



HAL
open science

**Fouille de données, Modélisation et Optimisation.
Applications à la mobilité intelligente et la santé
connectée.**

Romain Billot

► **To cite this version:**

Romain Billot. Fouille de données, Modélisation et Optimisation. Applications à la mobilité intelligente et la santé connectée.. Informatique [cs]. UCBL; Lyon1, 2017. tel-01760656

HAL Id: tel-01760656

<https://hal.science/tel-01760656v1>

Submitted on 6 Apr 2018

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.



MÉMOIRE D'HABILITATION À DIRIGER DES RECHERCHES

présentée au sein de
l'Université Claude Bernard Lyon 1

Discipline : Informatique

Manuscrit soumis en vue d'une soutenance le 16 Novembre 2017, par

Romain BILLOT

Maître de conférences, IMT Atlantique (Brest)

Fouille de données, Modélisation et Optimisation.

Applications à la mobilité intelligente et la santé connectée.

Devant le jury composé de :

M. Jérôme Azé, Professeur, Université de Montpellier Rapporteur
M. Flavien BALBO, Professeur, Ecole des Mines de Saint Etienne Rapporteur
M. Aziz MOUKRIM, Professeur, Université de Technologie de Compiègne Rapporteur
M. Nour-Eddin EL FAOUZI, Directeur de recherches, IFSTTAR Examineur
Mme. Salima HASSAS, Professeure, Université de Lyon 1 Examinatrice
Mme. Christine SOLNON, Professeure, INSA de Lyon Examinatrice

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Président du Conseil Académique

Vice-président du Conseil d'Administration

Vice-président du Conseil Formation et Vie Universitaire

Vice-président de la Commission Recherche

Directrice Générale des Services

M. le Professeur Frédéric FLEURY

M. le Professeur Hamda BEN HADID

M. le Professeur Didier REVEL

M. le Professeur Philippe CHEVALIER

M. Fabrice VALLÉE

Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de la Réadaptation

Département de formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur G.RODE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C. VINCIGUERRA

Directeur : M. X. PERROT

Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

UFR Sciences et Techniques des Activités Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Ecole Supérieure du Professorat et de l'Education

Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI

Directeur : M. le Professeur F. THEVENARD

Directeur : Mme C. FELIX

Directeur : M. Hassan HAMMOURI

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur G. TOMANOV

Directeur : M. le Professeur H. BEN HADID

Directeur : M. le Professeur J-C PLENET

Directeur : M. Y. VANPOULLE

Directeur : M. B. GUIDERDONI

Directeur : M. le Professeur E.PERRIN

Directeur : M. G. PIGNAULT

Directeur : M. le Professeur C. VITON

Directeur : M. le Professeur A. MOUGNIOTTE

Directeur : M. N. LEBOISNE

Remerciements

Table des matières

| | |
|---|-----------|
| Avant-propos | 1 |
| Liste des abréviations | 3 |
| 1 Contexte et démarche scientifiques | 7 |
| 1.1 Le <i>big data</i> , un changement d'époque. Une journée-type en 2017. | 7 |
| 1.2 Des Systèmes de Transports Intelligents (ITS) aux Systèmes de Transports Intelligents Coopératifs (C-ITS) | 9 |
| 1.2.1 Introduction | 9 |
| 1.2.2 Les systèmes de transports intelligents (ITS) | 10 |
| 1.2.3 Les systèmes coopératifs | 10 |
| 1.3 Les évolutions des systèmes d'information multimodale aux usagers | 12 |
| 1.4 Défis et motivations scientifiques | 14 |
| 1.4.1 Défi 1 : anticiper le déploiement de nouvelles technologies | 15 |
| 1.4.2 Défi 2 : extraire des connaissances pour la caractérisation temps réel et la prévision court terme du trafic routier | 16 |
| 1.4.3 Défi 3 : améliorer la planification des trajets pour les usagers des réseaux de transports | 17 |
| 1.5 Démarche méthodologique | 18 |
| 1.5.1 Cheminement logique : trois objectifs de recherche par montées en échelle successives | 18 |
| 1.5.2 Question de recherche Q1 : modélisation et simulation des C-ITS | 20 |
| 1.5.3 Question de recherche Q2 : approches statistiques pour la caractérisation temps réel et la prévision court terme du trafic routier | 25 |
| 1.5.4 Question de recherche Q3 : optimisation des itinéraires sous contraintes complexes | 30 |
| 1.5.5 Organisation de la suite du manuscrit | 36 |
| 2 Modélisation, simulation et contrôle des C-ITS | 39 |
| 2.1 Modélisation théorique des C-ITS : quand multi-anticipation rime avec stabilité . . . | 39 |
| 2.1.1 Un modèle multi-anticipatif bilatéral (BMA) | 39 |
| 2.1.2 Principales conclusions des études de stabilité | 41 |
| 2.2 Modèle multi-agent par couplage de dynamiques | 43 |
| 2.2.1 Motivation et présentation générale du modèle | 43 |
| 2.2.2 Couche physique | 44 |
| 2.2.3 Couche de communication | 46 |
| 2.2.4 Couche de confiance | 46 |

