édit royal du 23 avril 1823 (Duvergier, 1823) : pour les rues parallèles à la Seine, la numérotation suit le sens de l'écoulement de l'eau ; pour les rues perpendiculaires ou obliques, les premiers numéros commencent du côté de la rue qui est le plus proche de la Seine. Cette méthode nécessite toutefois d'identifier au préalable les premiers et derniers numéros de chaque rue, soit par un plan, soit par un dictionnaire

3. L'ensemble des données et des codes R utilisés dans ce papier seront publiés prochainement sur <https://gitlab.huma-num.fr/mhavret/>

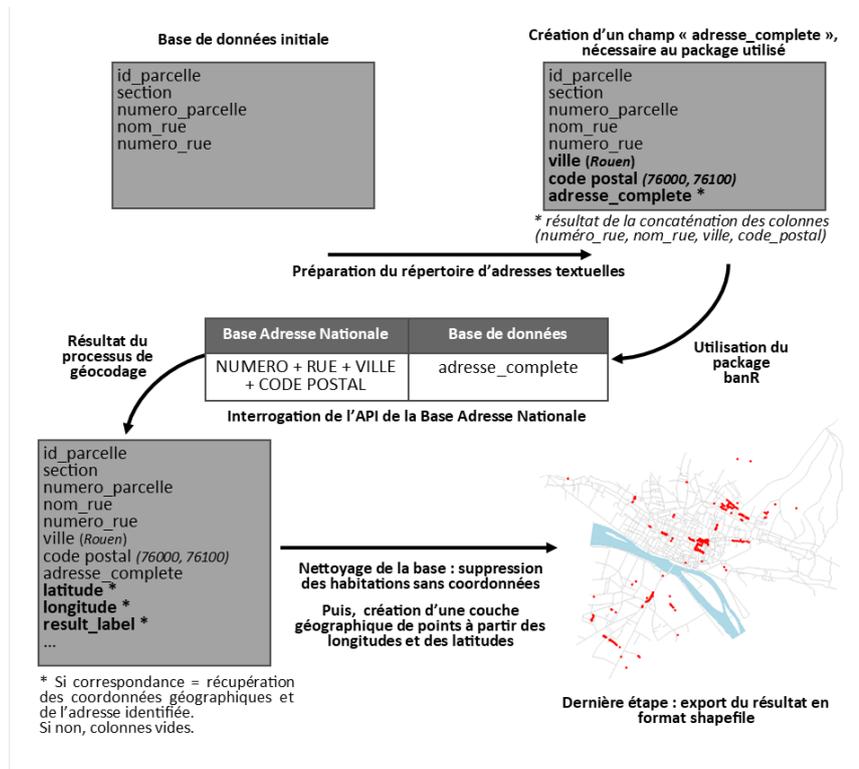


FIGURE 3 – Méthode d'attribution d'une localisation à partir du géocodage des adresses

recensant la localisation des numéros de rue comme dans le cas parisien (Varet-Vitu *et al.*, 2020), par exemple.

3.3. Critères d'évaluation

Trois critères sont pris en compte pour évaluer les méthodes :

- la complétude, mesurée par le nombre de maisons que l'algorithme parvient à positionner (correctement ou non) sur le plan de rues,
- la précision, mesurée par la distance entre la position calculée et la position réelle que nous étudions de façon agrégée et par la comparaison cartographique,
- la distorsion, mesurée par le différentiel analytique obtenu selon la méthode de localisation, que nous illustrons par la mesure de l'autocorrélation spatiale des valeurs foncières (prix au mètre carré) calculée à partir des positions réelles et des positions estimées.

Localiser le foncier en 1828

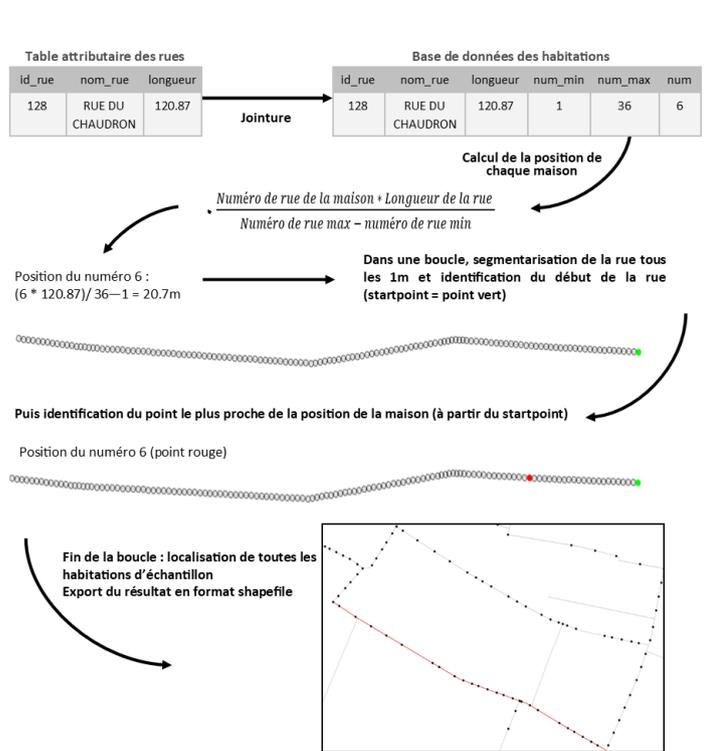


FIGURE 4 – Méthode de positionnement proportionnel

Les classes de distance choisies pour discrétiser les distances d'erreur se basent sur celles établies par Cura *et al* (2018) dans leur étude sur le géocodage des adresses parisiennes aux XIXe et XXe siècles afin de faciliter les comparaisons. A partir d'un échantillon d'adresses géocodées et de leurs positions réelles définies manuellement, les auteurs proposent les classes suivantes : entre 0 et 15m, le positionnement est correct ; entre 15 et 55m, la rue est correcte, mais les numéros des bâtiments sont légèrement mal placés ; entre 55 et 155m, la rue est correcte dans la plupart des cas, mais les numéros des bâtiments sont loin de leur position correcte ; au-dessus de 155m, les positions sont mauvaises.

4. Résultats et discussion

4.1. Complétude

A l'issue des traitements, les deux méthodes de localisation parviennent chacune à localiser 714 parcelles (97% de l'échantillon). La méthode de position-

nement proportionnel ne parvient pas à localiser les parcelles dont le numéro de rue est inconnu. Cette situation concerne les sections les plus récemment urbanisées en 1828 (5 parcelles pour la section 4, 14 parcelles pour la section 11). La méthode de géocodage ne parvient pas à localiser les parcelles de deux rues de la section 11, toutes deux disparues aujourd'hui.

4.2. Précision

Les modifications de toponymie et de réordonnement des numéros de rue ayant eu lieu entre 1828 et aujourd'hui expliquent l'importance des écarts de distance obtenus par la méthode de géocodage (FIG. 5). La méthode de positionnement proportionnel, n'étant pas vulnérable à ces changements, permet d'obtenir, globalement, une plus grande précision dans la localisation des parcelles historiques.

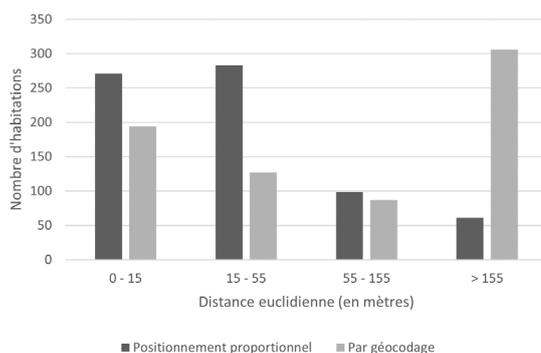


FIGURE 5 – Distance entre les positions réelles et les positions estimées selon les deux méthodes (n=713)

Le regroupement spatial des écarts de distance les plus importants renseignent sur les facteurs d'erreur de localisation (FIG. 6). On remarque ainsi que la méthode de positionnement proportionnel est moins précise lorsque l'urbanisation n'est pas homogène, notamment lorsque un tronçon de rue est très densément construit, alors que le reste ne l'est pas, à l'image de la rue Saint-Julien (section 11). C'est également le cas de la rue Eau de Robec (section 6) et du chemin aux Boeufs (section 4). Plus la rue est longue, plus le problème est exacerbé. Ces rues concentrent la quasi totalité des positions incorrectes (>155m).

Les résultats obtenus par la méthode de géocodage sont beaucoup plus hétérogènes spatialement. Cette méthode fonctionne mieux dans l'hypercentre de Rouen qu'en périphérie, notamment pour les rues anciennes dont la morphologie, le nom, et la numérotation sont stabilisés depuis deux siècles. Les résultats sont moins satisfaisants pour les sections 4 et 11, où plusieurs noms de rue

Localiser le foncier en 1828

actuels correspondent à ceux de rues disparues ou renommées depuis. C'est par exemple le cas de la "Rue du Barbet", anciennement située dans l'hypercentre de Rouen, associée à la "Rue Henri Barbet", actuellement située dans l'extrême ouest de la ville. Logiquement, les jointures par proximité sémantique partielle sont d'autant plus sources d'erreur que le seuil de ressemblance accepté est bas. Avec les paramètres par défaut de la fonction BanR, l'adresse "5 Rue du Corbeau" est par exemple associée à la "Rue du Fardeau".

Au regard de ce critère, la méthode d'attribution d'une localisation par géocodage n'apparaît pas satisfaisante, en l'état actuel. Ces résultats soulignent la nécessité d'associer cette méthode à une validation manuelle ou à une base de données historiques définie en amont.

4.3. Distorsion

Pour la méthode de positionnement proportionnel dont les résultats sont satisfaisants au regard des deux premiers critères d'évaluation, une évaluation thématique a également été conduite. La comparaison des valeurs locales d'autocorrélation spatiale⁴ du prix du foncier ramené au mètre carré (TAB. 2 et FIG. 7) montre que les résultats obtenus par une localisation manuelle ou automatique des parcelles sont relativement homogènes pour les sections 4 et 11, quelque soit la taille du voisinage pris en compte. Cependant, la section 6 fait apparaître de forts écarts entre les deux échantillons, notamment avec un voisinage restreint.

	1 voisin	5 voisins	10 voisins
Section 4	3.92	0.56	0.32
Section 6	100.5 (***)	11.04 (**)	5.41 (*)
Section 11	4.62 (.)	3.59	2.17

Les valeurs correspondent aux valeurs du Chi². Signif. p-value : 0.001 '***'; 0.01 '**'; 0.05 '*'; 0.1 '.'; 1 ' '.

TABLE 2 – Résultats du test du Chi² d'homogénéité.

5. Conclusion et perspectives

Comme nous le rappelons en introduction, le cadastre constitue une source unique pour l'analyse comparée du foncier dans le temps et dans l'espace. Cependant, le format des archives (papier ou image pixelisée) en rend la manipulation difficile. Nous avons comparé la qualité de l'information géographique obtenue avec deux méthodes de positionnement automatique des parcelles à

4. L'indice I de Moran local a été utilisé (Anselin, 1995).



FIGURE 6 – Distance euclidienne entre les positions réelles et estimées par la méthode de positionnement proportionnel (à gauche) et par géocodage (à droite)

Localiser le foncier en 1828

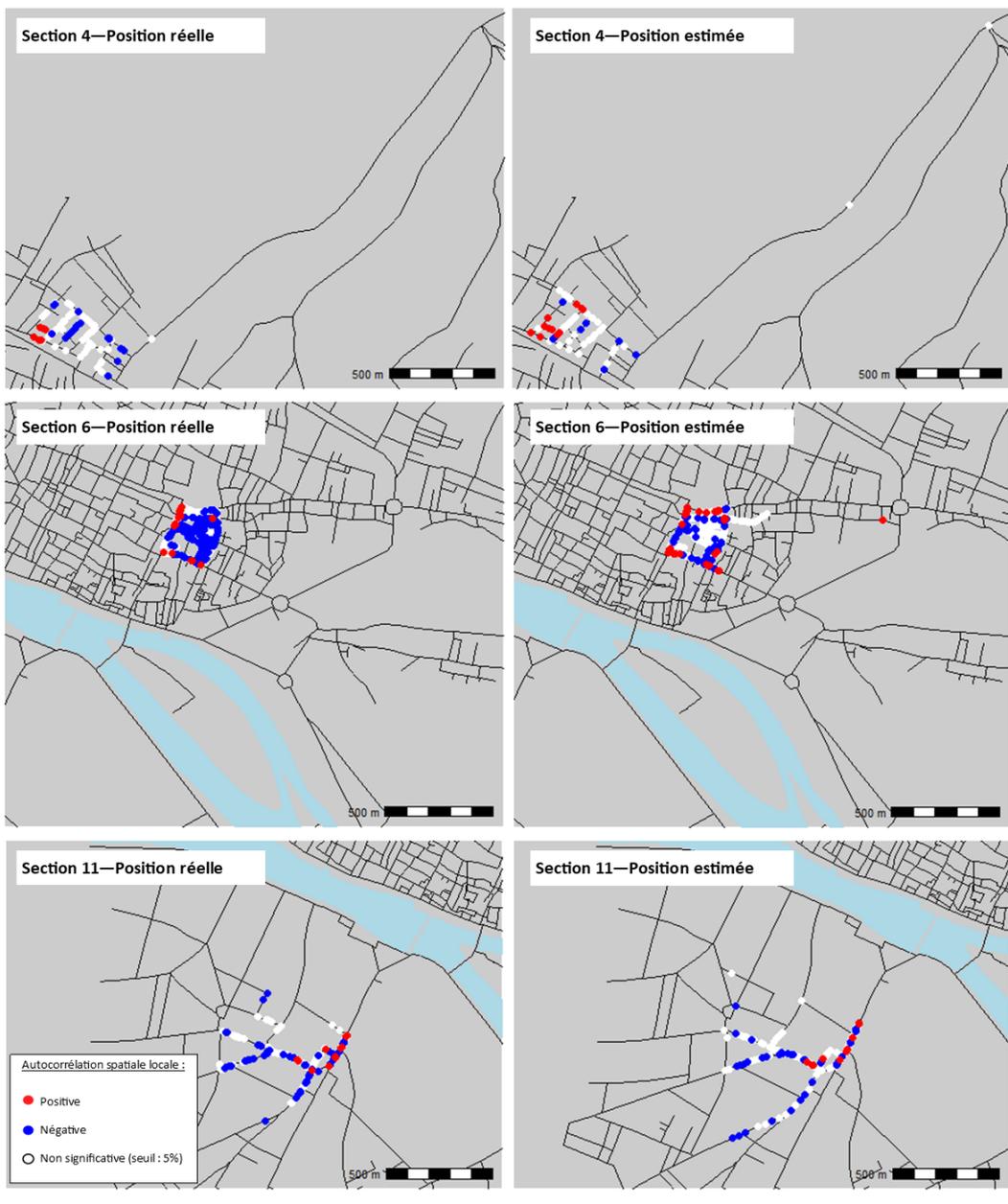


FIGURE 7 – Clusters issus des mesures d'autocorrélation spatiale de la variable "Prix au m²", avec les positions réelles (à gauche) et estimées (à droite) - 1 voisin

partir des adresses renseignées dans les matrices cadastrales (seul élément mis à jour pendant plus de 100 ans).

Nous avons montré qu'une évaluation basée sur des critères géométriques seuls (complétude, précision) conduirait à considérer que le positionnement proportionnel des numéros de rue -tel que défini dans cet article- produit d'excellents résultats. Cependant, l'utilisation des données ainsi définies dans une analyse thématique montre l'étendue des distorsions des résultats obtenus en comparaison de ceux issus d'une localisation manuelle des parcelles. Une des principales perspectives de ce travail est donc d'améliorer la méthode de positionnement proportionnel en définissant par exemple des points de contrôle (bâti remarquable) comme le font Anne Varet-Vitu *et al* (2020) pour affiner les localisations utilisées. La méthode de géocodage pourrait, quant à elle, être perfectionnée en constituant un répertoire d'adresses historiques en amont.

Bibliographie

- Anselin L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, vol. 27, n° 2, p. 93–115.
- Cura R., Dumenieu B., Abadie N., Costes B., Perret J., Gribaudo M. (2018). Historical collaborative geocoding. *ISPRS International Journal of Geo-Information*, vol. 7, n° 7, p. 262.
- Duvergier J.-B. (1823). *Collection complète des lois, décrets, ordonnances, règlements et avis du conseil d'état* (vol. 24). imprimé par Charles Noblet.
- Giraud T., Pecout H., Ysebaert R. (2019). Accessibilité et données openstreetmap sur r : exemple pratique sur les maternités de la nièvre et de la seine-saint-denis. *Séminaire CIST de juin 2019 - Atelier calculs d'accessibilités avec des données en libre accès*.
- Lan T., Longley P. (2019). Geo-referencing and mapping 1901 census addresses for england and wales. *ISPRS International Journal of Geo-Information*, vol. 8, n° 8, p. 320.
- Orford S., Dorling D., Mitchell R., Shaw M., Smith G. D. (2002). Life and death of the people of london: a historical gis of charles booth's inquiry. *Health & place*, vol. 8, n° 1, p. 25–35.
- Varet-Vitu A., Marraud M., Mermet É. (2020). Spatialités sociales à paris à la veille de la révolution. *Histoire urbaine*, n° 2, p. 157–186.
- Walford N. S. (2019). Bringing historical british population census records into the 21st century: A method for geocoding households and individuals at their early-20th-century addresses. *Population, Space and Place*, vol. 25, n° 4, p. e2227.
- Yu R., Luo Z., Chiang Y.-Y. (2016). Recognizing text in historical maps using maps from multiple time periods. In *2016 23rd international conference on pattern recognition (icpr)*, p. 3993–3998.

Communications par Session

Extraction de contenu pour le GeoWeb



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Constitution de corpus thématique : Pour un meilleur suivi du territoire de la Métropole de Montpellier Méditerranée

Rodrique Kafando^{1,4}, Rémy Decoupes¹, Lucile Sautot^{1,2},
Maguelonne Teisseire¹, Christiane Weber^{1,3}

1. TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE,
Montpellier, France
prenom.nom@inrae.fr

2. AgroParisTech, Montpellier, France
lucile.sautot@agroparistech.fr

3. CNRS, France
christiane.weber@cnrs.fr

4. Montpellier Méditerranée Métropole, France

RÉSUMÉ. Dans le cadre d'un partenariat avec la métropole de Montpellier (3M) et son service ville intelligente, nous établissons une plateforme permettant l'intégration et l'analyse de données hétérogènes massives pour une observation intelligente du territoire concerné. Dans cet article, nous décrivons un processus de récolte automatique de documents de sources diverses, selon trois thématiques choisies par 3M qui sont l'urbanisation, l'agriculture et l'hydrologie. A partir de termes graines à dire d'experts enrichies grâce à des articles Wikipédia puis Google, un vocabulaire de concepts est automatiquement constitué. Les documents obtenus à partir d'un moteur de recherche sont ensuite évalués à l'aide d'une mesure sémantique spécifique. Nous détaillons les premiers résultats obtenus et présentons les avantages du protocole proposé.

ABSTRACT. In collaboration with the metropolis of Montpellier (3M) and its smart city service, we are establishing a platform allowing the integration and analysis of massive heterogeneous data for intelligent observation of the territory concerned. In this article, we describe the process implemented to ensure an automatic collection of documents according to three themes chosen by 3M, which are urbanization, agriculture and hydrology. Results are detailed and the benefits of this scalable solution are highlighted.

MOTS-CLÉS : Observation du territoire, Constitution automatique de corpus

KEYWORDS: Land Planning, automatic corpus constitution

1. Introduction

La Métropole de Montpellier Méditerranée (3M) développe depuis plusieurs années une politique pro-active en faveur de développements numériques innovant par la mise en œuvre d'approches fondées sur l'hétérogénéité des sources, l'extraction de connaissances à partir de données massives dans le but de produire des outils d'aide à la décision. L'intégration de données hétérogènes, en prenant en compte l'écosystème « ville intelligente » de la Métropole (labos, start up, collectifs, etc.), se fonde sur des informations quantitatives (démographiques, économiques, climatiques, etc.) mais aussi sur des informations venant d'autres sources de données, officielles et non officielles. La finalité est de montrer le gain et l'efficacité de la mise en relation de données gérées jusqu'ici en silo et de l'organisation des interactions qui peuvent en découler. C'est dans ce contexte que se positionnent les travaux présentés dans cet article sur un processus de constitution de corpus thématique (ou corpus dédié) à partir du Web sur le territoire concerné. Cette étude rentre dans le cadre de travaux qui visent la mise en relation et l'analyse de données textuelles et hétérogènes. Le but principal est de pouvoir collecter par couple de thématique et territoire, des données textuelles pertinentes qui couvrent une grande diversité de sources. Le challenge et le verrou associé est de ne pas se limiter aux documents purement officiels (souvent disponibles dans l'open data de la métropole). Il devient indispensable de disposer d'une terminologie moins académique que celle utilisée par les différentes collectivités territoriales afin de pouvoir accéder à l'expression de la composante sociétale, indispensable à une véritable perception de la dynamique du territoire dans sa globalité.

Dans la suite de cet article, en Section 2, nous listons un panorama des travaux sur la constitution de corpus thématiques, en particulier ceux qui ont inspiré les bases de notre démarche. Ensuite, la méthodologie que nous proposons pour la constitution d'un corpus propre à notre cas d'étude est détaillée dans la Section 3. Dans la Section 4, nous présentons les premiers résultats quantitatifs suite aux différentes expérimentations réalisées. Enfin, nous dressons un bilan et les perspectives associées en Section 5.

2. État de l'art

La mise en oeuvre des concepts de la ville intelligente devrait permettre aux villes et métropoles d'améliorer l'efficacité de leurs services de gestion (Guéranger, Mathieu-Fritz, 2019), en offrant, notamment, davantage de transparence.

Protocole de constitution de corpus thématique

Cette efficacité peut être évaluée en analysant les opinions exprimées sur les différents projets et réalisations en lien avec l'aménagement du territoire (Kergosien *et al.*, 2014). Les ressources disponibles sur le web, et en particulier celles de la presse locale, permettent de suivre les débats que suscitent les projets d'urbanisme. Dans (Kergosien. *et al.*, 2015), les auteurs proposent, notamment, une méthodologie semi-automatisée pour enrichir les scènes issues d'une série temporelle d'images satellites, focalisées sur un projet d'aménagement.

Les genres, ou catégories textuelles, des ressources indexées par les moteurs de recherche sont très variés : blog, publicité, rapport officiels, articles scientifique etc. Plusieurs travaux proposent différentes hiérarchies de genre (Vidulin *et al.*, 2009; Santini, 2011) afin de faciliter la classification de nouvelles ressources. (Madjarov *et al.*, 2019) exploite ces hiérarchies pour renforcer les capacités d'un moteur de recherche à proposer des résultats pertinents selon les genres souhaités. L'évaluation de la pertinence des ressources collectées sur le web est, bien entendu, un enjeu majeur. Une approche classique vise à calculer les co-occurrences de termes entre la requête et les documents rapatriés pour ensuite classer les résultats en fonction du poids du terme dans le document, en utilisant, notamment la mesure TF-IDF (Aizawa, 2003). Cette méthode a cependant des limites. Cette approche par mot clé ne peut retranscrire la proximité sémantique entre deux termes comme, par exemple, vélo et mobilité douce.

Pour remédier ce problème, (Schaeffer. *et al.*, 2020) propose de filtrer les ressources pertinentes en calculant la proximité sémantique, via le modèle de prolongement lexical word2vec (Mikolov *et al.*, 2013), entre des termes qui composent le document et un thésaurus thématique. Dans un contexte de surveillance en épidémiologie animale, (Arsevska. *et al.*, 2016) montre que la construction de requête avec deux mots clés, symptôme et animal, permet de récolter des ressources avec davantage de pertinence. Il est possible, enfin, de combiner ces deux dernières approches, prolongement lexical et requête multi-termes, pour enrichir la sémantique du terme par l'ajout d'un contexte apporté par la phrase, ou le groupe de mots, qui le contient en utilisant le modèle BERT (Devlin *et al.*, 2018). De la même façon, nous proposons la mise en place d'un ensemble de techniques prenant en compte des stratégies d'extraction de concepts (officiels et non officiels) et de constitution automatique de corpus thématiques provenant de sources diversifiées, validées par des mesures de similarité sémantique.

3. Méthodologie

L'approche méthodologique que nous proposons porte sur deux grands points. Le premier est la constitution d'un vocabulaire de concepts thématique. Le second concerne la collecte de documents en utilisant l'ensemble de termes précédemment obtenus.

3.1. Constitution des vocabulaires de concepts thématique

Le processus pour la constitution de vocabulaire de concepts thématique est basé sur le principe de construction de corpus proposé dans (Kilgarriff, Grefenstette, 2003; Kilgarriff *et al.*, 2010; Sharoff, 2006). Il consiste à utiliser un ensemble de graines ou une liste de termes d'un domaine caractéristique pour requêter des documents sur le web. Les résultats sont ensuite utilisés pour étendre la liste des termes et donc le corpus, et ainsi de suite. L'approche proposée se décline en deux phases : la constitution de l'ensemble de termes graines noté *TG*, et la constitution du vocabulaire de concepts noté *VC*.

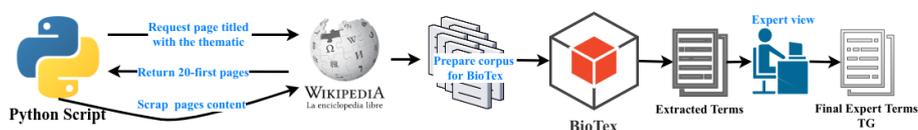


FIGURE 1 – Choix de termes initiaux - TG

3.1.1. Choix des termes graines *TG*

La proposition de l'ensemble *TG* comprend deux étapes (FIGURE 1). La première consiste à extraire un ensemble de termes à partir d'un mini corpus obtenu de Wikipédia¹ en utilisant l'outil BioTex de (Lossio-Ventura *et al.*, 2014). Ce corpus est construit en utilisant un mot clé thématique (ex : urbanisation), et les vingt premières pages renvoyées par Wikipédia sont utilisées pour former le corpus.

Pour BioTex, nous avons principalement utilisé deux mesures statistiques qui sont *C_Value* et *F-TFIDF-C_M*. *C_Value* donne de l'importance aux termes apparaissant plusieurs fois dans le même document et aux termes composés permettant de valoriser les expressions. *F-TFIDF-C_M* représente la moyenne harmonique de *C_Value* et *TF-IDF*, qui permet de classer les termes en fonction de leur pertinence vis-à-vis du document en prenant en compte l'ensemble du corpus. *F-TFIDF-C_M* présente l'avantage d'utiliser toutes les valeurs de la distribution et de réduire le bruit tout au long du processus d'extraction. *C-Value* et *F-TFIDF-C_M* sont complémentaires, car la première favorise l'extraction des termes composés pertinents et la seconde privilégie les termes discriminants. Discriminants dans le sens où elle permet de capturer les termes pertinents qui majoritairement ne sont pas reconnus par *C_Value* et *TF-IDF* prise individuellement.

La seconde étape consiste à faire choisir un ensemble de termes par des experts du domaine parmi les termes récupérés à partir de Wikipédia. L'ensemble

1. https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal

Protocole de constitution de corpus thématique

de termes ainsi constitué forme la liste de termes experts que nous appelons ici *TG*.

3.1.2. Constitution du vocabulaire de concepts *VC*

La constitution du vocabulaire de concepts est basée sur une première étape qui consiste à utiliser la liste *TG* comme une liste de graines permettant de collecter un grand nombre de documents via un moteur de recherche comme par exemple Google. L'objectif est d'obtenir des variantes de termes pour faire évoluer la terminologie de *TG*, afin de récolter des documents parlant de la thématique mais avec des mots et des termes qui ne seraient pas présents dans Wikipédia. Ensuite, grâce à BioTex, le corpus obtenu est utilisé pour extraire un grand ensemble de termes noté *TB* (termes BioTex), à partir duquel, nous allons réduire le bruit éventuel pour former le vocabulaire *VC*. Pour éviter de prendre en considération les termes généraux ou les termes qui ne permettent pas de définir la thématique étudiée, nous avons mis en place un filtre basé sur une approche sémantique en utilisant DistilBERT (Sanh *et al.*, 2019), reconnue pour l'évaluation de la Similarité Sémantique Textuelle (STS). La valeur de similarité varie entre -1 pour de très faible similarité et 1 lorsque les termes sont très proches ou identiques.

La mesure sémantique est évaluée à partir 1) des termes extraits *TB* qui sont les nouveaux termes, et 2) les termes graines ou experts *TG*. Pour chaque terme de *TB*, nous calculons sa similarité sémantique avec chacun des termes de l'ensemble *TG*. Ce qui nous donne un nombre de valeurs de similarité égal au nombre de termes dans l'ensemble *TG* pour chaque terme *t* de *TB*. Pour obtenir la valeur de similarité finale pour un terme, nous faisons la moyenne des TOP@n (n correspond aux n termes les plus pertinents) en tenant compte des plus grandes valeurs de similarités pour garder une certaine exhaustivité vis-à-vis de l'ensemble des termes experts. À la fin du processus, nous obtenons un nouvel ensemble ordonné de termes, constitué des mêmes termes de l'ensemble *TB*, mais avec des scores de pertinence différents. Les vocabulaires de concepts seront par la suite sélectionnés à partir de ces ensembles de termes pour chaque thématique, en partant de ceux disposant d'un plus grand score. La FIGURE 2 décrit les différentes étapes pour la construction de vocabulaire de concepts thématique dans son ensemble et le calcul de similarité entre termes est illustré sur la FIGURE 3.

3.2. Constitution des corpus thématiques

3.2.1. Collecte de documents

Une fois le vocabulaire de concepts thématique construit, nous avons pour chaque thématique, un ensemble de concepts permettant de la décrire au mieux. La phase de collecte (voir FIGURE 4) consiste à utiliser un module, qui prend en entrée un vocabulaire de concepts *VC* et une empreinte spatiale, pour retourner

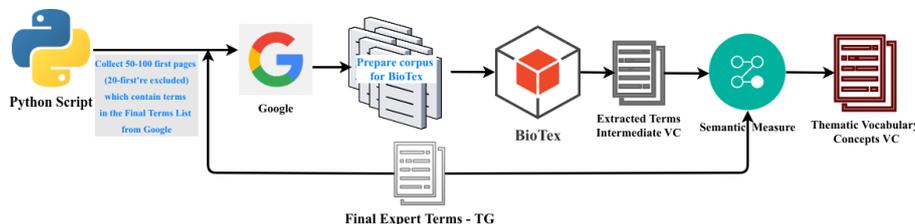


FIGURE 2 – Constitution de vocabulaire de concepts thématique - VC

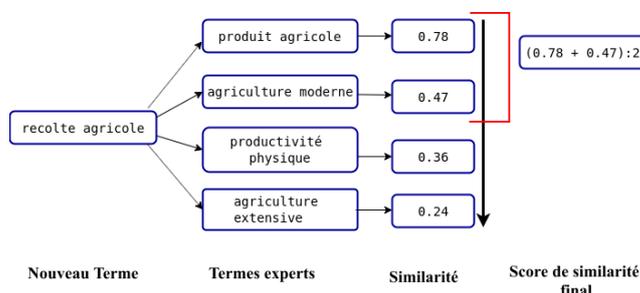


FIGURE 3 – Exemple de calcul de similarité sémantique - score final calculé en faisant la moyenne des 2 plus grandes valeurs

des documents en relation avec la thématique concernée pour former un corpus de taille importante. Nous pouvons ainsi constituer pour chaque thématique, un corpus le plus en cohérence possible avec son contenu. La prochaine étape permet de filtrer les documents les moins pertinents pour chaque corpus thématique. Lorsque le document obtenu est pertinent, une fiche de méta-données contenant les informations sur le document est automatiquement générée avant d’être stockée dans un lac de données (Kafando *et al.*, 2020).

3.2.2. Évaluation automatique des documents par mesure sémantique

Dans le but de ne considérer que les documents qui sont pertinents pour chaque thématique, nous avons introduit une évaluation par mesure de similarité. Identique à la précédente, l’évaluation des documents se fait entre un document collecté par rapport à l’ensemble des termes de son vocabulaire étendu de concepts thématique VEC obtenu avec WordNet (Miller, 1998). L’objectif principal de cette étape est de disposer 1) d’une couverture plus large sur les termes en introduisant leur synonyme grâce à WordNet, et 2) d’obtenir des termes moins experts et/ou administratifs qui sont importants pour l’évaluation des documents moins administratifs tels que les blogs, les annonces, etc. Cette technique permet de prendre en compte ces synonymes lors de l’évaluation des documents. Comme décrite précédemment, la valeur de similarité avec

Protocole de constitution de corpus thématique

DistilBERT varie entre -1 et 1. Un document de la thématique urbanisation sera évalué avec le vocabulaire étendu de concepts de la thématique urbanisation. Plus la valeur de similarité est forte, plus le document sera considéré comme pertinent vis-à-vis de la thématique, et vice-versa.

Le processus pour l'évaluation des documents est illustré sur la FIGURE 4. Afin d'éliminer les documents qui ne sont pas pertinents, nous définissons lors des expérimentations, une valeur seuil à partir de laquelle, nous décidons de garder le document ou de l'exclure du corpus.

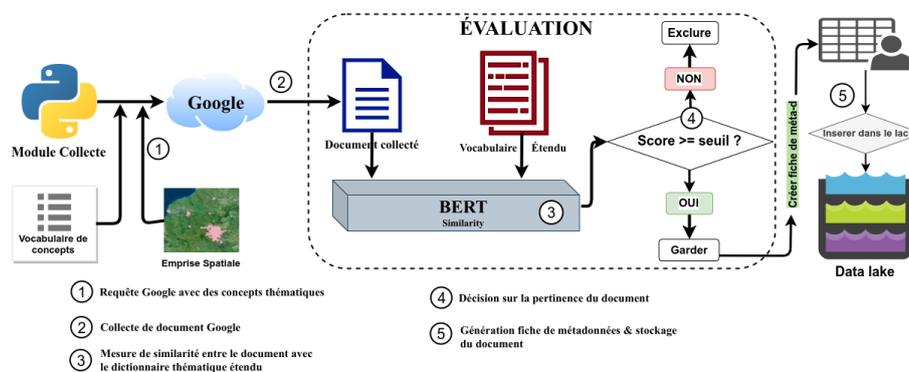


FIGURE 4 – Collecte et évaluation des documents

Les différentes étapes, illustrées sur la FIGURE 5, constituent le protocole de collecte proposé.

4. Expérimentation

Nous cherchons des documents traitant de la Métropole de Montpellier et relatifs à l'une des trois thématiques qui sont l'Agriculture, l'Urbanisation et l'Hydrologie. Les codes sources de la solution mise en place seront disponibles et accessibles dans notre *répertoire partagé*².

4.1. Vocabulaire de concepts thématique

La première étape est le choix de termes graines *TG* décrit dans la section 3.1. La seconde étape consiste à soumettre cet ensemble de termes à un groupe d'experts dont le rôle est de sélectionner un sous-ensemble (liste finale de *TG*) de termes permettant de décrire au mieux la thématique. Pour chaque requête, nous utilisons le libellé de la thématique comme mot clé principal de

2. <https://github.com/aidmoit>

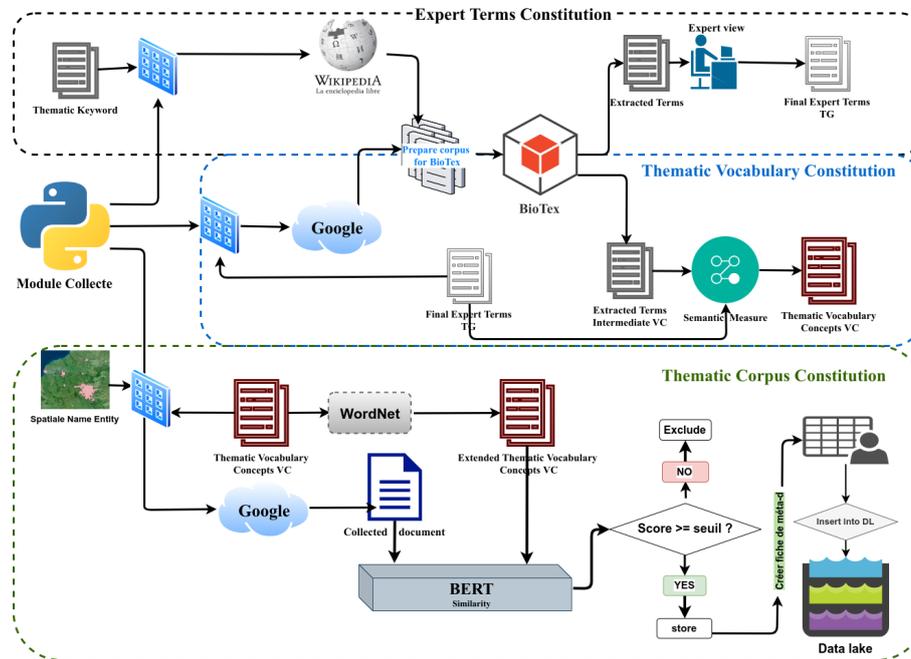


FIGURE 5 – Les étapes du protocole de collecte

la recherche sur Wikipédia, soit 'urbanisation', 'agriculture', ou 'hydrologie'. Les corpus ainsi constitués de chaque thématique sont ensuite utilisés pour l'extraction des termes parmi lesquels seront issus les termes graines.

TABLE 1 – Statistiques sur la constitution de la liste de termes graines TG

Termes requêtes	Urbanisation	Agriculture	Hydrologie
Taille Corpus (Nb docs)	20	20	20
Nombre de termes retenu	100	100	100

TABLE 2 – Extrait de termes experts par thématique

Urbanisation TG-1	Agriculture TG-2	Hydrologie TG-3
Politiques urbaines	Récolte agricole	Hydrogéologie
Aménagement territoriale	Agriculture	Eau pluviale
Aménagement foncier	Aménagement agricole	Mesure hydrologique
Aménagement touristique	Production agricole	Cycle de l'eau
Architectonique	Exploitation agricole	Réseau hydrologique

Les tableaux 1 et 2 présentent respectivement les statistiques et les TOP@5 de la liste de termes graines TG de chaque thématique.

Protocole de constitution de corpus thématique

Une fois les termes graines obtenus, nous procédons à la constitution des vocabulaires de concepts *VC*. Pour chaque thématique, la liste TG est utilisée pour servir de mots clés aux requêtes pour constituer de corpus plus importants, sur le web, sans distinction de site. Ceci nous permet d'avoir une forte diversité des données tant dans leurs contenus que dans leurs sources. Nous avons récolté respectivement 105, 120 et 117 documents pour former les corpus Urbanisation, Agriculture et Hydrologie à partir desquels seront extraits les ensembles de termes que nous nommons termes Biotex (TB). Afin de pouvoir constituer les vocabulaires de concepts thématiques finaux, nous avons utilisé la mesure sémantique proposée dans le Section 3.1.2 qui nous a permis de limiter le bruit, en donnant des poids très faibles aux termes dont la sémantique est éloignée de la thématique.

Dans le tableau 3, nous présentons les TOP@10 des termes de l'ensemble TB avant et après l'application de la mesure sémantique pour la thématique Agriculture. Tout d'abord, nous constatons que l'ordre de pertinence des termes

TABLE 3 – TOP@10 des termes de la thématique Agriculture avant et après la mesure sémantique (MS)

Thématique Agriculture	
TOP@10 avant la MS	TOP@10 après la MS
dégradation des terres	production agricole
agriculture urbaine	exploitation agricole
etats membres	agricole production
changement climatique	oeuvre agricole
google scholar	entreprise agricole
occupation du sol	service agricole
field sizes	techniques de production agricole
lutte contre la désertification	production agricole perdu
production agricole	agricole perdu
matière de dégradation des terres	gestion de la production agricole

a changé entre les deux étapes. Dans le tableau 3, première colonne, de la thématique *Agriculture*, nous remarquons la présence de terme comme '*google scholar*', qui n'a pas un sens sémantique proche à l'agriculture, mais qui occupe un rang important dans les classements de BioTex (c'est-à-dire dans l'ensemble TB). Après avoir appliqué le calcul de similarité sémantique entre ces termes et les termes experts (*conf* Table 2), nous constatons dans la seconde colonne, l'apparition de nouveaux termes plus pertinents dans les TOP@10. Cela signifie que les nouveaux termes (ex: production agricole au 9ème rang) présents dans la deuxième étape occupaient un rang de pertinence moins fort avant la mesure de la similarité. Cette remarque est aussi valable pour les thématiques Urbanisation et Hydrologie.

4.2. Constitution des corpus thématiques

4.2.1. Vocabulaire de concepts thématique final

Le choix du vocabulaire de concepts thématique final est effectué en fixant un seuil minimal spécifique à chaque thématique. Pour ce faire, nous regardons donc à partir de quelle valeur du score, les termes commencent à être éloignés de la thématique. Pour la suite de notre étude, nous avons considéré les TOP@1000 termes de chaque thématique. Le tableau 4 indique le seuil correspondant pour chaque thématique.

TABLE 4 – Statistiques sur la constitution du vocabulaire de concepts thématique VC

Thématique	Urbanisation	Agriculture	Hydrologie
Taille Corpus (Nb documents)	105	120	117
Taille du VC	TOP@1000	TOP@1000	TOP@1000
Valeur du seuil de similarité	0.80	0.75	0.77

Une fois le seuil fixé, la liste de termes est étendue en utilisant WordNet. Les synonymes obtenus permettent de sortir du cadre de langage officiel ou académique, et d'utiliser des expressions moins techniques, qui permettront par exemple d'évaluer des documents en relation avec des réseaux sociaux, des blogs ou encore des annonces telles que les offres d'emplois. En exemple, certains synonymes obtenus avec le terme "*zone d'aménagement concerté*" de la thématique Urbanisation sont : *district d'aménagement concerté, domaine d'aménagement concerté, aire d'aménagement concerté, quartier d'aménagement concerté, etc..*

4.2.2. Évaluation automatique des documents par mesure sémantique

Cette évaluation est faite par calcul de similarité sémantique entre chaque document collecté avec le vocabulaire étendu de concepts thématique de la thématique concernée.

Pour chaque thématique, nous avons considéré les TOP@1000 termes de son vocabulaire étendu de concepts thématique pour l'évaluation des documents la concernant. Ce choix est fixé suite aux différentes expérimentations qui montrent une très faible variation, voir nulle au delà des TOP@1000. À la fin du processus, pour chaque thématique, un score est affecté à chaque document, représentant sa proximité sémantique avec le vocabulaire étendu de concepts thématique précédemment obtenu. Nous avons défini une valeur seuil de 0.5 comme valeur minimale. Un document dont la valeur de la proximité sémantique est inférieur à 0.5 ne sera donc pas considéré dans le corpus, et ce pour l'ensemble des thématiques. Cette valeur est fixée après une vérification des résultats obtenus. Le tableau 5 indique les caractéristiques pour chaque corpus thématique.

Protocole de constitution de corpus thématique

TABLE 5 – Statistiques sur les corpus constitués par thématique

Thématique	Urbanisation	Agriculture	Hydrologie
Taille du VC utilisé	TOP@500	TOP@500	TOP@500
Taille du VEC utilisé	TOP@1000	TOP@1000	TOP@1000
Seuil de similarité	0.5	0.5	0.5
Taille Corpus	867	1.400	1380

5. Conclusion

Le protocole que nous avons élaboré dans cette étude vise à proposer une démarche générique permettant de constituer 1) des vocabulaires thématiques et 2) de constituer des corpus thématiques spécifiques (l'agriculture, l'urbanisation et l'hydrologie pour la métropole de Montpellier) pour permettre une meilleure observation du territoire. Elle s'appuie principalement sur le principe de faire évoluer une terminologie allant de la constitution de graines à la formation d'un vocabulaire de concepts thématique bien spécifique. Quant à la constitution des corpus, elle se fait en utilisant chaque vocabulaire thématique comme ensemble de mots clés de recherche et une spécification de l'emprise spatiale d'intérêt. Chacune des étapes est soumise à une évaluation sémantique, dans le but de limiter le bruit aussi bien dans le vocabulaire de concepts que dans les corpus thématiques. Nos travaux futurs porteront sur l'intégration de l'ensemble de ces documents au sein du lac de données de la métropole de Montpellier et leur mise en lien selon les dimensions spatiale, temporelle et thématique. Notre contribution s'inscrit en sciences de l'environnement avec le développement de méthodes permettant de gérer une grande masse de données pour assurer le suivi d'un territoire donné et d'en extraire une meilleure connaissance. Ces travaux ont été soutenus par l'Agence nationale française de la recherche dans le cadre du programme Investissements d'avenir #DigitAg, référencé ANR-16-CONV-0004.

Bibliographie

- Aizawa A. (2003, 01). An information theoretic perspective of tf-idf measures. *Information Processing Management*, vol. 39, p. 45-65.
- Arsevska. E., Roche. M., Hendrikx. P., Chavernac. D., Falala. S., Lancelot. R. *et al.* (2016). Identification of associations between clinical signs and hosts to monitor the web for detection of animal disease outbreaks. *International Journal of Agricultural and Environmental Information Systems*.
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Guéranger D., Mathieu-Fritz A. (2019, novembre). The smart city at work. *Rezeaux*, vol. No 218, n° 6, p. 41-75.