| | | |
|----------|---|-----------|
| 2.3 | D'un simulateur de trafic microscopique à un simulateur multi-agent des C-ITS | 48 |
| 2.3.1 | Architecture | 48 |
| 2.3.2 | Perception-Décision-Action | 49 |
| 2.4 | Résultats en simulation | 51 |
| 2.4.1 | Données et plan expérimental | 51 |
| 2.4.2 | Impact des C-ITS sur la stabilité du trafic et prise en compte de la fiabilité de l'information | 51 |
| 2.4.3 | Stratégies de contrôle locales | 53 |
| 2.4.4 | Discussion | 56 |
| 2.5 | Contrôle par apprentissage constructiviste | 57 |
| 2.5.1 | État de l'art | 57 |
| 2.5.2 | Construction d'une représentation par apprentissage concurrent | 58 |
| 2.5.3 | Application au trafic coopératif | 61 |
| 2.6 | Conclusion et perspectives | 68 |
| 3 | Approches statistiques <i>data driven</i> pour la caractérisation et la prévision du trafic routier | 71 |
| 3.1 | Présentation et qualification des données | 71 |
| 3.1.1 | Le Bluetooth | 71 |
| 3.1.2 | Lecture Automatique de Plaques d'Immatriculation (LAPI) | 74 |
| 3.1.3 | Véhicules traceurs ou <i>Floating Car Data (FCD)</i> | 75 |
| 3.2 | Modélisation statistique des temps de parcours | 76 |
| 3.2.1 | Présentation du système de Halphen | 76 |
| 3.2.2 | Le diagramme de ratio des moments δ -MRD | 77 |
| 3.2.3 | Méthodologie d'évaluation des distributions concurrentes | 79 |
| 3.2.4 | Résultats empiriques | 80 |
| 3.2.5 | Mesurer la fiabilité des temps de parcours avec le δ -MRD | 81 |
| 3.2.6 | Discussion | 83 |
| 3.3 | Analyse spatiotemporelle d'un réseau de capteurs Bluetooth | 86 |
| 3.3.1 | Préparation des données | 86 |
| 3.3.2 | Représentation et représentativité | 87 |
| 3.3.3 | Clustering spatiotemporel | 90 |
| 3.3.4 | Discussion | 94 |
| 3.4 | Prévision court terme du trafic par apprentissage statistique | 95 |
| 3.4.1 | Utiliser les données pour prévoir le trafic à l'échelle du réseau : verrous et synthèse des contributions | 95 |
| 3.4.2 | Présentation de l'approche globale par l'exemple des k plus proches voisins | 96 |
| 3.4.3 | Processus d'apprentissage | 99 |
| 3.5 | Application au cas d'étude de Brisbane | 100 |
| 3.5.1 | Performances générales | 100 |
| 3.5.2 | Résilience à l'horizon de prévision et aux incidents | 102 |
| 3.6 | Résumé des autres résultats | 103 |
| 3.6.1 | Comparaison à d'autres méthodes sur le cas nantais, données FCD. Vers une approche portfolio ? | 103 |
| 3.6.2 | Identification de sections critiques | 105 |
| 3.6.3 | Traitement des externalités | 105 |
| 3.7 | Conclusion et perspectives | 106 |

| | | |
|----------|--|------------|
| 4 | Optimisation des itinéraires sous contraintes complexes | 107 |
| 4.1 | La plateforme TEMPUS : quand les besoins industriels inspirent les travaux de recherche | 107 |
| 4.1.1 | Présentation de TEMPUS, calculateur d'itinéraires multimodaux | 107 |
| 4.1.2 | Passage en <i>open source</i> et rencontre avec Mappy. Retour d'expérience. . . | 109 |
| 4.2 | Outils pour l'électromobilité | 111 |
| 4.2.1 | Génération d'un graphe énergétique | 112 |
| 4.2.2 | Affectation des stations de recharge | 114 |
| 4.2.3 | Algorithmes de plus court chemin énergétique | 115 |
| 4.2.4 | Quelques résultats | 117 |
| 4.3 | Plus court chemin stochastique : sensibilité aux lois de probabilité et techniques d'accélération | 119 |
| 4.3.1 | Impact des distributions de probabilité sur le plus court chemin stochastique | 119 |
| 4.3.2 | Vers une accélération du PCC stochastique à l'aide des contractions hiérarchiques | 125 |
| 4.4 | Conclusion et perspectives | 129 |
| 5 | Chapitre d'ouverture à la santé connectée : fouille de données et aide à la décision pour la prévention du risque suicidaire. | 131 |
| 5.1 | Actualités de la prévention du risque suicidaire | 131 |
| 5.1.1 | Nouvelles technologies et prévention | 131 |
| 5.1.2 | Vers un système d'aide à la décision dynamique | 133 |
| 5.2 | Fouille de données pour la prévention du risque suicidaire | 134 |
| 5.2.1 | Recrutement des patients et évaluation clinique | 135 |
| 5.2.2 | Méthodologie statistique | 136 |
| 5.2.3 | Qualification des données brutes et description du jeu de données final . . . | 136 |
| 5.2.4 | Méthodes de fouille de données | 137 |
| 5.2.5 | Résultats | 139 |
| 5.3 | Discussion | 141 |
| 5.3.1 | Limites et perspectives du travail | 141 |
| 5.3.2 | Vers une application mobile française pour la prévention du risque suicidaire | 142 |
| 6 | Bilan et projet scientifique | 147 |
| 6.1 | Vision globale d'un projet de recherche à long terme | 148 |
| 6.2 | Recherches futures | 149 |
| 6.2.1 | Faire l'interface entre la fouille de données, l'aide à la décision et la modélisation | 149 |
| 6.2.2 | Méthodes non supervisées pour l'identification de pattern d'activités | 151 |
| 6.2.3 | Indices de qualité et leur typologie | 152 |
| 6.2.4 | Approches supervisées pour la prévision court terme | 152 |
| 6.2.5 | Plus court chemin multi-objectif et stochastique | 153 |
| 6.2.6 | Aide Multicritère à la Décision (AMCD) | 153 |
| 6.3 | Exemple d'un projet type ANR en santé connectée | 154 |
| 6.3.1 | Contexte | 154 |
| 6.3.2 | Verrous scientifiques | 155 |
| 6.3.3 | Axes de recherches | 156 |
| 6.3.4 | Caractère innovant du projet | 157 |
| | Bibliographie | 159 |

| | | |
|----------|--|------------|
| A | Annexe : Curriculum Vitae | 189 |
| A.1 | Données personnelles | 189 |
| A.2 | Expériences professionnelles | 189 |
| A.3 | Thèmes de recherche | 190 |
| A.4 | Activités d'enseignement | 190 |
| A.5 | Encadrement de la recherche | 192 |
| A.5.1 | Étudiants en thèse | 192 |
| A.5.2 | Étudiants de Master (Stage de Fin d'Études, 6 mois) | 192 |
| A.5.3 | Étudiants de Master 1 et Licence (Stages de 4 mois) | 193 |
| A.6 | Activités de valorisation | 193 |
| A.6.1 | Contrats de recherche | 193 |
| A.6.2 | Logiciels | 194 |
| A.6.3 | Organisation de conférence et comités de programme | 194 |
| A.7 | Rayonnement scientifique | 195 |
| A.7.1 | Séjours académiques en tant que chercheur invité | 195 |
| A.7.2 | Séminaires | 195 |
| A.7.3 | Conférencier invité | 196 |
| A.8 | Activités éditoriales | 196 |
| A.9 | Compétences linguistiques et informatique | 196 |
| B | Annexe : publications de l'auteur | 199 |
| B.1 | Chapitres dans ouvrage collectif | 199 |
| B.2 | Articles dans des revues internationales à comité de lecture | 199 |
| B.3 | Autres revues à comité de lecture | 201 |
| B.4 | Articles dans des conférences internationales avec comité de lecture | 202 |
| B.5 | Articles dans des conférences nationales avec comité de lecture | 206 |
| | Table des figures | 209 |
| | Liste des tableaux | 213 |

Avant-propos

Ce mémoire d'habilitation présente une synthèse de mes réalisations académiques depuis la thèse de doctorat [52] jusqu'à ma récente arrivée à l'IMT Atlantique de Brest. Après une année de post-doctorat en 2011 à l'Université de Technologie du Queensland (QUT, Australie), j'ai effectué la majorité de mes recherches entre 2011 et 2016 au laboratoire d'Ingénierie Circulation Transport (LICIT) de l'IFSTTAR et l'ENTPE de Lyon en tant que chargé de recherches (CR2 puis CR1) du ministère de l'écologie. Après ces cinq années lyonnaises, j'ai été recruté sur un poste de maître de conférences en *data science* à l'IMT Atlantique (ex Télécom Bretagne), au sein de l'équipe DECIDE du lab-STICC (UMR CNRS 6285). Mes thèmes de recherches s'inscrivent dans une application large des mathématiques, à la croisée de l'analyse de données et de la recherche opérationnelle. Pendant les cinq premières années de ma carrière post-thèse, le domaine applicatif de la **mobilité intelligente** a constitué le fil d'Ariane de mes recherches. Sur la dernière année, mon recrutement à l'IMT Atlantique m'a permis d'approfondir un nouveau champ d'application, la **santé connectée**, tout en continuant mes travaux précédents. Les mutations technologiques ont amené, dans ces deux domaines, une hyper-connectivité, des données ouvertes et de masse (*big data*), qui ouvrent de formidables perspectives mais aussi une exigence nouvelle de la part des différents acteurs qui souhaitent tout connaître, tout de suite : supervision en temps réel des activités, analyse et prévision des trajets, informations ciblées sur les loisirs, conseils pour leur bien-être, etc. Il s'ensuit un besoin sans précédent de recherche afin de transformer ces masses d'information en connaissances. Au cours des précédentes années, mes travaux ont tenté de contribuer aux domaines disciplinaires de l'informatique et des mathématiques appliquées afin de mieux comprendre et caractériser la mobilité des usagers, ou fournir des outils d'aide à la décision pour la prise en charge de patients à risque en santé mentale. Parmi les deux champs d'application cités, la mobilité intelligente a concentré l'essentiel de mes travaux et le manuscrit s'articule logiquement autour de ce thème prépondérant dans mes recherches. Le thème de la santé connectée fait l'objet d'un chapitre d'ouverture à part, qui fait la transition entre les contributions passées et le nouveau projet de recherche. Ce projet rassemble les deux domaines d'application dans une perspective plus globale de fouille de données et optimisation pour des données hétérogènes. Ainsi, l'organisation générale du mémoire est la suivante. Le chapitre 1 pose les motivations de nos travaux sur la mobilité intelligente et donne une présentation originale et synthétique de notre démarche. Ce positionnement global de nos travaux au sein du domaine applicatif de la mobilité intelligente se fonde sur trois questions de recherche complémentaires qui sont détaillées dans les chapitres suivants. Le cheminement logique entre les trois thèmes de recherche s'appuie sur des montées en échelle successives, du véhicule au réseau, en passant par les données générées par les véhicules. Le chapitre 2 rassemble nos contributions sur la modélisation et la simulation des véhicules connectés (C-ITS). Ce travail est relatif à la question de l'anticipation des nouvelles technologies, il est centré sur le véhicule à l'aide d'une approche de type *model driven*. Le chapitre 3 présente quelques résultats sur la caractérisation