- Kafando R., Decoupes R., Sautot L., Teisseire M. (2020). Spatial Data Lake for Smart Cities: From Design to Implementation. *AGILE: GIScience Series*, vol. 1, p. 1-15.
- Kergosien. E., Alatrística-Salas. H., Gaio. M., Güttler. F., Roche. M., Teisseire. M. (2015). When textual information becomes spatial information compatible with satellite images. In *Proceedings of the 7th international joint conference on knowledge discovery, knowledge engineering and knowledge management - volume 1: Kdir, (ic3k 2015)*, p. 301-306. SciTePress.
- Kergosien E., Laval B., Roche M., Teisseire M. (2014). Are opinions expressed in land-use planning documents? *International Journal of Geographical Information Science*, vol. 28, n° 4, p. 739-762.
- Kilgarrieff A., Grefenstette G. (2003). Introduction to the special issue on the web as corpus. *Computational linguistics*, vol. 29, n° 3, p. 333-347.
- Kilgarrieff A., Reddy S., Pomikálek J., Avinesh P. (2010). A corpus factory for many languages. In *Lrec*.
- Lossio-Ventura J. A., Jonquet C., Roche M., Teisseire M. (2014). Biotex: A system for biomedical terminology extraction, ranking, and validation. In *Proceedings of the 2014 international conference on posters demonstrations track - volume 1272*, p. 157-160. Aachen, DEU, CEUR-WS.org.
- Madjarov G., Vidulin V., Dimitrovski I., Kocev D. (2019). Web genre classification with methods for structured output prediction. *Information Sciences*, vol. 503, p. 551 - 573.
- Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, vol. 26, p. 3111-3119. Curran Associates, Inc.
- Miller G. A. (1998). *Wordnet: An electronic lexical database*. MIT press.
- Sanh V., Debut L., Chaumond J., Wolf T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Santini M. (2011). Cross-testing a genre classification model for the web. In A. Mehler, S. Sharoff, M. Santini (Eds.), *Genres on the web: Computational models and empirical studies*, p. 87-128. Dordrecht, Springer Netherlands.
- Schaeffer. C., Interdonato. R., Roche. M. (2020). Construction d'un corpus sur la problématique de la sécurité alimentaire guidée par un lexique et des approches de fouilles de textes, in : Toth 2020 - terminologie et ontologie : théories et applications. roche christophe (ed.). chambéry : Presses universitaires savoie mont blanc.
- Sharoff S. (2006). Creating general-purpose corpora using automated search engine queries. *WaCky*, p. 63-98.
- Vidulin V., Lustrek M., Gams M. (2009, 01). Multi-label approaches to web genre identification. *JLCL*, vol. 24, p. 97-114.

Géoweb de narration et de modélisation pour les territoires littoraux : la plateforme NAMO Géoweb

Marie Gradelier¹, Jean-Pierre Chery²

1. TETIS, AgroParisTech, Cirad, CNRS, INRAE, Univ Montpellier,
Montpellier, France
marie.gradelier@gmail.com

2. TETIS, AgroParisTech, Cirad, CNRS, INRAE, Univ Montpellier,
Montpellier, France
jean-pierre.chery@agroparistech.fr

RÉSUMÉ. Un géoweb concernant le littoral méditerranéen français est développé pour valoriser des travaux et résultats de recherche portant en particulier sur les changements d'occupation et d'usage de sols sur le temps long. La conception de cette plateforme géoweb, nommée NAMO (narration et modélisation), mobilise des fonctionnalités ouvertes, libres et gratuites suite à un état de l'art. C'est en particulier le développement de deux dimensions de valorisation et de positionnement en géoweb 2.0 qui est souligné : la cartographie narrative et la modélisation systémique. Cela permet de coupler les facteurs de temporalités variés des territoires considérés et la complexité des interactions entre sociétés et environnement, ainsi que de placer le dispositif en capacité d'assurer un croisement des connaissances scientifiques expertes et des usages citoyens.

ABSTRACT. To highlight the work and results of a research project on the French Mediterranean coastline, especially those regarding the changes in land occupation and use over time, a geoweb platform called NAMO (narration and modelling) is being developed. This platform's design mobilizes free and open-source functionalities following a state-of-the-art. This paper focuses on the development of two dimensions of valorisation and positioning in geoweb 2.0: narrative mapping and systemic modelling. This makes it possible to couple the various temporality factors of the considered territories with the complexity of the interactions between society and environment, as well as the right incorporation of the platform to ensure a crossover of expert knowledge and citizen use.

MOTS-CLÉS : Géoweb, Cartographie narrative, modèle systémique, DPSIR, littoral

KEYWORDS: Geoweb, Narrative cartography, system modeling, DPSIR, coastal area

1. Introduction

Le présent article porte sur le développement d'une plateforme géoweb dans le cadre du projet de recherche RestEAUr'Lag, qui traite de l'évolution de territoires littoraux méditerranéens, en particulier en Languedoc et en Corse. Il traite de la question de la durabilité de systèmes lagunaires, concentrant des enjeux de ressources, d'usages, de capacités de résilience face à des aménagements déployés sur le temps long, entre les XVIIIe et XXIe siècles. Partant d'un besoin générique de production et de représentation de l'évolution de l'occupation et de l'utilisation des sols de ces territoires sur une période longue, le projet retient l'option d'une plateforme de cartographie en ligne, relevant du géoweb. Le géoweb est « l'organisation de l'information d'Internet à travers un géoréférencement direct ou indirect sur la surface terrestre de tout contenu informationnel » (Joliveau, 2011).

Cette plateforme vise à répondre à deux besoins différents et complémentaires : (i) le partage des données et informations issues des travaux des membres du projet de recherche (géographes, historiens, politologues, écologues, etc.) sous une forme cartographique et (ii) la diffusion des résultats de la recherche auprès des acteurs des territoires concernés (gestionnaires et grand public), selon les principes prônés par le bailleur, la Fondation de France. La plateforme doit constituer un outil de sciences participatives permettant au grand public (associations, habitants, scolaires) une appropriation facilitée et de contribuer via des fonctionnalités, dans la logique du géoweb 2.0. Celui-ci peut être défini comme le geoweb qui « hérite des nouveaux principes du Web 2.0 tels que collaboration, participation, partage, échange, ouverture, interaction, communication, et plus encore » (Batita, 2019). Est attendu par exemple un usage en classes de collège ou de lycée afin de préparer des sorties de terrain sur le littoral, pour ensuite compléter ou enrichir les informations déjà présentes par celles récoltées durant ces sorties, avec des exercices narratifs qui assurent l'association du temps et de l'espace des observations. Pour les gestionnaires, il y a l'exemple de services chargés de la gestion des aménagements hydrauliques très intéressés par des modes de valorisations d'informations et d'apports argumentaires spatialisés dans des réunions de concertation, comme dans le cas de l'évaluation des chemins de l'eau en zones inondables et de chemins « oubliés » et redevenus fonctionnels lors d'événement de crues.

Du fait du double usage scientifique et grand public, la plateforme doit proposer une aide à la modélisation qui soit compatible avec un niveau d'exigence experte et un niveau d'approche de néophyte. La notion d'indicateur possède a priori ces deux dimensions afin de traiter des enjeux entre la société et l'environnement, et en particulier les indicateurs conçus dans la perspective du développement durable (Decamps et Vicard, 2010). Le géoweb offre des possibilités de communication et d'interactivité explorées et adoptées par de nombreuses plateformes d'institutions et de projets publics et privés (Laurini et Servigne, 2011). Beaucoup de sites de cartographie en ligne relèvent des technologies directement proposées dans les applications SIG et l'ergonomie en hérite directement : vue cartographique

NAMO Géoweb : narration et modélisation pour les territoires littoraux

principale, volet de liste et d'ordre des couches utilisées, paramétrage des propriétés de visibilité et d'apparence en symboles graphiques de ces dernières, etc. (Mericskay, 2016). Des outils de mesure, de sélection, de saisie accompagnent ces fonctionnalités génériques. La plupart des plateformes géographiques de collectivités territoriales, par exemple en France, donnent à voir tout ou partie de leur patrimoine de données géographiques associées aux couches de référentiels mises à disposition par les organismes assurant la couverture géographique standardisée comme l'IGN. Ces plateformes sont d'abord des interfaces de consultation des infrastructures de données géographiques (IDG) et permettent généralement des téléchargements de couches après la vérification de leurs caractéristiques topologiques et tabulaires, ainsi que des sorties cartographiques adaptées au besoin particulier de l'utilisateur.

Les choix méthodologiques de conception et de réalisation de la plateforme sont présentés, avec les principales fonctionnalités permettant d'assurer les usages liés aux questions de temporalité et de complexité du territoire littoral selon les possibilités d'usages de type géoweb 2.0.

2. Méthode : conception de la plateforme géoweb

Les besoins du projet de restituer des aspects de la complexité territoriale et des évolutions dans le temps appellent à dépasser le simple déploiement d'un service cartographique de type SIG en ligne. La méthodologie générale retenue s'appuie d'abord sur un état de l'art des fonctionnalités existantes dans différents sites proposant une cartographie en ligne. Les fonctionnalités attendues doivent traiter des deux besoins méthodologiques particuliers de valorisation des connaissances dans le temps, avec des cartes narratives, et de modélisation de la complexité des territoires littoraux. La sélection de fonctionnalités permet alors la conception d'une maquette pour validation.

2.1. Etat de l'art des critères de choix du type de plateforme

Un état de l'art a été réalisé sur la base d'un échantillon de 17 cartographies web interactives consultables en ligne en 2020¹. Il a permis de lister les principales caractéristiques de ces plateformes selon différents critères. Deux critères portent sur les aspects de complexité, via le traitement multi-thématique, et de temporalité. Trois autres critères relèvent d'aspects d'ergonomie, de partage liés au positionnement en sciences participatives. Ces cinq critères sont :

¹ L'Évolution de Paris du Bas Empire à 1900 ; OREME – Portail cartographique ; Picto – Occitanie – D6DAMC : Analyse multicritères ; DATA Portraits Paris / Grand Paris ; USGS Historical Topographic Map Explorer ; ALPAGE – AnaLyse diachronique de l'espace PARisien : approche GEomatique ; AHL – Atlas Historique du Limousin ; IGN – Remonter le temps ; Chronas ; CEREMA – Rivages ; Géolittoral – Carte dynamique (CEREMA) ; Lost Rivers Walks ; Observatoire des Territoires ; IGN – Géoportail ; SURFRIDER – Fondation Europe ; Global Forest Watch ; Mobiliscope.

– le critère de dimension thématique. Il porte sur la représentation cartographique de territoires de types régionaux, dans leurs aspects anthropiques et environnementaux, sur le sujet des littoraux particulièrement, avec l'entrée privilégiée de l'occupation et l'utilisation du sol et des ressources liées à l'eau ;

– le critère de dimension temporelle. Celui-ci vise à estimer les possibilités et contraintes pour la représentation diachronique, multi-dates et les évolutions, stagnations et changements dans le temps, avec importance marquée pour l'approche historique. La représentation sous forme de frises chronologiques et d'interactivité de la plateforme par la dimension temporelle sont particulièrement considérés ;

– le critère de la dimension fonctionnelle. Il relève d'une part du besoin assez générique d'une accessibilité, d'une gestion et d'une représentation cartographique de type SIG (multi-couches, vecteurs et rasters, sémiologie graphique adaptée), et d'autre part de choix technologiques de développement de type libre, ouvert, gratuit et référencé ;

– le critère d'interface. Il propose de s'intéresser à la capacité de la plateforme d'offrir un usage intuitif, une ergonomie dans l'accès et la manipulation des éléments de configuration, d'identité éditoriale et de charte graphique et d'équilibre général des fonctions et éléments affichés ;

– le critère d'accessibilité. Ce dernier concerne la capacité de faciliter l'utilisation de l'interface par un utilisateur non averti, en soulignant des fonctionnalités pédagogiques ;

Le choix des plateformes analysées est issu d'une recherche par mots-clés sur un moteur de recherche et de liens éventuels repérés en références dans les revues ou sites spécialisés dans le domaine. En l'absence d'un observatoire du géoweb, même si des travaux commencent à structurer un panorama, en considérant d'abord les IDG (Noucher et al., 2019), c'est donc une approche partielle qui est adoptée. La variété des fonctionnalités repérées permet de retenir celles considérées comme pertinentes pour le projet, et d'estimer les priorités de développement et d'ajouts selon les contraintes de temps, de sécurité et de gestion de la plateforme. La difficulté apparaît dans l'équilibre de ces fonctionnalités à retenir, afin d'assurer une ergonomie efficace. Ces fonctionnalités ont permis de définir l'architecture générale de la plateforme pour réaliser une maquette fonctionnelle à l'issue d'un travail de six mois dont trois de développement.

2.2. Architecture et fonctionnalités retenues

Suite à l'état de l'art, quatre volets sont retenus pour l'interface afin assurer un accès structuré à l'information géographique disponible selon les principes classiques de la cartographie en ligne, et une logique éditoriale issue des atlas : les zones d'études, les fonds de cartes, les couches et les graphiques. Les atlas généralistes édités sur papier au XXe siècle ou les atlas scolaires actuels possèdent ces éléments : les zones d'études relèvent du plan d'un atlas, en continents, pays, régions comme autant de cas particuliers à distinguer dans une page, une vue

NAMO Géoweb : narration et modélisation pour les territoires littoraux

cartographique. Les fonds de cartes de ces atlas sont généralement des cartes de relief et d'hydrographie, avec les toponymes et des villes et routes principales et des lieux remarquables. Les couches du géoweb sont autant de couches SIG permettant des combinaisons afin de produire des cartes thématiques et des sélections tandis que les graphiques sont des éléments d'accompagnement importants de la cartographie (diagrammes divers, histogrammes, etc.). Une partie de ces éléments sont ceux des SIG en ligne habituels : l'utilisateur gère l'information par sélection et paramétrages d'affichage (zoom, visibilité et transparence).

Ces quatre volets sont conçus via des méthodes et outils de structure et de style (HTML et CSS), d'interactivité (JavaScript, JQuery, JQuery UI, OpenLayers et D3.js) et de développement de la base de données et du serveur (QGIS, PostgreSQL et GeoServer).

2.3. Intégration des dimensions temporelle et systémique

Deux volets spécifiques ont été développés pour proposer deux options qui relèvent du géoweb 2.0 dans ses dimensions multimédia, dynamique et d'interaction avec l'utilisateur : un volet de cartographie narrative et un volet de modélisation selon un cadre conceptuel systémique. Ces volets ont été conçus avec les mêmes méthodes et outils que les quatre volets précédents.

2.3.1. Aborder la variété des temporalités : la carte narrative comme approche

Dans le temps d'un projet de recherche, une bonne partie des données sont souvent en cours de production et leur exploitation ne peut être envisagée que dans une seconde phase pour la cartographie en ligne. Les pistes méthodologiques et techniques pour traiter les dynamiques et les changements peuvent relever du domaine de la cartographie animée (Cunty et al., 2017). Depuis quelques années, des plateformes émergent dans les domaines de l'histoire et de l'archéologie, aux besoins marqués en bases spatio-temporelles (Costa, 2012). En France, différents programmes de standardisation de couches d'informations géographiques issues de sources des XIXe et XXe siècles en particulier, sont en cours². Ils donnent une perspective de mise à disposition de ces informations afin de les traiter selon les méthodes d'analyse et de représentation des changements. En attendant ces bases de données historiques et leur exploitation en cartographie animée des changements, une autre méthodologie est retenue : la cartographie narrative, approche assez souple pour alimenter la plateforme géoweb de manière autonome.

La plateforme doit valoriser les travaux d'historiens spécialisés dans l'environnement et les aménagements régionaux et locaux. Ces historiens produisent des récits qui reconstituent des situations temporelles, avec l'appui de sources primaires, issues d'archives et des travaux dans lesquels le texte est prédominant. La cartographie narrative possède des fonctionnalités appropriées pour mettre en valeur les dimensions spatiale et temporelle des récits (Caquard et Cartwright, 2014). Le

² Projet ANR COMMUNES à partir de 2020 ; Projet « Forêts de Cassini » depuis 2012 ; BD Carto Etat-major version 1.0 de l'IGN, depuis 2019.

géoweb renforce l'intérêt d'un cheminement narratif, depuis un récit dans le temps et l'espace avec l'association dynamique de différentes vues (de la carte, de textes et d'autres produits multimédias). Les éléments associés sont coordonnés à la vue cartographique principale affichée (Caquard et Dimitrovas, 2017). Le projet « Lost Rivers Walks »³ est un exemple de l'intérêt des cartes narratives pour aborder la question des relations entre les aspects anthropiques et environnementaux d'un territoire. Ce projet documente des cheminements dans la métropole de Toronto avec des affichages cartographiques s'adaptant aux étapes d'un itinéraire préalablement établi, avec des éléments significatifs d'aménagement, de règlements d'urbanisme et autres. Ceux-ci expliquent ou illustrent les changements dans les écoulements urbains de l'eau, et finalement la disparition des rivières du paysage, par des aménagements de couverture, dérivation ou comblement. Cela invite à des actions de type « sorties de terrain », préalables ou postérieures à une consultation de la plateforme. Dans notre projet, cette perspective de fonctionnalité est bien retenue comme élément de contribution aux sciences participatives. Cette même fonctionnalité, appliquée dans le cadre de réunions au sein des collectivités pour présenter la logique d'un projet d'aménagement ou de renaturation dans le temps long apparaît pertinente.

L'alimentation d'une cartographie narrative dans la plateforme consiste d'une part à fournir un service de dépôt de contenus que l'auteur de la narration souhaite combiner (textes du récit, images, cartes, schémas, photographies qui ponctuent la narration), avec les paramètres d'emprise cartographique du géoweb hébergeur (centre de la carte et niveau de zoom) et à intégrer d'autre part cette base de données dans le serveur pour son accès depuis le volet de carte narrative. Le service de dépôt permet de proposer des activités de contributions variées, avec des productions scientifiques d'historiens et autres experts qui replacent les étapes d'une évolution de relations homme-environnement selon les concepts, objets et données de leurs disciplines.

2.3.2. *Cartographier les éléments du système territorial via le cadre conceptuel DPSIR*

Dans le projet, la dimension systémique du territoire littoral est traduite sous forme cartographique. Chaque élément du système étudié disponible dans les couches géographiques du géoweb doit se voir affecté d'un attribut de sémiologie graphique (couleur, trame, symbole) qui renvoie à son statut dans les chaînes causales. La représentation cartographique vise à restituer les lieux selon leur rôle (localisation de phénomènes impactant ou localisation d'entités impactées). Par exemple, on peut différencier la distribution spatiale d'éléments émetteurs de pollution de l'eau (rejets d'eau usées de maisons individuelles non raccordées ou émetteur en mer d'une station d'épuration) et d'éléments impactés par ces pollutions (parties de lagunes ou portions du littoral maritime). Les notions de concentration ou de diffusion spatiale des conséquences d'un système d'eaux usées peuvent ainsi être représentées.

³ <http://www.lostrivers.ca/>

NAMO Géoweb : narration et modélisation pour les territoires littoraux

Une relation systémique oblige d'une part à déclarer un ensemble causal minimal (cause et effet) et nécessite d'autre part d'avoir une cohérence causale (une cause ne peut être qu'antérieure ou contemporaine à sa ou ses effets). Dans la plateforme, cette relation vise à être traitée par la constitution de tables de relations maillées, qui permettent d'utiliser les métadonnées des dates ou périodes de validité des informations contenues dans les couches mises en relation. Le volet de modélisation systémique doit avoir une cohérence avec les autres volets et aux usages qu'ils proposent.

Afin d'établir pratiquement les structures causales, le cadre conceptuel « *Driver, Pressure, State, Impact, Response* » (DPSIR) est retenu. Il est utilisé par la recherche et les politiques publiques pour identifier des indicateurs selon les principes systémiques causaux : c'est un cadre de modélisation à la fois répandu, régulièrement utilisé et bien documenté (Binder et al., 2013).

Les trois éléments centraux de la chaîne de causalité DPSIR sont (i) les pressions qui pèsent sur (ii) l'état du système que l'on souhaite représenter, et (iii) les réponses qui y sont apportées. Les pressions engendrent des modifications de l'état du système. On pense généralement à l'augmentation des activités humaines, la densification de la population, et dans le cas du projet de recherche, le développement de l'habitat pavillonnaire sur le littoral ou bien l'augmentation des rejets d'eaux usées qui en découle. La concentration en éléments polluants est ainsi plus importante. La société met alors en place des réponses correctives pour atténuer les effets non désirés des pressions sur l'environnement. Elles prennent généralement la forme de législations, d'aménagement du territoire, d'espaces de protection, etc. Dans le cas d'illustration, une des réponses apportées est le développement d'un réseau d'assainissement des eaux. En complément, les forces motrices permettent d'améliorer la notion de pression en renseignant celles qui sont indirectes ou globales. Elles correspondent généralement aux grandes tendances sociétales dont les effets sont soit d'accroître soit de mitiger les pressions exercées sur l'environnement. Dans le cas de l'exemple, une force motrice à l'origine de la densification de la population et du développement de l'habitat pavillonnaire sur le littoral est l'attrait pour le littoral méditerranéen et pour son climat. Les impacts permettent quant à eux de préciser l'état du système en lui affectant une notion d'enjeux : les concentrations importantes en polluants (comme état de système) dans l'eau des lagunes modifient entre autres la santé des êtres vivants qui consomment cette eau, directement ou indirectement. Ce risque ou ce dégât pour la santé et l'environnement conduit à d'éventuelles réponses correctives par la société.

La méthode retenue pour développer le volet de modélisation est la suivante : chaque attribut DPSIR (forces motrices ; pressions ; état ; impacts ; réponses) est considéré comme un conteneur pouvant accueillir un ou plusieurs indicateurs. Ces indicateurs sont dans notre cas représentés par les jeux de données géographiques, intégrés à l'application en tant que couches. Ces éléments, mis à disposition de l'utilisateur, lui permettent de constituer son propre modèle DPSIR à partir des données auxquelles il a accès, sur la base d'un graphe causal vierge préétabli. Il peut ainsi définir un état unique autour duquel construire son modèle. Cet état prend la forme d'une couche de données : pour reprendre l'exemple précédent, il peut

sélectionner une représentation de la concentration en polluants dans l'eau des lagunes. Il est libre de construire son modèle systémique autour de cet élément en sélectionnant d'autres couches de données et leur adjoignant un attribut DPSIR. Comme le modèle est produit par l'utilisateur, une fonctionnalité doit assurer qu'il ne peut pas définir plus d'une couche comme étant l'état du système, et qu'une contrainte temporelle s'applique sur son modèle. Une couche A ayant une date ou une période strictement plus récente que celle d'une couche B ne pourra pas être déclarée avec un attribut en amont dans la chaîne causale DPSIR. Par exemple, l'indicateur d'eutrophisation de 2000 de la couche des points de surveillance de l'étang de l'Or ne pourra pas être un impact si la quantité d'azote consommée par les exploitations agricoles du bassin versant de l'Or en 2010 est déclarée comme pression. La perspective est de permettre aux utilisateurs d'enregistrer leurs modèles en ligne dans le but de créer une bibliothèque distante. Elle pourrait fonctionner comme une base de données capable d'alimenter les travaux des chercheurs souhaitant proposer un modèle global par contributions d'acteurs pour le territoire littoral.

2.4. Base de données et collecte des données

La majorité des données concernées dans la plateforme géoweb sont celles des organismes publics français et de services et observatoires d'intérêt pour la thématique littorale⁴. Elles visent à être consultables en flux externes via les fonctionnalités retenues. Ces données peuvent être en formats tableurs, vectoriels et raster, flux externes WFS et WMS. Le stockage des différentes sources de données a été conçu en fonction de l'architecture serveur en place dans l'organisme qui accueille ce travail de développement et de recherche, à savoir une base de données PostgreSQL et GeoServer, en prenant en compte la migration des données lorsque la nouvelle IDG locale concernée sera mise en place.

Les données originales du projet de recherche n'ont pas encore été mises à disposition pour la plateforme, considérant que la plupart seront offertes en open data, selon les principes de la science ouverte. Elles portent par exemple sur la représentation des différentes classes d'occupation et utilisation du sol en mode vectoriel, à différentes dates. Par ailleurs, les informations du cadastre napoléonien sont traduites par la vectorisation et le géoréférencement de 18600 parcelles pour 11 communes littorales languedociennes.

⁴ On peut citer : IGN (BDCarto Etat-Major, BDOrtho Historique, BDOrtho, BDParcellaire, BDTopo, France Raster, OCS Grande Échelle, PVA Historiques, RPG, Scan 50 Historique, Scan25, Scans Cassini, Scans EM40K) ; SHOM (Litto 3D, MNT Bathy), Sandre (BDCarthage, ROE, Stations hydrographiques, Système de traitements) ; IFREMER (Photos anciennes, Surveillance littorale) ; Géolittoral (IPLI, LITTOMOS, Ortho Littoral) ; AD (Cadastre napoléonien, Plan terrier) ; ESA Copernicus (Corine Land Cover, Images Sentinel 2).

NAMO Géoweb : narration et modélisation pour les territoires littoraux

3. Résultat : la plateforme NAMO Géoweb, à dimensions narrative et de modélisation systémique

La plateforme Géoweb est appelée NAMO, pour « Narration et Modélisation ». La plateforme a été testée d'abord en accès local pour des raisons de sécurité. On ne détaille pas ici la solution déployée sur le serveur et la configuration serveur-client, qui relève des dispositions standards pour les services de cartographie en ligne et de bases de données distantes. NAMO Géoweb est développée en utilisant les langages HTML, CSS et JavaScript. L'emploi du JavaScript apporte de l'interactivité principalement via l'API DOM et la bibliothèque JQuery. Par ailleurs, la librairie D3 facilite l'affichage de graphiques en permettant d'écrire dynamiquement un SVG.

La figure 1 présente les cinq volets de l'interface géoweb, avec les éléments affichés pour la consultation, le choix et l'activation de traitements associés.

Le premier volet du choix de la zone d'étude est construit en proposant des miniatures d'emprises auxquelles sont associées un couple coordonnées / niveau de zoom dans le script de l'application.

Le deuxième volet, pour le choix du fond de carte, utilise les flux WMS pour chaque source retenue : OpenStreetMap et l'imagerie satellite d'ESRI accessibles librement, ainsi que les fonds IGN qui sont récupérés via une clé d'identification pour le bouquet de flux utiles au projet. Différents scripts permettent de récupérer ces flux cartographiques et, associés aux miniatures affichées dans le menu, de contrôler leur affichage. Le volet des couches géographiques permet la gestion de leur affichage et mobilise de nombreuses fonctionnalités ainsi que des scripts d'automatisation pour simplifier l'alimentation de l'application en données de sources diverses. L'approche orientée objet d'OpenLayers permet d'intégrer différents widgets, comme par exemple l'opacité des couches, la légende des couches issues de flux WMS ou le choix de dates pour afficher des séries temporelles.

Le volet des graphiques et diagrammes a été développé en permettant une lecture diversifiée des données présentées dans les couches affichées. Par exemple, une couche des populations communales par année de recensement couvre une trentaine de millésimes entre 1793 et 2016 et permet l'affichage d'un graphique représentant la variation temporelle de la population pour une commune donnée.

Le volet de la carte narrative a été testé en traitant un ensemble d'éléments fournis par un chercheur historien, à propos d'étapes de l'histoire de la construction des canaux sur les étangs palavasiens au XVIIIe siècle et des conséquences négatives qui ont pu en découler ultérieurement pour l'environnement et la société locale. Il est possible de modifier le fond de carte et les couches affichées pour chaque étape de l'histoire. Un script d'automatisation de l'affichage des histoires a été développé afin d'épargner une gestion lourde d'alimentation de contenus par les administrateurs de l'application.

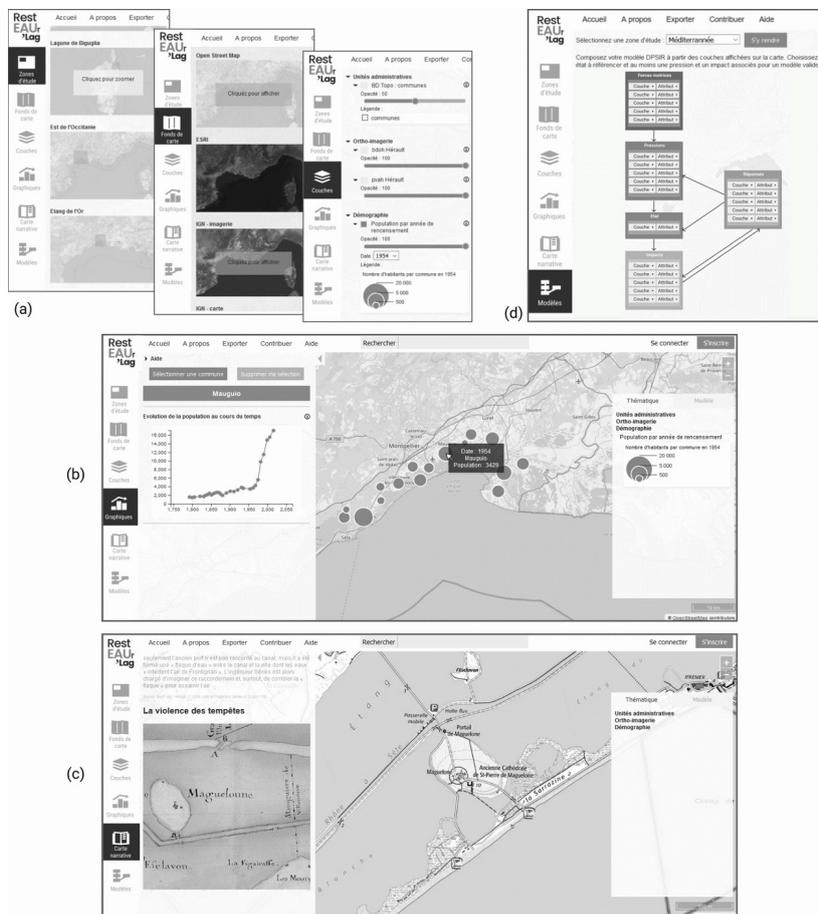


FIGURE 1. Principe multi-volets de NAMO géoweb : (a) volets « Zones d'études », « Fonds de cartes » et « Couches » ; (b) volet « Graphiques » avec la vue cartographique thématique associée ; (c) Consultation des informations d'une carte narrative (volet multimédia de textes et figures) avec vue cartographique associée et (d) volet de modélisation (source : auteurs)

Enfin, le volet de modélisation offre une interface permettant d'élaborer une carte thématique composée d'éléments selon les cinq catégories du modèle DPSIR. Ce volet est composé de deux parties. La première est un graphe causal du modèle DPSIR avec ses cinq catégories représentées par défaut avec des cases de couleurs différentes. L'utilisateur choisit des couches ouvertes en session, avec les champs des attributs associés, pour chaque case du modèle afin de produire la carte thématique DPSIR. La deuxième partie du volet est un bandeau d'affichage du test de la cohérence causale donne le résultat des affectations par un message de

NAMO Géoweb : narration et modélisation pour les territoires littoraux

validation ou d'invalidation du modèle avec le diagnostic des erreurs repérées. Ce volet n'est pas encore testé : il constitue une dimension exploratoire de la plateforme NAMO Géoweb.

4. Discussion

Les possibilités offertes par les méthodes et outils du géoweb, en termes de facilité de conception et d'interactions entre membres d'un projet de cartographie en ligne, soulignent l'intérêt du design géoweb. Ici, il semble important de souligner les possibilités du webdesign dans des conceptions libres, évolutives et adaptables pour conserver une même plateforme de cartographie en ligne en lui assurant des propriétés de capitalisation d'expériences d'usage.

Les propositions méthodologiques et techniques de notre plateforme visent à souligner l'importance du lien entre l'information et la connaissance. Les deux développements de cartographie narrative et de modélisation systémique suggèrent cette intégration, via les couches d'information généralement utilisées en mode « SIG en ligne ». Cela ouvre à notre avis des possibilités de travaux d'expérimentation et d'analyse pour des situations d'usages multi-acteurs, tant experts scientifiques que représentants de la société civile et du grand public en général. Les contributions qui peuvent être déposées sur la même plateforme alimentent la co-construction de la connaissance et de son partage. Des dispositifs d'expérimentation sont alors à engager dans le cadre des sciences citoyennes et pourront contribuer aux débats et enjeux de ce type de dispositifs engagés depuis le début de la décennie 2010 (Newmann et al., 2010 ; Mericskay et Roche, 2011).

Les enjeux de connaissances thématiques des milieux littoraux, et dans le cas de notre projet, des territoires lagunaires méditerranéens, constituent un domaine riche pour mettre en œuvre un géoweb des relations entre les sphères anthropiques et environnementales. Pour vérifier la robustesse conceptuelle et technique de la plateforme, celle-ci est en cours de réplification dans un autre domaine de recherche, concernant un système territorial marqué par les enjeux de pression de l'agriculture sur l'environnement. Des ateliers d'animation et de débat déjà réalisés pourront être mis en valeur et discutés en soulignant l'importance des localisations des parties prenantes et des éléments du territoire qui les placent en interaction.

5. Conclusion

Dans le projet de recherche, la question d'un passage à un niveau de mise à disposition d'une information construite par les différents axes scientifiques, sur des expertises et des exploitations particulières des données, ainsi que la question d'un meilleur accès aux informations par le plus grand nombre a permis de retenir une posture éditoriale contraignante. Elle vise à fournir des fonctionnalités qui permettent de cheminer dans les étapes d'un raisonnement causal, systémique, spatial et temporel, des thématiques associées aux enjeux des relations entre société et environnement, pour les territoires littoraux et applicables dans d'autres contextes.

Les perspectives offertes par le géoweb 2.0 sont celles de la contribution et de la science participative, entre des enrichissements au fil du temps, la mise en vis-à-vis de controverses et des possibilités d'animations au sein de collectifs variés.

Remerciements

Le projet RestEAUr'Lag, dans lequel se positionne ce travail, est soutenu financièrement par la Fondation de France, sur la période 2019-2021, dans le cadre de son programme Littoral.

Bibliographie

- Batita W. (2019). GeoWeb 2.0 applications could support the Geodesign process. *Revue Int. Geomat.*, vol 29, n° 3-4, p. 263-285.
- Binder C.R., Hinkel J., Bots P.W.G., Pahl-Wostl C. (2013). Comparison of frameworks for analyzing social-ecological systems. *Ecology and Society*, n° 18, p. 26.
- Caquard S., Cartwright W. (2014). Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping. *The cartographic Journal*, vol. 51, n° 2, p. 101-106.
- Caquard S., Dimitrovas S. (2017). Story Maps & Co. Un état de l'art de la cartographie des récits sur Internet. *Mappemonde*, n° 121, 17 p.
- Costa L. (2012). La construction de référentiels géohistoriques : un enjeu pour l'interdisciplinarité dans les sciences historiques. *L'Espace géographique*, vol. 41, n° 4, p. 340-351.
- Cunty C., Mathian H., Groupe Cartomouv. (2017). Les pratiques de cartographie animée pour représenter le changement. *Mappemonde*, n° 120, 21 p.
- Decamps M., Vicard F. (2010). Mesurer le développement durable : jeux d'indicateurs et enjeux locaux. *Revue d'Économie Régionale & Urbaine*, p. 749-771.
- Joliveau T. (2011). Le géoweb, un nouveau défi pour les bases de données géographiques. *L'Espace géographique*, vol. 40, n° 2, p. 154-163.
- Laurini R., Servigne S. (2011). Potentialité du géoweb : vers une nouvelle génération de Sig participatifs. *L'Espace géographique*, vol. 40, n° 2, p. 109-116.
- Mericskay B., Roche S. (2011). Cartographie 2.0 : le grand public, producteur de contenus et de savoirs géographiques avec le web 2.0. *Cybergeog : European Journal of Geography, Science et Toile*, document 552, mis en ligne le 20/10/2011, consulté le 21/02/2021.
- Mericskay B. (2016). La cartographie à l'heure du géoweb : retour sur les nouveaux modes de représentation spatiale des données numériques. *Cartes et géomatique*, n° 229-230, p. 37-50.
- Newman G., Zimmerman D., Crall A., Laituri M., Graham J., Stapel L. (2010). User-friendly web mapping: lessons from a citizen science website. *International Journal of Geographical Information Science*, vol. 12, n° 24, p. 1851-1869.
- Noucher M., Gourmelon F., Claramunt C. (2019). Pour un observatoire des données géographiques du Web. Expérimentation à partir des infrastructures de données géographiques françaises. *Revue Int. Geomat.*, vol. 29, n° 1, p. 9-30.

Grand débat national : une analyse automatique des contributions relatives aux transports

Jacques Fize¹, Lucile Sautot², Martin Lentschat³,
Ludovic Journaux⁴, Mohamed Hilal⁵

1. INSA Lyon - LIRIS - jacques.fize@insa-lyon.fr

2. AgroParisTech - UMR TETIS - lucile.sautot@agroparistech.fr

3. Univ. de Montpellier - UMR TETIS - martin.lentschat@umontpellier.fr

4. AgroSupDijon - Lab. Info. Bourgogne - ludovic.journaux@agrosupdijon.fr

5. INRAE - UMR CESAER - mohamed.hilal@inrae.fr

RÉSUMÉ. Le Grand Débat National, lancé par Emmanuel Macron début 2019 pour répondre au mouvement social des Gilets Jaunes, a permis de collecter les contributions de citoyens sur la transition écologique via une plateforme en ligne. Dans cet article, nous exploitons le corpus constitué par ces contributions pour identifier des zones où les participants demandent le développement de pistes cyclables et d'équipements ferroviaires. Pour cela, nous avons créé un modèle de classification permettant d'identifier les contributions traitant de la thématique du transport et proposé une méthode d'extraction de motifs traduisant les propositions des contributeurs. À l'aide des codes postaux donnés par les contributeurs, nous présentons les motifs extraits sur plusieurs cartes.

ABSTRACT. The Great National Debate, launched by Emmanuel Macron at the beginning of 2019 to respond to the Yellow Vests social movement, allowed the collection of citizens' contributions on the ecological transition via an online platform. In this article, we use the corpus constituted by these contributions to identify areas where participants are asking for the development of bicycle paths and railway facilities. For this purpose, we have created a classification model to identify contributions dealing with the theme of transportation and proposed a method for extracting patterns that reflect the contributors' proposals. We then represented these patterns on maps, using the contributors' postal codes.

MOTS-CLÉS : Grand Débat National, Motifs, Transport

KEYWORDS: Grand Débat National, Patterns, Transport

1. Introduction

En octobre 2018, une hausse du prix du carburant automobile déclenche en France un mouvement spontané de protestations sociales dénommé "mouvement des gilets jaunes". Le mouvement s'exprime par des blocages de routes, de ronds-points et de nombreuses manifestations. La première manifestation réunit 280 000 personnes le 17 novembre 2018 (acte 1) et surprend par son ampleur (Boyer *et al.*, 2020). Elle sera suivie par de multiples formes de mobilisations accompagnées de revendications sociales et politiques. En décembre 2018, face à ce mouvement, le gouvernement français saisit la Commission nationale du débat public (CNDP¹) afin de lancer une consultation citoyenne sur divers sujets : le *Grand Débat National*. Cette consultation s'est officiellement tenue du 15 janvier au 15 mars 2019 autour des quatre grands thèmes suivants : transition écologique, fiscalité, démocratie et citoyenneté, organisation de l'État et des services publics. La consultation s'est déroulée sous différentes formes : réunions publiques locales, cahiers de doléances en mairie, conférences avec des citoyens tirés au sort, contributions en lignes sur une plateforme dédiée, etc.

L'analyse et la synthèse des contributions en ligne posent plusieurs défis. Tout d'abord, elles sont nombreuses. Sur la plateforme en ligne, pour le questionnaire de la *Transition Écologique*, plus de 134 000 personnes ont rempli 12 champs de texte libre, soit un total de 1,6 millions de textes potentiels² à analyser. De plus, les textes recueillis sont de longueurs différentes et présentent une grande diversité (style, syntaxe, orthographe, grammaire, lexique). À la demande du gouvernement français, une première analyse des contributions en ligne est effectuée par l'entreprise *OpinionWay*³ immédiatement après la clôture du débat (Opinion Way, 2019). Les synthèses proposées offrent une vue d'ensemble du contenu des contributions mais présentent plusieurs limitations. Tout d'abord, la méthodologie utilisée n'est pas complètement documentée (Bellet *et al.*, 2020), ce qui rend difficile sa reproductibilité et la comparaison avec d'autres travaux. Ensuite, selon les questions analysées, un nombre important de réponses (entre 10 % et 30 %) est écarté. Enfin, la dimension territoriale inhérente tant au mouvement social à l'origine de la consultation qu'aux thématiques de la consultation a été très peu abordée.

Par ailleurs, peu de travaux scientifiques ont été publiés sur l'analyse de ces contributions. Parmi ces communications, Ploux *et al.* (2020) ont proposé une analyse automatique sur la base de motifs et d'une comparaison du lexique employé dans différentes consultations incluant celles du Grand Débat National. Cet article n'a a priori pas encore été publié dans une revue ou ouvrage à comité de lecture. Messer *et al.* (2019) ont également proposé une approche qualitative

1. <https://www.debatpublic.fr/>

2. Les contributeurs ne répondent pas nécessairement à l'ensemble des questions.

3. Les synthèses *OpinionWay* sont disponibles ici : <https://granddebat.fr/pages/syntheses-du-grand-debat>

Grand débat national et transports

des contributions relatives à la mobilité. Ce rapport, très riche sur la description des propositions formulées par les contributeurs, ne dit rien sur la dimension spatiale.

Notre objectif est de proposer une approche permettant de cartographier les opinions des contributeurs sur une thématique donnée. Cette approche doit être automatique (afin de traiter l'ensemble des contributions) et doit capter l'avis des personnes ayant proposé leur contribution. Enfin, cette approche doit générer des cartes permettant de localiser les opinions exprimées.

Pour répondre à ces problématiques, nous proposons un traitement des données reproductible et une analyse des résultats tenant compte de la dimension géographique. Dans cet article, nous présentons les travaux suivants:

- Focus sur les réponses de la "transition écologique" et sur une sous-thématique cruciale : les transports.
- Création d'un modèle de classification pour identifier les réponses parlant de cette sous-thématique.
- L'extraction de motifs dans les données traduisant les propositions émises par les contributeurs.
- Cartographie des motifs extraits au travers de deux objets : le train et les pistes cyclables.

Cet article s'organise de la manière suivante. La section 2, présente l'approche utilisée pour classifier les réponses et extraire les motifs d'opinions. La section 3, présente les motifs extraits selon leur fréquence et leur localisation géographique. Enfin, nous proposons plusieurs discussions en section 4 sur les résultats obtenus, avant de conclure en section 5.

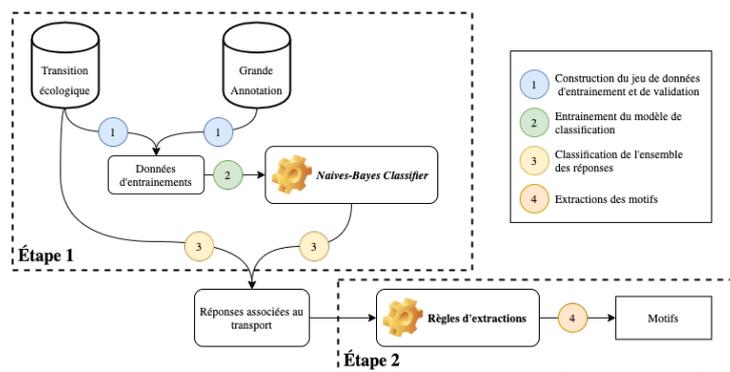


FIGURE 1. Processus d'extraction des motifs d'opinions

2. Matériel et méthodes

Dans ces travaux, nous réalisons une analyse des opinions sur la question du transport dans les réponses de la *transition écologique*. Ce processus s'effectue en deux étapes : la première (*c.f.* Section 2.2) identifie les réponses mentionnant le transport ; la seconde (*c.f.* Section 2.3) en extrait les motifs en s'appuyant sur des règles d'extraction et plusieurs lexiques. Le processus est illustré dans la Figure 1.

Avant de présenter les méthodes proposées pour chaque étape, nous présentons les données utilisées dans l'ensemble du processus.

2.1. Données

Dans cet article, nous utilisons les contributions de la plateforme web du *Grand Débat National*⁴. La plateforme proposait deux modalités de contribution : libre ou questionnaire à choix multiples. Seules les contributions libres sont exploitées dans cet article. Plus particulièrement, nous avons uniquement retenu les contributions au volet portant sur la *Transition Écologique*.

Concernant l'analyse géographique, nous exploitons le code postal indiqué par chaque contributeur, sachant que 97 % de ses codes sont exploitables. Les coordonnées géographiques (latitude, longitude) sont extraites de la base officielle des codes postaux⁵ et correspondent aux centroïdes.

Les réponses sont contextualisées avec des informations externes liées à l'offre de transport. Plus particulièrement, la présence des infrastructures⁶ (les pistes cyclables, les gares, les voies ferrées) et l'organisation du territoire par les déplacements domicile-travail (base communale des aires urbaines définies par l'*INSEE*⁷).

2.2. Classification des réponses liées aux transports

Dans ce travail, nous nous concentrons sur une thématique centrale dans le cadre de la transition écologique : le transport. Pour cela, nous isolons les réponses liées à ce thème en utilisant les annotations réalisées, parallèlement à l'appel à contribution du grand débat, par le collectif nommé *Grande Annotation*⁸. Les réponses du Grand Débat sont annotées à l'aide d'une liste de *tags*. Au moment du téléchargement des données⁹, 280 321 réponses étaient

4. <https://granddebat.fr/pages/donnees-ouvertes>

5. <https://www.data.gouv.fr/fr/datasets/base-officielle-des-codes-postaux/>

6. Données extraites sur *OpenStreetMap* <https://download.geofabrik.de/europe/france.html>

7. <https://www.insee.fr/fr/information/2115011>

8. Données accessibles ici : <https://grandeannotation.fr>

9. 31 mai 2020

Grand débat national et transports

annotées. Ce nombre d'annotations est insuffisant pour une analyse globale des réponses du Grand Débat, toutefois, elles sont en nombre suffisant pour entraîner un modèle de classification. L'objectif du modèle est d'indiquer si une réponse est associée ou pas à la thématique du transport. Pour entraîner ce modèle, nous avons construit un jeu de données d'entraînement/validation composé de deux types de réponses, l'une évoquant le transport et l'autre non. L'identification des réponses annotées parlant de transport s'opère en sélectionnant celles ayant un des *tags* faisant partie de la liste présentée dans la Table 1.