des états de trafic, à travers la modélisation des temps de parcours et de leur incertitude, l'analyse spatiotemporelle du réseau et la prévision court terme par des méthodes d'apprentissage statistique. Ce deuxième thème répond à la question de l'extraction de connaissances à partir des nouvelles sources de données, centré sur les nouvelles données générées par (entre autres) les nouveaux véhicules à l'aide d'approches de type *data driven*. Le chapitre 4 montre comment les résultats des deux premières approches peuvent offrir une meilleure planification des trajets à l'échelle d'un réseau. Il répond à la question de l'optimisation des itinéraires dans les réseaux de transports, il est centré sur l'utilisateur qui souhaite se déplacer d'un point A à un point B et le problème est résolu à l'aide d'algorithmes de recherche de plus court chemin. Le chapitre d'ouverture 5 synthétise nos premiers résultats sur le thème de la santé connectée, résultant de la collaboration avec le CHU Brest. Nous verrons que les méthodes d'apprentissage non supervisé et supervisé telles que celles utilisées au chapitre 3 permettent de faire avancer la recherche en santé mentale, par l'extraction de connaissances à partir de grandes bases de données patients. Ensuite, le chapitre 6 inscrira la mobilité et la santé connectées dans un projet de recherche global pour les années à venir, ayant trait à l'extraction de connaissances et l'optimisation sous contraintes complexes pour des données hétérogènes. Pour terminer ce mémoire, le lecteur trouvera en annexe un *curriculum vitae* complet et le détail de mes publications scientifiques. La lecture du manuscrit peut s'adapter à plusieurs types de consultation ou d'exigences. Le lecteur pressé, ou adepte des manuscrits d'habilitation très synthétiques, peut se contenter de la lecture des chapitres 1, 5, 6. Il y trouvera une synthèse de nos travaux sur la mobilité intelligente, un chapitre d'ouverture sur la santé connectée, et le projet de recherche. Le lecteur souhaitant plus de détails sur les contributions sur la mobilité intelligente pourra approfondir la lecture en consultant en sus les chapitres centraux 2, 3, 4.

ROMAIN BILLOT
Le Conquet, le 16 Août 2017

Liste des abréviations

Les abréviations suivantes sont classées par ordre alphabétique :

| | |
|--------------------------------------|---|
| A* | <i>A-star algorithm</i> |
| ACC | <i>Adaptive Cruise Control</i> |
| ACM | <i>Analyse des Correspondances Multiples</i> |
| ACP | <i>Analyse en Composantes Principales</i> |
| AICC | <i>Autonomous Intelligent Cruise Control</i> |
| ADAS | <i>Advanced Driver Assistance Systems</i> |
| ALT | <i>A-star search, Landmarks and Triangular Inequality</i> |
| AMCD | <i>Aide Multicritère à la Décision</i> |
| ANN | <i>Artificial Neural Networks</i> |
| ANOVA | <i>Analysis Of Variance</i> |
| ASIP | <i>Agence des Systèmes d'Information Partagés de santé</i> |
| B-FD | <i>Bluetooth Fundamental Diagram</i> |
| BGL | <i>Boost Graph Library</i> |
| BIS | <i>Barratt Impulsiveness Scale</i> |
| BMA | <i>Bilateral Multi-Anticipative model</i> |
| B-OD | <i>Bluetooth Origin Destination</i> |
| BT | <i>Buffer Time</i> |
| BTI | <i>Buffer Time Index</i> |
| C2C | <i>Car to Car (=V2V)</i> |
| C2I | <i>Car to Infrastructure (=V2I)</i> |
| C-ACC | <i>Cooperative Adaptive Cruise Control</i> |
| CAM | <i>Cooperative Awareness Message</i> |
| CDSS | <i>Clinical Decision Support System</i> |
| CH | <i>Contractions Hiérarchiques</i> |
| C-ITS | <i>Cooperative Intelligent Transportation Systems</i> |
| DENM | <i>Decentralized Environmental Notification Message</i> |
| δ-MRD | <i>Diagramme des ratios des moments δ_1-δ_2</i> |
| EMA | <i>Ecological Momentary Assessment</i> |
| EMD | <i>Enquête Ménages Déplacements</i> |
| ESP | <i>Electronic Stability Program</i> |
| FC | <i>Fréquence de Congestion</i> |
| FCD | <i>Floating Car Data</i> |
| GTFS | <i>General Transit Feed Specification</i> |
| GNSS | <i>Global Navigation Satellite System</i> |
| GPS | <i>Global Positioning System</i> |

| | |
|----------------------|--|
| GDM | <i>Global System for Mobile communications</i> |
| GNG | <i>Growing Neural Gas</i> |
| GRASP | <i>Greedy Randomized Adaptive Search Procedure</i> |
| HA | <i>Loi Halphen de type A</i> |
| Har | <i>Loi Harmonique</i> |
| HB | <i>Loi Halphen de type B</i> |
| HCPC | <i>Hierarchical Clustering on Principal Components</i> |
| HIB | <i>Loi Halphen de type B⁻¹</i> |
| I2C/I2V | <i>Infrastructure to Car/Vehicule</i> |
| IA | <i>Intelligence Artificielle</i> |
| IAD | <i>Intelligence Artificielle Distribuée</i> |
| IDM | <i>Intelligent Driver Model</i> |
| IF | <i>Indicateur Floridien</i> |
| ITS | <i>Intelligent Transportation Systems</i> |
| IVC | <i>Inter-Vehicular Communication</i> |
| kNN | <i>k Nearest Neighbours</i> |
| LAPI | <i>Lecture Automatique de Plaque d'Immatriculation</i> |
| LDA | <i>Latent Dirichlet Allocation</i> |
| LDWS | <i>Lane Departure Warning System</i> |
| LET | <i>Least Expected Time</i> |
| LIDAR | <i>Light Detection And Ranging</i> |
| LKS | <i>Lane Keeping System</i> |
| LWR | <i>Lighthill Whitham Richard</i> |
| MAC | <i>Media Access Control</i> |
| MFD | <i>Macroscopic Fundamental Diagram</i> |
| MI | <i>Misery Index</i> |
| MIMO | <i>Multi-Input Multi-Output</i> |
| MovSim | <i>Multi-model Open-source Vehicular-traffic SIMulator</i> |
| MPC | <i>Model Predictive Control</i> |
| ORP | <i>Optimal Routing Problem</i> |
| OSM | <i>Open Street Map</i> |
| PCC | <i>Plus Court Chemin</i> |
| PCCS | <i>Plus Court Chemin Stochastique</i> |
| PET | <i>Post-Encroachment Time</i> |
| PL | <i>Poids Lourd</i> |
| PMV | <i>Panneau à Message Variable</i> |
| PTI | <i>Planning Time Index</i> |
| RCL | <i>Restricted Candidate List</i> |
| RMSE | <i>Root Mean Square Error</i> |
| RRRS | <i>Risk-Rescue Rating Scale</i> |
| RSU | <i>Road Side Unit (= UBR)</i> |
| SFP | <i>Set of Feasible Paths</i> |
| SIM | <i>Système d'Information Multimodale</i> |
| SIS | <i>Suicide Intent Scale</i> |
| SMA | <i>Système Multi-Agent</i> |
| SOC | <i>State Of Charge</i> |

| | |
|----------------------|---|
| SOM | <i>Self-Organizing Maps</i> |
| SOTA | <i>Stochastic On-Time Arrival</i> |
| SPOTAR | <i>Shortest path Problem with On-Time Arrival Reliability</i> |
| SSPP | <i>Stochastic Shortest Path Problem</i> |
| SVM/SVR | <i>Support Vector Machine/Regression</i> |
| TIC | <i>Technologies de l'Information et de la Communication</i> |
| TMC | <i>Traffic Management Center</i> |
| TS | <i>Tentative de Suicide</i> |
| TTC | <i>Time To Collision</i> |
| TTS | <i>Total Time Spent</i> |
| UA | <i>Unité Application</i> |
| UBR | <i>Unité de Bord de Route (=RSU)</i> |
| UCB | <i>Upper Confidence Bound</i> |
| UEV | <i>Unité Embarquée Véhicule</i> |
| VANET | <i>Vehicular Ad Hoc Network</i> |
| V2I | <i>Vehicle to Infrastructure</i> |
| V2V | <i>Vehicle to Vehicle</i> |
| VSL | <i>Variable Speed Limit</i> |
| WAVE | <i>Wireless Access in Vehicular Environments</i> |

” [...] nous ne savons pas transformer ces data, c'est-à-dire les informations, en savoir. Nous faisons une incroyable confusion : les gens croient qu'en étant informés, ils savent quelque chose. Or, pas du tout ! Un savoir n'est pas seulement une information, mais la transformation de celui qui sait par ce qu'il apprend. L'information, elle, ne transforme rien. [...]