Une fois les données d'entraînement et de validation extraites, nous avons testé différents algorithmes de classification. Parmi ceux-ci, nous comparons des approches classiques : SUPPORT VECTOR MACHINE (SVM) (Vapnik, 2013); NAIVE-BAYES (Manning *et al.*, 2008); KNN (Altman, 1992); DECISION TREE (Loh, 2011) et RANDOM FOREST (Breiman, 2001). De plus, nous ajoutons des approches plus récentes venant du *deep learning* : BI-LSTM et BERT (Devlin *et al.*, 2019).

TABLE 1. *Tags de la Grande Annotation associée à la thématique du transport*

Transports doux ou moins nombreux	Pistes cyclables, sécurisées	Commerce de proximité, circuit-court
Déménagement ou changement travail	Obligation, incitation au transport collectif	Transport commun gratuit
Télétravail	Transport commun plus fiable, rapide	Covoiturage, auto-partage, stop
Pistes cyclables	Transport commun moins cher	Transport commun plus sûr
Transports en commun meilleurs	Télétravail, horaires flexibles	Parkings (vélo, auto) à proximité des gares
Transports en commun moins chers	Transport commun plus proche	Développer le réseau ferroviaire
Taxer les gros pollueurs	Transport commun plus fréquent	Ramassage scolaire
Vélo, trottinette (électrique)	Voiture électrique	Transports en commun (bus, tram)
Voiture plus propre	Marche à pied	Voiture à hydrogène
Voiture autonome	Train	Voiture hybride
A Déplaçons-nous moins	Transport fluvial	
Transport à la demande	Biocarburant (E85, ...)	

Pour la représentation des données, nous utilisons une représentation *bag of words* (sac de mots) pour les approches classiques. Pour résumer, une représentation *bag of words* encode un texte dans un vecteur $v = \{v_0, \dots, v_N\}$ de dimension N , où N correspond au nombre de mots uniques présents. Chaque valeur v_i du vecteur correspond à la présence d'un terme ou au nombre d'occurrences de celui-ci dans la réponse. Ici, une valeur correspond au nombre d'occurrences d'un mot pondéré par la formule TF-IDF de (Gerard Salton, 1983). Avant de transformer les textes en *bag of words*, ceux-ci sont *lemmatisés*. La lemmatisation consiste à remplacer chaque mot par son *lemme*, c'est à dire la forme canonique d'un mot (*e.g* aimé \rightarrow aimer).

Dans la Table 2, nous présentons les performances obtenues par les différents algorithmes de classifications sélectionnés. Si les valeurs d'exactitude

sont proches, ce sont les algorithmes NAIVE-BAYES et BERT qui présentent les scores les plus élevés. Une fois l'algorithme de classification ayant produit un modèle, nous utilisons celui-ci sur le reste des réponses des contributions sur la transition écologique.

TABLE 2. *Résultat de classification selon différents algorithmes*

Modèle	Paramètres	Exactitude
SVM	g=0.1 & kernel=rbf	0.86
NAIVE-BAYES	alpha = 1	0.87
SGD	loss=log & penalty = L1	0.86
KNN	n = 5 & p =2	0.76
DECISION TREE	criterion = gini	0.83
RANDOM FOREST	g=100 & kernel=rbf	0.84
Bi-LSTM	32 LSTM cells	0.69
BERT	BERT multilingual base model (cased)	0.87

2.3. Extraction de motifs associés à une opinion

Dans ces travaux, nous nous appuyons sur la définition d'une opinion de (Liu, Zhang, 2012). Une opinion o_i est définie selon le quintuplet $\{o_e, o_p, o_c, o_a, o_d\}$; $o_i \in O$, où O correspond à l'ensemble des opinions présentes dans une unité de texte (phrase, paragraphe, etc.), o_e correspond à l'émetteur de l'opinion, o_p est le polarisateur, o_c la cible de l'opinion, o_a est l'attribut de la cible de l'opinion et o_d est la date à laquelle l'opinion est émise. Par exemple, dans la phrase, "Je trouve que le prix du billet de train est hors de prix.", l'opinion o_1 est {"Je", "hors de prix", "billet de train", "prix", \emptyset }. Dans ces travaux, nous ajoutons à ce quintuplet la localisation géographique de l'émetteur, o_l .

2.3.1. Un format d'opinion : la proposition

Il existe plusieurs manières d'exprimer une opinion. Parmi celles-ci, on peut trouver le constat et la proposition. Par exemple, la phrase "J'utilise les transports en commun" exprime un constat et "Il faut augmenter le nombre de train" exprime une proposition. Dans la première phrase, la cible sont "les transports en commun", le verbe utiliser est le polarisateur. Dans la seconde phrase, la proposition est d'"augmenter"(polarisateur) le "nombre"(attribut de la cible) de "train"(cible de l'opinion). Le problème du constat réside dans l'ambiguïté du sentiment qu'éprouve un acteur concernant la cible de l'opinion. Reprenant la phrase d'exemple, le contributeur utilise peut-être les transports en commun par contrainte financière et souhaite dans un avenir proche passer à un moyen de transport individuel. Étant moins ambiguë, nous préférons nous focaliser sur la forme d'opinion qu'est la proposition.

Dans ces travaux, reprenant le quintuplet de (Liu, Zhang, 2012), l'émetteur o_e correspond à l'auteur de la contribution. Le polarisateur o_p est porté par un verbe (modifier, développer, augmenter, etc.) ou adverbe (plus, moins, etc.) de

Grand débat national et transports

changement. La cible de l'opinion o_c correspond à un objet appartenant à la catégorie de transport (moyen de locomotion, infrastructure, etc.). Si celle-ci peut être absente, l'attribut o_a de l'objet cible o_c est caractérisé par un nom (*e.g.* prix, fonctionnement, rapidité, etc.). La *date d'émission* o_d correspond à la date de publication sur la plateforme. La *localisation* o_l est indiquée par le code postal donné par le contributeur.

Afin de traiter de la thématique du transport, les objets cibles o_c , les attributs o_a mais aussi les polarisateurs o_p sont délimités au sein des trois lexiques suivants :

1. Le lexique du transport. Ce lexique contient les termes désignant des moyens de locomotions (train, vélo, bus, voiture, etc.), des infrastructures (gare, voie de chemin de fer).
2. Le lexique du changement. Ce lexique contient des termes impliquant une modification (plus de, changer, augmenter, etc.).
3. Le lexique d'attribut. Ce lexique contient les caractéristiques associées aux termes de transports (prix, accessibilité, quantité, etc.).

L'extraction de motifs de proposition s'effectue de la manière suivante. Dans un premier temps, nous lemmatisons l'ensemble du document. Ensuite, nous identifions les termes des différents lexiques ainsi que leur position dans le document. En s'appuyant sur ces mêmes positions, nous identifions les termes qui appartiennent à une même fenêtre dont la taille est définie par la constante t_{fen} . En conséquence, si un terme appartenant au lexique du changement et au lexique du transport apparaît dans une même fenêtre, une opinion suivant un premier motif est considérée (polarisateur + cible). Si un terme du lexique des attributs se trouve entre les deux termes formant ce motif, alors l'opinion suit un second motif (polarisateur + attribut + cible).

Dans la section suivante, nous présentons les résultats de l'extraction de motifs obtenus en s'appuyant sur notre méthode et les données utilisées.

3. Résultats

3.1. Répartition des motifs obtenus

La distribution des motifs selon le polarisateur et objet cible des opinions détectées est présentée dans la Figure 2.

Dans ces résultats, plusieurs tendances sont identifiables. La première concerne le développement des infrastructures telles que les pistes cyclables ou encore les réseaux de transport mais aussi de moyens de locomotions comme le covoiturage. Inversement, on observe une tendance négative envers les transports individuels et plus particulièrement la voiture, par exemple, associée très largement au verbe limiter, réduire et diminuer. Cependant, celle-ci est contrebalancée par une troisième tendance en lien avec le remplacement d'un véhicule souvent plus

économique (diesel) et/ou plus écologique (hydrogène, électrique). On peut observer cette dernière tendance avec les verbes tels que "acheter", "changer" et "remplacer".

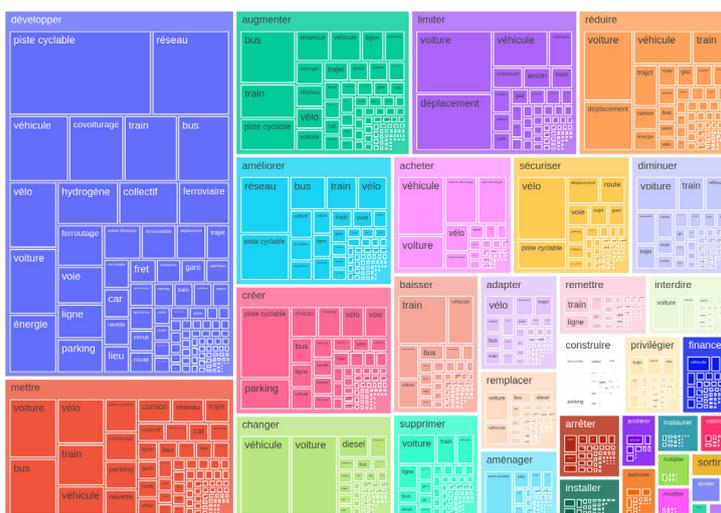


FIGURE 2. Répartition des motifs selon leur polarisateur et objet ciblé

3.2. Répartition géographique des motifs obtenus

Notre analyse permet de mettre en évidence la localisation des contributeurs qui demandent un développement des pistes cyclables, et des contributeurs qui demandent un développement du réseau ferroviaire et/ou une augmentation de la fréquence des trains.

Une échelle d'analyse plus locale semble plus appropriée pour exploiter au mieux ces données. La conférence ayant lieu à La Rochelle, nous avons choisi, à titre d'exemple, de présenter les mêmes données à l'échelle de la Charente-Maritime (Figure 3 et Figure 4).

Les contributeurs étant localisés grâce à leurs codes postaux, une analyse fine à l'échelle de la commune ou de la communauté de commune est impossible. En revanche, il est possible de faire des recommandations à l'échelle des aires urbaines et des départements. Ainsi, en *Charente-Maritime*, il semble que le développement de pistes cyclables soit demandé par des contributeurs dans les aires urbaines de *La Rochelle*, *Saintes* et *Saint Jean d'Angély*, ainsi que dans les communes de l'axe *Rochefort/Saintes/Cognac*. Concernant les demandes de développement du réseau ferroviaire et/ou une augmentation de la fréquence des trains, les besoins sont exprimés au centre des aires urbaines de *Sainte* et de

Grand débat national et transports

La Rochelle . On remarque également une demande dans les communes isolées au nord de *Saint Jean d'Angély* , entre les gares de *Saint Jean d'Angély* et de *Villeneuve-la-Comtesse* .

Les cartes produites par le croisement des contributions en ligne au Grand Débat National, des aires urbaines et des équipements recensés par *OpenStreetMap* permettent donc de proposer des zones de demandes pour certains équipements spécifiques. Les données et le code nécessaire à la réalisation des cartes sont mis à disposition en ligne.¹⁰

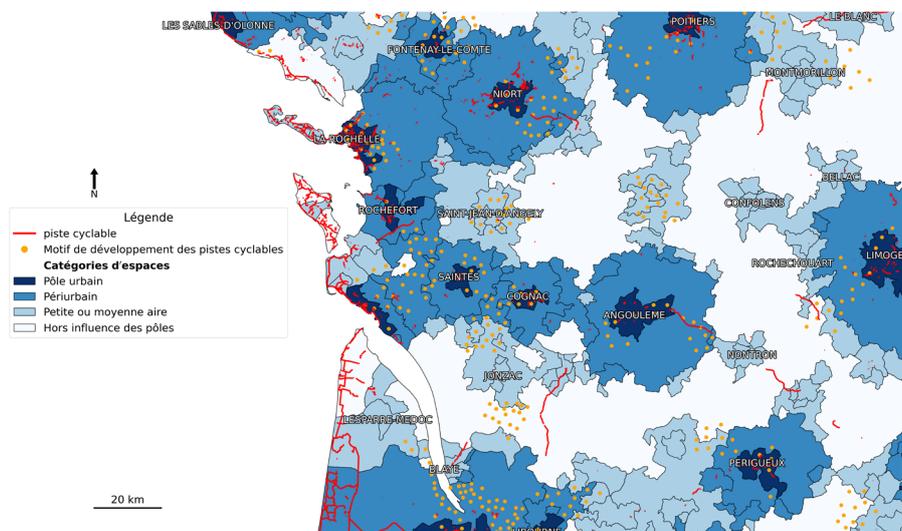


FIGURE 3. *Contributeurs demandant un développement des pistes cyclables en Charente-Maritime et en Charente*

4. Discussions

4.1. Représentativité des contributeurs

Les contributions du Grand Débat National ne sont pas représentatives de la population française. En effet, les résultats que nous présentons dans cet article reflètent l'opinion des contributeurs, mais ne peuvent en aucun cas être considérés comme un reflet fidèle de l'opinion de la population. La méthode d'enquête de ce débat le rend plus proche de la production participative (*crowdsourcing*) ou du réseau social, que du sondage. La différence entre les statistiques

10. Code. https://github.com/Make-the-Debat-Great-Again/map_generation; Données. <https://zenodo.org/record/4270184>

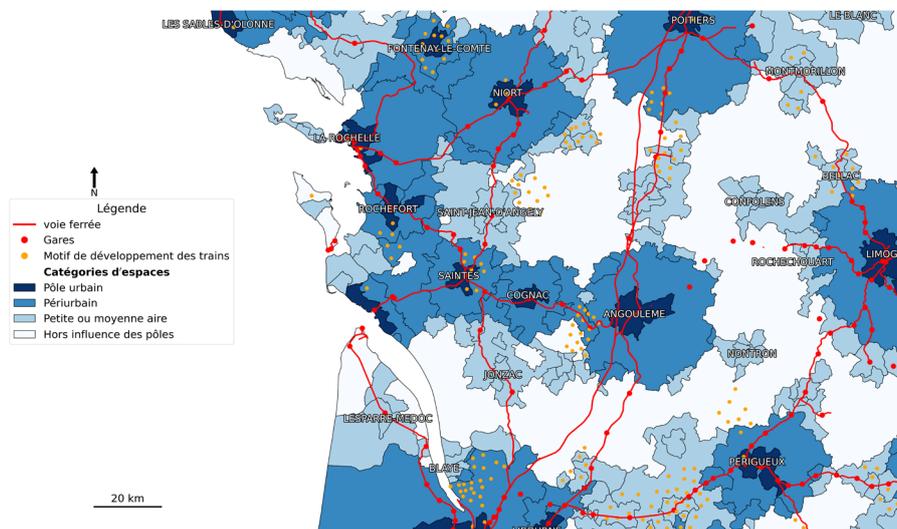


FIGURE 4. *Contributeurs demandant un développement du réseau ferroviaire et/ou une augmentation de la fréquence des trains en Charente et Charente Maritime*

classiques et l’analyse des réseaux sociaux (et plus généralement l’analyse de contenu généré par les utilisateurs) est bien connue, mais son importance a néanmoins été reconnue même en présence de différents types de biais (Mors-tatter, Liu, 2017). Ainsi, les caractéristiques sociologiques des contributeurs (âge, sexe, catégorie socio-professionnelle, ...) n’ont pas été demandées et sont absentes dans le jeu de données. Il n’est donc pas possible de s’assurer que les contributeurs constituent un échantillon représentatif de la population. Seul le code postal, renseigné par le contributeur, permet d’avoir une idée de la répartition spatiale des contributeurs. La participation n’a pas été homogène sur le territoire national (Bennani *et al.*, 2019). On constate par exemple une sous-représentation des grands pôles urbains (58,4 % de la population contre seulement 27 % des contributeurs) et une surreprésentation du périurbain (24,6 % contre 47,6 %) et des autres communes (17 % contre 25,5 %). De plus, il faut noter que les cartes qui ont été présentées ne montrent pas le nombre de contributions dans lesquelles les motifs ont été détectés.

4.2. *Couverture des motifs détectés dans les contributions*

Plusieurs limitations importantes du travail proposé dans cet article se rapporte à la couverture de la méthode de détection de motifs. Premièrement, notre méthode s’appuie sur un ensemble de lexiques non-exhaustifs. Deuxièm-

Grand débat national et transports

mement, l'utilisation d'une fenêtre glissante n'est pas suffisante pour détecter des motifs plus complexes. Troisièmement, nous ne détectons pas l'emploi de la négation. Au-delà de l'aspect de la méthode, le modèle de la "proposition" s'appuyant sur (Liu, Zhang, 2012) et les motifs choisis ne couvre pas l'ensemble des motifs disponibles. Par conséquent, un travail d'annotation et d'exploration des données plus profond sont nécessaires à l'amélioration de ce travail. Des travaux sont menés pour explorer deux nouvelles méthodes d'extractions. Les premiers s'appuient sur l'utilisation de la grammaire de dépendance (ou *dependency parsing*) (Kübler *et al.*, 2009) dans la formulation de règles d'extraction. Les seconds se concentrent sur la construction d'un corpus annoté pour la création d'un modèle de classification.

5. Conclusion

L'extraction de motifs d'opinion des contributions en ligne au *Grand Débat National* et la cartographie de ces opinions, enrichie de données publiques, permettent d'identifier des zones dans lesquelles une demande d'ajout, de développement d'équipements et d'infrastructures est proposée. Ces contributions recueillies à l'échelle nationale sont donc porteuses d'indications pour les décideurs locaux. Elles permettent en effet d'identifier des zones où l'investissement dans certaines infrastructures pourrait permettre de répondre aux attentes de la population. La précision spatiale de la localisation des contributeurs, les échelles spatiales les plus intéressantes nous semblent être l'aire urbaine et le département. Ainsi, ces premiers résultats sont encourageants et montrent le potentiel des contributions au Grand Débat National comme une source d'information pour les collectivités locales. Néanmoins, les limites inhérentes aux données (la non représentativité et le faible nombre de contributeurs) doivent nous conduire à utiliser les résultats avec prudence. Ainsi, les zones identifiées avec notre approche sont des zones où il y a apparemment une demande mais cette demande doit être confirmée par une enquête plus approfondie.

Par ailleurs, nous avons identifié des limites méthodologiques. Les futurs travaux porteront sur l'amélioration de l'extraction d'opinions. Plus particulièrement, par la mise en place d'une campagne d'annotation, puis par la construction d'une ressource lexicale (ontologie) plus complète sur la thématique du transport. L'objectif est de produire un corpus annoté permettant d'extraire d'identifier de nouveaux motifs et/ou de produire des modèles automatiques en s'appuyant sur des méthodes d'apprentissage automatique.

Bibliographie

- Altman N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, vol. 46, n° 3, p. 175–185.
- Bellet A., Denis P., Gilleron R., Keller M., Vauquier N. (2020, juin). *Pour plus de transparence dans l'analyse automatique des consultations ouvertes : leçons de la*

- synthèse du Grand Débat National*. Consulté sur <https://hal.inria.fr/hal-02860659> (working paper or preprint)
- Bennani H., Gandré P., Monnery B. (2019). *Les déterminants locaux de la participation numérique au grand débat national : une analyse économétrique*. Consulté sur https://economix.fr/pdf/dt/2019/WP_EcoX_2019-7.pdf?1.0
- Boyer P. C., Delemotte T., Gauthier G., Rollet V., Schmutz B. (2020). Les déterminants de la mobilisation des Gilets jaunes. *Revue économique*, vol. 71, n° 1, p. 109–138.
- Breiman L. (2001). Random forests. *Machine learning*, vol. 45, n° 1, p. 5–32.
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding*.
- Gerard Salton M. M. (1983). *Introduction to modern information retrieval*. McGraw-Hill.
- Kübler S., McDonald R., Nivre J. (2009). Dependency parsing. *Synthesis lectures on human language technologies*, vol. 1, n° 1, p. 1–127.
- Liu B., Zhang L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, p. 415–463. Springer.
- Loh W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, n° 1, p. 14–23.
- Manning C. D., Schütze H., Raghavan P. (2008). *Introduction to information retrieval*. Cambridge university press.
- Messer M. A., Utz S., Moreau D. (2019). *Analyse exploratoire des contributions au grand débat national, volet mobilité*. techreport. Forum Vies Mobiles, Paris (France). Consulté sur https://fr.forumviesmobiles.org/sites/default/files/editor/granddebat_rapport-final.pdf
- Morstatter F., Liu H. (2017, 1^{er} 6). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, vol. 1, p. 1–13.
- Opinion Way. (2019). *Traitement des données issues du grand débat national*. Rapport technique. Auteur. Consulté sur <https://granddebat.fr/pages/syntheses-du-grand-debat>
- Ploux S., Genay M., Ploux-Chillès L. (2020, janvier). *Les mots du Grand débat national : quelques résultats et réflexions préliminaires*. Consulté sur <https://hal.archives-ouvertes.fr/hal-02441477> (working paper or preprint)
- Vapnik V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Communications par Session

Modélisation spatio-temporelle



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Graph-based Pattern Detection in Spatio-Temporal Phenomena

Kamaldeep Singh Oberoi, Géraldine Del Mondo

Normandie Univ, INSA Rouen, LITIS, 76000, Rouen (France)

kamal.oberoi@insa-rouen.fr, geraldine.del_mondo@insa-rouen.fr

RÉSUMÉ. *Les modèles spatio-temporels (ST) sont souvent utilisés pour analyser les phénomènes ST. Une de ces techniques d'analyse consiste à détecter des motifs dans le phénomène pour comprendre son évolution et modéliser le comportement de ses entités dans l'espace et le temps. Dans cet article, nous nous concentrons sur l'utilisation d'une représentation fondée sur les graphes dynamiques pour la modélisation des phénomènes ST, au sein duquel des motifs structurels, également modélisés sous forme de graphes dynamiques, peuvent être détectés. Nous illustrons le concept de motif par deux exemples applicatifs - le trafic routier et les sports collectifs. Pour chacun d'eux, nous présentons une modélisation du graphe ainsi que des motifs associés. Ensuite, nous formalisons le problème de la détection de motifs comme celui de l'isomorphisme de sous-graphe pour les graphes dynamiques. Enfin, nous présentons les résultats de notre algorithme pour résoudre ce problème. Les premiers résultats décrits dans cet article, qui sont obtenus en utilisant les graphes aléatoires, présentent une base de référence pour les futurs tests de l'algorithme.*

ABSTRACT. *Spatio-temporal (ST) models are often used for analyzing ST phenomena. One such analysis technique is to detect patterns in the phenomenon to understand its evolution and model the behaviour of its entities over space-time. In this paper, we focus on using a dynamic graph-based representation for modeling ST phenomena, within which structural patterns, also modeled using dynamic graphs, can be detected. We illustrate the concept of pattern using two applications - road traffic and invasive team sports. For both these applications, we present the graph model as well as the corresponding patterns. Then we formalize the problem of pattern detection as that of subgraph isomorphism for dynamic graphs. Finally, we present the results of our algorithm to solve this problem. The initial results described in this paper, obtained using random graphs, present a baseline for the future tests of the algorithm.*

MOTS-CLÉS : *Modélisation spatio-temporelle, Graphes dynamiques, Isomorphisme de sous-graphe, Trafic routier, Sports collectifs*

KEYWORDS: *Spatio-temporal modeling, Dynamic graphs, Subgraph isomorphism, Road traffic, Invasive team sports*

1. Introduction

The field of Spatio-Temporal (ST) Modeling has seen a growing interest in the last decades, most notably in the development of Geographical Information Science (GIS). GISystems store and analyze data about real-world entities, highlighting their ST evolution as well as the evolution of the overall phenomenon. Taking Leibnizian point of view, the real-world entities act as the basic units of space. In fact, space is defined in terms of the relations between such entities (Chrisman, 1977). From a conceptual point of view, these distinct entities and their time-varying spatial relations represent the ST phenomenon. This point of view leads to the development of a graph model of the phenomenon (Batty, 2003), where the nodes of the graph correspond to the entities and its edges represent the spatial relations between them. With this approach, the evolution of the phenomenon and its entities is explained in terms of the evolution of the graph, i.e., by modeling change in the attributes of the nodes and the edges and by modeling change in its overall structure. This evolution of the graph makes the temporal dimension of the model explicit. Such a modeling approach has been used for modeling dynamics of street networks (Costes *et al.*, 2015), for describing topological relations between places and events (Maduako, Wachowicz, 2019), etc. in the literature.

Representing the structure of the phenomenon using a graph encourages the development of analysis techniques which exploit the structural properties of the phenomenon. In other words, using this methodology, graph algorithms can be directly applied to analyse the characteristics of the phenomenon. One such analysis technique is the graph-based pattern detection. Pattern detection, in general, is useful to analyse the evolution or the behaviour of the entities of the phenomenon. Depending on the context, patterns might represent the evolution of environmental data (Tsoukatos, Gunopulos, 2001), spatio-temporal proximity of trajectories of moving objects (Demšar, Virrantaus, 2010) etc. However, in these cases, since the representation of the phenomenon does not take into account its underlying structure, the patterns detected are based on the quantitative data acquired about the phenomenon. Such patterns are called *statistical patterns* since statistical techniques are used for their definition and detection. When the underlying structure of the phenomenon is modeled, the concept of *structural patterns* comes into picture. Such patterns take a qualitative modeling approach and are defined in terms of the distinct entities and their corresponding relations which make up the ST phenomenon.

In this paper, we consider the notion of structural patterns and describe a method for their detection in ST phenomena. First, we consider two applications - urban road traffic and invasive team sports (specifically Handball) - to demonstrate the conceptual idea behind the structural patterns within a phenomenon. In case of road traffic, the patterns represent the movement of dynamic entities, like vehicles and pedestrians, and in case of handball, they represent predefined *plays* of the attacking team. The considered ST phenomena

Graph-based pattern detection in ST phenomena

(road traffic and handball) as well as the structural patterns are modeled using dynamic graphs with time-varying edge sets. For now, we do not consider the change in the attributes of the nodes or the edges.

Next, we describe the problem of pattern detection in a ST phenomenon in terms of the well known problem of *subgraph isomorphism (SI)* in graph theory, where a given *pattern graph* has to be detected in a given *target graph*. SI has been well studied in case of static pattern and target graphs (cf. (Conte *et al.*, 2004)). However, we are currently developing an algorithm to extend it for dynamic pattern and target graphs. In this paper, without going into details of the algorithm, we will present some initial results in terms of its runtime, which were computed using random dynamic pattern and target graphs. By using random graphs, we were able to modify various graph parameters, for both pattern and target graphs, to formalize the performance of our algorithm. The results presented in this paper provide us with a baseline which will be helpful when the algorithm will be applied to real data about different ST phenomena in the future.

The paper is organised as follows. Section 2 presents the related work. In Section 3, we will describe the graph model for representing road traffic and discuss different kinds of structural traffic patterns which can be defined using this model. In Section 4, we follow the same plan and discuss structural patterns in a game of handball. In Section 5, we formalize the problem of SI in case of dynamic graphs and then discuss the initial results of our algorithm. Lastly, Section 6 concludes the paper and discusses the future work.

2. Related work

Graphs have been used to model different kinds of spatial and spatio-temporal phenomena for the last two decades. However, an interesting point of discussion is the type of analysis that has been done using such graph representations. For example, (B. Jiang *et al.*, 2000) proposed a graph representation, called connectivity graph, to model the spatial structure of a building, using which graph parameters, like *node degree*, were applied for developing a cognitive point of view of space. Similarly, graph representations have been used to model street networks to understand the topological structure of the network and the *centrality* or the importance of a given street (Porta *et al.*, 2006).

Incorporating temporal information in the graph model leads to the definition of temporal version of graph parameters. For example, in (Del Mondo *et al.*, 2010), the authors have proposed a spatio-temporal graph model for representing ST phenomena. Since time is incorporated in the graph in terms of its time-varying structure, the neighborhood of a node is time-dependent. The authors have used the term *spatio-temporal neighborhood* to describe the neighbors of a node at a given instant (spatial) or during an interval (temporal). Using this notion, the authors are able to model the propagation of

brambles through space-time. Furthermore, for understanding the propagation of a disease within a population, it is important to note the contacts between different people over time. This idea has been applied in (Shirani-Mehr *et al.*, 2012) where a dynamic contact network is used to analyze the propagation of a virus. The authors study *reachability queries* which verify if a path exists between two nodes of the graph. But since the graph is time-varying, the path between two nodes evolves over time.

Graph-based pattern detection is also a very well studied research field. Some existing research focuses on frequent pattern mining, where the pattern is not known *a priori* and a subgraph which exists a certain number of times is considered pertinent (C. Jiang *et al.*, 2013). Moreover, detecting a given pattern graph in a given target graph also has its particular applications (Conte *et al.*, 2004). However, detection of dynamic patterns in dynamic graphs is not very matured research area. With this paper, we want to, first, highlight the semantics behind dynamic graph-based patterns in different spatio-temporal phenomena and, second, formalize the problem of pattern detection as that of subgraph isomorphism for dynamic graphs. The experiments performed on random graphs outline the performance of our algorithm.

3. Structural patterns in urban road traffic

In this section, we consider the case of urban road traffic. We will first describe, in brief, the formalization of the graph model representing traffic and semantics behind its nodes and edges and then discuss some examples of structural traffic patterns.

3.1. Graph-based model of road traffic

In our previous work (Oberoi *et al.*, 2017), a graph for representing urban road traffic was proposed. One of the objectives of this work was to understand the behaviour of dynamic entities of traffic (vehicles and pedestrians) under different traffic situations. As a result, we developed a traffic model which takes into account not only dynamic but static entities (buildings, traffic signs, road markings, footpaths, road segments, intersections) as well which affect traffic flow. In addition, the interactions between these entities, in terms of spatial relations between them (like topological, orientation, relative trajectory, relative speed, qualitative distance etc.), were also considered. The entities constituting traffic were represented as the nodes and the relations between them were represented as its edges. Given that traffic is a dynamic phenomenon, the model incorporated the temporal dimension as well. The proposed model had the support for snapshot-based approach, where each graph snapshot represented the state of the road traffic at a given time instant. Hence, the evolution of the graph, and in turn of road traffic, was described by a set of distinct graph snapshots for a given period of time. The model represents the behaviour of dynamic

Graph-based pattern detection in ST phenomena

entities in terms of their movement and their interactions with other entities over time. Given that the evolution of traffic is described using a set of distinct graph snapshots, the movement of dynamic entities can also be modeled using the same methodology. This defines the semantics behind the structural traffic patterns discussed below.

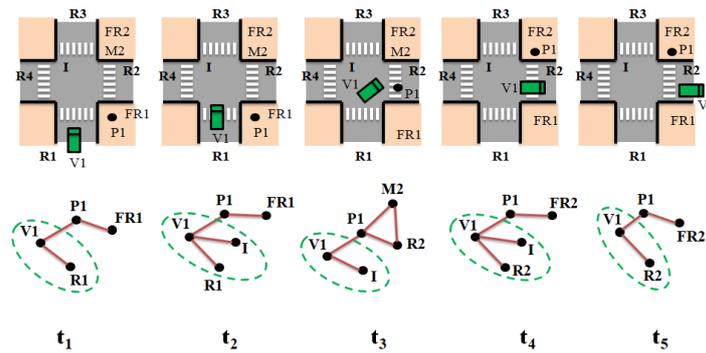


FIGURE 1. Pattern representing the movement of a vehicle and a pedestrian

3.2. Structural traffic patterns

Here, we present two examples of structural traffic patterns describing the movement of a vehicle and/or a pedestrian and their interactions with other entities over time. It is noteworthy that in the model proposed in (Oberoi *et al.*, 2017) edges represent spatial relations between entities but for defining the structural patterns we do not explicitly consider any spatial relations and focus only on the time-varying graph structure.

3.2.1. Vehicle crossing an intersection

A basic traffic situation is that of a vehicle crossing an intersection (the green encircled part of the graphs in Figure 1). In this situation, we consider five graph snapshots which represent the location of the vehicle $V1$ at a given instant. When the vehicle is on a road segment (at times t_1 , t_2 , t_4 and t_5), an edge between the vehicle and the corresponding road segment node ($R1$ or $R2$) exists in those graph snapshots. At t_3 , the vehicle is on the intersection I , hence, there is an edge between the vehicle and the intersection nodes. At times t_2 and t_4 , the vehicle is shared between the road segment and the intersection.

3.2.2. Vehicle crossing an intersection while pedestrian crossing a street

This situation combines the movement of the pedestrian $P1$ with that of the vehicle $V1$ (Figure 1). At time t_1 , the pedestrian is on the footpath $FR1$ and

the vehicle is approaching the intersection. The vehicle waits for the pedestrian at time t_3 and once the pedestrian has crossed the street, the vehicle continues moving at time t_4 . The road marking (zebra crossing) is represented as $M2$. This example highlights the fact that the movement of dynamic entities gets affected due to the presence of other entities around them and this effect can be represented using the structural representation of road traffic.

4. Structural patterns in team sports

In the literature, invasive team sports have been represented using graphs to model player interactions (Gudmundsson, Horton, 2017). In this section, while focusing on the game of handball, we will describe a graph modeling approach which takes into account the "role" of the players as well as their positions on the playing field. Then we will describe two examples of predefined plays for the attacking team, modeled using dynamic graphs. Such predefined actions could represent spatio-temporal patterns in other team sports as well.

4.1. Graph-based model for handball

We are working on developing a graph model to represent the game of handball while taking into account the role of each player and their corresponding positions for both defending and attacking teams. For the attacking team, there are six roles: Left and Right Backcourt, Center Backcourt, Left and Right Wingers and Pivot. For the defending team, there are seven roles (six defenders and one goal keeper) and we consider "6-0" defending formation. In addition, we consider two zones on the court which describe the position of each player. Free Throw ("FT") zone represents the area between the six meter line and the free throw line whereas Nine metre ("9m") zone represents the area beyond the free throw line towards the middle of the court. These two zones are added as nodes in the graph which has a total of seventeen nodes, fourteen corresponding to the player roles, one representing the ball and two for the zones.

The edges between "role" nodes represent the proximity relation, i.e., two players of same or opposite teams are connected if they are close to each other. The exact distance value to define the relation between two role nodes is not yet fixed. An edge between a role and the ball node represents the possession of the ball by the player having that role. The ball node could also be connected to two role nodes representing the action of pass between corresponding players. And finally, an edge between a role and zone node represents the topological relation between the player and the zone. It is added to represent the position of the player on the court. As in the case of road traffic, this graph represents the underlying structure of the game by describing player interactions. It is evident that this graph evolves with time and its evolution can be modeled in terms of its time-varying structure. The progress of the game is represented as

Graph-based pattern detection in ST phenomena

a set of discrete states of the graph. Such a graph model can be used to model and extract the predefined actions of the attacking team as discussed below.

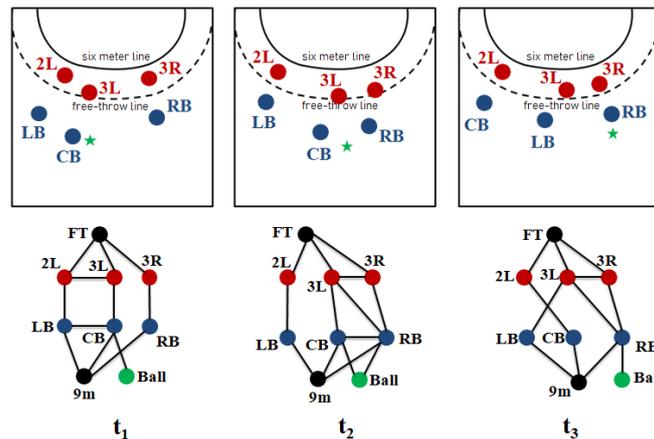


FIGURE 2. Switch between *CB* and *LB* players in the attacking team (in blue). The defending team is in red and the green star represents the ball

4.2. Structural patterns in handball

Similar to road traffic, structural patterns representing player actions in handball can be defined and modeled using dynamic graphs.

4.2.1. Switch between Center Backcourt and Left Backcourt

This tactic involves the players with roles of Center Backcourt *CB*, Left Backcourt *LB* and Right Backcourt *RB* in the attacking team (Figure 2). The tactic begins when Center Backcourt is in the possession of the ball at time t_1 . Then the ball is passed to the Right Backcourt at time t_2 . Then Center Backcourt takes the place of Left Backcourt and Left Backcourt takes the place of Center Backcourt leaving Right Backcourt with the ball possession at time t_3 . Then, according to the game situation, Right Backcourt may pass the ball or take the shot. Here, the roles of Right and Left Backcourt players can be interchanged.

4.2.2. Crossover between Center Backcourt and Right Backcourt

Here, the Center Backcourt and the Right Backcourt are involved in the action. The tactic begins when Center Backcourt is in the possession of the ball at time t_1 . He/she runs towards the direction of the Right Backcourt and Right Backcourt runs towards the middle of the playing field. When they are

close enough, the Center Backcourt makes a pass to the Right Backcourt at time t_2 , leaving Right Backcourt with the ball possession at time t_3 . Then, depending on the situation, Right Backcourt either shoots or makes another pass. The same action can also be performed by Center and Left Backcourt players as well.

5. Pattern detection in ST phenomena

In Sections 3 and 4, we described the graph-based approach for modeling road traffic and invasive team sports respectively. In both cases, the global phenomenon was modeled using a dynamic (target) graph within which some pertinent patterns, also modeled using dynamic (pattern) graphs, could be detected. In this section, we will formalize the problem of pattern detection as that of subgraph isomorphism (SI). First, we will formally define the problem and then present initial results of the algorithm under development to solve SI in case of dynamic pattern and target graphs.

5.1. Problem formulation

Consider a static graph $G = (X, E, \psi)$ with a set of nodes X and set of edges $E \subset X \times X$. Node classification function $\psi : X \rightarrow C$ assigns a class to each node from a set of classes C . It is possible for two nodes to belong to the same class. For example, in case of road traffic a node class could mean Vehicle, Road marking etc. whereas in case of handball a node class could mean Zone, Player of the attacking team etc. In addition, unique node labels are assigned to each node. Given a pair of static pattern graph $H = (X_H, E_H, \psi_H)$ and static target graph $G = (X_G, E_G, \psi_G)$, the problem of SI is to find an injective mapping $\mu \subset X_H \times X_G$ between the nodes of both graphs, while respecting the following conditions:

$$\forall u \in X_H, \exists v \in X_G \mid (u, v) \in \mu \quad (1)$$

$$\forall u, u' \in X_H, u \neq u', \exists \{(u, v), (u', v')\} \in \mu \Rightarrow v \neq v' \quad (2)$$

$$\forall (u, u') \in E_H, \exists (v, v') \in E_G \quad (3)$$

$$\forall u, u' \in X_H, (v, v') \in E_G, \{(u, v), (u', v')\} \in \mu \Rightarrow (u, u') \in E_H \quad (4)$$

$$\forall u \in X_H, \exists (u, v) \in \mu \Rightarrow \psi_H(u) = \psi_G(v) \quad (5)$$

Equations 1 and 2 state that all distinct nodes of the H must be mapped to some (but distinct) nodes of G . Equation 3 states that all edges of H are mapped to some edges of G . Equation 4 verifies the "induced" version of SI by checking that a mapped edge of G has a corresponding edge in H , i.e., the subgraph of G matched with H has no extra edges. Lastly, equation 5 states that only the nodes belonging to the same class in both H and G are mapped.

Graph-based pattern detection in ST phenomena

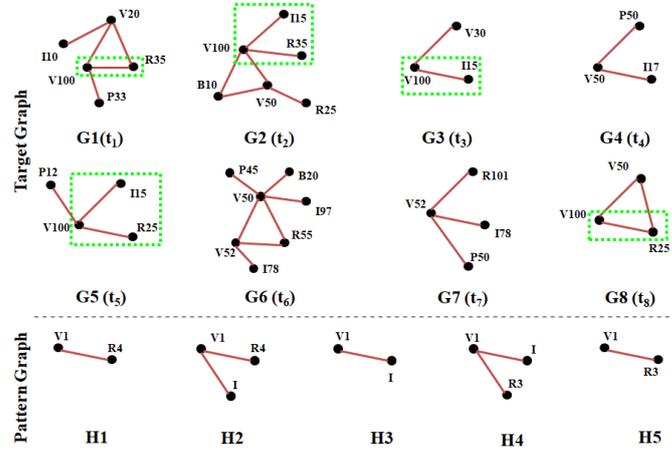


FIGURE 3. Snapshots of target graph \mathcal{G} and pattern graph \mathcal{H} with detected pattern highlighted in \mathcal{G} . Set of nodes in both graphs remains constant over time but the isolated nodes are not shown for better readability

Next, we define a simple dynamic graph $\mathcal{G} = (X_{\mathcal{G}}, E_{\mathcal{G}}, \mathcal{T}_{\mathcal{G}}, \psi_{\mathcal{G}})$ as a graph whose set of edges $E_{\mathcal{G}}$ is time-varying. Its discrete time domain $\mathcal{T}_{\mathcal{G}}$ is represented as a set of time instants $\mathcal{T}_{\mathcal{G}} = \{t_1, t_2, \dots, t_k\}$. We consider a snapshot-based model where a graph snapshot G_i at time instant t_i represents its state. Following this, a dynamic graph can be represented as a set of its discrete states, i.e., $\mathcal{G} = \{G_1, G_2, \dots, G_k\}$ (Figure 3).

Given a pair of dynamic pattern graph $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$, having m snapshots, and dynamic target graph $\mathcal{G} = \{G_1, G_2, \dots, G_n\}$, having n snapshots, with $m \leq n$, we extend the problem of SI for dynamic graphs, i.e., (1) Each snapshot of \mathcal{H} must be mapped to some snapshot of \mathcal{G} while respecting the conditions of subgraph isomorphism (Equations 1-5). (2) The time order of \mathcal{H} and \mathcal{G} must be respected, i.e., if a snapshot $H_i, 1 \leq i \leq m$ of \mathcal{H} is mapped to a snapshot $G_j, 1 \leq j \leq n$ of \mathcal{G} , then the next snapshot H_{i+1} of \mathcal{H} must be mapped to some snapshot G_f of \mathcal{G} where $f > j$. (3) To find a complete matching of \mathcal{H} , a node $v \in X_{\mathcal{H}}$ must be mapped to the same node $v' \in X_{\mathcal{G}}$ at every time instant where the matching is possible and both nodes v and v' exist. (4) It is possible to have some snapshots of \mathcal{G} to which no snapshots of \mathcal{H} are matched. Such snapshots represent *noise*. They are included to be realistic in terms of the sampling rate to generate the target graphs in real applications. In Figure 3, the node classes are represented by the first character of their unique labels, i.e., the node having label V20 belongs to the class V etc. The pattern and target graph snapshots are matched as follows: $H1 \rightarrow G1, H2 \rightarrow G2, H3 \rightarrow G3, H4 \rightarrow G5$ and $H5 \rightarrow G8$, while the nodes of the pattern graph are

matched as: $V1 \rightarrow V100$, $R4 \rightarrow R35$, $I \rightarrow I15$ and $R3 \rightarrow R25$. Some target graph snapshots (G_4 , G_6 and G_7) represent "noise".

5.2. Experiments and results

We are developing an algorithm which takes graphs \mathcal{H} and \mathcal{G} as input and detects all instances of \mathcal{H} in \mathcal{G} . Our algorithm is based on another algorithm, VF3 (Carletti *et al.*, 2018), which has been developed for static pattern and target graphs. We extend VF3 to include the temporal order for both graphs. We have tested our algorithm using 6050 random pattern and target graph pairs which were generated with different values of graph parameters - Size (N), Density (D), Number of node classes (C) and Number of snapshots (T), while making sure that at least one instance of the pattern graph exists in the corresponding target graph. We calculated the average runtime and standard deviation (SD) for fifty graph pairs with different values of each parameter, while considering the total time taken, excluding, however, the time required to load each graph from the files. Some of the results obtained are tabulated in Table 1. For denser target graphs, the average runtime increases. This is due to the fact that for denser target graphs, the number of possible node mappings to be verified is higher. Similarly, increasing the number of snapshots of the target graph, the runtime increases since algorithm performs more iterations to find the solution. Increasing, however, the number of node classes of the target graph does not have a significant effect on the average runtime. For denser pattern graphs, the runtime also increases since it takes more time to look for the corresponding edges in the target graph to satisfy the equations 3 and 4 (Section 5.1). Increasing the number of snapshots of the pattern graph shows a peculiar behaviour. The runtime keeps increasing with the increase in the number of pattern snapshots since more graphs have to be mapped to find the complete solution. However, when the number of pattern snapshots becomes equal to the number of target snapshots ($T_t = T_p = 100$ in this case) the algorithm terminates quickly. This is due to the fact that the steps required to look for a future snapshot of the target graph are not performed in this case since a snapshot H_i of the pattern graph gets mapped to the snapshot G_i of the target graph. Similarly to target graph, increasing the number of node classes of the pattern graph does not have any significant effect on the runtime.