— **Bernard Stiegler**
Philosophe

1.1 Le *big data*, un changement d'époque. Une journée-type en 2017.

Sept heures du matin, un lundi comme beaucoup d'autres, une jeune salariée d'une grande entreprise française se réveille pour entamer sa semaine de travail. Comme pour beaucoup d'entre nous, les technologies sont devenues omniprésentes dans son quotidien. Avec sa montre connectée, qui lui sert aussi de coach sportif pendant ses *joggings* ou sorties cyclistes, elle a la possibilité, dès son réveil, d'avoir un résumé de la qualité de son sommeil, accompagné d'un bilan des derniers mois. Les données de pulsations cardiaques mesurées par la montre sont transmises en wifi à l'application mobile installée sur son téléphone dernière génération, puis les cycles de sommeil sont analysés afin de construire des indicateurs de qualité faciles à apprivoiser, comme un score global, et ainsi encourager une supervision ludique et régulière de ses nuits. C'est sa meilleure nuit depuis bien longtemps et elle s'empresse de partager ses résultats, par SMS, avec son meilleur ami, puis sur les réseaux sociaux *via* Facebook et Twitter. Les messages de félicitations affluent pour saluer cette « performance » dans le monde de la santé connectée ! Pendant la douche, les services de musique en ligne de *streaming* comme Spotify ou Deezer proposent une playlist « réveil en musique », construite à partir des préférences et des commentaires de milliers d'utilisateurs. Il est possible d'indiquer son avis sur tel ou tel morceau pour que le logiciel adapte le planning musical en temps réel, à l'aide d'un puissant système de recommandation fondé sur les données historiques. Elle savoure son petit-déjeuner et se prépare à partir au travail lorsque le compte Twitter d'un grand opérateur de transports en commun, auquel elle est abonnée, signale un incident sur le réseau de transports, source de nombreuses perturbations. Elle décide alors d'affronter les bouchons matinaux en voiture, espérant ne pas arriver trop tard sur son lieu de travail. Elle planifie son itinéraire d'un point A à un point B sur le site de Mappy puis se met en route. Pour l'assister dans le choix de l'itinéraire, elle se connecte à une application d'information trafic et de navigation communautaire comme *Waze* au sein de laquelle les conducteurs créent de l'information routière à travers les traces GPS générées par leur appareil au cours du trajet pour mettre à jour une carte

d'information trafic temps-réel. Il est possible d'indiquer soi-même un événement particulier sur le réseau et notre conductrice signale un ralentissement causé par un accident. Elle décide de prendre l'itinéraire alternatif proposé par l'application. Arrivée au travail, elle partage sa frustration d'une matinée difficile dans les transports, sur les réseaux sociaux. Pendant sa journée de travail, en plus de son activité professionnelle, elle se sera connectée sur internet pour vérifier le solde de son compte bancaire et faire ses courses sur une application *drive* d'un grand supermarché qui permet de faire ses courses en ligne et de les récupérer en voiture. L'application de courses en ligne propose une vue historique et très bien résumée de son profil de consommatrice des derniers mois, ainsi qu'une suggestion de produits susceptibles de l'intéresser. Sur la route du retour, le coffre chargé de victuailles, une œuvre de *street art* peinte sur un mur attire immédiatement son attention. Elle s'arrête pour la prendre en photo, la retoucher à l'aide d'un filtre de couleurs et la partager sur un réseau social de type Instagram. La photo reçoit immédiatement une dizaine de mentions « j'aime ». Dans la soirée, un ami commente la photo, et ayant reconnu l'artiste, lui donne un lien sur un site de vidéos en ligne comme Youtube. Ce lien est une vidéo de la réalisation de l'œuvre *street art* mise en ligne par l'artiste pour accroître sa visibilité. Elle s'empresse de la visualiser puis, éreintée, dîne rapidement, branche son application sommeil et part se coucher.

Du lever au coucher, notre héroïne a généré un nombre significatif de données, volume qu'il aurait été difficile d'imaginer quelques années auparavant. Parfois en connaissance de cause, parfois à son insu, des centaines de mégaoctets de données transitent, comme autant de traces numériques de ses goûts, humeurs, envies, recherches, localisations, etc. Cette *homo sapiens*, devenue *homo numericus*, n'est pas seule, nous sommes des milliards à faire de même. Les chiffres sont éloquentes, et leur croissance vertigineuse : nous sommes entrés dans l'ère des *big data*. Depuis les débuts de l'informatique, les ingénieurs, chercheurs, ont bien sûr été confrontés à une forte croissance des volumes de données, stockées dans des bases de données de plus en plus grandes, appelées ensuite entrepôts de données, aux architectures toujours plus performantes pour garantir un haut niveau de service. Mais l'internet mobile et l'internet des objets, entre autres, ont entraîné depuis les années 2000 une véritable explosion des données, plus ou moins bien maîtrisée, qui nécessite de repenser les schémas classiques, aussi bien en termes d'architecture que de traitements de données. Trafic/navigation internet, sauvegardes sur le *cloud*, partages sur les réseaux sociaux, données ouvertes (*open data*), transactions d'achats, capteurs, historiques d'objets connectés, constituent un ensemble de marqueurs, spatiaux et/ou temporels, de l'activité humaine dans toutes ses dimensions. Nous produisons énormément de données et sommes en situation d'en produire depuis n'importe quel endroit en permanence (Internet est accessible depuis le bureau, la maison, les aéroports, les trains, la voiture, les restaurants, etc.) et en quelques clics, un individu, peut, par exemple, décrire, noter un repas et expédier une photo de son plat. Ces grandes masses de données posent certes des questions, éthiques et de sécurité entre autres, mais représentent également un fort potentiel pour la société [63]. Des jeux de données qui étaient auparavant obscurs ou réservés à une élite deviennent accessibles au plus grand nombre. Il en va de même du phénomène des données ouvertes qui s'installe à toutes les échelles des administrations. Pour les grandes entreprises, les enjeux sont multiples [87]. Au niveau du stockage de ces données, il s'agit de mettre en place une architecture souple et adaptée [484] permettant de traiter de façon parallèle et dynamique des données « volumineuses », « variées », « véloces », en évaluant leur « véracité », afin de tirer parti de la grande « valeur » de ces nouvelles masses de données [465]. C'est par le truchement de ces 5 V que les *big data* ou mégadonnées sont souvent présentées.