6. Conclusion and future work

The objective of this paper is to show the applicability of qualitative graph-based representations for modeling ST phenomena. Graphs represent the underlying structure of the phenomenon which can be exploited to analyse its spatio-temporal evolution. We focused on detecting patterns within different phenomena. Taking the examples of road traffic and handball, we highlighted the conceptual idea of structural pattern and then formalized the problem of

Graph-based pattern detection in ST phenomena

TABLE 1. Avg. runtime and standard deviation for different graph parameters

Pattern Graph				Target Graph				Runtime	SD
N_p	C_p	D_p	T_p	N_t	C_t	D_t	T_t	(secs)	(secs)
30	20	0.8	30	700	20	0.2	30	0.15	0.02
30	20	0.8	30	700	20	0.3	30	0.15	0.016
30	20	0.8	30	700	20	0.4	30	0.34	0.06
30	20	0.8	10	700	20	0.3	30	0.63	0.08
30	20	0.8	10	700	20	0.3	60	1.43	0.16
30	20	0.8	10	700	20	0.3	100	2.57	0.3
30	20	0.8	30	700	5	0.3	30	0.14	0.015
30	20	0.8	30	700	10	0.3	30	0.13	0.013
30	20	0.8	30	700	20	0.3	30	0.15	0.016
100	20	0.4	30	300	20	0.3	30	0.71	0.033
100	20	0.6	30	300	20	0.3	30	0.76	0.032
100	20	0.8	30	300	20	0.3	30	0.82	0.034
100	20	0.4	30	300	20	0.3	100	28.02	1.23
100	20	0.4	60	300	20	0.3	100	58.55	2.6
100	20	0.4	100	300	20	0.3	100	2.43	0.12
100	5	0.4	30	300	20	0.3	30	0.71	0.031
100	10	0.4	30	300	20	0.3	30	0.72	0.033
100	20	0.4	30	300	20	0.3	30	0.71	0.034

pattern detection as that of subgraph isomorphism for dynamic graphs. Finally, the results of our algorithm applied to random dynamic graphs were discussed. The next step would be to apply our algorithm on the real data. In fact, we have generated structural traffic patterns using a traffic simulator and we are in the process of applying our algorithm on the generated traffic model. For the case of Handball, the graph has been generated for a single game using video sequences and we are at the stage of defining pertinent structural patterns.

Acknowledgements

The authors would like to thank François Rioult, Quentin Bourgeais and Jeffrey Girault for providing the data for handball and for multiple discussions related to the detection of predefined plays in handball as part of NormaSTIC Federation.

Bibliographie

Batty M. (2003). Network geography: Relations, interactions, scaling and spatial processes in gis. *Centre for Advanced Spatial Analysis Working Paper Series*, p. 63:1-63:23.

- Carletti V., Foggia P., Saggese A., Vento M. (2018, April). Challenging the time complexity of exact subgraph isomorphism for huge and dense graphs with VF3. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, n° 4, p. 804-818.
- Chrisman N. (1977). Concepts of space as a guide to cartographic data structures. In *First international advanced study symposium on topological data structures for geographic information systems*.
- Conte D., Foggia P., Sansone C., Vento M. (2004). Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, n° 03, p. 265-298.
- Costes B., Perret J., Bucher B., Gribaudo M. (2015). An aggregated graph to qualify historical spatial networks using temporal patterns detection. In *18th agile international conference on geographic information science*.
- Del Mondo G., Stell J. G., Claramunt C., Thibaud R. (2010). A graph model for spatio-temporal evolution. *Journal of Universal Computer Science*, vol. 16, n° 11, p. 1452-1477.
- Demšar U., Verrantaus K. (2010). Space-time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, vol. 24, n° 10, p. 1527-1542.
- Gudmundsson J., Horton M. (2017, avril). Spatio-temporal analysis of team sports. *ACM Comput. Surv.*, vol. 50, n° 2.
- Jiang B., Claramunt C., Klarqvist B. (2000). Integration of space syntax into gis for modelling urban spaces. *International Journal of Applied Earth Observation and Geoinformation*, vol. 2, n° 3, p. 161 - 171.
- Jiang C., Coenen F., Zito M. (2013). A survey of frequent subgraph mining algorithms. *The Knowledge Engineering Review*, vol. 28, n° 1, p. 75-105.
- Maduako I., Wachowicz M. (2019). A space-time varying graph for modelling places and events in a network. *International Journal of Geographical Information Science*, vol. 33, n° 10, p. 1915-1935.
- Oberoi K. S., Del Mondo G., Dupuis Y., Vasseur P. (2017). Towards a qualitative spatial model for road traffic in urban environment. In *2017 IEEE 20th international conference on intelligent transportation systems*, p. 1724-1729. IEEE.
- Porta S., Crucitti P., Latora V. (2006). The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design*, vol. 33, n° 5, p. 705-725.
- Shirani-Mehr H., Banaei-Kashani F., Shahabi C. (2012). Efficient reachability query evaluation in large spatiotemporal contact datasets. *Proc. VLDB Endow.*, vol. 5, n° 9, p. 848-859.
- Tsoukatos I., Gunopulos D. (2001). Efficient mining of spatiotemporal patterns. In C. S. Jensen, M. Schneider, B. Seeger, V. J. Tsotras (Eds.), *Advances in spatial and temporal databases*, p. 425-442. Springer Berlin Heidelberg.

Une approche sémantique pour l'exploitation de données spatio-temporelles avec annotation Application à l'évolution d'assolement

Ba-Huy Tran¹, Alain Bouju²,
Christine Plumejeaud-Perreau³, Vincent Bretagnolle⁴

1. IRIT, Université de Toulouse

ba-huy.tran@irit.fr

2. Laboratoire L3i, Université de La Rochelle

alain.bouju@univ-lr.fr

3. Laboratoire LIENSs, U.M.R. 7266, CNRS & Université de La Rochelle

christine.plumejeaud-perreau@univ-lr.fr

4. Centre d'Etudes Biologiques de Chizé, U.M.R. 7372, CNRS & Université
de La Rochelle

breta@cebc.cnrs.fr

RÉSUMÉ. Cet article présente une approche ontologique pour l'exploitation des données d'assolement. Nous présentons une ontologie spatio-temporelle décrivant ces données en réutilisant une ontologie du temps et une ontologie de l'espace. Nous montrons ensuite comment incorporer le processus de fouille de données sémantique et l'annotation sémantique afin de fournir aux utilisateurs un outil d'analyse automatique et efficace. L'approche proposée a été appliquée aux évolutions des assolements de la Zone Atelier Plaine et Val de Sèvre.

ABSTRACT. This article presents an ontology-based approach for exploiting land use data. We present a spatio-temporal ontology describing these data by reusing an ontology of time and an ontology of space. We then show how to incorporate the process of data mining and annotation on semantic data to provide users an automatic and efficient analysis tool. The proposed approach was applied to the evolution of the land use of the Plaine and Val de Sèvre workshop area.

MOTS-CLES: ontologies, spatio-temporelle, assolement, ETL, fouille de données, annotation

KEYWORDS: ontology, spatio-temporal, land use, ETL, data mining, annotation

1. Introduction

Dans les zones rurales avec une prédominance d'activités agricoles, l'étude des questions environnementales peuvent bénéficier de l'analyse sur le long terme de la mosaïque des cultures résultant des pratiques agricoles. En effet, les paysages agricoles sont principalement le résultat des décisions des agriculteurs portant sur le choix des cultures et de leur répartition à l'échelle de l'exploitation. L'agencement, la forme et la nature des cultures structurent l'organisation spatiale du paysage qui a une incidence sur les processus écologiques à différentes échelles (Schaller *et al.*, 2012).

Cette organisation spatiale du paysage évolue dans le temps parce que les agriculteurs modifient l'assolement de leurs parcelles et éventuellement les limites de ces parcelles chaque année. En outre, les parcelles peuvent changer de propriétaire ou appartenir à des associations d'exploitants agricoles. Cette information peut être très utile lors de l'étude des liens entre l'environnement socio-économique, les pratiques agricoles et l'organisation spatiale du paysage.

L'article présente notre travail étendant celui de (Tran, 2017) réalisé pour le projet GEMINAT¹ visant à étudier ces pratiques agricoles et l'organisation spatiale du paysage d'une zone atelier, tout en exploitant ses données d'assolement. Nous présentons d'abord les ontologies spatio-temporelles qui servent au développement de notre ontologie montrée par la suite. La section 3 décrit les processus Extract-Transform-Load (ETL) sémantiques qui consistent à extraire, transformer et importer les données sources dans un triplestore RDF. La section suivante détaille deux techniques pour exploiter la base construite : la fouille de données sémantiques et l'annotation sémantique. Ensuite, nous démontrons l'utilisation du système appliqué à un cas d'étude sur la base d'assolement de la zone atelier. Enfin, la conclusion résume notre travail et ouvre des perspectives sur de futurs travaux.

2. Ontologies spatio-temporelles

La modélisation spatio-temporelle par une approche ontologique comporte l'ontologie du domaine, l'ontologie du temps et l'ontologie de l'espace pour la représentation des informations sémantiques, temporelles et spatiales respectivement. La modélisation spatio-temporelle consiste à coordonner ces trois composantes pour rendre compte des dynamiques spatiales. Nous présentons

1. <http://za-geminat.cnrs.fr/>

Une approche sémantique avec annotation

en premier une étude sur les ontologies existantes et proposons notre choix, ensuite, notre ontologie spatio-temporelle pour l'assolement.

2.1. Ontologies du temps

Les ontologies du temps visent à représenter des concepts et des relations temporelles. Plusieurs ontologies du temps ont fait leur apparition au sein de la communauté du Web sémantique. Les deux travaux majeurs du domaine sont l'ontologie OWL-Time (Hobbs, Pan, 2004) et SWRL Temporal Ontology (O'Connor, Das, 2011). Étant une recommandation du W3C, la première a été choisie pour notre modélisation.

L'ontologie repose sur les concepts temporels et les relations temporelles définies dans la théorie d'Allen et bénéficie d'une spécification précise formalisée en OWL. Elle a devenu une ontologie de référence pour représenter le temps dans plusieurs domaines, y compris le domaine de la recherche d'information. L'élément principal de l'ontologie est *TemporalEntity* qui comprend deux sous-classes : *Instant* et *Interval*. Les propriétés *hasBeginning* et *hasEnd* sont utilisées pour définir le début et la fin d'une entité temporelle *Interval*.

2.2. Ontologies de l'espace

De nombreux travaux ont proposé des ontologies pour définir les concepts et les relations spatiales. Ces ontologies sont majoritairement basées sur le modèle GML² de l'OGC³ et utilisent le langage de formalisation ontologique OWL. Nous nous intéressons à la partie déclarative et à la représentation des données spatiales par des ontologies. Les ontologies de l'espace les plus utilisées sont: Basic Geo Vocabulary, GeoRSS et GeoOWL. Il existe aussi des modèles minimalistes provenant des extensions du langage SPARQL comme le modèle stRDF (Koubarakis, Kyzirakos, 2010) et l'ontologie GeoSPARQL (Battle, Kolas, 2012). Il faut noter qu'il existe aussi d'autres ontologies comme NeoGeo Geometry Ontology, GeoNames Ontology ou Biological Spatial Ontology ou la transformation d'un modèle spatial en une ontologie spatiale comme utilisé dans (Mefteh *et al.*, 2012).

Parmi ces ontologies, nous avons choisi GeoSPARQL, qui est une recommandation du W3C, pour pouvoir profiter des fonctions fournies par les triplestore spatiaux pour effectuer du raisonnement spatial. L'ontologie GeoSPARQL possède trois classes : *geo:Feature* représentant des entités qui peuvent avoir une localisation, *geo:Geometry* décrivant un emplacement spatial, c'est-à-dire un ensemble de coordonnées, et *geo:SpatialObject* étant la classe racine des deux

2. Geography Markup Language

3. Open Geospatial Consortium

classes. La propriété *geo:hasGeometry* met en relation une entité à son emplacement. En séparant les entités réelles et leurs emplacements, GeoSPARQL permet à plusieurs emplacements d'être liés à une entité à des fins diverses. La ressource de la géométrie a alors un littéral qui est lié à une propriété en utilisant soit *geo:asWKT* ou *geo:asGML*.

2.3. Représentation des changements

Les ontologies de temps présentées ci-dessus permettent de représenter des concepts et des relations temporelles. Toutefois, elles ne s'intéressent qu'à décrire des éléments de données individuels plutôt qu'à construire un modèle temporel pour décrire systématiquement toutes les informations temporelles dans un système. Autrement dit, de telles ontologies ne suffisent pas à modéliser l'évolution des entités. Par ailleurs, comme les deux langages ontologiques du Web sémantique, OWL et RDFS, ne permettent que des relations binaires entre les entités, le traitement des informations qui changent au fil du temps devient un problème critique. Afin de pallier ces problèmes, plusieurs approches ont été proposées. Nous pouvons distinguer les approches qui ne modifient pas les ontologies comme le versionnement, ou les approches qui doivent introduire de nouveaux éléments dans l'ontologie comme les techniques de réification et l'utilisation des modèles des fluents (ou *4D fluents*).

Les modèles des fluents avec la notion de *TimeSlice* (Welty, Fikes, 2006) semblent les plus connus. On peut citer des travaux de (Batsakis, Petrakis, 2011) ou (Harbelot *et al.*, 2013) visant à améliorer la modélisation et à l'appliquer à différents domaines. En effet, le concept *TimeSlice* joue le rôle central en étant lié à une dimension spatiale, une dimension temporelle et éventuellement une dimension sémantique liée au domaine applicatif.

2.4. Ontologie pour l'assolement

Nous proposons une ontologie GEMINAT, préfixée *gem* dans les paragraphes qui suivent. Inspiré des modèles de fluents, nous considérons que les parcelles agricoles possèdent une ou plusieurs tranches de temps correspondant aux différents états à travers leur vie. Elles peuvent changer de forme (par des événements territoriaux) ou de culture suivant les saisons. Le changement peut être enregistré à travers des observations ou des relevés. Ainsi, la classe *gem:ParcelTimeSlice* est introduite. Elle se compose de trois composantes (Figure 1) :

- Composante sémantique : Cette composante représente l'ensemble des propriétés décrivant les caractéristiques de la parcelle durant la période de validité du timeslice. Dans un premier temps, seule la culture (*gem:LandUse*) est utilisée pour cette composante.

Une approche sémantique avec annotation

– Composante temporelle : Cette composante représente une unité de temps durant lequel le timeslice est valide. La propriété *gem:hasTime* associe un timeslice à une entité temporelle.

– Composante spatiale : Cette composante représente une description géométrique de l'entité. Pour cela, nous considérons la classe *gem:TimeSlice* comme une spécialisation de *geo:Feature*.

Afin de représenter ces associations spatiales dans le temps, nous utilisons la relation filiation qui permet de connecter deux timeslices consécutifs d'un même objet. Un changement sur la composante spatiale ou sur la composante sémantique génère un nouveau timeslice. D'une part, ce timeslice est obligatoirement lié par une relation de filiation avec le timeslice d'origine. D'autre part, l'intervalle d'existence du timeslice parent est contigu à l'intervalle du timeslice enfant.

La propriété *gem:hasEvent*, associant deux *gem:ParcelTimeSlice*, est utilisée pour représenter des événements territoriaux (Plumejeaud *et al.*, 2011) survenus entre ces timeslices.

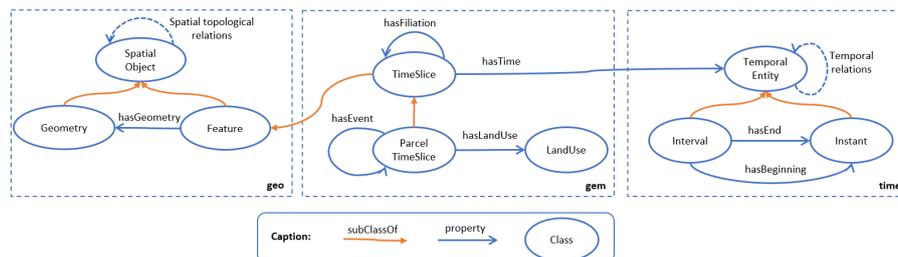


FIGURE 1. Un extrait de l'ontologie GEMINAT utilisée pour l'assolement et son articulation avec les ontologies GeoSPARQL et OWL-Time

3. ETL sémantique

Dans le peuplement de l'ontologie, lui-même, il existe des techniques pour exposer des données relationnelles en RDF : (i) la matérialisation de données et (ii) la traduction de données à la demande. Quelle que soit la technique choisie, un outil de traduction est nécessaire pour l'exposition de ces données. Nous pouvons distinguer les outils D2RQ⁴, Ultrawrap⁵, Morph⁶, TripleGeo⁷

4. <http://d2rq.org/>

5. <https://capsenta.com/ultrawrap/>

6. <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/technologies/315-morph-rdb/>

7. https://web.imsi.athenarc.gr/redmine/projects/geoknow_public/wiki/TripleGeo

ou GeoTriples⁸ aidant à transformer les données sources en RDF et les outils D2R⁹ visant à exposer un triplestore virtuel, certains outils comme Ontop¹⁰ réunissant les deux fonctionnalités.

Nous avons choisi la matérialisation de données dans un triplestore géospatial pour les raisons suivantes:

– Il n'est pas facile de réaliser la traduction à la demande parce que les sources de données sont disponibles dans différents formats, cela requiert une étape de pré-traitement. En effet, comme présenté, nous avons intégré d'autres sources de données (qui ne sont pas décrites ici).

– Un triplestore géospatial peut être utilisé comme un entrepôt de données pour stocker les données sémantiques. De cette manière, le raisonnement spatial sera pris en charge par le triplestore et l'enrichissement ou l'annotation de données peuvent être réalisés.

Actuellement, plusieurs triplestores prennent en charge le stockage et l'interrogation de données spatiales à l'aide de GeoSPARQL. Il existe des triplestores gratuits comme Parliament¹¹ (Battle, Kolas, 2012), Strabon¹² (Kyzirakos *et al.*, 2012) ou GraphDB¹³; et d'autres payants comme Oracle Spatial and Graph¹⁴ ou Stardog¹⁵. Parmi les triplestores gratuits, Strabon a été choisi car il a de bonnes performances globales grâce à des techniques d'optimisation particulières qui permettent aux opérations spatiales de bénéficier des fonctionnalités de PostGIS au lieu de s'appuyer sur des bibliothèques externes (Patroumpas *et al.*, 2014). Pour les applications complexes qui incluent à la fois des jointures spatiales ou des agrégations spatiales, Strabon est le seul triplestore qui fonctionne bien (Ioannidis *et al.*, 2019).

4. Exploitation de données

Une fois les données importées dans l'entrepôt, elles peuvent être exploitées de différentes manières. Nous allons présenter la fouille de données sémantiques et l'annotation sémantique de données. La première permet de détecter des erreurs au sens statistique du terme par exemple, tandis que la seconde fournit un support sémantique pour l'enregistrement des résultats de la fouille dans le modèle.

8. <http://geotriples.di.uoa.gr/>

9. <http://d2rq.org/d2r-server>

10. <http://ontop.inf.unibz.it/>

11. <http://parliament.semwebcentral.org/>

12. <http://strabon.di.uoa.gr/>

13. <http://graphdb.ontotext.com/>

14. <https://www.oracle.com/database/technologies/spatialandgraph.html>

15. <https://www.stardog.com/>

Une approche sémantique avec annotation

4.1. Fouille de données sémantiques

Au cours de la dernière décennie, de nombreuses approches ont été proposées pour appliquer des algorithmes de fouille aux données sémantiques. Nous pouvons distinguer trois approches : (i) Approches basées sur les logiques de description visant à adapter les algorithmes existants au nouveau format de représentation des données, comme dans (Józefowska *et al.*, 2010). (ii) Fouille des graphes consistant à extraire des sous-graphes (Kuramochi, Karypis, 2001) ou des sous-arbres fréquents (Chi *et al.*, 2004). (iii) Approches statistiques visant à transformer les données RDF en structure de données classique utilisée par des algorithmes de fouille comme dans (Nebot, Berlanga, 2012). Comme nous souhaitons implémenter un système simple et efficace pour faciliter l'analyse des données, il est préférable de réutiliser des algorithmes existants. Dès lors, les approches statistiques conviennent à nos besoins car elles ne consistent pas à introduire de nouveaux algorithmes comme dans le cas de la fouille de graphes ou à modifier des algorithmes existants comme dans le cas des approches basées sur la logique de description.

4.2. Annotation sémantique

Une annotation attache des données à d'autres données : elle établit, dans un certain contexte, une relation (typée) entre les données annotées et les données d'annotation. Nous distinguons trois types d'annotations (Oren *et al.*, 2006) : informelles, formelles (formellement définie) et ontologiques (utilisant uniquement des termes ontologiques). Il existe de nombreuses définitions de l'annotation sémantique mais le point commun de ces définitions est qu'une annotation sémantique est le processus de liaison d'une ressource électronique à une ontologie spécifique. Les ressources électroniques peuvent être des contenus textuels, des images, des vidéos, des services, etc. L'ontologie n'est ici qu'un des moyens possibles pour fournir une sémantique formelle (Liao *et al.*, 2011).

Les modèles d'annotation les plus connus sont Annotea (Kahan *et al.*, 2002), Annotation Ontology (Cicarese *et al.*, 2011), ou Open Annotation Data Model (Sanderson *et al.*, 2013). Le dernier, présenté par le W3C, est utilisé pour annoter les ressources de notre système. Le modèle décrit les annotations par la classe *Annotation* qui est une association entre deux éléments, une ressource de contenu, appelée *Body* et une ou plusieurs ressources appelée *Target*. L'annotation fournit des informations sur le *Target* via la connexion au *Body*. Les annotations peuvent être complétées par des méta-données, par exemple, l'auteur ou la date de création de l'annotation.

5. Exploitation de données d'asselement par l'approche ontologique

Notre pipeline pour l'exploitation de données d'asselement par l'approche ontologique se compose de cinq modules:

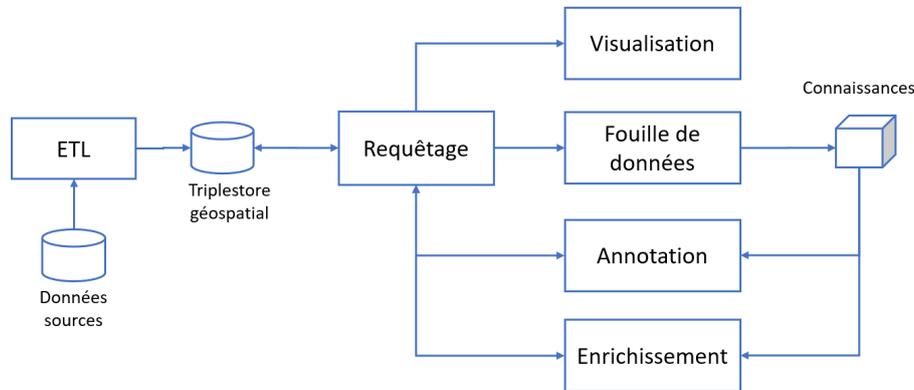


FIGURE 2. Le pipeline pour l'exploitation de données sémantiques.

1. ETL : Les données sources sont extraites, transformées en RDF et importées dans le triplestore Strabon.

2. Requêtage : Le module sert à traiter les requêtes SPARQL 1.1. Elles peuvent être des requêtes de mise à jour comme dans le cas de l'annotation ou l'enrichissement; ou des requêtes normales pour les autres cas.

3. Visualisation : Le résultat d'une requête peut être affiché sur une carte grâce à la librairie OpenLayers.

4. Fouille de données : La fouille de données peut être enchaînée après une requête. Pour cela, une sélection d'attribut peut être réalisée pour éliminer les attributs non pertinents comme l'identifiant ou la géométrie contenus dans chaque transaction. Le format de l'entrée du processus est tabulaire. Trois algorithmes de fouille sont implémentés à l'aide de l'outil Weka: (i) *Apriori* pour la recherche des règles d'association, (ii) *Part* pour la recherche des règles de classification. (iii) *K-moyennes* pour la recherche des partitions.

5. Annotation: Basé sur les connaissances obtenues lors de l'étape précédente ou sur le résultat des requêtes, on peut annoter des doutes ou anomalies. Par exemple, on peut annoter des *timeslices* ayant une culture douteuse relativement à une règle de succession ou à la période donnée.

6. Enrichissement: Les connaissances ou les résultats obtenus peuvent servir à enrichir la base de connaissances. Un cas d'étude sur le regroupement des cultures visant à améliorer le taux de confiance des règles d'association a été démontré dans (Tran, 2017).

Une approche sémantique avec annotation

6. Application

Pour démontrer la faisabilité de notre proposition, nous avons développé un démonstrateur¹⁶. Celui-ci a été appliqué à la base d'assolement de la Zone Atelier Plaine et Val de Sèvre. Nous présentons d'abord la zone atelier et les données, ensuite, nous résumons le résultat obtenu par une méthode de fouille de données et enfin nous décrivons notre annotation des données.

6.1. La base d'assolement

Appartenant au réseau national des Zones Ateliers, la Zone Atelier Plaine et Val-de-Sèvre couvre 450 km² au sud de Niort, Deux-Sèvres. Il s'agit essentiellement d'une plaine de cultures céréalières intensives où les activités d'élevage sont encore présentes mais en forte baisse. Les parcelles agricoles sont encore de taille modeste (4-8 ha) et 15% d'entre elles sont occupées par des prairies.

L'organisation spatiale du paysage évolue dans le temps parce que les agriculteurs modifient l'assolement de leurs parcelles chaque année, mais également recomposent parfois les parcelles entre elles, changeant ainsi leurs formes. Depuis 1994, les occupations du sol sont donc relevées annuellement sur le terrain et numérisées sur les 19 000 parcelles agricoles. Ces données sont centralisées dans une base de données nommée «Assolement».

Dans cette base, une parcelle agricole est considérée comme une unité de gestion, un polygone entouré par des entités ayant différentes cultures au cours des années successives. Une parcelle est délimitée par des limites physiques et ne contient qu'un seul type de culture. Elle diffère de la parcelle cadastrale, mais également des blocs stockés dans le registre parcellaire graphique (RPG).

6.2. Règle d'association sur la succession de cultures

Afin d'examiner les pratiques agricoles, nous avons pris le cas d'étude simple sur la succession de culture de deux années. En formulant une requête SPARQL et en paramétrant l'algorithme *Apriori* avec le *support* = 0.01 (plutôt petit pour réserver les successions rares) et la *confiance* = 0.8, nous avons obtenu une liste des règles ci-dessous (Tableau 1). Les résultats d'analyse détaillés ont été présentés dans (Tran, 2017).

6.3. Annotation des valeurs douteuses

À partir des règles obtenues, on peut examiner des observations (ou transactions) douteuses. On peut identifier et annoter les instances concernées (par

16. Son code source est accessible sur <https://github.com/tbhuy/stRDFMining>

TABLE 1. Les règles d'association trouvées par l'Apriori pour une succession de cultures sur deux années

N°	Culture année T	Culture année T+1	Confiance
1	Urbain	Urbain	0,97
2	Forêt et friches	Forêt et friches	0,91
3	Colza	Blé	0,87
4	Tournesol	Blé	0,82
5	Prairies et luzernes	Prairies et luzernes	0,73

exemple les *timeslices*) afin de les faire ré-examiner et corriger par les experts ou les éliminer dans les futures analyses.

Par exemple, pour la règle Bâti -> Bâti, il existe 9177 transactions qui la satisfassent (98%). On pourrait donc ensuite examiner celles (183 transactions équivalent 2%) qui ne la satisfont pas. Le tableau 2 représente un extrait des observations (ou transactions) ne satisfaisant pas la première règle.

Le *timeslice* correspondant peut être marqué comme douteux. Figure 3 en décrit un exemple dans lequel, une annotation est créée pour associer le *Timeslice/337* à la *Suspicion/01* pour marquer un doute sur la culture quand la règle d'association a été lancée sur une succession de cultures sur deux ans.

TABLE 2. Un extrait des transactions ne satisfaisant pas la règle choisie

N°	Timeslice T	Timeslice T+1	Culture T	Culture T+1
1	Timeslice/209	Timeslice/210	Péri-village	Urbain
2	Timeslice/337	Timeslice/338	Blé	Urbain
3	Timeslice/1585	Timeslice/1586	Blé	Urbain

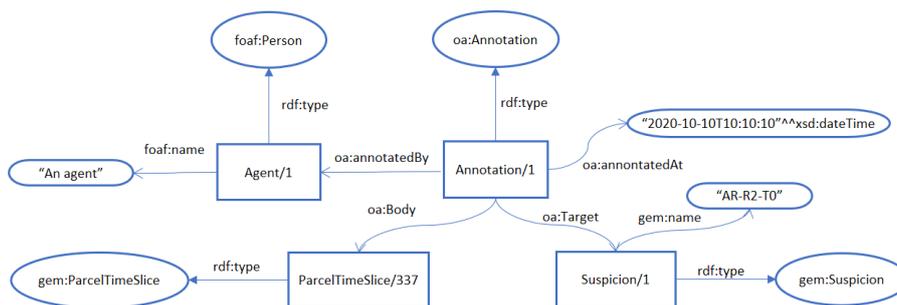


FIGURE 3. Exemple d'une annotation sémantique sur un timeslice.

Une approche sémantique avec annotation

7. Conclusion et perspectives

Nous avons présenté dans l'article une approche sémantique pour l'exploitation de données d'assolement de la zone atelier Plaine de Val de Sèvre. Pour décrire les données sources, nous avons développé une ontologie spatio-temporelle en réutilisant des modèles bien connus. Les processus ETL ont été appliqués pour matérialiser les données dans un triplestore spatial à partir duquel la fouille de données sémantique et l'annotation sémantique de données peuvent être effectuées. Nous pensons appliquer le système sur d'autres jeux de données de la zone atelier. À cet égard, nous souhaitons étendre l'ontologie développée et la rendre la plus générique possible pour que le système s'adapte au mieux à ces sources.

Bibliographie

- Batsakis S., Petrakis E. G. M. (2011). SOWL: A Framework for Handling Spatio-temporal Information in OWL 2.0. In *Proceedings of the 5th international conference on rule-based reasoning. and Applications, RuleML'2011, Programming*.
- Battle R., Kolas D. (2012). Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web*, vol. 3, n° October 2012, p. 355–370.
- Chi Y., Muntz R. R., Nijssen S., Kok J. N. (2004, novembre). Frequent subtree mining - an overview. *Fundam. Inf.*, vol. 66, n° 1–2, p. 161–198.
- Ciccarese P., Ocana M., Castro L. J. G., Das S., Clark T. (2011). An open annotation ontology for science on web 3.0. In *Journal of biomedical semantics*, vol. 2, p. S4.
- Harbelot B., Arenas H., Cruz C. (2013). Continuum: A spatiotemporal data model to represent and qualify filiation relationships. In *Proceedings of the 4th acm sigspatial international workshop on geostreaming*, p. 76–85. New York, NY, USA, ACM.
- Hobbs J. R., Pan F. (2004). An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing*, vol. 3, p. 66–85.
- Ioannidis T., Garbis G., Kyzirakos K., Bereta K., Koubarakis M. (2019). Evaluating geospatial RDF stores using the benchmark geographica 2. *CoRR*, vol. abs/1906.01933.
- Józefowska J., Lawrynowicz A., Lukaszewski T. (2010). The role of semantics in mining frequent patterns from knowledge bases in description logics with rules. *CoRR*, vol. abs/1003.2700.
- Kahan J., Koivunen M.-R., Prud'Hommeaux E., Swick R. (2002). Annotea: an open rdf infrastructure for shared web annotations. *Computer Networks*, vol. 39, n° 5, p. 589 - 608.
- Koubarakis M., Kyzirakos K. (2010). Modeling and querying metadata in the semantic sensor web: The model strdf and the query language stsparql. In *Proceedings of the 7th international conference on the semantic web: Research and applications - volume part i*, p. 425–439. Berlin, Heidelberg, Springer-Verlag.

- Kuramochi M., Karypis G. (2001). Frequent subgraph discovery. In *Proceedings 2001 IEEE international conference on data mining*, p. 313-320.
- Kyzirakos K., Karpathiotakis M., Koubarakis M. (2012). Strabon: A semantic geo-spatial dbms. In *Proceedings of the 11th international conference on the semantic web - volume part i*, p. 295-311. Berlin, Heidelberg, Springer-Verlag.
- Liao Y., Lezoche M., Panetto H., Boudjlida N. (2011, octobre). Semantic Annotation Model Definition for Systems Interoperability. In OTM (Ed.), *OTM 2011 Workshops 2011 - 6th International Workshop on Enterprise Integration, Interoperability and Networking (EI2N)*, vol. LNCS 7046, p. 61-70. Hersonissos, Crete, Greece, Springer-Verlag.
- Meftah W., Bouju A., Malki J. (2012, mars). Une approche ontologique pour la structuration de données spatio-temporelles de trajectoires. Application à l'étude des déplacements de mammifères marins. *Revue Internationale de Géomatique*, vol. 22, n° 1, p. 55-75.
- Nebot V., Berlanga R. (2012, février). Finding association rules in semantic web data. *Know.-Based Syst.*, vol. 25, n° 1, p. 51-62.
- O'Connor M. J., Das A. K. (2011). A method for representing and querying temporal information in owl. In A. Fred, J. Filipe, H. Gamboa (Eds.), *Biomedical engineering systems and technologies*, p. 97-110. Berlin, Heidelberg, Springer Berlin Heidelberg.
- Oren E., Möller K., Scerri S., Handschuh S., Sintek M. (2006). What are semantic annotations. *Relatório técnico. DERI Galway*, vol. 9, p. 62.
- Patroumpas K., Giannopoulos G., Athanasiou S. (2014). Towards geospatial semantic data management: Strengths, weaknesses, and challenges ahead. In *Sigspatial '14: 22nd sigspatial international conference on advances in geographic information systems, dallas, texas, november, 2014*, p. 301-310. New York, USA, ACM.
- Plumejeaud C., Mathian H., Gensel J., Grasland C. (2011). Spatio-temporal analysis of territorial changes from a multi-scale perspective. *International Journal of Geographical Information Science*, vol. 25, n° 10, p. 1597-1612.
- Sanderson R., Ciccarese P., Sompel H. Van de. (2013). Designing the w3c open annotation data model. In *Proceedings of the 5th annual acm web science conference*, p. 366-375. New York, NY, USA, Association for Computing Machinery.
- Schaller N., Lazrak E., Martin P., Mari J.-F., Aubry C., Benoît M. (2012). Combining farmers' decision rules and landscape stochastic regularities for landscape modelling. *Landscape Ecology*, vol. 27, n° 3, p. 433-446.
- Tran B.-H. (2017, novembre). Une approche sémantique pour l'exploitation de données environnementales : application aux données d'un observatoire.
- Welty C., Fikes R. (2006). A reusable ontology for fluents in owl. In *Foils 2006: The 4th international conference on formal ontology in information systems, november 9-11, 2006, baltimore, maryland, usa*, p. 226-236. Amsterdam, NL, IOS Press.

Towards a semantic web representation from a 3D geospatial urban data model

Diego Vinasco-Alvarez¹, John Samuel², Sylvie Servigne³, Gilles Gesquière⁴

1. Univ. Lyon, Université Lumière Lyon 2, LIRIS
5 av. Pierre Mendès-France, 69622, Villeurbanne, France
diego.vinasco-alvarez@univ-lyon2.fr
2. Univ. Lyon, CPE Lyon, LIRIS
43, bd du 11 novembre 1918, 69616, Villeurbanne, France
john.samuel@cpe.fr
3. Univ. Lyon, INSA Lyon, LIRIS
7 Avenue Capelle, 69621, Villeurbanne, France
sylvie.servigne@insa-lyon.fr
4. Univ. Lyon, Université Lumière Lyon 2, LIRIS
5 av. Pierre Mendès-France, 69622, Villeurbanne, France
gilles.gesquiere@univ-lyon2.fr

RESUME. L'urbanisation est un processus d'évolution continue qui est actuellement étudié par nombre de chercheurs. Les modèles d'information de la ville, multi-sources et multidimensionnels, sont souvent utilisés pour comprendre le paysage urbain, qui est en constante évolution. Ces modèles peuvent cependant poser des problèmes d'interopérabilité et de perte de données lors de conversions, souvent nécessaires pour permettre leur intégration. Aujourd'hui, nous pouvons nous baser sur des modèles conceptuels qui peuvent aider à traiter les pertes de données lors de conversions, et peuvent également contribuer à préserver l'interopérabilité des données. Ce type d'approche modèle-centrée permet de définir une représentation commune. Récemment, un mouvement vers la représentation de données basée sur les graphes et la sémantique a gagné en popularité pour répondre aux problèmes d'interopérabilité. Dans un premier temps, pour illustrer notre approche, nous considérons CityGML, un standard de représentation des informations urbaines en 3D. Nous proposons une stratégie pour convertir la sémantique du modèle conceptuel de CityGML vers des ontologies et, ensuite, en formats web sémantiques pour faciliter l'intégration. Nous proposons également une méthode pour convertir et stocker les instances CityGML en individus RDF qui respectent l'ontologie générée. L'approche proposée permet de pallier le déficit d'information sémantique résultant de la traduction directe de différents types de données en graphe, tels que RDF.

ABSTRACT. Urbanization is a continuous evolution process that is currently studied by a number of researchers. Multi-source and multidimensional city information models are often used to

understand the ever-changing urban landscape. These models may encounter issues with interoperability and data-loss if conversion is required for integration. Today we can base ourselves on conceptual models which can help deal with data losses during conversion and may also help preserve data interoperability. This kind of model-driven approach can be useful as common representation. Recently, a movement towards graph and semantic based data representations has also grown in popularity to respond to these issues. As a first step, we consider CityGML, a common standard that can be used to represent 3D urban information. We propose a strategy for converting the semantics of CityGML conceptual model into ontologies and later to semantic web formats to facilitate integration. In addition, we propose a method for converting and storing CityGML instances into RDF individuals that respect the generated ontology. This proposed approach overcomes the loss of semantic information resulting from the direct translation of different types of data into graphs such as RDF.

MOTS-CLES : Modèle conceptuel de données urbaines 3D, Web sémantique, Interopérabilité, CityGML, RDF, GeoSPARQL

KEYWORDS: 3D urban data conceptual models, Semantic Web, Interoperability, CityGML, RDF, GeoSPARQL

1. Introduction

The urban landscape is a complex and heterogeneous source of information. The anthropization and urbanization of modern cities have brought various actors, issues, and systems that all have complex interconnections and dependencies both on local and global scales. The data-driven approaches used to represent and model these urban environments often depend on multi-source and multidimensional urban data, which consists of a variety of different information coming from multiple actors and organizations (Barbosa et al., 2014). For example, city governments have different departments for handling different subjects or issues like water and sewage, pollution, traffic, energy, etc. (Rochet and Pinzon, 2016). Sometimes these organizations release open datasets using their own internally defined business formats or using formats that conform to a specific use-case as opposed to more generic or widely used standards, thus making it difficult to get an integrated view of the data from multiple sources.

In response to the limits of using internal formats, we are now seeing the growing use of geospatial and 2D/3D urban data standards. Different international organizational bodies (OGC, ISO, W3C...) propose standards and release conceptual data models for their standards to ensure that the data is represented uniformly by the users, thereby giving open datasets an aspect of interoperability. Take for example, the CityGML standard¹ released by the Open Geospatial Consortium (OGC), a city

¹ <https://www.ogc.org/standards/citygml>

A semantic web representation of urban data

information model used for 3D urban modeling to represent city objects at different levels of detail.

Another aspect of urban data is the underlying n-dimensional structure. For a long time, 2D maps, aerial views, and cadastral views were used for the management of the urban lifecycle. With the growing availability of 3D data, thanks to new data acquisition processes (LIDAR, photogrammetry...), numerous applications are now possible with 3D urban data (Biljecki 2015). Recently, some urban data models also take into account the changes of cities as they evolve over time in addition to spatial information (Chaturvedi and Kolbe, 2019; Jaillot et al., 2020; Samuel et al., 2020), thereby adding the temporal dimension (4D). These models help historians and city planners visualize and contextualize the impacts previous key projects have had on the development of a city. Such studies require an integration of the urban models with document corpus consisting of project plans, newspaper articles, archives etc. However, it is challenging to integrate data from different sources with different standards for providing an integrated urban view (Biljecki et al., 2018; Ohori et al., 2018).

Semantic web models are now being used increasingly to ensure interoperability among multiple data sources (Malinverni et al., 2020; Nuninger et al., 2020; Tran et al., 2016). Due to its increasing popularity, a number of tools are now available to transform data in legacy formats to semantic web formats but, as said by C. Claramunt in (Claramunt, 2020) “3D models and Building Information Models should be fully integrated.” However, any transformation must conform to the original conceptual model used for developing the standard or risk inconsistent or semantically incomplete representations (Bohring and Auer, 2015).

In this work, we take into consideration a model-centric approach for transformation of urban data into ontological and graph formats. We demonstrate our approach with CityGML, particularly making use of open data from the metropole of Lyon. Our approach is detailed in this article. Section 2 presents the problem statement of model-centric urban data transformation in detail. Section 3 presents the existing state of the art. Taking the case study of CityGML, we detail our approach in Section 4. However, the focus of our work is not limited to transforming CityGML, but to link data from other sources, especially those based on open standards. These perspectives are discussed in section 5. Section 6 concludes the article.

2. Problem Statement

Managing and analyzing urban data is complex because different stakeholders and different entities produce these data, sometimes in silos. City administrators and urban planners often require an integrated view of these data for analyses and for obtaining meaningful insights for daily routine tasks and future planning. Take for example, urban planners who wish to construct a new high-rise building in a particular sector, need to have information on the other structures present around the proposed building. Such information may help them to study the impact on visibility of historical buildings of interest, analyze the reduced or increased sunshine or shadows because

of the new construction. All these studies require data from multiple sources, such as the terrain information, 3D structures of urban objects, position of celestial bodies at different points of times etc.

However, the task of integrating urban data from multiple sources remains very difficult. Additionally, the various standards used to model this data evolve based on the growing requirements of the domain users. New versions of standards need to be released. Sometimes, the data released under previous versions is incompatible with the newer versions. Hence, any semantic web representation solution for an integrated view of urban data must deal with the heterogeneous data from multiple sources as well as their evolution.

Concerning 3D geospatial urban data models, many approaches propose directly translating different data formats into graphs, such as RDF, without taking into account the underlying conceptual model (Bohring and Auer, 2015), and consequently omit semantic information during translation, weakening the initial model. The question is therefore to propose an approach that allows the preservation of the semantics of a 3D urban model when converting into a semantic web representation.

3. State of the art

There are growing efforts to ensure semantic interoperability across heterogeneous data from diverse multiple sources. To obtain an integrated view, one commonly used approach is to transform the data from one data format to a common format by making use of data transformers. Stylesheets like XSLT, JSONT are commonly used for this purpose. However, some of these stylesheets are written on demand basis and may not be available on the internet for use. This means that if the users want to reuse and reproduce some of the results based on the transformed data, they may not be able to achieve it, especially if transformers follow different methodologies. In addition, information loss is another major problem during data transformation (Levina, 2012). It is important to take into consideration the initial conceptual models used for the conception of the modeling.

Several information modeling techniques exist in the literature (Bork et al., 2020). One possible approach is to ensure that a common modeling language is used across domains, like the semantic web ontology (OWL) language. However, UML models are often used by domain experts (De Paepe et al., 2017) to conceive and represent information models instead of semantic web technologies like OWL, RDFS etc. Another possible approach is to build ontologies from existing standards, i.e., transforming models conceived using different modeling languages to a semantic web ontology. Researchers are currently exploring how to make the most of the familiar modeling languages like UML (De Paepe et al., 2017) and automatically generate ontologies in OWL from these models. Works like (Bohring and Auer, 2015; Kramer et al., 2015; Usmani et al., 2020) use similar approaches based on XML Schema as a model towards OWL ontologies.

Another possible solution to this problem is if a standards body releases the ontologies for their conceived models. Take for example, the GML 3.2 geometries

A semantic web representation of urban data

ontology² made available by the OGC. Currently, an official ontology for CityGML 2.0 is missing from the standards body. Nevertheless, it is important to state some works like (Brink et al., 2014; Métral and Falquet, 2018) proposed ontologies for CityGML 2.0. Additionally, there is a GIS working group^{3,4} looking to ensure interoperability with semantic web technologies such as RDF and OWL for future releases of CityGML (Kutzner et al., 2020).

As discussed above, we are looking for integrating heterogeneous data from multiple sources by proposing an approach that preserves the semantics of a 3D urban model when converting it to semantic web representations.

Hence, it is important to have standard procedures or guidelines for ensuring the mapping between different models so that it can be easily reproduced by other users. Our approach is to take into consideration the conceptual model of information related to cities and transform the model using these standard procedures so that the data converted is conformance to the original model and can also be reproduced by other users following these guidelines.

4. Proposed approach illustrated with the CityGML 2.0 Conceptual Model

As previously stated, the information produced by cities and the urban lifecycle can be broken down into hierarchies of concrete and abstract subdomains of information - such as energy, transportation, infrastructure, and many others - and be represented and visualized by 3D urban data models. CityGML is an international OGC standard that is commonly used for this purpose. Taking the CityGML conceptual model as an example, we see frameworks for describing the spatial and thematic information of the more 'material' subdomains within CityGML's modules, e.g. water, buildings, terrain, etc. More immaterial subdomains such as education or public administration are not currently represented by these modules but as a first step the existing conceptual model is more than sufficient to begin transforming open urban data into semantic web formats.

To efficiently take advantage of the benefits of semantic web formats, a direct transformation of CityGML instances is not enough, the CityGML conceptual model itself must also be transformed and be used to 'guide' instance transformations. This results in meaningful RDF graphs that respect the resulting urban data conceptual model.

² http://schemas.opengis.net/gml/3.2.1/gml_32_geometries.rdf

³ <https://github.com/opengeospatial/CityGML-3.0Encodings>

⁴ <https://github.com/opengeospatial/geosemantics-dwg>

4.1. Proposed Approach Design

In order to transform the CityGML conceptual model and CityGML instances into semantic web representations, our proposal uses several XSLT-based transformations based on previous works on XML to RDF, RDFS, and OWL conversion (Bedini et al., 2011; Bohring and Auer, 2015; Brink et al., 2014; Kramer et al., 2015; Métral and Falquet, 2018; Vinasco-Alvarez et al., 2020). The general proposed pipeline is described in Figure 1, which can be broken down into 3 main activities:

- Creation of an urban data model ontology from the CityGML conceptual model (as described by its application schema)
- Generation of a CityGML instance to RDF transformation stylesheet from the CityGML conceptual model (as described by its application schema)
- Transformation of CityGML instances to RDF graphs using the stylesheet generated in step 2.

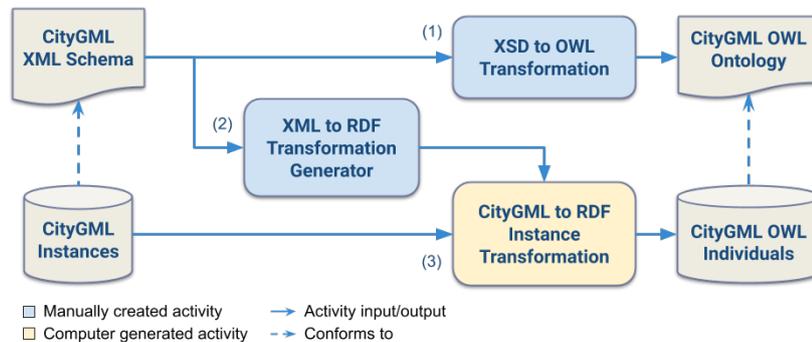


FIGURE 1. Overview of proposed pipeline applied to the CityGML conceptual model and instances

During each transformation, the resulting information must be logically consistent and maintain its interoperability from CityGML and GML. To ensure this, several challenges need to be overcome. For instance, the generated RDF graphs and types must conform to the model described by the ontology. How can mappings be created to ensure this? The CityGML application schema often implements elements that do not have a direct equivalent in OWL or RDFS; how should these elements be represented to best describe CityGML with semantic web formats? In addition, CityGML schema often draws from elements, types, functionality from external schema such as xLinks and xAL addresses. How should these imported elements be addressed to preserve their original functionality? Finally, how can we make use of existing semantic web standards to make geospatial queries? These issues are addressed in sections 4.2 and 4.3.

In addition, to create holistic transformations, the CityGML conceptual model must be assimilated or imported in its entirety into our pipeline. Because the model is

A semantic web representation of urban data

divided into several modules which are sometimes interdependent and dependent on the GML standard, the complete model is defined by 43 different XML schema documents: 29 to define the GML 3.1 standard⁵ and 14 to define the CityGML 2.0 standard⁶ (not including xAL). Schema cannot be processed in isolation or we risk certain loss of data in the transformation (for example in distinguishing when to generate an *owl:ObjectProperty* from an *owl:DatatypeProperty* when the XML element type in question is declared outside the current schema).

Our proposed approach chooses to consolidate these components in a composite XML schema, created from the definitions within each relevant schema in the conceptual model. This prerequisite step also requires that all definitions in the composite schema use normalized namespace prefixes to differentiate between references to one module or another. Once this is complete, the composite schema can be passed into the transformation pipelines.

4.2. Creating a 3D Urban Data Ontology

As an initial approach, we consider that the conceptual model can be expressed by XML schema and in a general simplification, XML schema can use *xs:elements*, *xs:complexType*s, and *xs:simpleTypes* to define the concepts and semantics of domain specific urban. In addition, when creating the transformation from the CityGML conceptual model into an OWL ontology, it is important to reuse as many existing concepts as possible in the semantic web to enrich and render the model interoperable. Using the mapping transformations proposed in the previously mentioned works (Bedini et al., 2011; Bohring and Auer, 2015; Kramer et al., 2015; Vinasco-Alvarez et al., 2020), this can be efficiently achieved.

Once the XML schema is transformed into a domain ontology, we can integrate the GeoSPARQL standard⁷. In GeoSPARQL, the *geo:SpatialObject* class represents any real world object with thematic and geospatial properties. Geometric properties of these classes are represented by *geo:Geometry*, a subclass of *geo:SpatialObject*, and can be serialized by *geo:gmlLiterals*. It is disjoint with the *geo:Feature* class, which represents the thematic properties of a *geo:SpatialObject* and can be linked to related geometric classes via the *geo:hasGeometry* property. As suggested in (Battle and Kolas, 2012), to integrate the thematic classes of our generated ontology with GeoSPARQL, we must declare our classes as a subclass of the *geo:Feature* class. Since our ontology is partially generated from the GML schema, any class which is derived from the *gml:_Feature* type is implicitly declared an *rdfs:subClassOf* *geo:Feature*. This also means that all of the geometry class declarations defined from the GML schema can be removed from the ontology as they are already defined in the

⁵ <http://schemas.opengis.net/gml/3.1.1/>

⁶ <http://schemas.opengis.net/citygml/>

⁷ <https://www.ogc.org/standards/geosparql>

GeoSPARQL standard and by ISO 19136:2007⁸ and references to these classes can be changed to <http://www.opengis.net/ont/gml>.

These proposed transformation strategies to create OWL ontologies from XML schema are an amalgam of previous transformation approaches with several specializations for the CityGML application schema. However, they are largely generalized and can be applied to schema outside of the main GML and CityGML schema, such as the external xAL addressing schema used by CityGML.

4.3. Creating Meaningful Transformations of Urban Data Instances

When generating the CityGML instance to RDF transformation, it is important that the patterns created are general enough to be reused and can take advantage of the OWL/RDFS and GeoSPARQL vocabulary whenever possible. Several techniques suggested in (Brink et al., 2014) are implemented to achieve this. In addition, this process uses the general structure of the approaches proposed in (Bohring and Auer, 2015; Kramer et al., 2015; Vinasco-Alvarez et al., 2020) with consideration taken so that these mappings work with the ontology to be generated alongside the RDF data.

In general, the following mappings in table 1 are applied to the composite schema to generate the CityGML to RDF transformation stylesheet. The generated CityGML to RDF template also uses *gml:id* attributes as *rdf:ID* or *rdf:about* whenever possible, as proposed in (Brink et al., 2014). If no *gml:id* is available, a unique id is generated and appended to the local name of the element. Also as suggested in (Brink et al., 2014), the *rdf:type* of each individual should be generated from its local name in order to reference the generated ontology from section 4.2, or the official GML ontology in the case of geometric elements.

Table 1. General XSD to 'XML to RDF XSLT' mapping transformations

XML Schema Concept	Resulting XML to RDF XSLT
global <i>xs:element</i> elements based on a <i>xs:complexType</i>	<i>xsl:template</i> for generating <i>owl:NamedIndividual</i> of type <i>xs:element</i> that calls the <i>template</i> for its <i>xs:complexType</i>
<i>xs:element</i> which are children of <i>xs:complexType</i> or <i>xs:group</i>	<i>xsl:template</i> for <i>owl:ObjectProperty</i> or <i>owl:DatatypeProperty</i> based on the type of the <i>xs:element</i>
<i>xs:attribute</i> of <i>xs:complexType</i> or <i>xs:attributeGroup</i>	<i>xsl:template</i> for <i>owl:DatatypeProperty</i>
<i>xs:complexType</i> , <i>xs:attributeGroup</i> , and <i>xs:group</i>	<i>xsl:template</i> that calls relevant <i>templates</i> for possible properties or text the element could have

⁸ <https://def.isotc211.org/ontologies/iso19136/>

A semantic web representation of urban data

xs:simpleType with an xs:restriction or xs:extension of a native XML datatype	xsl:template that selects the text of an element
xs:simpleType with an xs:restriction or xs:extension of another xs:simpleType	xsl:template that calls the relevant template for the xs:simpleType it is based on

These transformations also map the geometry of a CityGML feature to an RDF triple using GeoSPARQL's *geo:asGML* datatype property. This is applied whenever a branch of GML elements is detected which has a CityGML parent and is composed of only GML nodes. Such a branch represents a valid *geo:gmlLiteral* if all of its nodes are in the substitution group *gml:_Geometry*. This is ensured by placing the *geo:gmlLiteral* transformation within the template for *gml:AbstractGeometryType*, as any element within the same substitution group as this type can be serialized as a GML literal.

This script also scans the transformed instances for malformed RDF triples and fully qualifies any RDF attributes that contain namespace prefixes such as *rdfs:Resource*, *rdf:type*, and *rdf:about*. These transformations follow the “garbage in, garbage out” concept that poorly formed data input into a program, will produce nonsensical results, and thus assume that the GML and CityGML instance documents provided are well structured and conform to their application schema. If this assumption is met, the resulting data should conform to the ontology transformation discussed in the previous section.

4.4. Leveraging the GeoSPARQL Standard

Once the dataset was generated, a GeoSPARQL endpoint was installed to perform geospatial queries. The Parliament triple-store⁹ was used for this purpose as proposed in (Battle and Kolas, 2012) since it contains a SPARQL endpoint with GeoSPARQL support based on the Apache-Jena libraries¹⁰. We noted that if a particular coordinate reference system (CRS) is used, it must be declared in the literal value as a *gml:srsName* attribute in order for Parliament to parse and index the coordinates and geometry as the default reference system is `<http://www.opengis.net/def/crs/OGC/1.3/CRS84>`. Thus any GML data that uses a different reference system must include it in the GML instances themselves during transformation or manually add it to the output *geo:gmlLiterals* after transformation. This implementation adds these CRS declarations to the data through a Python script after the transformation since they are not declared in the source CityGML data. Through these practices, the geometry of CityGML instances can effectively be generated, stored, and queried through GeoSPARQL.

⁹ <https://github.com/SemWebCentral/parliament>

¹⁰ <https://jena.apache.org>

5. Discussion

The CityGML conceptual model is a useful starting point for creating a 3D urban data ontology as a city information model. Its domain specific, modular structure provides a solid base for describing several hard domains of the urban ecosystem, which can be easily extended and integrated once in ontological form (Métral and Falquet, 2018). Even ADE extensions of CityGML could be applied to this approach to describe existing sub-domains of information if an XML schema is provided - such as the NoiseADE¹¹ provided by the OGC. Another benefit of this modular approach is that the ontology can easily be pruned into domain specific sub-ontologies as each domain can be referenced by a unique URI namespace. In addition, enriching the ontology after transformation with existing geospatial standards is straightforward.

However, a limitation of this approach is its reliance on the semantic limitations of XML Schema. For example, the recently declared ISO 19150-2 standardizes guidelines for UML to OWL transformation of geographic information standards. For GML, (Brink et al., 2014) proposes using Shapechange¹², which takes advantage of this standard, to generate ontological models and acknowledges that ontology generation from UML and XML schema conceptual models are not semantically equivalent. When comparing the resulting ontologies between Shapechange and this proposed approach, still focusing on CityGML, initial explorations indicate that the ISO 19150-2:2015¹³ ontology mappings are more expressive and direct than XML Schema to OWL. For example, an aggregation in UML can be converted into an OWL object property, while XML Schema expresses these entities as complex types which may be converted into extraneous OWL classes and properties to represent the same semantic information. Yet, Shapechange does not facilitate the transformation of CityGML or XML data into graph formats, and thus can only improve the city information model generated by our approach without supplemental transformations.

6. Conclusion

In this article, we proposed a model centric approach towards semantic 3D urban data representations, where we consider that any data transformation from one data format to another format must conform to the equivalent underlying conceptual model. We implemented this proposed approach to convert the CityGML 2.0 conceptual model into an OWL ontology and to convert CityGML 2.0 data to RDF based on the conceptual model. Subsequently, geospatial queries were made on the resulting data using a GeoSPARQL endpoint. This open-source solution is available

¹¹ <http://schemas.opengis.net/citygml/examples/2.0/ade/noise-ade/CityGMLNoiseADE.xsd>

¹² <https://shapechange.net/>

¹³ Geographic information - Ontology - Part 2: Rules for developing ontologies in the Web Ontology Language (OWL), <https://www.iso.org/standard/57466.html>, Publication date: 2015-07

A semantic web representation of urban data

online and has been tested with the data from the 1st borough of Lyon (see the data and code availability statement below). Our future works will explore integrating other information sources - taking into consideration their conceptual models - and integrating a UML based approach. This approach will lead us to facilitate integration of 3D data in city information models and building information models. We are working to test this approach on CityGML 3.0. This will also demonstrate the application of this approach across different versions of standards.

Acknowledgements

The authors would like to thank University Lumière Lyon 2 for funding the PhD Thesis of Diego Vinasco-Alvarez and the LIRIS UMR 5205 Laboratory for funding his internship as part of a transversal effort with the GOAL team. This work has been done within the Vcity project (<https://projet.liris.cnrs.fr/vcity/>). Special thanks to Eric Boix for fruitful discussions and kind help.

Data and code availability statement

Detailed notes and the code for reproducing the results of section 4.2 and 4.3 can be found on software heritage¹⁴ under the folders 'CityGML_Transformations/XSD-to-OWL' and 'CityGML_Transformations/XML-to-RDF' respectively.

Bibliographie et références

- Barbosa, L., Pham, K., Silva, C., Vieira, M.R., Freire, J., 2014. Structured Open Urban Data: Understanding the Landscape. *Big Data* 2, 144–154.
- Battle, R., Kolas, D., 2012. Enabling the geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web* 3, 355–370. <https://doi.org/10.3233/SW-2012-0065>
- Bedini, I., Matheus, C., Patel-Schneider, P.F., Boran, A., Nguyen, B., 2011. Transforming XML Schema to OWL Using Patterns, in: 2011 IEEE Fifth International Conference on Semantic Computing. Presented at the 2011 IEEE Fifth International Conference on Semantic Computing, pp. 102–109.
- Biljecki, F., Kumar, K., Nagel, C., 2018. CityGML Application Domain Extension (ADE): overview of developments. *Open Geospatial Data Softw. Stand.* 3, 13. <https://doi.org/10.1186/s40965-018-0055-6>
- Bohring, H., Auer, S., 2015. Mapping XML to OWL ontologies. Presented at the Leipziger Informatik-Tage (LIT 2005), Gesellschaft für Informatik e. V., pp. 147–156.
- Bork, D., Karagiannis, D., Pittl, B., 2020. A survey of modeling language specification techniques. *Inf. Syst.* 87, 101425. <https://doi.org/10.1016/j.is.2019.101425>
- Brink, L., Janssen, P., Quak, W., Stoter, J., 2014. Linking spatial data: automated conversion of geo-information models and GML data to RDF. *Int. J. Spat. Data Infrastruct. Res.* 9, 59–85. <https://doi.org/10.2902/1725-0463.2014.09.art3>

¹⁴<https://archive.softwareheritage.org/swh:1:dir:7a2459481066a4f77a433ec9a220ddfe5eaf279;origin=https://github.com/VCityTeam/UD-Graph;visit=swh:1:snp:d95e3eaaa6422e77d63f1531be3901e71fe6744b;anchor=swh:1:rev:22b63103c50d21234961d1854c03cb3abd4b2abc/>

- Chaturvedi, K., Kolbe, T., 2019. A Requirement Analysis On Extending Semantic 3d City Models For Supporting Time-Dependent Properties. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. IV-4/W9*, 19–26.
- Claramunt, C., 2020. Ontologies for geospatial information: Progress and challenges ahead. *J. Spat. Inf. Sci.* <https://doi.org/10.5311/JOSIS.2020.20.666>
- De Paepe, D., Thijs, G., Buyle, R., Verborgh, R., Mannens, E., 2017. Automated UML-Based Ontology Generation in OSLO2. pp. 93–97.
- Jaillot, V., Servigne, S., Gesquière, G., 2020. Delivering time-evolving 3D city models for web visualization. *Int. J. Geogr. Inf. Sci.* 25 p. <https://doi.org/10.1080/13658816.2020.1749637>
- Kramer, T.R., Marks, B.H., Schlenoff, C.I., Balakirsky, S.B., Kootbally, Z., Pietromartire, A., 2015. Software Tools for XML to OWL Translation.
- Kutzner, T., Chaturvedi, K., Kolbe, T.H., 2020. CityGML 3.0: New Functions Open Up New Applications. *PFG – J. Photogramm. Remote Sens. Geoinformation Sci.* 88, 43–61.
- Levina, O., 2012. Assessing Information Loss in EPC to BPMN Business Process Model Transformation, in: 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops. Presented at the 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops, pp. 51–55.
- Malinverni, E.S., Naticchia, B., Lerma Garcia, J.L., Gorreja, A., Lopez Uriarte, J., Di Stefano, F., 2020. A semantic graph database for the interoperability of 3D GIS data. *Appl. Geomat.* <https://doi.org/10.1007/s12518-020-00334-3>
- Métral, C., Falquet, G., 2018. Extension And Contextualisation For Linked Semantic 3d Geodata, in: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Presented at the 13th 3D GeoInfo Conference 2018 (Volume XLII-4/W10) - 1-2 October 2018, Delft, The Netherlands, Copernicus GmbH, pp. 113–118.
- Nuninger, L., Verhagen, P., Libourel, T., Opitz, R., Rodier, X., Laplaige, C., Fruchart, C., Leturcq, S., Levoguer, N., 2020. Linking Theories, Past Practices, and Archaeological Remains of Movement through Ontological Reasoning. *Information* 11, 338. <https://doi.org/10.3390/info11060338>
- Ohuri, K., Diakité, A., Krijnen, T., Ledoux, H., Stoter, J., 2018. Processing BIM and GIS Models in Practice: Experiences and Recommendations from a GeoBIM Project in The Netherlands. *ISPRS Int. J. Geo-Inf.* 7, 311. <https://doi.org/10.3390/ijgi7080311>
- Rochet, C., Pinzon, J.D., 2016. Urban lifecycle management: A research program for smart government of smart cities. *Rev. Gest. E Secr.* 7, 1–20. <https://doi.org/10.7769/gesec.v7i2.531>
- Samuel, J., Servigne, S., Gesquière, G., 2020. Representation of concurrent points of view of urban changes for city models. *J. Geogr. Syst.* 22, 335–359. <https://doi.org/10.1007/s10109-020-00319-1>
- Tran, B.-H., Plumejeaud, C., Bouju, A., Bretagnolle, V., 2016. Towards a semantic framework for exploiting heterogeneous environmental data. *Int. J. Metadata Semant. Ontol.* 11. <https://doi.org/10.1504/IJMSO.2016.081586>
- Usmani, A.U., Jadidi, M., Sohn, G., 2020. Automatic Ontology Generation Of BIM And GIS Data, in: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Presented at the XXIV ISPRS Congress, Commission IV (Volume XLIII-B4-2020) - 2020 edition, Copernicus GmbH, pp. 77–80. <https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-77-2020>
- Vinasco-Alvarez, D., Samuel, J.S., Servigne, S., Gesquière, G., 2020. From CityGML to OWL (Technical Report). LIRIS UMR 5205.

Posters



SAGEO 2021, 16th Spatial Analysis and Geomatics Conference

Etude des pratiques de déplacement dans une ville moyenne algérienne : le cas de Mila

Heythem Adjeroud^{1,2}, Adrien Lammoglia³, Laurent Chapelon⁴

*1.Univ Paul Valéry Montpellier 3, LAGAM, F34000, Montpellier, France
haythem@hotmail.es*

2.IGTU, Université Salah Bounider Constantine 3, Algérie

*3.Univ Paul Valéry Montpellier 3, LAGAM, F34000, Montpellier, France
lammoglia.adrien@gmail.com*

*4.Univ Paul Valéry Montpellier 3, LAGAM, F34000, Montpellier, France
lammoglia.adrien@gmail.com*

*MOTS-CLES : Ville moyenne, Algérie, Transport artisanal, Enquête
déplacement, Planification urbaine*

RESUME ETENDU :

Le paysage du transport urbain dans la plupart des villes algériennes s'est transformé après le désengagement de l'état par la libéralisation du marché des transports en 1988 (Boubakour, 2013), puis son réengagement à travers les EPIC dans les années 2000, il est aujourd'hui caractérisé par la coexistence d'un réseau public structurellement organisé, et d'un réseau artisanal auto-organisé (Lammoglia, 2013).

Ce travail fait partie d'une étude portant sur Mila, ville moyenne du nord-est algérien, qui a pour finalité de mettre en exergue des particularités du système de transport très diversifié (Cf. Figure 1) dans cette ville et l'évaluation de sa performance, ainsi que l'identification des enjeux et des perspectives en termes de gouvernance et de planification.

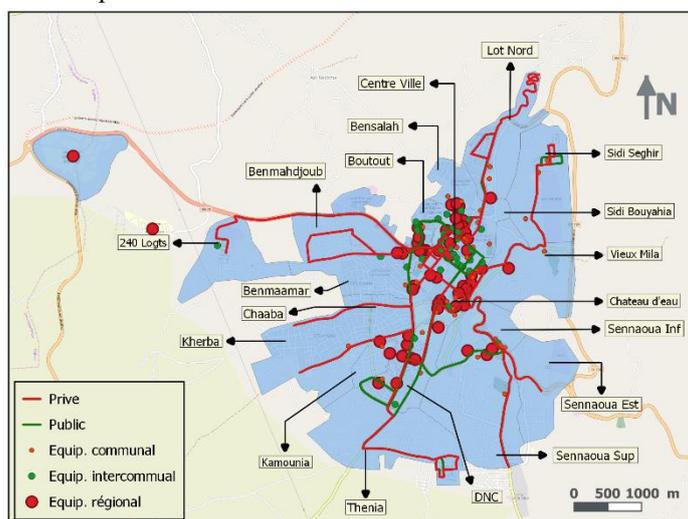


FIGURE 1. Réseaux de transport à Mila et les équipements générateurs de flux (Adjeroud 2020)

Cette partie de l'étude, qui vient à la suite d'une analyse territoriale préalable, sert à la construction d'une base de données géolocalisées pour la ville de Mila, se trouvant dans un contexte où l'accès à la donnée et à l'information est plutôt compliqué, et cherche à déterminer, de manière localisée les besoins de mobilité et de déplacement dans la ville, ainsi qu'à évaluer les niveaux de satisfaction des usagers par rapport au service fourni, et leur connaissance de l'offre.

Ceci est concrétisé à partir d'une enquête réalisée auprès de 923 citoyens, à travers un échantillonnage empirique par quotas, obtenu après avoir réalisé des estimations de la démographie des différents quartiers à l'aide de chiffres officiels (très incomplets), de travail de terrain et d'exploitation des images satellitaires.

Les niveaux de mobilité, de connaissance et de satisfaction, ainsi que la répartition des flux et le taux d'usage des transports en commun montrent les limites du système

Etude des pratiques de déplacement dans une ville moyenne algérienne : le cas de Mila

de transport de la ville, constitué d'un réseau privé artisanal dense, à caractère atomisé (Godard, 2008), et relativement flexible, régi par des logiques de rentabilité, et d'un réseau public récent et peu développé, qui peine à s'élargir.

In fine, malgré l'existence d'une complémentarité, notamment sur le plan spatial, ce double fonctionnement assure une desserte et une qualité de service qui n'est pas en adéquation avec les besoins des usagers. Il semble indispensable de repenser la forme de coexistence public/privé et les modes de gouvernance des transports urbains à Mila.

Bibliographie :

Adeniji K. (1987). Para-transit modes in Nigeria: problems and prospects. *Cities* n° 4 (4), 290-392.

Adjeroud H., Lammoglia A., Chapelon L., & Kebiche A., (2019). Offre de transport et besoins de mobilité à Mila : les particularités d'une ville moyenne algérienne. *Géotransports*, n°10, p. 21-39.

Bavoux J.J., Beaucire F., Chapelon L., Zembri P., (2005). *Géographie des transports*, Armand Colin, 232 p.

Boubakour F., et Bencherif H., (2013). Évolution du transport urbain en Algérie : du mode artisanal à la régulation par l'État, *Géotransports. Transport et développement des territoires*, (1-2), p. 91-104.

Boubakour F., Merzaga A., (2006). Planification du transport urbain et développement durable : Quelques réflexions à propos de la ville de Batna (Algérie). *Revue Transports*, n° 437. Éditions techniques et économiques. Paris.

Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques. (2013). Les enquêtes déplacements « standard CERTU », Note de présentation.

Cervero R., Golub A., (2007). Informal transport : A global perspective. *Transport Policy*, n°14 (6), p. 445-457.

EMA/BETUR & CNEAP (2004). Enquête ménages sur les déplacements dans l'agglomération d'Alger.

Godard X. (2002). *Les transports et la ville en Afrique au sud du Sahara : le temps de la débrouille et du désordre inventif*, KARTHALA Editions.

Godard X., (2008). *Transport artisanal, esquisse de bilan pour la mobilité durable*, Codatu Hô Chi Minh Ville (Vietnam).

Lammoglia A. (2013). *Analyse et modélisation multi-agents de transports flexibles. Comparaison de services français et sénégalais*. Thèse de doctorat, Université d'Avignon.

Observations d'évolutions socio- écosystémiques en haute montagne : l'apport de la géomatique au dispositif Refuges Sentinelles

**Balzarini Raffaella¹, Ployon Estelle¹, Marcuzzi Mélanie¹,
Bourdeau Philippe¹**

*1. Laboratoire Pacte, Univ. Grenoble Alpes, CNRS, Sciences Po Grenoble, 38000
Grenoble, France*

raffaella.balzarini@univ-grenoble-alpes.fr

estelle.ployon@univ-grenoble-alpes.fr

melanie.marcuzzi@umrpacte.fr

philippe.bourdeau@univ-grenoble-alpes.fr

*MOTS-CLES : RECHERCHE-ACTION, SOCIO-ECOSYSTEME, MONTAGNE PEU AMENAGEE, REFUGES,
PRATIQUES SPORTIVES ET TOURISTIQUES, TRANSITION, INNOVATION TERRITORIALE, OBSERVATIONS IN
SITU, BASE DE DONNEES SPATIALISEE, SIG, CARTOGRAPHIE INTERACTIVE, GEO-VISUALISATION.*

Les refuges de montagne constituent des observatoires privilégiés des effets croisés des changements environnementaux et culturels. Au-delà de leur fonction première d'hébergement, ils deviennent des laboratoires récréatifs dans lesquels des innovations de niche contribuent à l'émergence de modèles de développement, en phase avec les enjeux de diversification et de transition du tourisme dans « la montagne peu aménagée » (Bourdeau *et al.*, 2019).

Or, face à un déficit de connaissances majeur sur la fréquentation et la gouvernance de la montagne « peu aménagée », un programme de recherche-action « Refuges Sentinelles (RS) », a été développé depuis 2017, par le Parc National des Écrins (PNE) et le LabEx Innovation et Transitions Territoriales en Montagne de l'Université Grenoble Alpes (UGA), en partenariat avec l'Agence Française de la Biodiversité, dans le cadre du programme Sentinelles des Alpes (Girard, 2019). Ce dernier vise à co-construire une vision alpine des relations climat-homme-biodiversités sur la base de suivis sur le long-terme. Dans ce contexte, RS met au cœur de son suivi les transformations environnementales et culturelles dont les refuges sont témoins. À l'échelle d'un panel de 16 refuges du PNE destiné à couvrir des configurations diversifiées en matière d'altitude, de processus naturels et de contextes touristiques, les principaux axes de recherche de RS portent sur les fréquentations et pratiques, en lien avec les métiers touristiques et sportifs, la météorologie, la biodiversité, la géomorphologie, les risques et la sécurité (Mourey et Ravel, 2017a). L'objet d'étude est « le pratiquant » qui est observé à travers la mise en place de protocoles adaptés au contexte de la haute montagne déconnectée (Bourdeau *et al.*, 2018). Les protocoles comportent la collecte de données quantitatives (i.e. : nombre de nuitées par saison dans un refuge) et qualitatives (i.e. : profils sociodémographiques issus d'enquêtes, questionnaires), d'observations in situ (i.e. : carnet de récit), de photo-constats, d'ateliers collaboratifs.

Si la démarche scientifique implique l'ensemble des parties prenantes professionnelles et territoriales concernées (fédérations sportives, syndicats professionnels, offices du tourisme), qui participent à l'orientation du programme, les gardien.ne.s, de refuge sont les protagonistes du dispositif car le déploiement des protocoles s'appuie sur leur implication (Bourdeau, 2018). L'exploitation des données et les indicateurs de fréquentation qui en découlent (i.e. : quota de remplissage des refuges, nombre de pratiquants par randonnée et/ou par sommet à partir d'un refuge, % des typologies des profils des pratiquants) sont restitués aux gardiens à la fin de chaque saison lors du débriefing interprofessionnel qui permet de mutualiser les observations et de croiser le regard des chercheurs et des parties prenantes.

RS est caractérisé par une configuration éparse, dans l'espace et dans le temps : des refuges localisés dans des zones très diverses, des saisonnalités différenciées, ce qui implique des protocoles de suivi qui produisent des données multi-source, multi-format et multi-échelle. Ceci explique l'importante approche exploratoire, nécessaire à la validation de la méthodologie qui a abouti au suivi saisonnier des flux de fréquentations et de destinations (présences dans les refuges des pratiquants et leurs randonnées effectuées) et au suivi pluriannuel (pas de temps de 4 ans) des typologies des usagers (des refuges).

Observations d'évolutions socio-écologiques en haute montagne

Initialement fortement exploratoire, le principal défi du dispositif Refuges Sentinelles a été de passer d'une logique d'expérimentations de terrain, ponctuelles et locales, à la co-construction d'une méthodologie d'observation pérenne et reproductible dans d'autres massifs. C'est grâce au rôle central de la géomatique que ce défi a pu être relevé. La mise en place d'une base de données (PostGres-PostGis) et d'un atlas cartographique interactif (ArcGIS Online) a permis de mettre en cohérence les données hétérogènes pour qu'elles puissent être corrélées aux autres infrastructures de suivi existantes, notamment celles de Sentinelles des Alpes.

La logique 'géomatique' a eu un impact conceptuel et structurel dans le dispositif RS. Conceptuellement, elle a amené à redéfinir les questions de recherche autour de la dimension géographique de l'objet 'refuge' considéré comme 'point' cartographique, camp de base et laboratoire avancé d'observations, mais aussi comme catalyseur des flux et d'interactions (les itinéraires empruntés, les zones de divagations, les croisements homme-biodiversité, etc..) et part d'un maillage territorial d'hébergements semblables, à l'échelle d'un massif de montagne. La logique 'géomatique' a été révélatrice de l'importance d'une approche spatiale des protocoles d'observation. Structurellement, la base de données et l'atlas interactif de RS ont permis de cadrer les nouveaux protocoles de terrain en standardisant des modèles descriptifs des données, de faciliter l'interopérabilité des données multi-sources et de fédérer les différentes disciplines qui se croisent dans son contexte.

Notre poster présentera le processus de cette démarche d'intelligence collective entre chercheurs et praticiens, que Refuges Sentinelles a initié dans la durée ; il mettra en évidence comment les outils de la géomatique qui ont supporté cette action, induisent une valeur ajoutée informationnelle et stratégique à l'échelle du dispositif, en contribuant à la co-construction, participative et pédagogique, de connaissances. Particulièrement, le poster montrera comment les outils de la géomatique ont joué un rôle cognitif dans la visualisation d'informations agrégées (Balzarini *et al.*, 2018) permettant aux gardiens de mieux comprendre comment leur refuge contribue et s'adapte aux changements, à l'échelle locale et du massif.

Bibliographie non numérotée et références

Balzarini R., Ployon E., Marcuzzi M. (2018). La montagne en été : une approche empirique pour étudier spatialement le changement des pratiques, *Conférence SIG ESRI 2018*, Paris.

Bourdeau Ph., Marcuzzi M., Dentant C., Charron J., Bonet R. (2019). Mountain huts: laboratories for environmental and cultural change, *International Mountain Conference*, Innsbruck.

Bourdeau Ph. (2018). Guides et gardiens de refuge, pionniers de l'adaptation ! In Impacts du changement climatique et transition(s) dans les Alpes du Sud, *Les cahiers du GREC-SUD*, édités par l'Association pour l'innovation et la recherche au service du climat (AIR), p. 36.

Girard V. (2019). "Recherche action" Sentinelles des Alpes, des scientifiques impliqués sur le terrain. *Montagne Magazine*, Numéro Spécial #469 – Octobre 2019.

Mourey J. et Ravanel L. (2017a). Évolution des itinéraires d'accès aux refuges du bassin de la Mer de Glace (massif du Mont Blanc, France). *Revue de Géographie Alpine*.

Apport des images ALSAT-2A pour la cartographie de l'occupation des sols - le cas de la ville d'Oran

Rabia Sarah Chériguène^{1,2}, Anne Puissant¹ et Habib Mahi²

1. Laboratoire Image, Ville et environnement, UMR 7362 CNRS / Université de Strasbourg, 3 rue de l'Argonne, 67000, Strasbourg, France.

2. Centre des Techniques Spatiales, 1 Avenue de la Palestine, 31200, Arzew, Algérie.

rabia-sarah.cheriquene@live-cnrs.unistra.fr

MOTS-CLES : OCCUPATION DES SOLS, HAUTE RÉOLUTION SPATIALE, CLASSIFICATION SUPERVISÉE, ALSAT-2A.

La croissance démographique rapide et incontrôlée ainsi que le développement économique et industriel, en particulier dans les pays en développement, ont multiplié à plusieurs reprises le taux de changement d'utilisation des sols/couverture des sols. (Talukdar et al., 2020), la ville d'Oran fait partie des espaces dans lesquels la problématique du développement démographique est très pertinente dans la mesure où les atteintes à l'environnement y sont généralisées. L'objectif de ce travail est d'exploiter les images ALSAT-2A et de cartographier par une approche supervisée les modes d'occupations des sols afin de quantifier les changements entre les années 2010 et 2018. Différents tests ont été effectués sur l'image 2018, d'une part en faisant varier les données en entrée de la classification et d'autre part en comparant deux algorithmes d'apprentissage supervisé : (1) les Forêts Aléatoires (RF) et les Supports à Vecteur de Machines (SVM) introduites par (Birmann, 2001) et (Vapnick, 1995). Les résultats de classification sont évalués à l'aide de différentes métriques quantitatives et qualitatives. Les mêmes tests seront appliqués sur l'image de 2010.

1. Données utilisées et prétraitements

Les images satellites utilisées sont des images ALSAT-2A couvrant la région d'Oran de l'année 2018, avec une résolution spatiale de 10m en mode multispectrale (Bleu, Vert, Rouge et Proche Infrarouge) et de 2.5 m de résolution spatiale en mode panchromatique. Les images sont corrigées au niveau 2A et ont été orthorectifiées. Le choix de l'agglomération d'Oran comme terrain de recherche a été fait sur la base d'une connaissance préalable du terrain ainsi que sur la disponibilité des images.

L'enjeu de cette étude est de discriminer les différents modes d'occupation des sols à 10 m de résolution spatiale à partir des bandes spectrales et de différents indices spectraux de végétation (NDVI), de brillance (IB) ou texturaux. L'intérêt de l'utilisation des indices de ce type pour caractériser l'état des sols particulier dans les zones urbaines, a été démontré dans plusieurs travaux (Schaaf et al., 2003 ; Pelletier, 2017) L'analyse de texture est fondée sur deux paramètres l'homogénéité et l'entropie, calculés à partir de l'indice NDVI sur une fenêtre de 5 x 5 en raison de la résolution spatiale de 10 m (Puissant et al., 2005). Une Analyse en Composantes Principales (ACP) a été aussi appliquée sur les 7 bandes /indices afin de réduire la corrélation statistique entre les bandes.

2. Classification supervisée, protocoles de test et résultats

La première étape consiste à définir la typologie des classes thématiques à extraire. Huit classes ont été définies basées sur les connaissances de terrain et des échantillons ont été sélectionnés par photo-interprétation pour les classes d'eau (en 3 classes – Mer, Etang et Sebkha), la forêt, le jachère, le maraîchage, la céréaliculture, la végétation, le bâtis/sol nu ainsi que la classe route issue des données OpenStreetmap. Une analyse de la séparabilité de ces échantillons a permis d'évaluer la qualité de notre jeu de données destiné à l'entraînement et à la validation des modèles de classification et de vérifier leur pertinence. Dans un second temps, les deux algorithmes RF et SVM ont été appliqués sur les résultats de l'ACP. Dans le premier test, la première CP est utilisée en entrée de la classification avec un pourcentage de valeurs propres de 80%. Les deuxièmes et troisièmes tests ont consisté à prendre en entrée de la classification, les deux premières, puis les trois premières avec un taux de 16% et 4%. La CP1 est composée des bandes proche infrarouge, rouge, vert et de l'indice de brillance, la CP2 est composée de la bande proche infrarouge et la CP3 est composée de la bande vert et rouge.

Les résultats ont montré que le RF donne de meilleurs résultats (93.22%) par rapport au SVM (77.16%). La performance globale de la classification en appliquant l'algorithme RF est de 92.98% sur les deux composantes comparativement 93.22% lorsque la classification utilise les trois composantes en entrée. (Table 1).

TABLE 1. Résultats de classification RF

	CP1	CP1+CP2	CP1+CP2+CP3
Précision globale	66.96%	92.98%	93.22%
F-score	0.67	0.93	0.93

3. Conclusion et perspectives

Les résultats de différents tests ont montré que le RF donne de bons résultats en utilisant trois composantes principales 93.22% à partir de 7 bandes spectrales/indices. Cette méthode sera appliquée sur l'image de 2010 et les changements d'occupation des sols seront analysés et présentés dans le poster proposé. La méthode sera également appliquée sur notre deuxième ville test

Apport des images ALSAT-2A pour la cartographie

Tlemcen, en utilisant le même modèle d'apprentissage, afin de vérifier la transférabilité de la méthode sur une ville aux caractéristiques morphologiques assez proche.

Références

- Breiman L. (2001). *Random Forests*. Machine Learning, 45, 5-32.
- Pelletier C. (2017). *Cartographie de l'occupation des sols à partir de séries temporelles d'images satellitaires à hautes résolutions*. Thèse de doctorat, Université de Toulouse.
- Puissant A., Hirsh J., Weber C., (2005). *The utility of texture analysis to improve per pixel classification for high spatial resolution imagery*. International Journal of Remote Sensing, 26 (4), pp. 733-745.
- Schaaf C. Strahler A. Hodges J. Gao F. Reed B. Huete A. Zhang X., Friedl M. *Monitoring vegetation phenology using modis*. Remote Sensing of Environment.84, 471-475, 2003.
- Talukdar S., Singha p., Mahato S., Shahfahad., Pal S., Liou Y., Rahman A., (2020). *Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations*, Remote sensing review, 12, 1135; doi:10.3390/rs12071135.
- Vapnick V N. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag New York.

Détecter le processus de dépoldérisation par l'observation satellite à très haute résolution temporelle

Antoine COLLIN^{1,2}, Tony DUROZIER³, Gwenal HERVOUET⁴, Stéphane COSTA⁵

1. EPHE-PSL Université, CNRS LETG

15 boulevard de la mer, 35800 Dinard, France

antoine.collin@ephe.psl.eu

2. LabEx CORAIL

58 avenue Paul Alduy, 66860 Perpignan, France

3. Conservatoire du Littoral et des Rivages Lacustres – Délégation

Normandie 5 avenue Tsukuba, 14200 Hérouville-Saint-Clair, France

t.durozier@conservatoire-du-littoral.fr

4. Conservatoire du Littoral et des Rivages Lacustres – Délégation Bretagne

8, quai Gabriel Péri, 22194 Plérin, France

g.hervouet@conservatoire-du-littoral.fr

5. Université Caen Normandie, CNRS LETG

Esplanade de la Paix, 14 000, Caen, France

stephane.costa@unicaen.fr

MOTS-CLES : HAUSSE DU NIVEAU MARIN, STRATEGIE DE REcul, DEPOLDERISATION, LANCIEUX, SATELLITE TRES HAUTE RESOLUTION TEMPORELLE, PLANETSCOPE.

RESUME ETENDU.

L'interface littorale est le siège de processus complexes, où se combinent les hausses du niveau marin, de la démographie et des enjeux socio-économiques. L'augmentation de ces facteurs contribue à l'aggravation du risque de submersion marine. Plusieurs stratégies existent pour lutter contre ce risque : la plus résiliente est le recul des enjeux, de la mer vers la terre. Ainsi, des zones terrestres gagnées sur la mer au fil des siècles font désormais l'objet de réflexions de dépoldérisation (Bazin et Olivri, 2017). Le Conservatoire du littoral anime le projet « Adapto » en accompagnant sur 10 sites naturels, dont la baie de Lancieux, une gestion souple du trait de côte, pouvant aboutir à des recompositions territoriales significatives. En mars 2020, ce site breton, incluant un polder doté d'une digue érigée au XVIIIème siècle, a connu un basculement notable dans sa gestion avec l'apparition d'une brèche dans l'ouvrage, laissant les hautes mers de vive eau pénétrer sur les parcelles poldérisées, acquises par le Conservatoire.

Suivie depuis quatre ans, la baie ne bénéficie pas d'une couverture spatiale continue. En effet, les photographies au sol, dotées d'une résolution millimétrique, ne permettent pas d'embrasser la dynamique du site. D'autre part, le suivi par imagerie aérienne ou satellitaire traditionnelle, pourvue d'une résolution, respectivement, centimétrique et décimétrique, est entaché par une faible résolution temporelle, empêchant l'acquisition durant les marées de vive eau, (Collin *et al.*, 2017). Ce travail propose le premier suivi de dépoldérisation par l'utilisation d'imagerie satellitaire à très haute résolution temporelle.

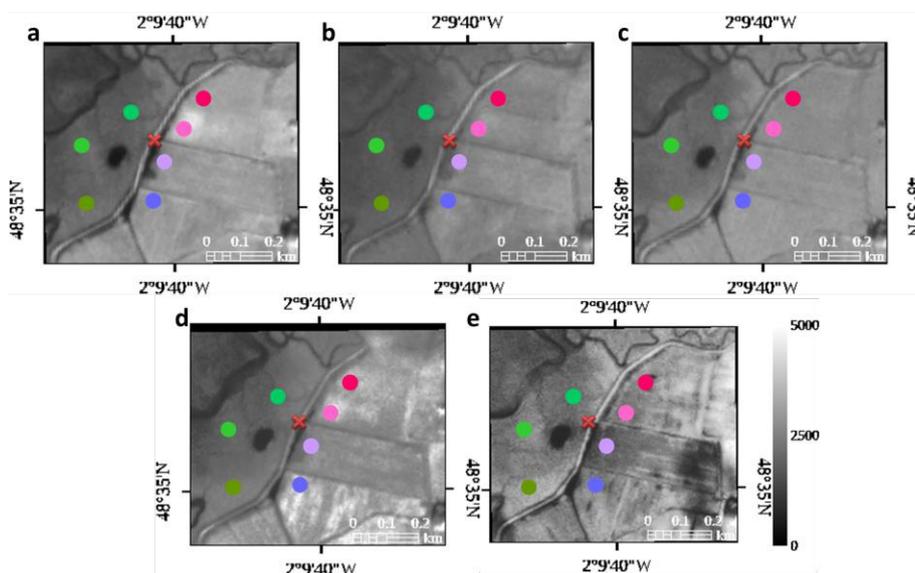


FIGURE 1. Image de réflectance infrarouge du site d'étude pour octobre (a) 2016, (b) 2017, (c) 2018, (d) 2019, et (e) 2020. Croix : brèche ; cercles : stations suivies.

Dépolderisation par satellite THR temporelle

La partie méridionale du polder de la Baie de Lancieux a ainsi été suivie après chaque grande marée d'équinoxe, lors de la première quinzaine d'octobre, entre 2016 et 2020 (Figure 1). L'imagerie infrarouge PlanetScope, dotée d'une résolution de 3 m, a été géométriquement et radiométriquement corrigée, résultant en une réflectance de surface comparable pour les cinq années. La variance élevée de la réflectance infrarouge du site d'étude (Figure 2a) indique une évolution notable de la station du polder au nord-est de la brèche (cercle rose) en 3 phases : fort déclin de 2016 à 2017, plateau de 2017 à 2019, et faible déclin de 2019 à 2020 (Figure 2b).

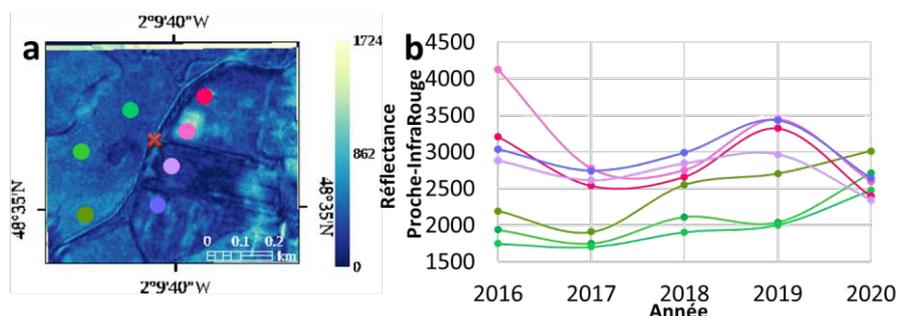


FIGURE 2. Variance spatiale (a) et évolution pluriannuelle (b) de la réflectance infrarouge sur sept stations d'intérêt, dont la rose est la plus proche de la brèche.

Cette tendance de réduction de la réflectance infrarouge témoigne de l'augmentation de la saturation en eau (Collin *et al.*, 2019) de la station, avant que la brèche ne soit apparue en mars 2020. Il est important de souligner que la brèche a accentué ce processus. La variabilité spatio-temporelle de cette dépolderisation fera l'objet d'analyses via PlanetScope sur d'autres stations d'intérêt, comme les zones d'accumulations sédimentaires et de colonisation des communautés d'habitats intertidaux.

La très haute résolution temporelle, idoine à la constellation PlanetScope, atteint la journée. Couvrant toutes les surfaces terrestres à la résolution de 3 m, cette constellation civile est la seule à offrir cette résolution temporelle, permettant de répondre adéquatement aux exigences de la saisonnalité de la variation de hauteurs d'eau dans des sites mégatidaux. Cet apport technologique est aussi très prometteur pour d'autres suivis littoraux d'intérêt, comme l'évolution des falaises, dunes, prés salés, mangroves, herbiers marins, algueiraie, ou encore récifs coralliens.

Collin A., Dubois S., James D., Houet T. (2019). Improving intertidal reef mapping using UAV surface, red edge, and near-infrared data. *Drones*, vol. 3, n° 3, p. 67-.

Collin A., Etienne S., Feunteun E. (2017). VHR coastal bathymetry using WorldView-3 : colour versus learner. *Remote Sensing Letters*, vol. 8, n° 11, p. 1072-1081.