Mais revenons à l'exemple préliminaire et à cette journée-type d'une personne hyperconnectée de notre siècle. Alors que le temps se contracte irrésistiblement pour un grand nombre d'activités, il reste un moment pour "éprouver le temps dans tout sa longueur"¹ : le trafic routier, les bouchons ! Comme l'évoque souvent le professeur Markos Papageorgiou dans ses conférences plénières, l'homme est capable d'aller sur la lune mais n'est toujours pas capable de résoudre ses problèmes de congestion. Si des milliards de contenus peuvent s'échanger à la seconde, la téléportation n'existe toujours pas et les hologrammes, aussi brillants soient-ils, ne suffisent pas. De plus en plus d'êtres humains cherchent à se déplacer physiquement dans des réseaux de plus en plus saturés. L'explosion du volume d'information disponible rend les usagers des réseaux de transports toujours plus exigeants sur les services proposés. Les chercheurs et ingénieurs en gestion du trafic routier s'inscrivent dans ces grandes mutations technologiques pour proposer des solutions innovantes.

1.2 Des Systèmes de Transports Intelligents (ITS) aux Systèmes de Transports Intelligents Coopératifs (C-ITS)

1.2.1 Introduction

Qu'il semble loin le temps où la gestion du trafic s'appuyait essentiellement sur des comptages de véhicules à l'aide de boucles électromagnétiques ! Les systèmes de transports intelligents (ITS – *Intelligent Transportation Systems*), thématique prégnante à l'INRETS, au LCPC puis à l'IFSTTAR depuis les années 1990, s'appuient sur les technologies de l'information et la communication comme autant d'outils d'aide à la gestion du trafic. Ils ont connu une évolution très rapide ces dernières années, grâce à un équipement croissant des véhicules, qui peuvent fournir leurs traces GPS ou devenir des véhicules traceurs (FCD – *Floating Car Data*), mais sont également équipés de systèmes d'aide à la conduite qui renseignent le conducteur sur la stabilité de son véhicule ou les aspects latéraux et longitudinaux. En sus des conditions météorologiques dégradées comme une visibilité faible, les véhicules sont maintenant capables de détecter des obstacles ou des zones où l'adhérence est réduite. Avec cette hyper-informatisation du véhicule par rapport à son environnement extérieur, et les véhicules autonomes prototypes qui impressionnent dans les médias (comme la *google car*), il est tentant de penser que le véhicule « sans conducteur » est pour très bientôt. Pourtant, si la technologie est presque mature, ces véhicules sont encore bien trop chers et trop de questions restent en suspens quant à leur acceptabilité d'un point de vue public ou juridique. C'est plutôt une étape intermédiaire vers l'automatisation qui se joue actuellement, celle des systèmes coopératifs, deuxième génération des systèmes ITS. Ils intègrent les technologies de communication C2C (*car to car*), C2I (*car to infrastructure*) et I2C (*infrastructure to car*). En communiquant entre eux mais aussi avec l'infrastructure à l'aide d'unités de bord de route (UBR ou RSU *road side units*), des cas d'usage précis se développent (ex : freinage d'urgence) et la gestion de trafic peut s'envisager d'une façon nouvelle, par exemple en prévenant les véhicules d'une perturbation en aval ou en diffusant d'autres informations ciblées. Une vraie dynamique s'est créée autour des véhicules connectés, comme en témoigne le projet étandard Scoop@F porté par le ministère de l'écologie.

1. un clin d'oeil à l'épigraphe de la thèse de Raphaël Delhome, citation d'Albert Camus