Bazin P., Olivri D. (2017). Le Conservatoire du littoral: un acteur face au changement climatique. *Sciences Eaux Territoires*, vol. 3, p. 50-53.

Détermination semi-automatique de l'occupation du sol grâce au logiciel GOURAMIC : évaluation des performances sur données historiques et opérateur-dépendance

Matthieu DUBUIS¹², Rémi RATAJCZAK³, Lény GRASSOT¹²,
Laure TOUGNE³, Béatrice FERVERS¹²

1. Centre Léon Bérard, Département Prévention Cancer Environnement – Lyon

2. INSERM UA8 Radiations : Défense, Santé, Environnement – Lyon

3. CNRS UMR 5205 LIRIS, Université Lumière Lyon 2 – Lyon

MOTS-CLÉS : OCCUPATION DU SOL, PRISE DE VUES AÉRIENNES HISTORIQUES, DÉTERMINATION SEMI-AUTOMATIQUE, LOGICIEL SIG

1. Introduction

L'étude ci-présentée s'intègre au sein du volet environnement de « l'étude prospective multicentrique de type cas-témoins visant à étudier l'association entre expositions environnementales – domestiques – professionnelles et le risque de tumeur germinale du testicule à l'âge adulte » dit TESTIS (Centre Léon Bérard / OMS CIRC / UMR LIRIS). Elle a pour objectif l'analyse de l'association entre différentes expositions, au premier rang desquelles les pesticides, tout au long de la vie et particulièrement pendant les périodes de sensibilité critique du développement du testicule et le risque de développer un cancer à l'âge adulte. Le logiciel GOURAMIC (1) intervient dans le volet évaluation des expositions environnementales afin de déterminer l'occupation du sol autour d'une adresse à une date donnée. Il doit permettre de reconstituer tout au long des parcours de vie des sujets de l'étude, les occupations du sol environnant les lieux de résidence successivement habités.

L'objectif ici est d'évaluer les performances du logiciel GOURAMIC tant dans la finesse spatiale de ses résultats que dans la variabilité inhérente à une utilisation qui en est faite par différents utilisateurs.

I – la première analyse consiste à comparer les performances du logiciel GOURAMIC au regard d'autres méthodes de détermination de l'occupation du sol. Cet aspect vise à comparer les résultats traitement avec les différentes sources disponibles d'occupation du sol sur des périodes anciennes à l'échelle du territoire national ou d'espaces particuliers.

II – la seconde analyse consiste à mesurer la variabilité induite par les utilisateurs du logiciel GOURAMIC : quantifier les différences d'interprétation et donc des résultats spatialisés induites par le traitement par les différents opérateurs d'un même espace dans GOURAMIC.

2. MÉTHODOLOGIE

Plusieurs contraintes ont été identifiées : ancienneté des données, non disponibilité d'une vérité terrain, diversité des sources et des méthodes. Considérant ces difficultés, nous avons intégré les différentes sources disponibles à une date relativement ancienne et commune : Corine Land Cover (COPERNICUS, éd.1990), HILDA (HISTORIC Land Dynamics Assessment, éd.1990), OCS Picardie (édition 1994). Nous avons de plus, ajouté une version automatisée de GOURAMIC – appelée DeepGOURAMIC – utilisant un réseau de neurones pour reconstituer l'occupation du sol sans intervention opérateur à partir des jeux de données générées par les utilisateurs sur plus de 1000 images. Se rajoutent enfin les résultats issus de GOURAMIC de différents utilisateurs ainsi que des images dites de référence travaillées manuellement par un utilisateur expérimenté.

Pour toutes les sources, un travail préliminaire d'harmonisation des items d'occupation du sol a été fait pour agencer 7 classes distinctes : Arable – Eau – Forêt – Prairie – Urbain – Verger – Vignoble. L'interprétation des espaces en eau étant difficile sur les photographies aériennes, notamment dans GOURAMIC et deepGOURAMIC, et le fait que cette classe ne soit pas stratégique dans l'étude des pesticides (non ciblée par l'épandage) il a été décidé de la retirer de la comparaison, à l'aide d'un masque (*BDTopo* hydrologie de l'IGN) pour l'ensemble des couches étudiées.

L'étude croisée des résultats s'intéressera à l'analyse des accords entre sources ou utilisateurs grâce aux kappa de Cohen. Pour ce faire, toutes les couches résultats ont été intersectées pour découper l'ensemble des territoires analysés en polygones disposant d'une unique occupation du sol par source ou utilisateur.

La performance de GOURAMIC au regard des autres sources utilisera comme jeu de données initial 206 zones tampon d'environ 7 km² (GOURAMIC – deepGOURAMIC – HILDA – CLC) ; et à cause du périmètre très restreint de cette couche locale seulement 4 avec la source OCS Picardie.

La variabilité inter-opérateur aura pour données les résultats issus d'un test d'utilisation par différents opérateurs, préalablement formés sur le logiciel et devant

Détermination semi-automatique de l'occupation du sol grâce au logiciel GOURAMIC

9 images classée par difficulté d'interprétation. Ces résultats sont croisés avec les images de référence : l'analyse des écarts permettant dès lors de quantifier l'impact de l'interprétation dans le processus de détermination de l'occupation du sol.

2. RÉSULTATS

2.1. Comparer les performances du logiciel GOURAMIC

Les matrices de confusion nous ont permis de comprendre quelles étaient les catégories les plus difficiles à repérer et sur lesquelles un travail supplémentaire doit être réalisé. Ainsi, HILDA présente de nombreuses difficultés dues notamment à sa précision (mailles très importantes) ; GOURAMIC et deepGOURAMIC disposent de bonnes tendances par rapport à CLC mais divergent sur certaines catégories (vergers et vignobles) où la finesse de CLC (25Ha minimum) tend à limiter la détection de ces parcelles habituellement de faible surface.

2.2. Mesurer la variabilité induite par les utilisateurs

Comme pressenti, les résultats confirment la faible variabilité générale à l'échelle d'un territoire mais des écarts néanmoins forts pour certaines classes d'occupation du sol induite par des interprétations différentes au sein d'une même image par différents utilisateurs. Ainsi, les catégories *Prairies* et *Arable* sont souvent confondues, qui plus est dans les tests sur les images disposant de tous les types de surface et donc où la place à l'interprétation est la plus importante.

(1) Ratajczak R. (2020). *Analyse automatique d'images aériennes historiques : application à une étude épidémiologique*. Thèse de doctorat en Informatique, École doctorale en informatique de Lyon.

Application d'une approche semi-quantitative pour estimer les impacts cumulés des interventions anthropiques sur les systèmes fluviaux : l'exemple de la Charente entre Angoulême et Saintes (Ouest de la France)

Amélie Duquesne¹, Christine Plumejeaud-Perreau¹, Jean-Michel Carozza¹

*1. Littoral Environnement et Sociétés (LIENSs), La Rochelle Université
2 rue Olympe de Gouges, 17000 La Rochelle, France
amelie.duquesne1@univ-lr.fr*

MOTS-CLES : ANALYSE MULTICRITERE, ANALYSE SPATIALE, SYSTEME D'INFORMATION GEOGRAPHIQUE, CARTOGRAPHIE D'IMPACTS ANTHROPIQUES, RESTAURATION DES COURS D'EAU.

La quantification des impacts anthropiques cumulés sur les milieux « naturels » a fait l'objet de travaux en particulier pour les milieux estuariens, côtiers et marins (e.g., Halpern *et al.*, 2007 ; Pereira *et al.*, 2019). Ces approches restent fortement discutées en raison des limites intrinsèques liées à la méthode employée (Toro *et al.*, 2010).

Alors que l'anthropisation des cours d'eau est une thématique en vogue (Brown *et al.*, 2018), ils n'ont pas fait l'objet de tentatives d'application de ces approches. Ils ont en effet été sujets à un aménagement massif pour la navigation commerciale, la force motrice de l'eau ou encore la subsistance des populations. Certains de ces usages sont à l'heure actuelle sur le déclin ou à l'arrêt, mais les paysages fluviaux en conservent les traces et impacts. A cela s'ajoute, de nouveaux usages (navigation de plaisance, activités récréatives et sportives de plein air, etc) dont les impacts s'inscrivent dans un milieu déjà très anthropisé.

De ce fait, la Directive cadre européenne sur l'eau déclare que 52% des cours d'eau français sont sujets à une dégradation générale de leur état écologique en 2009. La pollution des eaux, la faible continuité écologique et sédimentaire et les altérations morphologiques et hydrologiques sont les principaux obstacles à l'atteinte du bon état écologique des masses d'eau. Les gestionnaires de cours d'eau sont alors à l'origine de projets de restauration dont l'objectif est de rétablir les fonctionnalités des milieux fluviaux ou de les recréer, ou de reconstituer un état antérieur de l'écosystème fluvial dans une optique de gestion durable. Cependant, ces pratiques de restauration se heurtent à un déficit de connaissances liées aux incertitudes sur le fonctionnement écologique, à la faible présence de programmes interdisciplinaires et au manque de retours d'expérience. Les actions de restauration écologique sont priorisées par les acteurs car elles sont plus faciles à mettre en œuvre que des actions de restauration hydromorphologique ou sur la qualité de l'eau (logistique, financement, accès au foncier, etc).

Le fleuve Charente ne fait pas exception. L'état des lieux initial du bassin versant de la Charente effectué dans le cadre de l'élaboration du Schéma Directeur d'Aménagement et de Gestion des Eaux (SDAGE 2010-2015) conclue à une dégradation générale de l'ensemble des cours d'eau charentais. Seulement 14% des masses d'eau superficielles du périmètre du Schéma d'Aménagement et de Gestion des Eaux (SAGE) de la Charente sont déclarées en bon état écologique contre 47% pour le bassin Adour-Garonne. A l'échelle même de la Charente, 18% des masses d'eau sont en bon état écologique dans la partie en amont et 6% dans la partie aval. Les principales causes de dégradation de l'état écologique du cours d'eau sont la pollution des eaux, la faible continuité écologique et sédimentaire en raison de la présence de multiples ouvrages en travers, l'altération du régime hydrologique et la prolifération des espèces invasives tant animales que végétales. Elles sont liées aux interventions humaines directes et indirectes sur le système fluvial. Il est par conséquent fondamental de proposer aux gestionnaires de cours d'eau une synthèse des connaissances sur l'organisation spatiale et l'intensité des usages anciens et

Impacts anthropiques cumulés sur le fleuve Charente

récents et de leurs impacts cumulés afin de les aider dans la mise en œuvre d'actions adaptées et durables de gestion et de restauration sur la Charente.

Ce travail propose alors d'analyser et de quantifier les impacts cumulés des interventions anthropiques sur le fleuve Charente entre Angoulême et Saintes sur la période 1700-2020. Un inventaire des interventions humaines a été réalisé sur la période d'étude à partir de données géohistoriques récupérées auprès des Archives départementales de Charente-Maritime et de Charente et municipales de Saintes et Cognac, et intégré dans une base de données. Les interventions humaines ont ensuite été appréhendées à partir d'une approche globale semi-quantitative et spatialisée. Elle s'appuie sur le calcul d'un indice d'impacts cumulés unitaire estimé pour chaque intervention humaine (ex : dragage...), et d'un indice d'impacts cumulés total évalué pour chaque type d'intervention humaine (ex : entretien du lit...). Le modèle d'impacts cumulés est estimé via une méthode multicritère basée sur une série d'attributs qualitatifs (zone impactée, impact spatialisé, impact fonctionnel...) déterminés par dire d'expert.e.s. Les données obtenues sont spatialisées sur un maillage couvrant 241 km² disposé sur l'ensemble de la plaine alluviale afin d'évaluer les impacts anthropiques selon un double gradient longitudinal et latéral.

Les résultats permettent 1) de caractériser et d'analyser l'organisation spatiale et l'intensité de l'anthropisation sur le fleuve Charente ; 2) de proposer une hiérarchisation des interventions humaines les plus impactantes sur le système fluvial et 3) de cartographier les impacts cumulés des interventions humaines et leurs principaux effets sur le fonctionnement du cours d'eau. Au final, ce travail de recherche au caractère exploratoire contribue à mieux identifier l'organisation et l'intensité des impacts anthropiques sur le système fluvial, et à mieux prioriser les enjeux de restauration et le type d'action à mener sur la Charente.

Bibliographie

- Brown A. G., Lespez L., Sear D. A., Macaire J.-J., Houben P., Klimek K., Brazier R. E., Van Oost K., Pears B. (2018). Natural vs anthropogenic streams in Europe : History, ecology and implications for restoration, river-rewilding and riverine ecosystem services. *Earth-Science Reviews*, vol. 180, p. 185-205.
- Halpern B. S., Selkoe K. A., Micheli F., Kappel C. V. (2007). Evaluating and Ranking the Vulnerability of Global Marine Ecosystems to Anthropogenic Threats. *Conservation Biology*, vol. 21, n°5, p. 1301-1315.
- Pereira C. I., Madrid D. A., Correa I. D., Pranzini E., Botero C. M. (2019). An evaluation of human interventions in the anthropogenically disturbed Caribbean Coast of Colombia. *Anthropocene*, vol. 27, p. 1-11.
- Toro J., Requena I., Zamorano M. (2010). Environmental impact assessment in Colombia : Critical analysis and proposals for improvement. *Environmental Impact Assessment Review*, vol. 30, p. 247-261.

Résolution de toponymes par apprentissage profond à partir de cooccurrences et de relations spatiales

Jacques Fize¹, Ludovic Moncla¹, Bruno Martins²

1. LIRIS UMR 5205, INSA Lyon, France

jacques.fize@gmail.com, ludovic.moncla@insa-lyon.fr

2. Instituto Superior Técnico - INESC-ID, University of Lisbon, Portugal

bruno.g.martins@tecnico.ulisboa.pt

1. Introduction

En traitement automatique du langage, le processus de geoparsing qui permet d'identifier les noms de lieux (toponymes) dans un texte et de leur associer des coordonnées géographiques non-ambiguës (latitude-longitude) (Gritta *et al.*, 2018) est une tâche complexe qui s'effectue en deux étapes : (1) l'identification des toponymes (ou geotagging) et (2) la résolution des toponymes (ou geocoding). La plupart des approches de résolution de toponymes utilisent des index géographiques (i.e. *gazetteer*) associant à un nom de lieu plusieurs données : alias, coordonnées géographiques, catégorie (e.g. ville, pays), etc. Plusieurs types d'ambiguïtés rendent complexe la tâche de résolution des toponymes (i) plusieurs noms peuvent exister pour un même lieu, (ii) plusieurs lieux peuvent avoir le même nom, (iii) un même nom peut faire référence à un lieu ou à un autre type d'entité. Par ailleurs, pour diverses raisons, certains lieux existants ne sont pas répertoriés et si les gazetteers peuvent être suffisants pour l'analyse de documents contemporains, ce n'est pas le cas pour des documents plus anciens ou historiques. Pour contourner ces problèmes, des travaux récents (DeLozier *et al.*, 2015 ; Cardoso *et al.*, 2019) proposent d'entraîner des modèles de *geocoding* afin de ne pas avoir à utiliser de gazetteers lors de l'étape de résolution des toponymes. Ces approches sont basées sur des techniques d'apprentissage profond avec en entrée les toponymes et leur contexte (e.g. mots cooccurrents).

2. Méthodologie

Dans ce travail, nous proposons une architecture¹ basée sur des réseaux de neurones récurrents pour l'entraînement d'un modèle de résolution des toponymes (Fig. 1). Cette architecture prend en entrée des paires de toponymes et les coordonnées géographiques du premier. Le modèle entraîné a pour rôle de prédire la localisation d'un premier toponyme auquel nous associons un deuxième toponyme comme contexte. La phase d'entraînement du modèle se décompose en trois étapes. La première consiste à transformer les toponymes en séquences de n-grammes de caractères (e.g. Paris → par, ari, ris). Cela permet au modèle de s'adapter à différentes variations des noms de lieux ainsi qu'à la régionalité de certains affixes (préfixe, suffixe)². L'ensemble des n-grammes est ensuite encodé selon un embedding pré-entraîné avec la méthode Word2vec. Les vecteurs obtenus pour chaque n-gramme des toponymes d'entrée sont donnés en entrée de deux réseaux Bi-LSTM (Bidirectional Long Short-Term Memory) (Graves, Schmidhuber, 2005) chargés de l'extraction des caractéristiques puis l'entraînement pour la prédiction des coordonnées est réalisé par des réseaux de neurones.

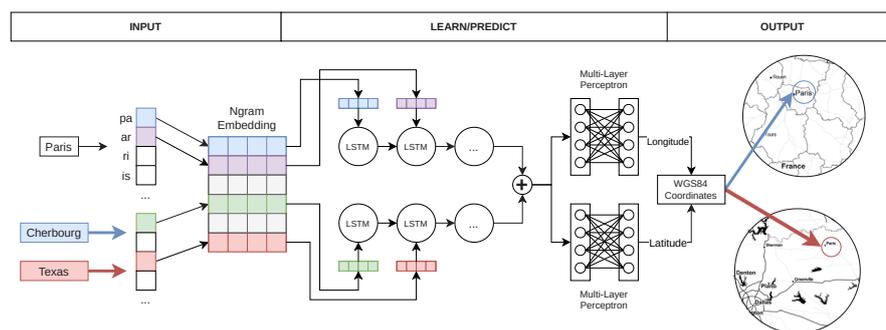


FIGURE 1. Architecture pour l'entraînement d'un modèle de geocoding.

Dans le cadre de nos expérimentations nous avons considéré trois manières différentes de constituer les paires de toponymes : la première basée sur les cooccurrences de toponymes dans des textes et les deux autres basés sur leurs relations spatiales (relations de proximité et relations d'inclusion). Pour le contexte "cooccurrence", nous générons les paires de toponymes à partir des liens exprimés entre les noms de lieux dans les articles Wikipedia. Pour le contexte 'spatial' nous construisons les paires de toponymes à partir de relations d'inclusion et de proximité calculées à partir des données de Geonames. L'inclusion

1. Une implémentation en Python de ce modèle est disponible en ligne : <https://git.liris.cnrs.fr/HExtGEO/toponym-geocoding>

2. <https://jacobe2169.github.io/mapthetoponymsim/>

Résolution de toponymes par apprentissage

indique que le premier toponyme est contenue dans le second et la proximité est calculée selon une distance maximale entre les deux lieux (100km dans le cadre de nos expérimentations). Pour éviter des temps de calcul trop élevés, les relations de proximité et de cooccurrences sont échantillonnées aléatoirement.

3. Expérimentations et résultats

Les trois contextes décrits dans la section précédente constituent trois jeux de données pouvant être combinés (cooccurrences, proximité et inclusion) pour l'entraînement d'un modèle. Nous avons constitué des jeux de données sur différentes zones géographiques. Nous utilisons la mesure d'exactitude (avec une précision de 100km) et obtenons les résultats suivants : 0.79 pour la France, 0.88 pour la Grande-Bretagne, 0.21 pour les Etats-Unis, et 0.90 pour Texas + Île de France. Par ailleurs, nous observons que la combinaison des différents contextes (spatiaux et textuels) permet d'améliorer les résultats et que les cooccurrences de toponymes jouent un rôle très important pour cette tâche de résolution des toponymes. Nous observons des scores très faibles pour les États-Unis contrairement aux autres zones géographiques qui peuvent s'expliquer par une toponymie particulière sur ce territoire. Ces premiers résultats sont encourageants et plusieurs pistes, telle que l'utilisation de modèles de langue pré-entraînés (e.g. BERT), sont envisagées pour étendre la couverture et améliorer l'exactitude du modèle.

4. Remerciements

Ce travail a été réalisé grâce au soutien financier du Projet IDEXLYON de l'Université de Lyon dans le cadre du programme Investissement d'Avenir (ANR-16-IDEX-0005).

Bibliographie

- Cardoso A. B., Martins B., Estima J. (2019). Using Recurrent Neural Networks for Toponym Resolution in Text. In *Proceedings of the EPIA Conference on Artificial Intelligence*, p. 769–780. Springer International Publishing.
- DeLozier G., Baldrige J., London L. (2015). Gazetteer-Independent Toponym Resolution Using Geographic Word Profiles. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.
- Graves A., Schmidhuber J. (2005). Framewise phoneme classification with bidirectional LSTM networks. In *Proceedings of the IEEE International Joint Conference on Neural Networks.*, vol. 4, p. 2047-2052 vol. 4.
- Gritta M., Pilehvar M. T., Limsopatham N., Collier N. (2018). What's missing in geographical parsing? *Language Resources and Evaluation*, vol. 52, n° 2, p. 603–623.

FarmSIG : une application web pour l'analyse de l'évolution des espaces agricoles en lien avec les politiques publiques environnementales

Follin Jean-Michel¹, Voniarinjanahary Rovaso¹, Thommeret Nathalie¹, Fournier Marie¹

1. Laboratoire Géomatique et Foncier, Cnam-ESGT

1 boulevard Pythagore, 72000 Le Mans

jean-michel.follin@lecnam.net

RESUME. Le projet FARMaine, mené dans le cadre du programme national PSDR4 a pour objectif d'éclairer les modalités de mise en œuvre des Politiques Publiques d'Environnement (PPE) dans les territoires de fonds de vallée du bassin versant de la Maine, et leurs conséquences sur le fonctionnement des exploitations agricoles et la gestion du foncier agricole. Pour comprendre les caractéristiques foncières, spatiales et agricoles des fonds de vallée étudiés ainsi que leurs évolutions en lien avec les PPE, des analyses spatiales et statistiques ont été menées. Elles combinent des données abordant différentes thématiques sur la période 2009 à 2018 : le RPG niveau 2, les Mesures Agro-Environnementales et Climatiques (MAEC), et les Fichiers fonciers.

Un web-SIG baptisé FarmSIG, encore en cours de développement, permet de partager et diffuser les données rassemblées et produites, en cohérence avec la démarche partenariale du projet impliquant des chercheurs et acteurs des territoires de fond de vallée.

MOTS-CLES : WEB MAPPING, DATAVIZ, MESURES AGROENVIRONNEMENTALES, RPG, FICHIERS FONCIERS, ANONYMISATION PAR CARROYAGE, ANALYSE STATISTIQUE

Problématique et motivations

La question du suivi des territoires agricoles où sont mises en œuvre des politiques publiques d'environnement (PPE), en particulier les conséquences de celles-ci sur le fonctionnement des exploitations agricoles (EA) et la gestion du foncier agricole, est primordiale pour les acteurs publics qui en ont la charge.

Le projet FARMaine (Foncier et Régulations dans le bassin de la Maine) mené dans le cadre du programme national PSDR4 (Pour et Sur le Développement Régional) décliné à l'échelle du Grand Ouest, s'est fixé pour objectif d'éclairer les modalités de mise en œuvre des PPE dans les territoires de fonds de vallée du bassin versant de la Maine. Quatre territoires d'étude ont été sélectionnés pour mener nos études en fonction de leurs caractéristiques (enjeux en matière d'environnement, existence de PPE depuis plusieurs années, variété dans les systèmes agricoles observés) : les Basses Vallées Angevines, les Alpes Mancelles, la Vallée du Loir et le bassin versant de l'Oudon.

Dans ce cadre, différents traitements (combinaisons, analyses statistiques et anonymisations) ont été menés sur des données spatiales portant sur différentes dates (entre 2009 et 2018), thématiques (foncier, agricole, réglementaire) et échelles (bassin versant et territoires d'études). Ils poursuivent un double objectif qui est d'une part, d'assurer la confidentialité des données produites avant diffusion et d'autre part d'apporter des éléments de réponse aux questions suivantes : observe-t-on des évolutions du profil agricole des zones étudiées ? peut-on établir un lien entre structures spatiales des EA et présence de PPE ?

Un web-SIG, baptisé FarmSIG, a été développé pour rassembler et diffuser les données et analyses produites dans le cadre du projet FARMaine. Il vise à proposer une vision cartographique synthétique des zones de fonds de vallée étudiées et des clés pour comprendre leurs caractéristiques et leurs évolutions en lien avec les PPE. FarmSIG s'adresse à des profils variés (exploitants agricoles, agents de Chambres d'Agriculture, chercheurs en agronomie, aménagement et géographie).

Données, méthodes et résultats

Les principales données mobilisées sont le Registre Parcellaire Graphique (RPG) niveau 2 qui comporte, pour chaque îlot de parcelles agricoles, l'identifiant de son EA (numéro PACAGE), les Mesures Agro-Environnementales et Climatiques (MAEC) qui associent à chaque parcelle les dispositifs MAEC adoptés par son EA, et les Fichiers fonciers contenant des informations sur les propriétaires des parcelles.

Dans une première analyse diachronique, les changements ont été abordés en croisant les Fichiers fonciers avec le RPG. Cela a abouti à la construction d'une typologie des évolutions foncières des parcelles agricoles dans les Basses Vallées Angevines entre 2009 et 2016. Celle-ci a été réalisée par ACM et classification ascendante hiérarchique. Trois catégories ont été mises en évidence : 1) les parcelles agricoles qui ont fait l'objet d'un engagement en MAEC sans changement lié au propriétaire et ont évolué vers une exonération Natura 2000, 2) les parcelles agricoles

sans engagement MAEC qui ont conservé une exonération des terres agricoles, et 3) les parcelles forestières qui ont connu un changement lié à leurs propriétaires.

Puis une analyse synchronique a été menée en confrontant, pour 2015 et 2018,

- l'intensité de l'adoption des MAEC par EA, calculée en cumulant, après les avoir normalisés, les nombres, longueurs et surfaces totales de MAEC ponctuelles, linéaires et surfaciques par EA,

- et l'organisation spatiale des parcellaires agricoles des EA, appréhendée par les indicateurs de dispersion et de morcellement (Latruffe et Piet, 2014 ; Renard, 1972).

Ces informations ont été rapportées aux centroïdes des exploitations (obtenus par le barycentre géométrique des îlots de même exploitation). Elles ont ensuite été anonymisées par la méthode de carroyage (agrégation dans des grilles de carreaux) pour rendre impossible l'identification de données relatives à une EA (Loonis, 2018).

L'application FarmSIG (<http://farmaine.fr/index.php/farmsig/>), bâtie sur les bibliothèques javascript Leaflet et Chart.JS, propose une restitution des analyses réalisées avec des fonctionnalités i) de navigation thématique et temporelle dans les données cartographiques aux deux échelles du bassin versant et des territoires d'étude ; et ii) d'exploration des données via un panneau latéral proposant des éléments synthétiques caractérisant les secteurs sélectionnés.

Conclusion et perspectives

FarmSIG est en cours de développement mais permet déjà une exploration temporelle et thématique des territoires étudiés, proposant des éléments de réponse aux questions posées sur les liens entre PPE et caractéristiques agricoles et foncières de ces secteurs. Une restitution 3D est explorée afin de mieux appréhender certaines informations comme les liens entre morcellement ou dispersion des exploitations (perçu via la couleur) et intensité d'adoption des MAEC (perçue via la hauteur). Par ailleurs FarmSIG a vocation à être alimenté avec d'autres résultats d'analyses qui restent à mener comme celles portant sur les dynamiques d'évolutions des EA.

Bibliographie

- Renard M. (1972). Recherches méthodologiques sur le degré d'émiettement parcellaire des exploitations agricoles des bocages de l'ouest. *Bulletin de l'Association de géographes français*, n° 397, p. 83-94.
- Puech T., Durpoix A., Barataud F., & Mignolet C. (2020). Une méthode pour caractériser l'organisation spatiale des parcellaires d'exploitations agricoles. *Cybergeo: Revue européenne de géographie/European journal of geography*, Document 938.
- Latruffe L., Piet L. (2014), "Does land fragmentation affect farm performance? A case study from Brittany, France", *Agricultural Systems*, vol. 129, p. 68-80.
- Loonis V. (dir), Bellefon M.P. (coord) (2018), *Manuel d'analyse spatiale. Théorie et mise en oeuvre pratique avec R*, Insee Méthodes n° 131, 406 p.

Étude spatio-temporelle d'un écosystème littoral : le pré-salé de Sainte-Anne en baie du Mont-Saint-Michel

Dorothee JAMES¹, Antoine COLLIN^{1,2}, Mathilde LETARD¹

1. EPHE, PSL Université Paris, CNRS UMR 6554 LETG
15 Boulevard de la mer, 35800 Dinard, France
dorothee.james@ephe.psl.eu

2. LabEx CORAIL, Moorea, French Polynesia

MOTS-CLES : MOSAÏQUE, METASHAPE, EVOLUTION, PRE-SALE, CLASSIFICATION, SVM

Face aux changements globaux, l'étude des milieux et des paysages est essentielle pour comprendre leur fonctionnement. L'environnement littoral est sujet à une modification intense aussi bien par l'activité anthropique que par les processus météo-marins. Afin de comprendre les trajectoires socio-écologiques, l'étude rétrospective du milieu littoral permet d'appréhender les mécanismes pour mener de manière éclairée une gestion prospective.

Les données historiques, qui nous renseignent sur l'état du littoral à un moment précis, sont issues de diverses sources telles que les photographies aériennes ou les cartes anciennes. Ces données sont donc définies par des caractéristiques spatiales, spectrales, radiométriques et temporelles spécifiques à chaque campagne aérienne, ce qui pose problème pour établir une série temporelle cohérente.

Cette étude vise, dans un premier temps, à mettre en place une méthodologie innovante pour l'utilisation et la standardisation de données anciennes massives (à Très Haute Résolution, THR, et à une échelle régionale) issues de sources diverses (Figure 1). Dans un second temps, grâce la classification supervisée, les cartes créées permettent d'obtenir une comparaison de l'évolution de l'occupation sol à partir d'une classe d'intérêt (pré-salé) ayant une large emprise et à THR.

Notre site d'étude, composé d'un vaste pré-salé évoluant le long de la digue Saint-Anne, se localise dans le golfe normano-breton, entre la chapelle Sainte-Anne (48.60°N ; 1.67°W) et le Mont-Saint-Michel (48.63°N ; 1.51°W). Les spartines, l'obione, la salicorne sont les principales communautés végétales qui composent ce pré-salé. Au-devant de celui-ci, s'étend une vaste vasière composée de sédiment vaso-sableux.

Les données historiques sont au cœur de cette étude. Fastidieux, le téléchargement des images aériennes est la première étape de ce processus. Ainsi, 9 images ont été retenues pour 1966, 19 pour 1979, 6 pour 1995 et 4 pour 2014. Les images ont ensuite été découpées automatiquement *via* XnConvert, puis mosaïquées, pour chaque millésime, par l'intermédiaire du logiciel de photogrammétrie Metashape (Rault *et al.*, 2020). À partir de cette compilation d'images, le géoréférencement en RGF93/Lambert93 a été réalisé à l'aide du logiciel SIG ArcGIS. L'Ortholittorale V2 (2014) a servi de référence. La création de filtres texturaux de co-occurrence (moyenne, contraste, homogénéité, variance, dissimilarité, second moment, corrélation et entropie) a été implémentée sur chacune des reconstructions photogrammétriques panchromatiques (logiciel ENVI). *In fine*, un algorithme de classification supervisée, *Support Vector Machine* (SVM) a été appliqué sur les 8 bandes de textures. Deux classes ont été identifiées : le pré-salé et la vasière, avec 15000 pixels de validation et 15000 pixels de calibration, extraits aléatoirement.

Étude spatio-temporelle d'un écosystème littoral

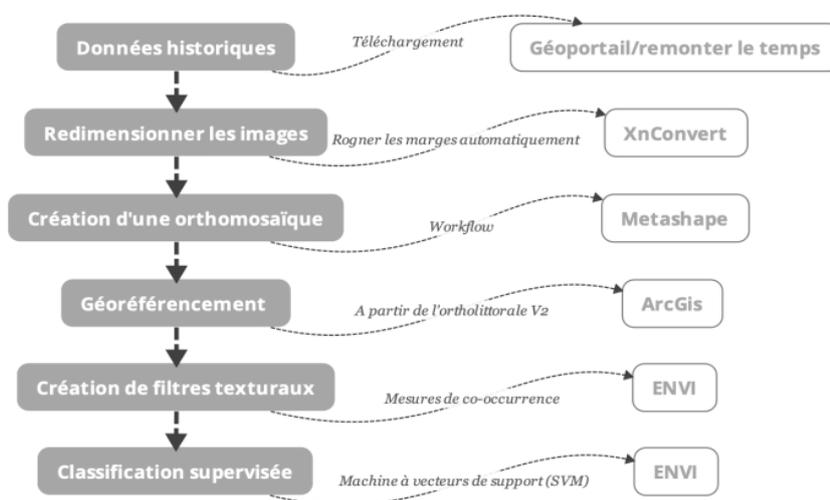


FIGURE 1. Processus méthodologique pour l'étude rétrospective du littoral.

Les résultats de classification permettent d'analyser la superficie de l'occupation du sol pour les années 1966, 1979, 1995 et 2014. Ainsi, respectivement, l'enveloppe du pré-salé était de 9,92 km², de 12,69 km², de 9,69 km² et 15,19 km². La saisonnalité, la composition floristique du pré-salé (Valéry et Radureau, 2014) et le développement des activités anthropiques ont contribué à la vitesse de progradation du pré-salé d'environ 1,1% par an en 48 ans. L'utilisation d'algorithme de *deep learning* (Letard *et al.*, 2020) permettrait d'affiner la détection de la composition du pré-salé.

Bibliographie

- Letard, M., Collin, A., et Pastol, Y., (2020). Improving mangrove and salt marsh mapping using superpixel-based deep classification of VHR superspectral satellite imagery and topobathymetric LiDAR data, *Sea Tech Week 2020, 12-15 Oct. 2020, Brest*.
- Rault, C., Dewez, T., et Aunay, B., (2020). Structure-from-motion processing of aerial archive photographs: sensitivity analyses pave the way for quantifying geomorphological changes since 1978 in la réunion island. *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 773-780.
- Valéry, A., Radureau, A., (2014). *Evolution de la végétation des marais salés de la baie du Mont-Saint-Michel-Analyse cartographique 1984-2013*. Rapport Convention CREBS/Université de Rennes1/Conservatoire du Littoral/Inter-SAGE Baie du Mont-Saint-Michel/Agence de l'Eau Seine-Normandie/Agence de l'eau Loire-Bretagne ; Rennes, France, 15p.

Adaptation de domaine par transport optimal pour la cartographie à large échelle des peupleraies

Yusra Hamrouni¹, Olivia Bernardoff¹, Nicolas Karasiak¹, David Sheeren¹

*1. Université de Toulouse, INRAE, UMR DYNAMFOR, Castanet-Tolosan, France
yusra.hamrouni@inrae.fr david.sheeren@ensat.fr*

MOTS-CLES : TELEDETECTION, TRANSPORT OPTIMAL, ADAPTATION DE DOMAINE, TRANSFERT SPATIAL, PEUPLERAIES, CLASSIFICATION, LARGE ÉCHELLE, SENTINEL-2

La cartographie des peuplements forestiers et leur composition en essences est une donnée incontournable pour les gestionnaires réalisant des opérations sylvicoles. Bien qu'il existe des bases de données forestières à l'échelle nationale, les connaissances sur ces essences sont encore lacunaires en raison, soit d'une caractérisation peu exhaustive des peuplements, soit d'une marge d'erreur importante ou d'une mise à jour trop longue. En particulier, les estimations de surfaces des peupleraies françaises s'avèrent peu précises en raison de leur court cycle de rotation (15 ans en moyenne).

La disponibilité d'images à haute résolution spatiale, spectrale et temporelle provenant des nouveaux satellites Sentinel-2 a offert de nouvelles opportunités pour le suivi des plantations de peupliers sur de grandes étendues. Plusieurs travaux ont démontré le potentiel de la télédétection pour la cartographie des plantations telles que le caoutchouc et le palmier à huile mais seulement quelques-uns se sont intéressés au peuplier. Malgré les résultats prometteurs rapportés, la plupart des études se limitent à des zones d'études relativement restreintes (petites forêts connues) et les performances des modèles de classification dépendent largement des données utilisées pour l'entraînement et la validation des algorithmes. En effet, lorsque ces modèles, calibrés sur une zone, sont appliqués sur de vastes étendues, possiblement très éloignées de la zone initiale, les performances prédictives s'effondrent. Les modèles locaux sont souvent inadaptés au transfert d'une région à l'autre (*transfert spatial*) car ils ne tiennent pas compte de toute la diversité des spectres de réflectance des classes et de leur caractère non stationnaire dans l'espace (problème du "*data shift*" en apprentissage automatique). La variabilité spatiale est souvent due à des variations dans les conditions d'acquisition des images, dans la phénologie de la végétation ou dans les conditions stationnelles. Les techniques d'adaptation de domaine et en particulier le transport optimal (TO) sont proposées dans cette étude pour répondre au problème de transfert évoqué. Elles permettent d'adapter les distributions des données de départ sur lesquelles des connaissances sont disponibles (domaine source) à des données d'arrivée pour lesquelles il y a peu ou pas d'informations (domaine cible). L'hypothèse qui sous-tend le transport optimal est que si les données sources subissent une phase d'adaptation pour ressembler davantage aux données cibles, le *data shift* est réduit, et un classifieur entraîné sur les données sources adaptées, donne de meilleures performances qu'un classifieur s'appuyant uniquement sur les données du domaine source ou encore, sur un nombre restreint d'échantillons disponibles dans le domaine cible (Courty et al., 2017).

Le transport optimal a été appliqué entre deux tuiles Sentinel-2 de 100 km² chacune. Elles sont situées dans deux bassins populicoles contrastés dans le Nord-Est (T31UEQ) et le Sud-Ouest (T31TCJ) de la France. Les séries temporelles Sentinel-2 datent de 2017 et comportent 34 dates ré-échantillonnées au même pas temporel et dans lesquelles les pixels nuageux ont été corrigés. Le transport optimal a été appliqué depuis la tuile Sud-Ouest (source) vers la tuile Nord-Est (cible) selon quatre stratégies : TO non supervisé (utilise les distributions de la source et de la cible), TO non supervisé régularisé (avec la contrainte de l'entropie), TO non supervisé avec les étiquettes de la source et TO semi-supervisé (utilise les distributions et les étiquettes disponibles de la source et de la cible). Pour chacune des stratégies, les résultats de

Transport optimal sur les peupleraies

classification avec TO ont été comparés à un modèle entraîné sur la source non adaptée. *Random Forest* a été utilisé comme algorithme de classification et trois classes ont été considérées : peupleraies, robiniers et chênes avec, 3850 échantillons étiquetés par classe pour la source et 1250 pour la cible. Les paramètres optimaux du TO (coefficients de régularisation par l'entropie et par classe) et du classifieur (nombre d'arbre, profondeur maximale) ont été fixés à partir d'une grille de recherche.

Les résultats obtenus montrent que l'adaptation de la distribution du domaine source par TO permet d'améliorer la précision globale de la classification de 3 à 35% et la précision producteur (mesure de rappel) des peupleraies de 8 à 27% (TABLE 1). L'amélioration est particulièrement marquée lorsque des étiquettes de la source et de la cible ont été incorporées. Il est important de noter que dans la dernière stratégie nous avons utilisé toutes les étiquettes disponibles du domaine cible et qu'il serait intéressant d'étudier à termes l'impact du nombre d'étiquettes rajoutées sur les performances de la classification. Une des contraintes de l'approche envisagée dans ce travail est la nécessité d'avoir les mêmes classes entre les domaines source et cible. Ainsi, seules les classes communes ont pu être conservées. Par ailleurs, afin de garantir de bonnes performances, les proportions des classes entre les deux domaines doivent être équivalentes ($\pm 15\%$). D'autres approches d'adaptation de domaine qui se focalisent non pas sur l'adaptation des données mais plutôt du classifieur peuvent être envisagées pour contourner ces limites. C'est notamment le cas de l'apprentissage actif qui permet de garder les deux distributions inchangées et de modifier le classifieur en lui incorporant un nombre limité d'échantillons du domaine cible (Hamrouni et al., 2021).

TABLE 1: Résultats des différentes stratégies de transport optimal

	Avant transport optimal	TO non supervisé	TO non supervisé régularisé	TO semi-supervisé + étiquettes de la source	TO semi-supervisé + étiquettes (source + cible)
Précision globale	61%	64%	66%	69%	96%
Précision producteur (peupleraies)	70%	78%	78%	88%	97%

Courty N., Flamary R., Tuia D. and Rakotomamonjy A. (2017). Optimal Transport for Domain Adaptation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, n° 9, p. 1853-1865.

Hamrouni Y., Paillassa, E., Chéret V., Monteil C., Sheeren D. (2021). From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, p. 76-100.

Le carrefour dont vous êtes le héros : description de carrefours pour les personnes déficientes visuelles

Jérémy Kalsron¹, Jean-Marie Favreau¹, Guillaume Touya²

*1. LIMOS, UMR CNRS 6158, Université Clermont Auvergne
1 rue de la Chebarde, 63178 Aubière, France
jeremy.kalsron@uca.fr, j-marie.favreau@uca.fr*

*2. UMR LASTIG, IGN
73 avenue de Paris, 94160 Saint-Mandé, France
g.touya@ign.fr*

MOTS-CLES : déficience visuelle, carrefour, modèle de données, génération automatique de texte

Pour les *personnes en situation de déficience visuelle* (PSDV), le déplacement en milieu urbain est un défi. Celles-ci doivent en effet comprendre l'agencement des infrastructures nécessaires à une traversée en sécurité, une information absente des outils de guidage mobiles actuels. Les cartes tactiles, notamment mobilisées par les instructeurs de locomotion pour enseigner les techniques de déplacement en autonomie, ne permettent pas de représenter une haute densité d'informations ou des détails fins à petite échelle (Touya *et al.*, 2019). Elles nécessitent pour cela d'être complétées d'une information supplémentaire, textuelle ou sonore.

Parmi les travaux ayant étudié la représentation sonore de l'information géographique à destination des PSDV, (Boularouk *et al.*, 2017) proposent d'audio-décrire des lieux spécifiques en exploitant la sémantique d'une base de données comme OpenStreetMap. Plus proches de notre problématique, (Guth *et al.*, 2019) réalisent un tableau de données permettant de décrire l'enchaînement des entités qui composent une traversée. Ces travaux présentent l'information sous la forme d'un enchaînement de données brut, et ne proposent pas de modèle de données structuré permettant de formaliser un carrefour et les éléments qui le composent. Nous développons ces aspects dans le travail que nous présentons ici, en proposant un processus de génération automatique de description de carrefour.

Ce processus s'appuie sur un modèle de donnée original dédié (Figure 1). Le modèle est pensé pour permettre la génération d'un ensemble de textes que l'utilisateur pourra consulter suivant deux modes : une description générale du carrefour, autosuffisante et statique, et une série de descriptions cohérentes permettant une visite interactive du carrefour.

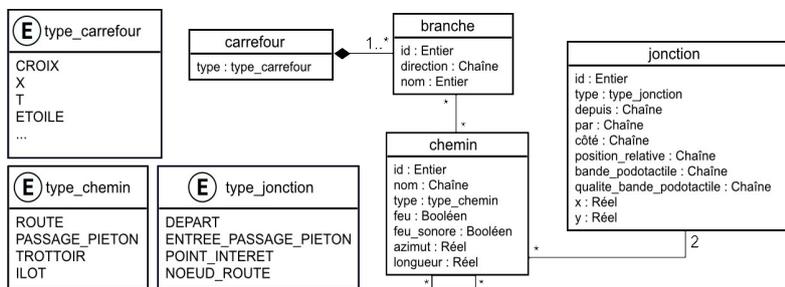


FIGURE 1. Le modèle de données proposé représente un graphe piéton du carrefour.

La description générale repose sur une description par branche, mobilisée par la nouvelle génération d'instructeurs de locomotion (Ratelle et Couturier, 2018). Une branche correspond à un segment du carrefour, et comprend un nombre de voies et un mode de régulation du trafic. La description interactive place l'utilisateur au sein du carrefour et permet de détailler chaque traversée en proposant de choisir une direction lorsqu'une bifurcation est possible. Ces directions sont matérialisées par un point d'intérêt distant dont le choix, aujourd'hui manuel, pourrait être automatisé par des techniques de fouille de données, selon sa prégnance au sein de la ville. Le texte de la description est généré depuis le modèle présenté à l'aide de patrons, des textes

Le carrefour dont vous êtes le héros

à trous complétés procéduralement, faciles à mettre en œuvre, et permettant d'implémenter les phrases telles que nous les avons conçues. Nous proposons une implémentation sous licence libre GPLv3¹, accessible en démonstration en une preuve de concept interactive par hyperliens depuis un navigateur web² (Figure 2).

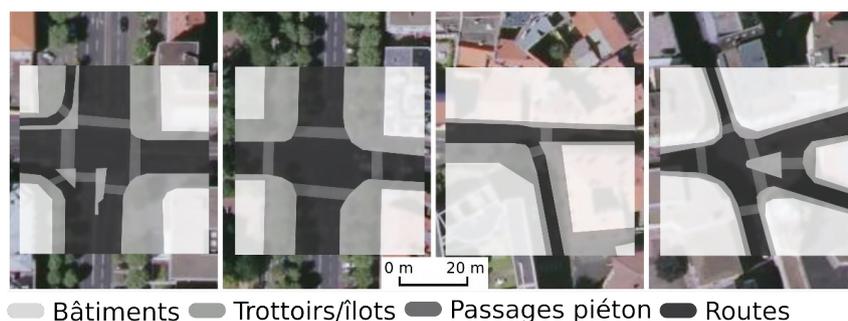


FIGURE 2. Les carrefours décrits au sein de notre preuve de concept. Il nécessitent respectivement 72, 64, 31, et 82 pages de description.

Nous avons conçu ces descriptions en conduisant une série d'entretiens avec une dizaine de PSDV et d'instructeurs de locomotion, et les formulations employées sont issues de leurs propositions. Ceux-ci, familiers avec le territoire que nous avons exploré, ont exprimé l'intérêt de pouvoir visiter un carrefour à distance par le biais d'une description textuelle. Le modèle étant aujourd'hui complété manuellement, nous envisageons pour la prochaine étape de le générer depuis des données ouvertes telles qu'OpenStreetMap.

Ce travail est financé dans le cadre du projet ACTIVmap (ANR-19-CE19-0005).

Bibliographie

- Boularouk *et al.*, (2017). Ontology for a voice transcription of OpenStreetMap data: the case of space apprehension by visually impaired persons, *World Academy of Science, Engineering and Technology 2017*, London, United Kingdom.
- Guth *et al.*, (2019). An Intersection Database Facilitates Access to Complex Signalized Intersections for Pedestrians with Vision Disabilities, *Transportation Research Record*, vol. 2673, p. 698-709.
- Ratelle A., Couturier J.-A., (2018). *Manuel d'intervention en orientation et mobilité*, Les Presses de l'Université de Montréal.
- Touya *et al.*, (2019). Automatic derivation of on-demand tactile maps for visually impaired people: first experiments and research agenda, *International Journal of Cartography*, vol. 5, p. 67-91.

1. <https://gitlab.limos.fr/jeremyk6/carrefourV2/-/tree/sageo2021>
 2. <https://carrefours.activmap.limos.fr/publications/sageo2021>

Intégrer des données géospatiales dans un modèle multi-agents : cas de LittoSIM-GEN

Cécilia PIGNON-MUSSAUD¹, Ahmed LAATABI¹, Marion AMALRIC², Brice ANSELME³, Élise BECK⁴, Nicolas BECU¹, Nicolas MARILLEAU⁵

1. 5. UMR 7266 LIENSs, La Rochelle Université
2 rue Olympe de Gouges, 17000 La Rochelle, France
cecilia.pignon-mussaud@univ-lr.fr
2. UMR 7324 CITERES, Université de Tours
33, Allée Ferdinand de Lesseps, 37204 TOURS CEDEX 03, France
3. UMR PRODIG, Université Paris 1
2, rue Valette, 75005 Paris, France
4. UMR Pacte, CNRS, Université Grenoble Alpes
1030 avenue Centrale 38100 GRENOBLE, France
5. Institut de Recherche pour le Développement,
32 avenue Henri Varagnat, 93143 Bondy, France

MOTS-CLES : SYSTEME MULTI-AGENTS, DONNEES GEOSPATIALES, LITTOSIM-GEN, GENERICITE.

Aujourd'hui, un des défis majeurs de la modélisation multi-agents spatialisée porte sur l'exploitation et l'assimilation grandissante de données hétérogènes géolocalisées. A travers le dispositif LittoSIM¹ (Becu *et al.*, 2017), nous présenterons, comment l'automatisation de l'intégration des données dans le système multi-agent (SMA) a été géré. Développé suite à la tempête Xynthia de 2010, le dispositif LittoSIM a pour objectif de sensibiliser les élus et les agents communaux, aux risques de submersion marine et accompagner la concertation sur les stratégies de prévention du risque et d'aménagement du territoire.

LittoSIM est une plateforme de simulation participative qui combine, à la fois, un jeu multi-acteurs (SMA distribué sous la plateforme GAMA), un modèle hydrodynamique de submersion (LISFLOOD-FP) et une mise en situation pour les acteurs locaux. Initialement développé pour une application sur l'île d'Oléron, le dispositif est désormais déployé sur d'autres territoires littoraux pour évaluer sa généralité et son efficacité en matière d'apprentissage (projet LittoSIM-GEN). Cette ouverture vers la généralité du dispositif s'accompagne d'une re-formalisation et d'une automatisation des traitements des données d'entrée du modèle. LittoSIM nécessite en entrée une grande diversité de données géospatiales issues de sources nombreuses et variées. Par exemple, les données relatives aux ouvrages de défense des côtes proviennent des services de l'État, de collectivités territoriales, de bureaux d'études ou encore de laboratoires de recherche. Elles sont souvent livrées dans des formats et des structures différentes. Ainsi, la transformation, l'agrégation, l'homogénéisation de ces données sont des opérations incontournables et chronophages.

Nous commencerons par présenter rapidement l'utilisation du protocole ODD (Overview, Design concepts, and Details) (Laatabi *et al.*, 2018), qui décrit la structure du modèle LittoSIM. Ce protocole ODD constitue un fil d'Ariane pour la compréhension et la réutilisation des modèles multi-agents à travers la description détaillée de l'ensemble de leurs composantes (variables d'état, comportements, ordonnancement, ...).

Puis, une large part de la présentation portera sur les données et les traitements nécessaires pour le modèle. Nous présenterons alors l'extension ODD+2D qui permet d'enrichir la description du modèle, notamment pour les données géospatiales, à travers la présentation des sources de données et l'ajout de liens de correspondance/mapping.

¹ <https://littosim.hypotheses.org/> ; https://github.com/LittoSim/LittoSim_model

LittoSIM has been financed by Foundation of France.

Intégrer des données géospatiales dans un modèle multi-agents

En termes de données géospatiales, les liens de mapping représentent l'ensemble des opérations sur les composantes spatiales (création, intersection, fusion, reprojection, contrôle topologique...) et textuelles/attributaires (création, conversion, agrégation, remplacement...). Ce mapping des données permet alors d'extraire, de transformer et d'homogénéiser les données d'entrée (Laatabi *et al.*, 2020).

Le protocole ODD+2D permet ainsi de faciliter la réutilisation du modèle, tout en contribuant fortement à améliorer la généricité du dispositif LittoSIM-Gen.

Bibliographie

- Becu N., Amalric M., Anselme B., Beck E., Bertin X., Delaye E., Long N., Marilleau N., Pignon-Mussaud C., Rousseaux F. (2017). Participatory simulation to foster social learning on coastal flooding prevention. *Environmental Modelling & Software*. vol 98. p.1-11.
- Laatabi A., Marilleau N., Nguyen-Huu T., Hbid H., Ait Babram M. (2018). ODD+ 2D: an ODD based protocol for mapping data to empirical ABMs. *Journal of Artificial Societies and Social Simulation*, 21 (2) 9. p.1-24.
- Laatabi A., Becu N., Marilleau N., Pignon-Mussaud C., Amalric M. (2020). Mapping and Describing Geospatial Data to Generalize Complex Models: The Case of LittoSIM-GEN. *International Journal of Geospatial and Environmental Research*. Vol 7. p.1-21.

Mesure de la vulnérabilité des constructions résidentielles face au risque de submersion marine : proposition pour un outil opérationnel

Louisa LE ROUZIC¹, Bruno GERVAIS², Jean-François BREILH³ et Axel CREACH⁴

1. Université de Bretagne Occidentale, stagiaire au SDIS 17 de mai à août 2020, louisa.lerouzc@gmail.com

2. Service Départemental d'Incendie et de Secours (SDIS 17), Pôle moyens généraux, Bruno.Gervais@sdis17.fr

3. Union des Marais de la Charente-Maritime (UNIMA), Cellule prévention des Inondation, jf.breilh@unima.fr

4. Sorbonne Université, Laboratoire de Géographie Physique (UMR 8591-CNRS), axel.creach@sorbonne-universite.fr

Ce poster présente un outil de mesure de la vulnérabilité des maisons pour leurs occupants face au risque de submersion marine en Charente-Maritime.

Ce projet a comme point de départ la tempête Xynthia, à l'origine de 41 décès par noyade en 2010. En plus du niveau marin exceptionnel engendré par la conjonction d'une marée haute de vives-eaux et la surcote associée à la tempête, les retours d'expérience ont pointé l'urbanisation des zones basses comme facteurs explicatifs : tous les décès sont survenus à l'intérieur de constructions résidentielles où les occupants ont été piégés, sans possibilité de s'échapper (Vinet et al., 2010).

Plusieurs initiatives ont émergé pour anticiper au mieux les conséquences éventuelles de tels phénomènes dans le futur.

A l'échelle des constructions résidentielles, un travail a été réalisé pour pouvoir mesurer la vulnérabilité que représente chacune des constructions pour ses occupants : l'indice VIE (Creach et al., 2015). Cet indice se base sur les caractéristiques de localisation (hauteur d'eau potentielle, proximité aux digues) et la configuration des constructions (type architectural, possibilité de se réfugier). Il permet de définir le niveau de danger qu'elles peuvent représenter pour les occupants. Cet indice permet de dresser des cartes de la vulnérabilité des constructions et ainsi savoir où il serait nécessaire d'évacuer de manière préventive en cas de submersion marine.

A l'échelle du département de la Charente-Maritime, le projet SURVEY 17 a comme objectif de mieux anticiper les conséquences de submersions. En premier lieu, le projet s'attache à modéliser les submersions potentielles associées à 96 scénarios de tempêtes aux caractéristiques différentes (vent, houle, marée, etc.). Pour chacun de ces scénarios, une carte des zones inondables et des hauteurs d'eau associées est

proposée. L'établissement de ces scénarios sont facilités par une connaissance des événements historiques ayant impactés le département (Breilh et al., 2014). Dans un second temps, le projet cherche à modéliser directement la submersion potentielle associée à des conditions météo-marines prévues afin d'anticiper les zones pouvant être impactées. Ces deux objectifs permettront de prédire avec un temps d'avance les zones pouvant être inondées lorsqu'une tempête est annoncée. Pour les services de secours, l'objectif est de traduire une alerte météo « risque submersion » de niveau rouge ou orange, en zone d'impact pour prépositionner des moyens d'intervention ou anticiper une évacuation.

Une réflexion a été engagée pour inclure les données de l'indice VIE dans le projet SURVEY 17 afin de pouvoir indiquer les constructions présentant le niveau de danger le plus élevé pour chaque scénario de tempête. Cela permettrait d'ajuster finement et localement les dispositifs d'intervention et d'évacuation et ainsi optimiser le déploiement des services de secours. Cette réflexion s'est néanmoins heurtée à des questions opérationnelles auxquelles ce poster apporte de premières réponses :

- Alors que l'indice VIE n'a été appliqué qu'à 7 communes tout en nécessitant un temps de collecte des données important, comment l'appliquer aux 80 000 constructions potentiellement concernées en Charente-Maritime ?
- Par ailleurs, pour conserver la valeur de l'outil, celui-ci doit être mis à jour régulièrement pour être fiable et utile pour les services de secours. Comment faciliter cette mise à jour en proposant une méthode facilement reproductible et peu chronophage ?

Ces réflexions aboutissent à la proposition d'un indice VIE alternatif permettant une collecte plus rapide des données (mise à jour plus aisée) et la possibilité de couvrir un territoire plus important. Cet indice VIE alternatif est tout d'abord simplifié dans sa structure. Ne sont conservés que deux critères sur les quatre : la hauteur d'eau potentielle et le type architectural. Le premier critère est aisé à extraire et à mettre à jour via un processus de jointure spatiale. Le second nécessite une phase de collecte de terrain qui peut être lourde. Pour faciliter la mise à jour de ce second critère, une méthodologie alternative à l'indice VIE est proposée : plutôt que de collecter systématiquement l'information sur le type architectural pour chaque construction, il est proposé de généraliser la donnée selon une maille de 250 m de côté, telle que déjà utilisée par le SDIS. A l'intérieur de ces mailles, est réalisée une estimation de la proportion de chaque type architectural. Cette estimation peut être réalisée via Google Street View et complétée par une vérification de terrain. En croisant ce coefficient avec le niveau de vulnérabilité associé à chaque catégorie architecturale, cela permet d'obtenir une note globale de la vulnérabilité architecturale des bâtiments situés au sein de la maille. Cette note est ensuite croisée avec le niveau d'eau moyen au sein de la maille pour obtenir un nouveau résultat de l'indice VIE présentant le niveau de vulnérabilité à l'échelle de la maille. La figure 1 présente un exemple de résultat de l'indice à cette échelle.

Cette méthode peut ainsi être transposée aisément aux 96 scénarios modélisés dans le projet SURVEY 17 et être mise à jour régulièrement. Elle permet ainsi de connaître, à l'échelle d'une maille de 250 m, les secteurs où il faudrait intervenir en priorité en cas de submersion marine.

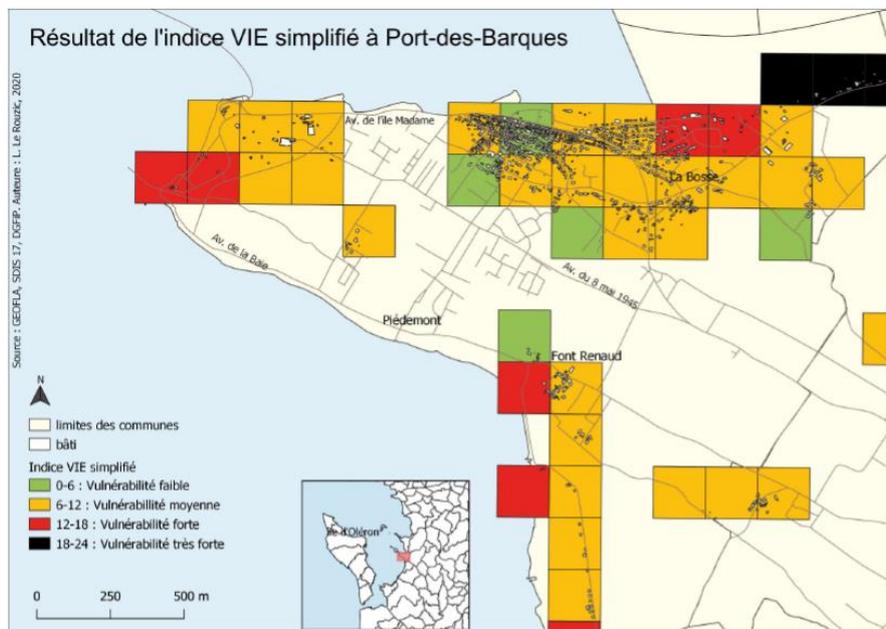


FIGURE 1. Exemple de résultat de l'indice VIE alternatif pour la commune de Port-des-Barques (17).

Finalement, ce travail permet, en associant sapeurs-pompiers et géomaticiens/chercheurs de proposer un outil cartographique opérationnel pour une meilleure appréhension des littoraux urbanisés dans un contexte où le risque de submersion marine devrait se renforcer.

Bibliographie

- Breilh J-F., Bertin X., Chaumillon É., Giloy N., Sauzeau T. (2014). How frequent is storm-induced flooding in the central part of the Bay of Biscay? *Global and Planetary Change*. Vol. 122, p. 161-175.
- Creach A., Pardo S., Guillotreau P., Mercier D. (2015). The use of a micro-scale index to identify potential death risk areas due to coastal flood surges: lessons from Storm Xynthia on the French Atlantic coast, *Natural Hazards*, vol 77, no 3, p. 1679-1710.
- Vinet F., Boissier L., Defossez S. (2011). La mortalité comme expression de la vulnérabilité humaine face aux catastrophes naturelles: deux inondations récentes en France (*Xynthia*, Var, 2010). *VertigO- la revue électronique en sciences de l'environnement*. Vol. 11, n°2, p. 1-28.

Construction d'une méta-base de ports à l'échelle mondiale

Arnaud Ménard¹, Vincent Lambert de Cursay²,
Cyril Ray³, Charles Guenois², Maryam Maslek Elayam³,
Maxime Dréau¹

1. Fabrique Numérique de la Défense

arnaud@swail.io, maxime@dreau.fr

2. Etat Major de la Marine

vcursay@laposte.net, charles.guenois@intradef.gouv.fr

3. Ecole Navale

cyril.ray@ecole-navale.fr, maryam.maslek_elayam@ecole-navale.fr

1. Introduction

L'étude du trafic maritime requiert d'identifier les escales des navires (Carlini *et al.*, 2020; Ducruet, 2017). De nos jours, de nombreuses bases de données listant les localisations des ports du monde existent mais l'identification des zones portuaires et de leur extension géographique reste un problème ouvert (Cotteleer, Koldenhof, 2013). Plusieurs travaux ont abordé ce problème par des mesures de densité extraites des positions de navires (Millefiori *et al.*, 2016). Dans ce travail, nous proposons l'intégration d'une indexation spatiale hexagonale pour permettre une identification de zones portuaires à différentes échelles et permettre la construction d'une méta base de données des zones portuaires.

2. Méthodologie pour la définition d'une méta-base des zones portuaires

Identification de zones portuaires

L'identification et la délimitation des espaces portuaires peut être réalisée à partir des traces historiques des navires (Millefiori *et al.*, 2016). L'objectif de cet article est de proposer une méthode de construction et d'identification des zones portuaires sur la base de plusieurs listes de ports et du trafic maritime constaté (fourni par le système AIS (Bole *et al.*, 2014)) sur plusieurs mois afin d'assurer une grande exhaustivité ainsi qu'un relevé précis des zones portuaires.

Source de données

Notre méthodologie pour la construction d'une méta-base de zones portuaires repose sur les listes (et localisations) des ports du monde issues de 4 bases de données: le *World Port Index*¹, la *World Food Programme*², les mouillages de la *Global Fishing Watch*³ et la base de port de l'*Ifremer*⁴. En associant ces listes, nous pouvons garantir une couverture mondiale sur les ports maritimes ainsi que les fleuves principaux.

Méthodologie

Afin de construire les zones géographiques contenant des navires stationnaires et de les rapprocher des points de référence, nous nous appuyons sur une discrétisation de l'espace maritime en utilisant plusieurs résolutions de l'index H3 (Brodsky, 2018). Cet index utilise une grille globale discrète basée sur une hiérarchie d'hexagones dite d'ouverture 7. Une telle approche a montré son efficacité dans plusieurs applications de recherche dans le domaine maritime, (Wang *et al.*, 2019), (Maslek Elayam *et al.*, 2021), (Tsatcha, 2014). Notre méthodologie est basée sur quatre grandes étapes :

1. Construction des zones géographiques contenant des navires stationnaires : L'analyse de plusieurs mois de trafic maritime (données AIS) a permis d'identifier les hexagones où des navires ont été observés comme stationnaires.

2. Association des zones dites stationnaires aux ports des listes de référence : Les zones contenant des navires stationnaires sont rapprochées des points de référence afin de les associer ou non à un port de chaque liste de port en entrée, par recherche successive dans les hexagones voisins d'un hexagone étudié jusqu'à une certaine distance définie en fonction de chaque liste de port.

3. Création d'une entité port dans la méta-base : Un port dans la méta-base correspond à un tuple contenant la clé identifiant un port dans chacune des listes de référence.

4. Nomme des entités port construites par association : Pour chacun de ces tuples, un nom est construit à partir des noms de port de chaque liste de référence. Ce modèle peut induire un nom composé quand deux listes de référence décrivent plus ou moins précisément un port.

Résultats

Notre méthode a été implémentée et testée à l'échelle mondiale. Une méta-base contenant environ 21000 entités a été construite. Cette structuration permet d'ajouter d'autres listes de port de référence sans altérer les données existantes ou de prendre en compte les évolutions des listes utilisées.

1. <https://msi.nga.mil/Publications/WPI>

2. <https://data.humdata.org/dataset/global-ports>

3. <https://geo.data.gouv.fr/fr/datasets/387e3b19086e5d9ca050a68adc01bd4c7acd1842>

4. <https://data.humdata.org/dataset/global-ports>

3. Conclusion

La construction d'une méta-base de ports exhaustive à partir de données disparates a été proposée pour l'identification des zones portuaires à l'échelle du globe. Cela facilite l'analyse des routes maritimes empruntées et la compréhension du trafic maritime mondial (Maslek Elayam *et al.*, 2021). Plusieurs perspectives sont actuellement considérées. Une annotation sémantique prenant en compte le type ou le pavillon des navires ainsi que leur domaine de navigation permettraient d'identifier plus précisément le type de port (marchand, pétrolier, pêche, plaisance) ainsi que leur rayonnement (international, national, régional). L'intégration de données de ligne de côte permet d'éviter les associations entre deux entités de port proches mais séparées par une langue de terre.

Bibliographie

- Bole A., Wall A., Norris A. (2014). Chapter 5 - automatic identification system (ais). In A. Bole, A. Wall, A. Norris (Eds.), *Radar and arpa manual*, Third Edition éd., p. 255 - 275. Oxford, Butterworth-Heinemann. Consulté sur <http://www.sciencedirect.com/science/article/pii/B9780080977522000052>
- Brodsky I. (2018). H3: Uber's hexagonal hierarchical spatial index. *Available from Uber Engineering website: <https://eng.uber.com/h3/> [22 June 2019]*.
- Carlini E., Lira V. M. de, Soares A., Etemad M., Machado B. B., Matwin S. (2020). Uncovering vessel movement patterns from ais data with graph evolution analysis. In *Proceedings of the 22nd International Conference on Extending Database Technology (EDBT), March 30-April 2*.
- Cotteleer A., Koldenhof Y. (2013). Ship's travelling time in approaching the port of rotterdam. In *Workshop on moving objects at sea*, p. 1-4.
- Ducruet C. (2017). *Advances in shipping data analysis and modeling: Tracking and mapping maritime flows in the age of big data*. Routledge.
- Maslek Elayam M., Ray C., Claramunt C. (2021). Modèle de graphe pour l'analyse des structures de trajectoires maritimes. In *21ème édition de la conférence extraction et gestion des connaissances (egc)*.
- Millefiori L. M., Zissis D., Cazzanti L., Arcieri G. (2016). Scalable and distributed sea port operational areas estimation from ais data. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, p. 374-381.
- Tsatcha D. (2014). *Contribution à l'extraction et à la représentation des connaissances de l'environnement maritime: Proposition d'une architecture dédiée. aux applications de navigation*. Thèse de doctorat non publiée.
- Wang X., Li J., Zhang T. (2019). A machine-learning model for zonal ship flow prediction using ais data: A case study in the south atlantic states region. *Journal of Marine Science and Engineering*, vol. 7, n° 12, p. 463.

Implications des infrastructures de drainage des versants dans la dynamique de transferts de sédiments et de produits phytosanitaires, le cas du Beaujolais viticole

Jessica PIC¹, Etienne COSSART¹, Mathieu FRESSARD²

*1. UMR 5600 – Environnement Ville Société, Université Jean Moulin Lyon 3
18 rue Chevreul, 69362 Lyon Cedex, France
Correspondance : jessica.pic@univ-lyon3.fr*

*2. UMR 5600 – Environnement Ville Société, CNRS
5 avenue Pierre Mendès-France, 69676 Bron Cedex, France*

MOTS-CLES : STRUCTURE PAYSAGERE, MODELISATION SPATIALE, CONNECTIVITE, BEAUJOLAIS

L'utilisation par l'agriculture de produits phytosanitaires est à l'origine d'une pollution généralisée des sols et des cours d'eau et des enjeux d'ordre environnementaux, économiques et sanitaires (Mateo-Sagasta *et al.*, 2018). Parmi les différents types de cultures, la vigne mobilise d'importantes quantités de produits phytosanitaires. Dans une perspective de limitations des transferts de sédiments et de produits phytosanitaires associés des parcelles agricole au cours d'eau, une prise en compte de la dynamique de transferts à l'échelle du bassin versant – unité hydrologique et de gestion – paraît nécessaire à être mise en œuvre.

Toutefois, les transferts de produits phytosanitaires des parcelles agricole au cours d'eau restent difficilement appréhendés en dehors d'une considération stationnaire et bien que diverses études scientifiques (Reichenberger *et al.*, 2007 ; Gascuel-Oudou *et al.*, 2011) aient souligné l'importance de l'interaction entre des objets de différente nature et des processus agissant à des échelles variables. Trois facteurs influencent les transferts de produits phytosanitaires :

- les pratiques culturales conditionnent les quantités de produits pouvant être transférées : quantités et des calendriers d'application ;
- la dynamique hydrologique : le ruissellement de surface constitue le vecteur préférentiel des transferts par le biais des de l'eau et des sédiments issus de l'érosion des sols ;
- la structure paysagère (éléments du paysage et leurs interrelations), intervient dans le schéma d'organisation spatiale et temporelle des transferts (Gascuel-Oudou *et al.*, 2011 ; Reulier, 2015). Elle conditionne la connectivité des parcelles, c'est-à-dire leur capacité à transférer les sédiments et produits phytosanitaires au cours d'eau ;

Dans une perspective d'étude de la connectivité structurelle, c'est-à-dire de la prise en compte des caractéristiques des éléments du paysage et leur configuration spatiale, nous nous intéressons à l'évaluation de l'influence des éléments du paysage et de leur agencement spatial sur la dynamique des flux de produits phytosanitaires associés aux particules de sol pour deux petits bassins versants (8km²) du Beaujolais viticole.

Dans ce cadre, trois types d'éléments du paysage doivent être pris en compte sur le territoire : les modes d'occupation du sol, les infrastructures linéaires de drainage des versants et les bacs de collecte de sédiments. Si la parcelle constitue la maille de référence de la structure paysagère et le lieu de production du ruissellement, ce sont les infrastructures de drainage et leur structuration en réseau qui régissent les potentialités de transferts de sédiments et de produits phytosanitaires d'une parcelle au cours d'eau (fig.1). Nous nous concentrerons ici sur ces éléments et sur la méthode développée pour les envisager.

Implications des infrastructures de drainage des versants dans la dynamique de transferts de sédiments et de produits phytosanitaires, le cas du Beaujolais viticole

Un inventaire exhaustif des infrastructures de drainage a été réalisé par travail de terrain, dans le but d'identifier les secteurs de forte connectivité, l'influence de la structuration des sous-réseaux et de construire une méthodologie d'analyse reproductible à différents contextes paysagers. Les réseaux de drainage sont ensuite intégrés et analysés sous SIG sous la forme d'un graphe orienté en fonction de l'écoulement de surface. L'idée étant d'identifier l'ensemble des points de départ et les exutoires des sous-réseaux, ainsi que les séquences empruntées, c'est-à-dire l'enchaînement d'infrastructures de drainage entre un point de départ et un exutoire.



FIGURE 1. Influence des infrastructures de drainage sur les flux hydro-sédimentaires, secteur de la commune de Villié-Morgon (Beaujolais, France)

La densité d'infrastructures linéaires au sein des zones d'étude révèle une forte anthropisation des versants. L'analyse des séquences – longueur de réseaux, nombre d'éléments dans la séquence, fréquence d'occurrence de séquence – est très variable en fonction des caractéristiques des zones d'exutoire, appuyant la pertinence de l'analyse de la structure du réseau dans la connectivité des parcelles au cours d'eau. Toutefois, une compréhension approfondie de la connectivité structurelle ne pourra avoir lieu que par l'intégration dans le graphe des bacs de collecte de sédiments et du réseau parcellaire.

Mateo-Sagasta, J., Marjani Zadeh, S., Turrall, H. (2018). *More people, more food, worse water? a global review of water pollution from agriculture*. Rome, Italy: FAO. 225p.

Gascuel-Oudou, C., Arousseau, P., Doray, T., Squidant, H., Macary, F., Uny, D., & Grimaldi, C. (2011). Incorporating landscape features to obtain an object-oriented landscape network representing the connectivity of surface flow pathways over rural catchments. *Hydrological Processes*, 25, 3625–3636

Reichenberger, S., Bach, M., Skitschak, A., & Frede, H. G. (2007). Mitigation strategies to reduce pesticide inputs into ground- and surface water and their effectiveness; A review. *Science of the Total Environment*, 384(1–3), 1–35.

Reulier, R. (2015). *Structure paysagère et dynamiques spatiales des transferts hydro-sédimentaires Approche par simulation multi-agents*. Université de Caen Normandie.

Exposition à la pollution atmosphérique et risque de cancer du sein : vers une prise en compte des mobilités quotidiennes dans les études épidémiologiques

**Marie Ramel-Delobel¹²³, Thomas Coudon¹²³, Lény Grassot¹²,
Delphine Praud¹², Béatrice Fervers¹², Pietro Salizzoni³**

*1. Département Prévention, Cancer, Environnement
Centre Léon Bérard
Cheney C, 28 rue Laennec, 69008, Lyon, France
Marie.ramel-delobel@lyon.unicancer.fr*

*2. INSERM UA8 Radiations : Défense, Santé, Environnement
Centre Léon Bérard, Ministère des Armées, Service de Santé des Armées (SSA)
Cheney A, 28 rue Laennec, 69008, Lyon, France*

*3. CNRS UMR 5509 Laboratoire de Mécanique des Fluides et d'Acoustique
Ecole Centrale de Lyon, Université Claude Bernard Lyon I
36 avenue Guy de Collongue, 69132 Ecully, France*

MOTS-CLES :

GEOGRAPHIE, POLLUTION ATMOSPHERIQUE, CANCER DU SEIN, MOBILITES QUOTIDIENNES, NO₂, PM

Contexte : Le cancer du sein est le cancer le plus fréquent chez la femme en France et au niveau mondial. L'augmentation de son incidence ces 30 dernières années est associée au dépistage généralisé, aux traitements hormonaux de la ménopause et aux changements de mode de vie. Cependant, le rôle des facteurs environnementaux est aussi suspecté. Plusieurs études épidémiologiques suggèrent en effet un lien entre l'exposition à long terme à la pollution de l'air ambiant et le risque de cancer du sein (White et al., 2018).

De nombreuses études épidémiologiques utilisent l'historique des adresses résidentielles pour évaluer l'exposition aux polluants de l'air, ce qui peut entraîner une imprécision des estimations. En effet, l'exposition « chronique » survient tout au long de la journée, au domicile, mais également sur le lieu du travail et d'autres activités ainsi que pendant les mobilités. Or, bien que les déplacements représentent une faible partie du temps circadien, ils peuvent représenter une part importante de la dose globale qu'une personne inhale dans la journée (de Nazelle et al., 2012).

Ce travail, effectué dans le cadre d'un stage de Master 2 qui se poursuit désormais en thèse s'inscrit dans le projet APoPCo¹ - collaboration Département Prévention Cancer Environnement du Centre Léon Bérard et Laboratoire de Mécanique des Fluides et d'Acoustique de l'Ecole Centrale de Lyon (CNRS UMR 5509) ; financé par l'ANSES et la Fondation de France. APoPCo vise à estimer l'association entre le risque de cancer du sein (étude cas-témoins nichée dans la cohorte nationale E3N) et l'exposition à la pollution atmosphérique en prenant en compte les expositions chroniques, à la fois résidentielles, professionnelles, et durant les déplacements domicile-travail (DDT).

Objectif : L'objectif est ici de développer une méthode de calcul des expositions pendant les déplacements – et en premiers desquels les DDT – en vue d'évaluer leur contribution dans l'exposition totale de la journée. La part de ce type d'exposition est analysée au regard de l'influence des différents modes de déplacement, des itinéraires et des types de polluants (e.g. NO₂ et PM₁₀).

Matériel et méthode : Ce projet s'appuie sur des données préalablement recueillies et produites par l'équipe dans le cadre du projet XENAIR (2016-2021 ; financement par l'Institut National du Cancer et de la Fondation ARC) :

- un questionnaire sur l'histoire résidentielle et professionnelle, et les modes des DDT des sujets d'étude (10 444 femmes, 15 865 adresses résidentielles entre 1990 et 2010, adresses professionnelles en cours de finalisation) ;
- des données de concentrations de NO₂ et de PM₁₀ à l'échelle locale (modèle SIRANE, 10x10m, horaire) (Soulhac et al., 2017), et nationale (modèle Land Use Regression (LUR), 50x50m, annuel), sur une période de 21 ans (1990-2010) (Amadou et al., 2020) ;

La méthode de quantification de l'exposition, développée via R, se découpe en 4 étapes qui correspondent à différentes échelles d'analyse, allant de l'échelle globale à celle de l'individu. En premier lieu, les concentrations selon le type de polluant sont estimées à l'échelle de la rue (1). En second lieu, il est possible de déduire les expositions induites par les microenvironnements des modes de déplacement (2). Ensuite, les itinéraires calculés pour chaque individu permettent de définir les concentrations auxquelles il est exposé (3), pour enfin, calculer la dose qu'il inhale selon l'intensité du volume respiratoire (4). En dernier lieu, les résultats sont analysés afin d'identifier la variation de l'exposition qu'induit le choix du mode, de même que la part de l'exposition quotidienne totale attribuable aux déplacements pendulaires.

Résultats : Les premiers résultats intermédiaires² – à confirmer avec toute la population d'étude³ – montrent que la dose de polluant inhalée pendant les DDT participe entre 5 à 10% en moyenne à la dose quotidienne totale. Pour chaque polluant

¹ *Atmospheric Pollution and Physical activity linked with Commute*

² Test de la méthode appliquée à partir d'un couple d'adresses résidentielles issues de E3N et d'adresses professionnelles aléatoires sur la Métropole lyonnaise (partie Master)

³ Application de la méthode sur les adresses professionnelles géocodées (partie Thèse)

Mobilités et pollution atmosphérique : risque de cancer du sein

étudié, les DDT représentent entre 3 et 6.6% de la dose en NO₂, et 5.7 à 11.2% en PM₁₀ pour une part du temps relativement faible (3% du temps circadien), en comparaison des parts du temps et des expositions estimés en moyenne au domicile (respectivement 60% et 70%) et au travail (30% et 25%).

Ensuite, l'exposition attribuable aux DDT varie selon les caractéristiques propres à l'individu. Le premier facteur est le choix du mode (motorisé, vélo, marche) : le vélo est celui qui contribue le plus à augmenter la dose inhalée dans la journée, suivi de la marche, et enfin de la voiture, dont les valeurs sont en moyenne 2 fois inférieures à celles des modes actifs. Cette hiérarchie des modes s'explique par le fait que les modes actifs induisent une activité physique qui accroît le volume respiratoire. Enfin, la dose inhalée est influencée par le choix de l'itinéraire dont ceux exposant l'individu à de plus fortes concentrations empruntent les axes les plus directs et les plus rapides sur lesquels la vitesse autorisée varie entre 80 à 130 km/h.

Conclusion : L'exposition aux polluants de l'air pendant les DDT varie selon les caractéristiques propres à l'individu (volume respiratoire), celles de sa mobilité (mode, itinéraire), de même que selon les caractéristiques de l'environnement urbain (structuration du réseau routier, vitesse limite autorisée, trafic). La méthode présentée, toujours en cours de développement, fournit une première estimation de l'exposition pendant les déplacements et elle est amenée à être approfondie pour affiner ces résultats intermédiaires. Ces précisions concerneront notamment la classification des modes les plus ou moins exposés, en intégrant plus de paramètres pour définir l'exposition en automobile. En outre, la pratique d'une activité physique, même dans un environnement pollué, permettrait de réduire les risques de mortalité. La question des bénéfices de l'utilisation des modes actifs sera donc également à appréhender. Enfin, une des perspectives sera de déterminer si la prise en compte des mobilités dans les études épidémiologiques influence les conclusions antérieures sur les effets sanitaires de la pollution.

Références bibliographiques

- Amadou, A., Praud, D., Coudon, T., Danjou, A. M. N., Faure, E., Leffondré, K., Le Romancer, M., Severi, G., Salizzoni, P., Mancini, F. R., & Fervers, B. (2020). Chronic long-term exposure to cadmium air pollution and breast cancer risk in the French E3N cohort. *International Journal of Cancer*, 146(2), 341-351. <https://doi.org/10.1002/ijc.32257>
- de Nazelle, A., Fruin, S., Westerdahl, D., Martinez, D., Ripoll, A., Kubesch, N., & Nieuwenhuijsen, M. (2012). A travel mode comparison of commuters' exposures to air pollutants in Barcelona. *Atmospheric Environment*, 59, 151-159. <https://doi.org/10.1016/j.atmosenv.2012.05.013>
- Soulhac, L., Nguyen, C. V., Volta, P., & Salizzoni, P. (2017). The model SIRANE for atmospheric urban pollutant dispersion. PART III: Validation against NO₂ yearly concentration measurements in a large urban agglomeration. *Atmospheric Environment*, 167, 377-388. <https://doi.org/10.1016/j.atmosenv.2017.08.034>
- White, A. J., Bradshaw, P. T., & Hamra, G. B. (2018). Air pollution and Breast Cancer : A Review. *Current Epidemiology Reports*, 5(2), 92-100. <https://doi.org/10.1007/s40471-018-0143-2>

Vers une détermination automatique du nombre d'usagers sur une plage urbaine De l'observation au modèle à base d'intelligence artificielle pour la plage du Prophète à Marseille

Samuel ROBERT¹, Briec CABIOCH¹, Eric CARROLL¹, Marie-Laure TREMELO¹

1. ESPACE, UMR 7300, CNRS, Aix Marseille Université, Avignon Université, Université Côte d'Azur
Technopole de l'environnement Arbois-Méditerranée F-13545 Aix-en-Provence
Cedex 4, France
samuel.robert@univ-amu.fr

MOTS-CLES : PLAGE, FRÉQUENTATION, ÉVALUATION, MODÈLE, MARSEILLE

Contexte

Dans les grandes villes littorales, le bord de mer et les plages forment un espace public et récréatif majeur. Les flux de visiteurs qui s'y rendent nécessitent d'être évalués et suivis pour permettre la conception et l'ajustement des mesures de gestion : surveillance de la baignade, nettoyage, travaux d'entretien, services aux usagers, gestion de l'affluence, ou surveillance des faits de délinquance, pour ne citer que quelques exemples. Pourtant, si des expérimentations diverses ont vu le jour (Balouin et al., 2014 ; Guyonnard et al., 2020 ; Guyonnard et Vacher, 2018 ; Guyonnard, 2017), rares sont les villes disposant de dispositifs d'évaluation et de suivi fin de la fréquentation de leurs plages.

Objectif

Ce poster évoque comment cette problématique est travaillée à Marseille, France. Depuis 2013, plusieurs plages de la ville sont l'objet d'observations des fréquentations et des pratiques au cours de l'été. Ces travaux ont permis de saisir les motivations des usagers, mieux appréhender les plages en termes de qualité environnementale (Labille et al., 2020 ; Tournoud et al., 2019), et révéler les flux de personnes (Robert et Trémélo, 2018 ; Robert et Trémélo, 2020a, 2020b, 2019). Sur

ce dernier point, de nombreuses données ont été produites pour mettre en évidence une voire des fréquentations-types. Actuellement, les travaux visent à définir un modèle pour l'une des plages les plus étudiées : le Prophète.

Résultats

Des observations exploratoires répétées sur diverses plages de la ville ont permis de saisir l'hétérogénéité de la fréquentation. Pendant l'été, une grande variabilité a été mise en évidence d'un jour à l'autre sur une même plage, et d'une plage à l'autre lors d'un même jour. Des effets de sites ont également été mis en relief, de même que l'influence d'évènements locaux pouvant affecter une plage à un moment donné.

Pour une étude plus approfondie, le Prophète a été choisi comme site expérimental. On y dispose de données variées : des effectifs d'usagers dénombrés en continu aux entrées et sorties de la plage entre 8h et 20h ; des photographies obliques prises d'un point haut à chaque heure (8h, 9h, 10h, etc.), permettant de dénombrer les usagers par photo-interprétation ; et des questionnaires sur les pratiques. L'objectif est de concevoir un protocole de dénombrement automatique des usagers à partir de photos qui seraient prises par un système installé sur site. D'ores et déjà, les photos et les fichiers de semis de points localisant les usagers sur chacune d'elles constituent un jeu de données conséquent, qui permet d'engager une expérimentation en intelligence artificielle (IA), en entraînant un réseau de neurones convolutifs (CNN).

Ces algorithmes d'IA en apprentissage profond (deep learning) s'inspirant du cortex humain, sont conçus pour effectuer des tâches descriptives et génératives sur des images numériques. Leurs applications sont nombreuses notamment dans la médecine, la reconnaissance faciale, la transcription manuscrite ou les véhicules autonomes. Leur élaboration nécessite un entraînement sur de nombreuses images pour lesquelles on dispose des données qui font l'objet de l'apprentissage afin de produire un système capable d'inférer les données souhaitées sur de nouvelles images. Les spécificités des photos de plage dont nous disposons nous orientent vers des algorithmes spécifiques du champ du comptage de foule dont l'objectif est de prédire des densités, des effectifs (Zhang et al., 2016, Li et al., 2018), plutôt que ceux communément mobilisés dans le comptage par reconnaissance de forme (Balouin et al., 2014). Les premiers résultats des modèles entraînés sur des photos de foules appliqués sur les photos de la plage du Prophète sont encourageants, mais des biais subsistent. La réussite de l'expérimentation pourrait profondément influencer sur le suivi à venir des plages du littoral marseillais.

Vers une détermination automatique du nombre d'usagers sur une plage urbaine

Bibliographie non numérotée et références

Balouin Y. et al., 2014. Automatic assessment and analysis of beach attendance using video images at the Lido of Sète beach, France. *Ocean and Coastal. Management*, 102. 10.1016/j.ocecoaman.2014.09.006

Guyonnard V. 2017. *Dimensions cachées et attentes spatiales dans un espace de pratique de tourisme et de loisir : une analyse géographique de la plage en Charente-Maritime (France)*. Th. de géographie, Univ. La Rochelle.

Guyonnard V. et al., 2020. REMEPP: a simulation model of beach filling as a tourism and leisure practice. *Cybergeogeo*. <https://journals.openedition.org/cybergeogeo/35378>

Guyonnard V., Vacher L., 2018. Penser et mesurer les distances de l'interaction sociale dans l'espace de la plage. *L'Espace géographique*, 47, 2, pp. 159-181. <10.3917/eg.472.0159>

Labille J. et al., 2020. Assessing UV filter inputs into beach waters during recreational activity: A field study of three French Mediterranean beaches from consumer survey to water analysis. *Science of the Total Environment*, 706. <10.1016/j.scitotenv.2019.136010 >

Li Y. et al., 2018. CSRNet : Dilated convolutional neural networks for understanding the highly congested scenes. In : Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp.1091-1100.

Robert S., Apers-Trémélo M.-L., 2020a. *Données de fréquentation de la plage du Prophète (Marseille) le 17 juillet 2019*, 26 p. <hal-02461322>

Robert S., Apers-Trémélo M.-L., 2020b. *Données de fréquentation de la plage du Prophète (Marseille) du 10 au 16 juillet 2017*, 110 p. <hal-02907086>

Robert S., Apers-Trémélo M.-L., 2019. *Données de fréquentation de la plage du Prophète (Marseille) le 18 juillet 2018*, 26 p. < hal-02438256v2>

Robert S., Trémélo M.L., 2018. Analyse spatiale des pratiques des usagers de trois plages urbaines de Marseille. *merIgéo*, Aix-en-Provence, Fr. <hal-01746087>

Tournoud M.-G. et al., 2019. Microbiological quality of bathing waters versus sanitary regulation and beach users' practices. A case study in Marseilles. *Séminaire annuel du Labex DRIIHM*, Oct 2019, Lyon, France. <hal-02409637>

Zhang Y. et al., 2016. Single-image crowd counting via multi-column convolutional neural network. In : Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 589-597.

Apport de l'utilisation du drone et de transects écopaysagers pour l'échantillonnage spatial d'habitats naturels : le cas des digues de la vallée de la Dives (Calvados)

Nicolas ROBINET¹, Gaëtan JOLLY², Guy LEMPERIERE³, Yves PETIT-BERGHEM³

1. *Cermosem (Institut d'Urbanisme et de Géographie Alpine – Université Grenoble Alpes / UMR Pacte)*

1064 chemin du Pradel 07170 Mirabel, France

Nicolas.robinet@univ-grenoble-alpes.fr

2. *Diplômé du Master Expertise et Gestion de l'Environnement Littoral, Institut Universitaire Européen de la Mer, Technopôle Brest-Iroise*

rue Dumont d'Urville - 29280 Plouzané, France

jolly.gaetan@gmail.com

3. *Larep (Laboratoire de recherche en projet de paysage), École Nationale Supérieure du Paysage*

10 avenue du Maréchal Joffre 78000 Versailles, France

guy.lempriere@ird.fr

y.petitberghem@ecole-paysage.fr

MOTS-CLÉS : drone, écologie du paysage, transects, indices, méthode d'échantillonnage, habitats, digues, milieu littoral, hausse du niveau marin, stratégie de gestion territoriale

KEYWORDS: drone, landscape ecology, transects, metrics, sampling method, habitats, dikes, coastal environment, sea level rise, territorial management strategy

1. Introduction, contexte, problématique et objectif du poster

Dans le contexte des changements globaux, la digue est amenée à se transformer et à faire face à un certain nombre de risques (inondation, submersion...) susceptibles de se manifester. Accompagner les acteurs locaux dans ce contexte changeant revient à entrecroiser de multiples enjeux et à réfléchir à des orientations variées de gestion des digues, parfois nouvelles en France, oscillant entre fermeture et ouverture, rehaussement et abaissement, artificialisation et renaturation. L'objectif majeur du programme est l'établissement de propositions nouvelles relatives au devenir des digues et l'évaluation des conditions de faisabilité et d'acceptabilité de ces propositions. Pour y parvenir, l'un des enjeux est de pouvoir connaître précisément l'état des digues et leur capacité à accueillir des habitats naturels dont la spatialisation à échelle fine s'avère indispensable.

Dans le cadre du site des digues de la Dives, l'accès aux quatre kilomètres de berges est problématique. De nombreuses portions sont des propriétés privées. De plus, l'entretien des digues a été délaissé depuis une cinquantaine d'années, ce qui a entraîné le recouvrement progressif de ces infrastructures par une végétation haute, rendant la prospection sur site difficile. Cette problématique écarte l'emploi de méthodes terrestres traditionnelles de relevés de végétation et d'habitats, comme par exemple un échantillonnage de type Braun-Blanquet. En effet, cela nécessiterait de défricher des zones d'accès et risquerait d'endommager les habitats pour parvenir au site. Nous avons mobilisé différents moyens pour élaborer une méthode de relevé non destructif, permettant de cartographier les habitats des deux rives.

2. Matériel, méthodes et outils

Pour analyser dans sa globalité le site d'étude, nous utilisons une approche écopaysagère afin de qualifier le paysage et les habitats des digues aval de la Dives.

Pour pallier l'impossibilité d'accéder aux berges, nous avons réalisé deux missions de survol aéroporté qui ont permis de couvrir 7 secteurs pour une surface d'environ 47 hectares. Le traitement numérique des données et la photo-interprétation ont permis la caractérisation des habitats, de la végétation, l'identification des ruptures de pente, l'exposition, la présence de fossés et la réalisation de profils altimétriques. En complément, un parcours aquatique effectué le long de la Dives a permis de réaliser un transect utilisant une méthode dite du quadrat par Point Centré (PCQM) modifiée dans le but d'échantillonner les différents habitats placés le long du cours d'eau sur un parcours aquatique de 4 km et 17 points d'échantillonnage. Cette technique est utilisée pour estimer la surface des différents types d'habitats ou d'objets naturels. Couplée aux relevés aéroportés, elle a permis de déterminer les limites des différents habitats classés selon le Code Corine Biotope, leur taille, leur densité, la fréquence de leur répartition et leur arrangement spatial contagieux ou aléatoire.

3. Résultats et discussion

Cet ensemble de données permet d'établir une carte des habitats à une définition scalaire précise (de l'ordre de 0,20 m) et permet de réaliser par la suite des calculs d'indices caractérisant les diversités bêta et gamma.

Les indices de Jaccard et Sorensen montrent des paysages aux habitats très différents entre les deux rives avec notamment, en aval du cours d'eau, des enrochements facilitant l'installation d'espèces pionnières, des plantations de peupliers en rive droite, des milieux ouverts en rive gauche essentiellement (schorres). L'indice de Shannon indique une hétérogénéité plus élevée pour la rive droite ($H=1,67$) en comparaison de la rive gauche ($H=1,46$). Cette disparité s'explique en grande partie par l'usage des sols de part et d'autre de la rivière. En partie amont, nous retrouvons des zones de pâturage extensif. Les digues, surmontées de chemins de halage, ont été abandonnées, ce qui a conduit à un enrichissement progressif de ces structures. Des brèches sont disposées tout au long de la rivière, ce qui profite aux bovins pour s'abreuver et accentue la détérioration des digues. En partie, les parcelles attenantes à la Dives s'urbanisent et les berges constituent des espaces publics. Elles sont encore surmontées par des sentiers de halage, ce qui a permis de conserver les digues.

Quelques limites ont été observées concernant l'usage du drone. En effet, les survols de certaines zones font l'objet d'autorisations spécifiques, souvent longues à obtenir et à concilier avec les aléas météo. Ce facteur est à prendre en compte, car il peut rallonger, de façon non négligeable, les délais d'étude, sachant qu'un refus n'est pas à exclure. Les espaces boisés représentent eux aussi un frein à la photo-interprétation (étalement de la végétation). Les contraintes techniques peuvent également être un frein : coût, logistique, stockage de la donnée, temps de traitement. L'observation à partir du cours d'eau montre elle aussi quelques limites. Le champ visuel étant restreint par les digues, les habitats écologiques situés de l'autre côté du cours d'eau (face terre) ne peuvent être observés. On ne dispose alors que d'une perception partielle des habitats repérés.

4. Contenu du poster et présentation générale

Le poster présente les éléments suivants :

- 1^{ère} partie : un résumé rapide du contexte de l'étude et de la problématique, accompagné par une carte de localisation du site d'étude.
- 2^e partie : un exposé de la méthode employée, appuyée par une carte de localisation des points d'échantillonnage et les zones de survols par drone.
- 3^e partie : présentation des résultats qui comportera :
 - * une carte des habitats et tableau comparatif des indices,
 - * un bloc-diagramme,
 - * des photographies,
 - * un encart de définition des indices utilisés.
- Dernière partie : ouverture sur les limites de la recherche entreprise accompagnée d'une discussion.

Processus de modélisation et optimisation pour la collecte de déchets plastiques en mer

Loic Salmon¹, Nicolas Maslov¹, Christophe Claramunt¹

Institut de recherche de l'Ecole Navale, 29240 Brest Cedex 9, France
loic.salmon@ecole-navale.fr

1. Introduction

La pollution plastique est considérée comme un problème environnemental majeur, avec plus de cinq trillions de déchets plastiques répandus dans les océans (Eriksen *et al.*, 2014). Cette accumulation croissante dans les mers et les océans constitue un risque majeur pour les écosystèmes marins, notamment du fait de son introduction dans la chaîne alimentaire. Les déchets plastiques peuvent être classés en fonction de leurs dimensions, et en trois types: macro-plastique (plus de 2,5 cm), méso-plastique (entre 5 mm et 2,5 cm) et micro-plastique (taille inférieure à 5 mm).

L'objectif de recherche de cet article est d'introduire une méthodologie générale pour observer, modéliser, visualiser les déplacements spécifiquement des macro-plastiques en mer afin d'en faciliter la récupération.

2. Méthodologie générale pour le suivi et la collecte de macro-déchets plastiques en mer

La chaîne de processus allant de l'observation des macro-déchets plastiques au calcul d'une route optimale permettant d'aller les collecter est décrite dans la figure 1. La première étape consiste à développer et formaliser (1) un modèle de données pour la récupération et la gestion des données géolocalisées des macro-déchets. La deuxième étape concerne (2) les aspects de fusion et de mise à jour alors que différentes sources d'observation sont disponibles et que les zones de déchets se déplacent constamment en raison des courants et des vents. Une fois que la base de données est alimentée et que les mécanismes de mise à jour sont effectifs (3) l'extraction d'une carte de densité probabiliste des déchets plastiques permet de visualiser l'emplacement des amas. Enfin la

Méthodologie pour la collecte de plastiques en mer

mer a été introduit et permet de générer des cartes de densité de probabilité de répartition des déchets en mer.

(4) *Optimisation pour trouver la route optimale de collecte de déchets en mer.* A partir de la carte de densité de probabilité construite précédemment, il s'agit ensuite de trouver la ou les routes optimales pour le ramassage de déchets plastiques. Une série de paramètres sont à prendre en compte tels que la quantité de plastique collectée, le temps de navigation ou l'énergie consommée.

3. Premiers résultats

Des données de positions d'amas plastiques ont été simulées aux abords de la Bretagne, et des données courant fournies par le SHOM ont été utilisées pour estimer la dérive de ces déchets en mer. A gauche (Figure 2), le dégradé de couleur correspond à la probabilité de densité de déchets plastiques pour 1000 simulations de dérive grâce à un modèle probabiliste en fonction des courants sur 12h depuis les reports de positions initiales. A droite (Figure 2), un exemple de trajectoire pour le ramassage de déchets en mer, tenant compte de ces courants.

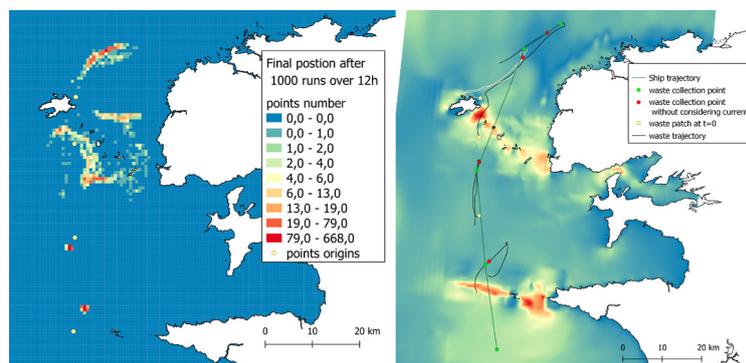


FIGURE 2. Carte de densité et simulation de ramassage de déchets plastiques

Bibliographie

- Eriksen M., Lebreton L. C. M., Carson H. S., Thiel M., Moore C. J., Borerro J. C. *et al.* (2014, 12). Plastic pollution in the world's oceans: More than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. *PLOS ONE*, vol. 9, n° 12, p. 1-15.
- Ghamisi P., Rasti B., Yokoya N., Wang Q., Hoffe B., Bruzzone L. *et al.* (2018). *Multisource and multitemporal data fusion in remote sensing*.
- Shen Y.-T., Lai J.-W., Leu L.-G., Lu Y.-C., Chen J.-M., Shao H.-J. *et al.* (2019). Applications of ocean currents data from high-frequency radars and current profilers to search and rescue missions around taiwan. *Journal of Operational Oceanography*, vol. 12, n° sup2, p. S126-S136.

Effets de la planification urbaine sur les températures de surface du sol (Montpellier, France)

Magalie Técher¹, Hassan Ait Haddou², Rahim Aguejdad^{3,4}

1. Laboratoire LIFAM, École Nationale Supérieure d'Architecture de Montpellier
179, Rue de l'Esperou, 34093 Montpellier, France
magalie.techer.au@gmail.com

2. Laboratoire LIFAM, École Nationale Supérieure d'Architecture de Montpellier
179, Rue de l'Esperou, 34093 Montpellier, France
haithaddou@montpellier.archi.fr

3. USR 3330, UMIFRE 21 CNRS-MAE, IFP, 11 Saint-Louis street, Pondichery –
605 001, Inde

4. UMR TETIS, CNRS, 500 rue Jean-François Breton, 34090 Montpellier, France
rahim.aguejdad@cnr.fr

RESUME. L'accroissement de la population dans les villes conduit à une expansion urbaine importante et à une pression sur les milieux urbains. Cette expansion s'accompagne d'une augmentation des températures urbaines plus connu sous le nom d'îlot de chaleur urbain et exacerbée par les effets du réchauffement climatique global.

Dès lors, il devient primordial d'apporter des pistes de réflexion pour lutter contre ce phénomène d'ICU et permettre un développement contrôlé de cette expansion urbaine.

À travers une étude menée sur la métropole de Montpellier Méditerranée, ce travail présente une analyse des effets de la planification urbaine, à travers le règlement du plan local d'urbanisme (PLU) français, sur l'îlot de chaleur urbain.

Dans un premier temps, grâce à l'intégration des différentes règles du PLU au sein d'une base de données SIG, des cartes de zonage sont générées, et intègrent pour chaque zone, l'ensemble des règles s'y appliquant. Dans un second temps, grâce aux données satellitaires du landsat-8, des cartes de températures du sol (LST) sont générées pour quatre dates en 2019. Enfin, par un outil de géotraitement, nous obtenons par zonage, une température du sol moyenne, ce qui permet de mettre en relation, une règle du PLU et ses effets sur les températures du sol.

Les résultats trouvés sont d'une part, conformes à la littérature scientifique et fournissent d'autre part, des leviers d'action pour les décideurs politiques, urbanistes et maîtres d'ouvrage urbain, dans le cadre de l'élaboration de documents d'urbanisme français sur des moyens d'adaptation des territoires aux problématiques d'ICU et de changement climatique.

ABSTRACT. Population growth in cities leads to significant urban expansion and pressure on urban areas. Along with this expansion, is an increase in urban temperatures better known as the Urban Heat Island (UHI) and exacerbated by the effects of global warming. Therefore, it becomes essential to provide avenues for reflection to fight against this phenomenon of UHI and allow a controlled development of this urban expansion.

Through a study carried out on the metropole of Montpellier Méditerranée, this works aims to assess the effects of urban planning policies, through the regulation of French local urban plan (PLU), on urban heat island.

First, thanks to the integration of the various PLU rules into a GIS database, zoning maps are generated, and include for each zone all the rules that apply to it. Second, thanks to landsat-8 satellite data, land surface temperature (LST) maps are generated for four dates in 2019. Finally, using a geoprocessing tool, we obtain, by zoning, an average land surface temperature, which shows how PLU rules interact with land surface temperatures.

The results found are, on the one hand, consistent with the scientific literature and, on the other hand, provide levers of action for decision makers and urban planners, within the framework of the development of French urban planning documents on mean of adaptation of territories to problems of UHI and climate change.

MOTS-CLES : planification urbaine, forme urbaine, températures de surface de sol, îlot de chaleur urbain, France

KEYWORDS: urban planning policies, urban form, land surface temperature, urban heat island, France

Bibliographie non numérotée et références

- Bahi H., Mastouri H, et Radoine H. (2020). Review of methods for retrieving urban heat island. *Materials today : proceedings*, 2020, 6p.
- Benmecheta A. (2016) *Estimation de la température de surface à partir de l'imagerie satellitale; validation sur une zone côtière d'Algérie*. Thèse de doctorat en Géographie, Université Paris-Est Créteil Val de Marne, 118p.
- Chen H.C., Qi H., et De Vries B. (2020). Modeling the spatial relation between urban morphology, land surface temperature and urban energy demand. *Sustainable Cities and Society*, 35 p.
- Chen Y., Wu J., Yu K., et Wang D. (2020) Evaluating the impact of the building density and height on the block surface temperature. *Building and environment*, n° 168, 14p.
- Jusuf, S.K., Ignatius M., Hien W.N., et Akbari H. (2019). Editorial : urban heat island (UHI) and its mitigation through urban planning, design, and landscaping. *Architectural science review*, vol. 1, n° 62, 2 p.

Analyse des effets du règlement d'urbanisme sur les températures de surface du sol

- Melot, R, et Bransiecq M. (2016). Règles d'urbanisme et choix politique : les observations de l'état sur les projets locaux. *Revue d'économie régionale et urbaine*, Armand Colin, n° 4, p 767-98.
- Mehebab S., Raihan A., et Sajjad H. (2016). Analyzing land surface temperature distribution in response to land use/land cover change using split window algorithm and spectral radiance model in suburban biosphere reserve, India. *Modeling Earth Systems and Environment*, vol. 2, n° 81, 12 p.
- Oguz H. (2013). LST Calculator : a program for retrieving land surface temperature from landsat TM/ETM + Imagery. *Environmental Engineering and Management Journal*, vol 12, n° 3, p 549-55.
- Prévost A. (2013). *Inférence(s) des documents d'urbanisme sur le territoire: application à la ville de Toulouse : modélisation multicritère et évaluation durable*. Thèse de doctorat en Architecture et Aménagement de l'Espace, Université de Technologie de Compiègne, 398p.
- Salvati A., Palme M., et De la Barrera F. (2018). Urban morphology parametrization for climate modelling in urban planning. *10th International Conference on Urban Climate / 14th Symposium on the urban environment*, New York, USA.
- Yiannakou A., et Salata K.D., (2017). Adaptation to climate change through spatial planning in compact urban areas: a case study in the city of Thessaloniki. *Sustainability*, vol. 9, n° 271, 18 p.

L'état des piscines, construction d'un indicateur de la reprise du tourisme en situation post-catastrophe. L'île de Saint-Martin après le passage du cyclone Irma (sept 2017)

Thommeret N.¹, Grancher D.², Dousseron A.², Brunstein D.³, Virmoux C.²

1. Laboratoire Géomatique et Foncier, Cnam-ESGT, 1 bd Pythagore, F- 72000, Le Mans, France ; nathalie.thommeret@lecnam.net

2. Laboratoire de Géographie Physique (CNRS – UMR 8591), 1 place Aristide Briand, F-92195 Meudon Cedex, France

3. UMR Lieux, Identités, Espaces et Activités (UMR 6240), Université de Corse Pascal Paoli

MOTS-CLES : CYCLONE, TELEDETECTION, PISCINES, TOURISME, NDWI

Les images à très haute résolution spatiale telles que celles des satellites Pléiades permettent de suivre l'évolution du territoire à des échelles spatiales et temporelles très fines. Dans le cadre de la gestion des catastrophes naturelles, la chartre internationale des risques majeurs permet d'obtenir des images pendant l'événement ou juste après l'événement pour évaluer les dégâts et diriger les secours et aides humanitaires. Pour le suivi à plus long terme, on peut avoir accès des images régulières pour le suivi de la reconstruction du territoire ou de la reprise de l'activité économique et sociale. Ces images sont notamment utiles pour caractériser la résilience du territoire et proposer une analyse spatialisée permettant de différencier les quartiers.

En 2017, suite au passage du cyclone Irma (5-6 septembre) sur l'île de Saint Martin, les études d'endommagements et de destructions ont estimé que 95% des bâtiments et infrastructures ont été touchés (Desarthe *et al.*, 2020). Dans le cadre de l'ANR TIREX, plusieurs indicateurs sont explorés pour le suivi de la reconstruction territoriale (Duvat *et al.*, 2021). Parmi les indicateurs spatialisés qui peuvent être considérés, nous nous intéressons ici au suivi de l'état de l'eau des piscines des propriétés sur la partie française de l'île. En effet, sur un territoire tel que Saint Martin dont l'économie repose essentiellement sur le tourisme, une grande partie des logements (individuels ou collectifs – hôtels, résidences) disposent de piscine. En étudiant la réponse spectrale de la surface d'un même bassin sur des images prises à différentes dates successives, on peut identifier la date de remise en service de la piscine. Cette dernière caractérise au minimum une intervention dans la propriété pour redémarrer le moteur ou une intervention plus importante de nettoyage du bassin et de restauration de l'équipement, correspondant à un usage de la propriété. On part de l'hypothèse que la piscine est remise en état lorsque la maison elle-même est de nouveau habitable. Par contre dans le cas d'une piscine mutualisée entre plusieurs logements il est possible qu'il reste des logements non habitables ou non habités et suivant le coût de la réparation par exemple certaines piscines peuvent être laissés en l'état (*i.e.* non fonctionnelles) alors que la maison est remise en location ou ré-occupée. De manière générale, on peut considérer que la chronologie correspondant à la remise en eau propre des piscines est un bon indicateur pour le suivi de la reconstruction et le retour de l'activité touristique après l'événement majeur qu'est le cyclone Irma.

La méthode proposée pour établir cette chronologie repose sur : 1) la localisation des piscines pour un état de référence avant cyclone ; 2) l'identification des piscines fonctionnelles à plusieurs dates après cyclone à partir des images Pléiades.

Pour localiser les piscines avant cyclone, nous avons complété par photo-interprétation les données OpenStreetMap (OSM), elles-mêmes issues du cadastre français (Direction Générale des Impôts - Cadastre, 2015). Toutefois, du fait d'un décalage en planimétrie entre les images, la position des piscines varie jusqu'à plusieurs mètres localement, empêchant la superposition des polygones. Nous avons ainsi fait le choix de travailler à l'intérieur d'une zone tampon autour des centroïdes des polygones OSM. Le rayon de la zone tampon a été fixé à partir des distances minimales inter-piscines permettant de travailler sur un voisinage avec une superposition minimale entre les zones des différentes piscines.

Pour évaluer l'état de l'eau de chacune des piscines, nous avons utilisé l'indice NDWI (*Normalized Difference Water Index* ; McFeeters, 2013 ; Kim et al., 2011) qui repose sur la différence de réflectance des surfaces dans les longueurs d'onde du vert et du proche-infrarouge (équation 1).

$$NDWI = \frac{(Vert - PIR)}{(Vert + PIR)} \quad (1)$$

Cet indice nous permet de repérer les piscines fonctionnelles (*i.e.* en eau claire avec un fond réfléchissant ; figure 1). L'absence de pixel « eau claire » dans la zone tampon est associée à une piscine en eau sale ou vide. Une analyse par arbre de classification (CART) a permis d'une part de montrer que le NDWI était pertinent pour identifier les zones d'eau claire et d'autre part de déterminer un seuil objectif et valable sur les images aux différentes dates. Les principales raisons de mauvaise classification sont liées d'une part aux images elles-mêmes (présence de nuages ou de zones d'ombre) et d'autre part à un effet de taille (moins bonne détection des piscines < 25m²).

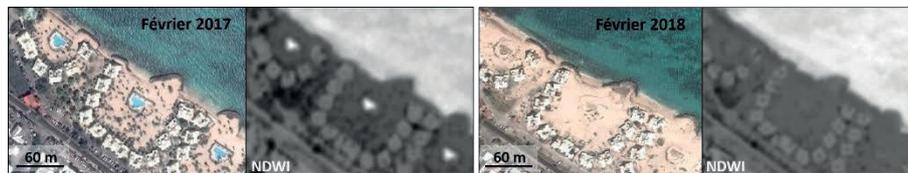


FIGURE 1. Piscines ensablées après le cyclone Irma sur images Pléiades et NDWI

L'approche par le NDWI nous a ainsi permis d'obtenir une première cartographie des piscines fonctionnelles pour neuf dates de 2017 à 2020 et d'établir une chronologie par quartier de remise en fonction des piscines sur la partie française de Saint Martin. Les résultats vont maintenant être confrontés à d'autres indicateurs de la dynamique touristique permettant une analyse plus globale de la reprise sur ce territoire depuis trois ans.

Desarthe J., Bauduceau N., Quantin A. (2020). Les défis soulevés par la reconstruction post-Irma, *Annales des Mines - Responsabilité et environnement*, 36-40.

Duvat V., Volto N., Stahl L., Moatty A., Defossez S., Desarthe J., Grancher D., Pillet V. (2021), Understanding interlinkages between long-term trajectory of exposure and vulnerability, path dependency and cascading impacts of disasters in Saint-Martin (Caribbean), *Global Environmental Change*, 67, 102236.

Kim M., Holt J., Eisen R., Padgett K., Reisen W., and Croft J. (2011), Detection of swimming pools by geographic object-based image analysis to support west nile virus control efforts, *Photogrammetric Engineering and Remote Sensing*, 77, 1169–1179.

McFeeters, S. K. (2013). Using the normalized difference water index (NDWI) within a geographic information system to detect swimming pools for mosquito abatement: A practical approach, *Remote Sensing*, 5, 3544-3561.

Sensibilisation des gestionnaires à la prévention des risques liés à l'eau : modélisation, simulation et participation dans le Delta du Mékong

Thi Huyen Tran¹, Marion Amalric¹, Frédéric Rousseaux², Benoit Gaudou³

*1. UMR CITERES CNRS, Université de Tours, 35 Allée Ferdinand de Lesseps,
37200 Tours, France
huyenrosietran@gmail.com, marion.amalric@univ-tours.fr*

*2. UMR LIENSs CNRS, Université de La Rochelle, 2, rue Olympe de Gouges,
17000 La Rochelle – France
frederic.rousseau@univ-lr.fr*

*3. IRIT, Université de Toulouse, Cours Rose Dieng-Kuntz, 31400 Toulouse – France
benoit.gaudou@gmail.com*

MOTS-CLES : SIMULATION PARTICIPATIVE, GESTION DES RISQUES, APPRENTISSAGE SOCIAL, DELTA DU MEKONG.

Notre étude s'inscrit dans le thème de la simulation participative. Elle présente un jeu sérieux, qui a été réalisé en France « LittoSIM », et qui est en cours d'expérimentation et d'adaptation dans le Delta du Mékong au Vietnam. Cette simulation intitulée « LittoKONG » se base sur une plateforme participative et un système multi-agents.

L'objectif principal de notre étude est de sensibiliser les autorités locales à l'adaptation au changement climatique. Plus précisément, il s'agit de favoriser l'apprentissage social des mesures de prévention des risques liés à l'eau avec les autorités et les gestionnaires locaux.

En raison du changement climatique, il est probable que l'intensité des submersions marines augmentent sensiblement, tel que l'élévation du niveau de la mer et les événements météo-marins intenses, voire extrêmes (IPCC, 2014). Il est donc essentiel de questionner et adapter la gouvernance des littoraux urbanisés afin de limiter les conséquences liées à la vulnérabilité des sociétés concernées. La démarche collaborative entre les disciplines et les technologies diverses (environnement, société, économie, urbanisation, gouvernance, etc.) est donc indispensable pour la gestion durable du littoral (Lacroix *et al.*, 2019). Dans cette perspective, le jeu sérieux de simulation participative aide les acteurs concernés (communautés affectées, décideurs, parties prenantes) de poursuivre une démarche d'apprentissage et de sensibilisation dans la gestion des risques (Marion Amalric *et al.*, 2017; Solinska-Nowak, 2018).

« LittoKONG » est constitué de différents scénarios face à la submersion dans le contexte mondial du changement climatique. A travers les ateliers du jeu, ce dispositif permet aux acteurs locaux de coopérer, « tester » et envisager les solutions de prévention et de gestion des risques littoraux à l'aide d'un ensemble d'actions d'aménagement et une représentation de leur territoire.

L'étude présentée dans ce poster a permis de répondre à la question de faisabilité de LittoKONG vis-à-vis du système juridique au Vietnam. L'adaptation du modèle conceptuel initial du jeu sérieux dans un nouveau contexte et les différents défis de traitement des données spatiales sont également abordés.

Ce travail de recherche a reçu le soutien financier de l'Ambassade de France au Vietnam, au titre d'une bourse d'excellence de thèse et par du projet CNRS-IRD RISQUES 2019 pour le financement de la mission.

Modélisation, simulation et participation à la prévention des risques liés à l'eau

Références bibliographiques

- IPCC. (2014). *Climate Change 2014: Synthesis Report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change.
- Lacroix, D., Mora, O., De Menthiere, N., & Bethinger, A. (2019). *La montée du niveau de la mer: Conséquences et anticipations d'ici 2100, l'éclairage de la prospective*. Rapport d'étude, Alliance nationale de recherche pour l'Environnement, 172.
- Marion Amalric, Anselme, B., Bécu, N., Etienne Delay, Marilleau, N., Pignon, C., & Rousseaux, F. (2017). Sensibiliser au risque de submersion marine par le jeu ou faut-il qu'un jeu soit spatialement réaliste pour être efficace ? *Sciences du jeu*, 8.
- Solinska-Nowak, A. (2018). An overview of serious games for disaster risk management – Prospects and limitations for informing actions to arrest increasing risk. *International Journal of Disaster Risk Reduction*, 17.

Modélisation spatiale et positionnement de zones de dépôt pour la collecte de déchets post-ouragans

Quy Thy Truong¹, Anne Ruas², Serge Lhomme³

1. *Université Gustave Eiffel, COSYS/LISIS
truong.quythy@gmail.com*
2. *Université Gustave Eiffel, COSYS/LISIS
anne.ruas@univ-eiffel.fr*
3. *Université Paris Est Créteil, Lab'URBA
serge.lhomme@u-pec.fr*

MOTS-CLES : COLLECTE DE DÉCHETS POST-OURAGANS, SIMULATION, MÉTHODE DE POSITIONNEMENT

Les ouragans génèrent d'importants volumes de déchets qu'il faut évacuer d'urgence pour permettre au territoire sinistré de se reconstruire. Pour mieux anticiper les futures catastrophes naturelles de ce genre, le projet ANR DéPOs vise à étudier le processus de collecte des déchets post-ouragans à partir d'événements passés et à proposer des pistes d'amélioration à partir de simulations de collecte de déchets post-ouragans en se concentrant sur les Antilles Françaises. Le travail présenté dans ce poster est centré sur l'algorithme de collecte et est appliqué à Saint-Martin à partir de données réelles après le passage de l'ouragan Irma en septembre 2017. Des campagnes de télédétection post-ouragans sur Saint-Martin réalisées par le SERTIT permettent en effet de disposer de l'inventaire des bâtiments détruits et endommagés très utile pour la mise au point des méthodes proposées. Une fois finalisé, l'algorithme peut être appliqué sur un autre territoire et pour un ouragan simulé. L'objectif de la recherche est d'estimer au mieux les temps de collecte selon les scénarios, les aménagements et les moyens de collecte (acteurs, nombre et types de camions) afin d'aider à minimiser ce temps en situation réelle grâce à une planification éclairée.

Le processus de collecte des déchets post-ouragans consiste en un transfert successif des déchets de leur source vers différentes zones de dépôt temporaire, puis de ces zones de dépôts temporaire vers les zones de traitement ou d'enfouissement des déchets. D'après les témoignages recueillis après Irma, les habitants de Saint-Martin rassemblent les déchets post-ouragans devant ou à proximité des bâtiments détruits ou endommagés sur des lieux appelés des aires de dépose (Beraud *et al.*, 2019). Ces déchets sont alors ramassés par des entrepreneurs vers des installations temporaires de transit, avant d'être à nouveau déplacés vers des centres de stockage ou de traitement de déchets. Le processus de collecte est donc composé de deux étapes. Une modélisation a été proposée pour décrire ce processus (Ruas et Lhomme, 2019). Cette modélisation permet de distinguer différents types de déchets mais nous ne traitons ici qu'un type de déchet, ce qui était le cas pour Saint Martin. Nous présentons dans ce poster de la première étape qui s'implémente comme un algorithme de déplacement de stock de déchets entre des origines (proche des habitations endommagées) et des destinations (les installations temporaires de transit). Pour réaliser ce processus, nous proposons des méthodes d'identification des origines, ce qui est une vraie difficulté car ces lieux sont éphémères et nous ne disposons pas d'information pour les localiser puisque ces aires sont informelles donc non planifiées et disparaissent lorsque les déchets ont été collectés. Dans notre recherche, nous proposons plusieurs méthodes de localisation de ces aires de dépose et comparons l'impact de ces choix sur les temps de collecte calculés.

L'algorithme proposé est un algorithme de plus court chemin classique calculé sur le réseau routier entre des origines à proximité des bâtiments détruits et endommagés et des zones de transit temporaire, puis des zones de transit aux exutoires, ici la décharge de Saint Martin. Les bâtiments détruits et endommagés ont été identifiés grâce à la campagne de télédétection réalisée par le SERTIT. Le calcul du volume a été estimé grâce à des photos sur place et des connaissances de volumes total en fin de chaîne à l'unique décharge de l'île (Lamour, 2019). Le temps de

Modélisation spatiale et positionnement de zones de dépôt...

collecte est la somme du chargement, du déplacement et du déchargement des déchets avec une flotte de camions ayant une certaine capacité de chargement, jusqu'à livraison des déchets dans la zone de transit. Dans notre recherche quatre modes de calcul des origines sont proposées : à la porte des bâtiments, à l'intersection la plus proche, projeté sur la rue la plus proche avec regroupement de déchets en fonction des volumes et aux grandes intersections. La méthode basée sur les regroupements part du constat que les habitants regroupent les déchets sur une même aire, lorsque le volume de déchet n'est pas trop important et que la distance n'est pas trop grande afin de minimiser le nombre de tas de déchets dans les rues. Nous calculons des zones tampons autour des aires de dépose potentielles, avec un rayon variable en fonction du volume de déchets présents. Au niveau expérimental nous présentons ci-dessous les résultats des temps sur le quartier de Grand Case, avec 3 camions de 32m³ et avec des journées de travail de 8 heures (table 1).

TABLE 1. Durées de collecte selon quatre scénarios

Méthode de positionnement	Nombre d'aller-retour	Durée de la collecte en jour
Au niveau des gisements de déchets	157	71
Intersections proches des gisements	70	65
Projection et regroupement des gisements	46	63.5
Intersections des routes principales	32	62

L'analyse des données montre de façon non surprenante que la méthode de projection et regroupement des gisements proches permet de simuler une collecte plus rapide que celle du 'porte à porte'. On sait aussi que cette méthode est plus réaliste que la méthode de positionnement aux intersections proches. Elle est cohérente avec les témoignages rapportés. Ce qui est intéressant c'est que le temps en jours varie relativement peu d'un scénario à l'autre : c'est en effet la durée de chargement et de déchargement qui est la plus pénalisante dans le processus : dans le scénario de collecte en porte à porte, la durée de collecte résulte à 97% du temps de chargement/déchargement et 3% du temps de transport des déchets. C'est donc bien sur ce point qu'il faut chercher les améliorations pour réduire le temps de collecte.

Bibliographie

- Beraud H., Nithart C., Durand M. (2019). Le difficile suivi des déchets postcatastrophe : le cas de l'Ouragan Irma à Saint-Martin. *Risques urbains*, vol. 3, no 1. Consulté sur <https://www.openscience.fr/Le-difficile-suivi-des-dechets-post-catastrophe-le-cas-de-l-Ouragan-Irma-a>
- Lamour M., (2019) *Système d'information pour la gestion des déchets post-ouragan*, Rapport de Master Télédétection et Géomatique appliqué à l'environnement, Université Paris & Gustave Eiffel, 2019
- Ruas A., Lhomme S. (2019) L'analyse et l'amélioration de la gestion des déchets après les ouragans. Le cas des Antilles françaises. *Cartes et Géomatique*. N° 241-242, p 59-64.

La planification littorale spatialisée de Balagne à l'épreuve de son opérationnalité : renouveler les méthodes d'identification des espaces vides

Véronique Venturini¹, Caroline Tafani²

1. Doctorante en Géographie, Urbaniste

UMR CNRS 6240 LISA Università di Corsica Pasquale Paoli

Campus Mariani, Bâtiment Edmond Simeoni, Avenue Jean Nicoli, 20250 CORTE

veronique.venturini@outlook.com

2. Maître de Conférence en Géographie

UMR CNRS 6240 LISA Università di Corsica Pasquale Paoli

Campus Mariani, Bâtiment Edmond Simeoni, Avenue Jean Nicoli, 20250 CORTE

Tafani c@univ-corse.fr

*MOTS-CLES : PLANIFICATION, LITTORAL, URBANISME, CONFIGURATIONS
SPATIALES DU FONCIER, CORSE*

Planifier en Corse : une problématique incontournable ?

La planification littorale spatialisée, en tant que traduction réglementaire des évolutions du territoire (Demazière, Desjardins 2016), apparaît comme un sujet complexe en Corse. Sa conception, qui se définit entre la précision et l'adaptation des normes nationales, semble rencontrer des problématiques de mise en œuvre à l'échelle locale. En effet, la gouvernance territoriale, qui s'était heurtée à plusieurs reprises aux difficultés d'élaborer le Plan d'Aménagement et de Développement Durable de la Corse (PADDUC), peine aujourd'hui à le rendre opérationnel. Dans ce contexte, la Balagne ne fait pas figure d'exception. Cette microrégion située au Nord-Ouest de l'île, apparaît comme un territoire attractif concentrant les enjeux en matière de développement territorial (Tafari 2012). Cependant, on ne retrouve pas de véritable dynamique de mise en œuvre de la planification stratégique spatialisée. Sur 36 communes, on ne recense que deux documents compatibles avec le PADDUC. Face à ce constat, nous soulevons l'idée que les méthodes d'analyse des territoires, et notamment celles liées à l'identification des espaces non bâtis, ne semblent pas tenir compte de la situation foncière (DDTM 2017 ; Orsini 2014). L'hypothèse soulevée soumet l'idée qu'une analyse foncière plus précise, intégrée à la démarche planificatrice, pourrait permettre d'améliorer son opérationnalité à travers une évaluation plus approfondie des capacités foncières disponibles dans les espaces non bâtis.

Approfondir les méthodes d'analyse des espaces bâtis et appréhender la thématique foncière dans la démarche planificatrice

En dressant un état des lieux du terrain au regard des méthodes d'analyse des territoires inhérentes à la démarche d'élaboration des documents d'urbanisme nous avons pu constater que la pratique ne tient pas compte de la situation foncière des interstices non bâtis et notamment des configurations spatiales, au sens « *d'agencement et de structuration* » (Pumain 2004) du foncier. Identifier les configurations spatiales du foncier permet de caractériser ce dernier au regard de sa situation géographique, de son agencement, de son accessibilité, sa structure et enfin de ses usages et qualités (Marraccini et al. 2017). L'ensemble de ces éléments apparaissent comme des indicateurs spatialisés qui vont permettre d'identifier un stock de foncier libre réellement exploitable en traitant sa dimension relationnelle (droits et usages), son accessibilité et les fonctions attribuées (urbanisation, coupure urbaine...). Cette approche, appliquée à l'échelle des interstices non bâtis et intégrée aux méthodes d'analyse, doit permettre de faciliter la démarche planificatrice.

Les configurations spatiales du foncier, un indicateur pertinent

Cette méthode d'analyse des configurations spatiales du foncier doit avoir

La planification littorale spatialisée de Balagne à l'épreuve de son opérationnalité

plusieurs effets pour la pratique. D'une part, elle doit permettre de résoudre certaines approximations liées aux ratios démographie/logements pouvant proposer des objectifs de développement parfois inatteignables. D'autre part, cette méthode tient compte des phénomènes de tensions foncières pouvant se traduire par une forme de bicéphalisme entre les noyaux villageois originels situés dans l'arrière-pays avec les formes d'urbanisation littorales récentes. Dès lors, identifier le potentiel foncier au regard de ses configurations spatiales permettrait de garantir une meilleure prise en compte de l'organisation des formes urbaines. Dans le cas du terrain étudié, l'intégration des configurations spatiales du foncier peut exprimer un différentiel de l'ordre de 3 à 4 hectares selon les types de formes urbaines et leur localisation.

Révéler les potentialités de développement des territoires littoraux

L'analyse des configurations spatiales du foncier, intégrée aux méthodes d'analyse des territoires apparaît comme un indicateur pertinent pouvant améliorer l'opérationnalité de la planification. Ce concept, présenté sous la forme d'indicateurs visant à identifier le foncier réellement valorisable, peut permettre ainsi d'éclairer les choix urbanistiques à réaliser. Par conséquent cette proposition méthodologique fondée sur le caractère multidimensionnel du foncier et son niveau de mutabilité se place également comme une première étape d'action à mener dans le cadre de politiques foncières.

Bibliographie non numérotée et références

- Demazière C., et Desjardins X. (2016). La planification territoriale stratégique : une illusion nécessaire ? *Revue internationale d'urbanisme*. N° 2.
- Direction Départementale des Territoires et de la Mer. (2017). *La pression foncière en Corse*. Rapport d'étude. Prefecture de Corse.
- Marraccini E., Lardon S., et Sabbatini T. (2017). Les configurations spatiales agri-urbaines - Pour intégrer l'agriculture dans l'aménagement des territoires urbains. *Spatial Analysis and GEomatics*. Novembre 2017, Rouen.
- Orsini L. (2014). Les aspects civils et fiscaux de l'indivision en Corse. *Droit et Ville*, Vol. 77, n° 1, pp. 73-99.
- Pumain D. (2004). Les configurations. *Hypergeo*. 2004.
- Tafari C. (2012). Dynamiques du foncier agricole et développement territorial en Balagne (Haute-Corse). *Conférence inaugurale de l'OHM Littoral méditerranée*. Septembre 2012, Marseille.

Méthodes de Machine Learning pour l'analyse de l'occupation du sol des zones exposées au risque d'inondation : la ville côtière d'Ain Témouchent (Ouest Algérien)

Narimane ZAABAR^{1,2}, Simona NICULESCU², Mustapha Kamel MIHOUBI¹, Soumia BENGOUFA^{1,2,3}

1. Ecole Nationale Supérieure de l'Hydraulique (ENSH), Laboratoire de Mobilisation et de Valorisation des Ressources en Eau (MVRE), 29 route de Soumaa, BLIDA, ALGERIE, n.zaabar@ensh.dz

2. UBO, Université de Bretagne Occidentale, Laboratoire LETG Brest
UMR 6554 CNRS, Brest, France

3. Ecole Nationale Supérieure des Sciences de la Mer et de l'Aménagement du Littoral (ENSSMAL), Dely Ibrahim Bois des Cars, Alger, Algérie

MOTS-CLES : Inondation, Occupation Du Sol, Ain Témouchent, Classification, Sentinel-1(S1), Sentinel-2(S2), Random Forest (RF), Support Vecteur Machine (SVM)

Dans le contexte de l'étude du risques d'inondation, l'analyse de l'occupation du sol permet d'identifier les enjeux socio-économiques et quantifier les dommages provoqués (Pottier, 2005). En effet, l'évolution de la technologie spatiale en matière de capteurs satellitaires, ouvre des possibilités inédites pour l'identification, l'analyse et la surveillance de l'occupation du sol. Dans cette optique, les produits de l'imagerie satellitaire Sentinel-2 (S2) de haute résolution spatiale (HRS) offrent une analyse de haute précision. De même, les images radar Sentinel-1 (S1) à synthèse d'ouverture (SAR), permettent de s'affranchir du problème de la couverture nuageuse. Par conséquent, la fusion des images S1 et S2 offre une meilleure discrimination des éléments de l'occupation du sol, ce qui a été prouvé dans de nombreuses études (Niculescu et al., 2020 ; Mustak et al., 2019), notamment par le biais de approches de *Machine Learning* et de *Deep Learning*. En effet, l'objectif de cette étude est d'identifier des méthodes de machine Learning qui permettent de détecter le type d'occupation du sol dans le contexte de la caractérisation du risque d'inondation de la ville de Ain Témouchent (nord-ouest de l'Algérie).

Deux approches de classification ont été testées : pixel par pixel et orientée objet, par le biais de deux algorithmes de *Machine Learning* *Random Forest* (RF) et *Support Vector Machine* (SVM). Pour cela, une série temporelle de 12 images S1 et S2,

acquises durant l'année 2019 a été utilisée. Le choix de l'année est justifié par les évènements des inondations que la région d'étude a connu. Après des prétraitements des images S1, une fusion avec les bandes de S2 et des couches d'indices spectraux a été effectuée. Cinq classes d'occupation du sol ont été définies : (1) eau, (2) sols nus, (3) terres agricoles, (4) bâti, (5) végétation. Le choix de ces classes a pris en considération les enjeux situés dans la zone exposée au risque d'inondation. Pour l'approche orientée objet, l'algorithme de Multirésolution (MRS) a été appliqué pour la segmentation de l'image fusionnée. Pour l'approche pixel par pixel, les échantillons créés ont permis d'entraîner directement le classificateur. Tandis que, pour la deuxième approche, des objets ont été extraits à partir de la segmentation, et utilisés en tant qu'échantillons d'entraînement. Afin d'obtenir de meilleures classifications, un processus d'optimisation a été effectué pour choisir les meilleures valeurs des paramètres de segmentation et des algorithmes RF et SVM. La performance des résultats a été évaluée par le calcul d'une matrice de confusion, un indice Kappa, et un indice de précision globale.

Les deux approches de classification ont permis de détecter 5 classes de type d'occupation du sol ce qui correspond au niveau 2 d'une nomenclature adaptée au contexte d'inondation selon Maurel et al (2003) (Figure 1). Les résultats obtenus par les deux algorithmes sont satisfaisants, néanmoins la performance de l'algorithme RF a surpassé celle de l'algorithme SVM. Ceci rejoint les constats d'autres études (ex. Htitiou, et al., 2019 ; Niculescu, et al., 2020 ; Inglada, et al., 2016). Selon les résultats, l'approche orientée objet révèle plus de pertinence, en atteignant les valeurs de 96% pour l'indice de précision globale (OA) et 0.95 pour l'indice Kappa (Table 1). Cette approche ainsi que l'avantage de libre droit des produits de HRS, S1 et S2, offrent la possibilité d'analyser des enjeux socio-économiques, dans le cadre de l'évaluation de la vulnérabilité face aux risques d'inondations.

Comme perspective de cette étude, il est recommandé de tester ces méthodes sur d'autres types de données afin de détecter les classes du niveau supérieur (niveau 3), ce qui est requis pour l'élaboration des Plans de Prévention des Risques d'Inondation.

TABLE 1 : Indices kappa et de précision globale (OA) de RF et SVM appliqués pour les deux approches ; pixel par pixel et orientée objet (OOB)

	OA(%)	Kappa
RF-pixel	92.3	0.90
SVM-pixel	91.0	0.89
RF-OOB	96.8	0.95
SVM-OOB	92.0	0.90

Méthodes de Machine Learning pour la modélisation de l'occupation du sol

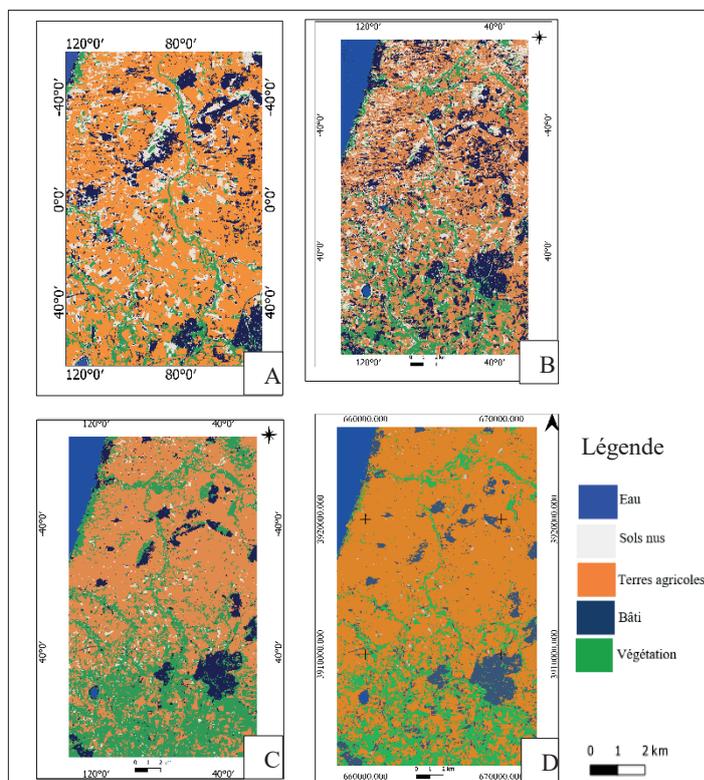


FIGURE 1 : Résultats de la classification; A) RF-pixel B) SVM-pixel C) RF-OOB
D) SVM-OOB

Bibliographie

Maurel, P, Raclot D, Puech C (2003). *Apports de la télédétection à très haute résolution spatiale à l'étude du risque d'inondation*. <https://hal.inrae.fr/hal-02582784>

Niculescu, S.; Boissonnat, J.-B.; Lardeux, C.; Roberts, D.; Hanganu, J.; Billey, A.; Constantinescu, A.; Doroftei, M., Synergy of High-Resolution Radar and Optical Images Satellite for Identification and Mapping of Wetland Macrophytes on the Danube Delta. *Remote Sensing* 2020, 12, 2188, doi.org/10.3390/rs12142188.

Pottier N., Penning-Rowsell E., Tunstall s., ET Hubert G. (2005). Land Use and Flood Protection: Contrasting Approaches and Outcomes in France and in England and Wales. *Applied Geography* 25 (1) : 1-27. 10.1016/j.apgeog.2004.11.003.

Responsables de la publication :

Paule-Annick DAVOINE, Didier JOSSELIN, François PINET

Éditeur : UMR 7300 ESPACE / Avignon, Case 41, 74 rue L. Pasteur, 84029 Avignon Cedex, France

Collection : Actes Avignon - ISBN : 978-2-910545-12-1, Mai 2021

Design et Mise en page : N. Brachet (UMR ESPACE – CNRS)

<http://sageo2021.univ-lr.fr>