1.2.2 Les systèmes de transports intelligents (ITS)

Les systèmes de transports intelligents (STI) ou *Intelligent Transport Systems* (ITS) constituent un terme chapeau utilisé initialement pour caractériser l'avènement des technologies de l'information et de la communication dans les systèmes avancés de gestion du trafic. Tout le monde connaît Bison Futé, initié dès 1976 comme un centre d'information routière à destination des usagers *via* les différents médias (radio, télévision). À l'époque, les technologies de recueil étaient rares et l'information pouvait être directement communiquée par des opérateurs en charge du comptage manuel des véhicules. Dans les années 1980, les panneaux à message variable (PMV) affichaient déjà aux automobilistes une information temps-réel sur les conditions de circulation. Les mesures de trafic ont ensuite su se développer pour répondre aux besoins d'exploitation et de régulation des flux, d'optimisation du réseau et représenter pour l'exploitant un moyen efficace d'évaluation et de régulation. Au début des années 1990, l'État a mis en place le réseau SIREDO (Système Informatisé de REcueil de DONnées), vaste ensemble de stations destinées à recueillir des informations sur le réseau routier et autoroutier français. Eu égard aux problématiques de gestion du trafic, ces données trafic permettent la connaissance des conditions de circulation et l'évaluation des performances et des mesures de gestion. Les capteurs utilisés étaient des boucles électromagnétiques implantées dans la chaussée. Si les premiers besoins se satisfaisaient de données moyennes journalières annuelles, les problèmes de saturation du réseau ont vite soulevé l'importance d'un stockage plus massif en vue d'une exploitation en temps réel et mené à la collecte de données agrégées toutes les 6 minutes voire à la minute. Au même moment sont apparus des systèmes de régulation et de gestion du trafic. Ils se focalisent d'abord sur les transports en commun avant de s'intéresser aux véhicules particuliers. Des systèmes comme la régulation d'accès [343], la gestion dynamique des voies ou des vitesses, se mettent en place pour réduire les congestions. Les années 2000 ont été le théâtre de la démocratisation des récepteurs GPS dans les véhicules particuliers. Les données temporelles en un point fixe ont ainsi été complétées par des données trajectoires qui retracent la position des véhicules sur une certaine zone d'étude. Mais la décennie précédente a surtout été celle de l'équipement du véhicule et de la multiplication des capteurs intra-véhiculaires à travers les systèmes embarqués. Cette catégorie regroupe les systèmes d'aide à la conduite ou systèmes avancés d'assistance au conducteur (ADAS : *Advanced Driver Assistance Systems*). Ce sont des dispositifs automatiques intégrés aux véhicules destinés à assister le conducteur dans ses tâches de conduite et le cas échéant corriger d'éventuelles erreurs pour apporter un bénéfice en termes d'anticipation ou de perception. Il existe des systèmes actifs qui agissent directement sur la dynamique du véhicule comme le maintien de trajectoire (LKS : *Lane Keeping System*), et des systèmes passifs qui informent le conducteur ou les autres usagers comme le système d'avertissement de franchissement de ligne (LDWS : *Lane Departure Warning System*). D'autres systèmes notables reposent sur une automatisation partielle du comportement longitudinal du véhicule. On peut citer le régulateur de vitesse à contrôle de distance (ACC : *Adaptive Cruise Control*). Ce système se compose d'un capteur de proximité permettant au véhicule de percevoir la distance et la vitesse relative avec son véhicule prédécesseur (*leader*). L'étape suivante consiste à introduire une forme de communication qui ouvre la voie à une coopération entre véhicules.

1.2.3 Les systèmes coopératifs

Les systèmes coopératifs sont un parfait exemple d'un système complexe composé d'entités fixes et mobiles connectées les unes aux autres. Ils constituent la deuxième génération des ITS. Les

informations dynamiques à propos des conditions de circulation et des paramètres véhiculaires (position, vitesse) sont échangées à travers des protocoles de communication *vehicle to vehicle (V2V)*, *infrastructure to vehicle (I2V)* et *vehicle to infrastructure (V2I)*, comme l'illustre la figure 1.1. Les UBR collectent l'information sur une section de route couverte. Ces équipements peuvent avoir des rôles plus ou moins raffinés, de la simple collecte de données à la régulation du trafic. Pour étendre le champ de perception des usagers, il est envisagé d'utiliser les UBR comme des relais d'information. Les informations échangées (I2V) sont alors des notifications concernant des événements dangereux [157] ou des conditions de trafic dense [341]. Pour le moment, les UBR sont déployées sur des sites pratiques car équipés ou stratégiques (présence récurrente de congestion) mais la question du déploiement optimal se pose face aux coûts d'installation [33].

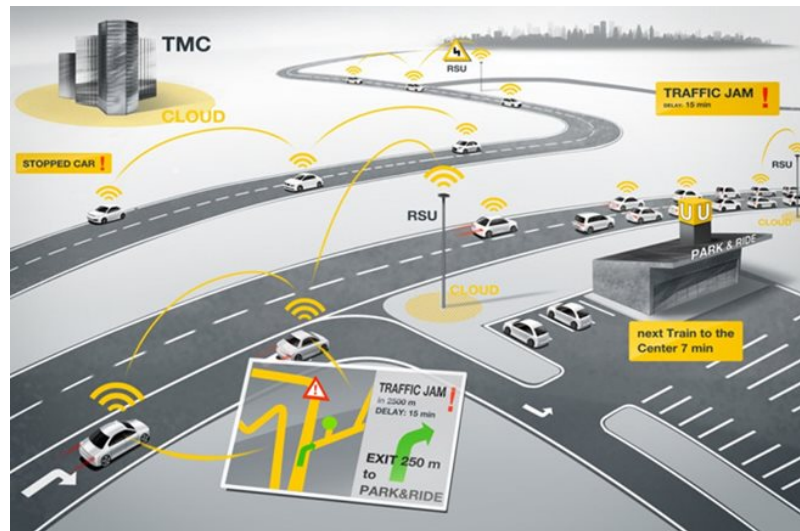


Fig. 1.1.: Echanges d'information dans les C-ITS.
(Source : <http://www.kapsch.net>)

Au niveau architecture, un véhicule dit coopératif est composé d'une unité embarquée véhicule (UEV) et d'une unité d'application (UA). L'UA est le dispositif qui exécute les applications et tire profit des capacités de communication de l'UEV. L'UA et l'UEV peuvent être reliées par connexion sans fil (Bluetooth) ou filaire. Les C-ITS peuvent être vus comme une extension de réseaux ad hoc ou *Vehicular Ad Hoc Network (VANET)*. Il repose avant tout sur la capacité de communication entre les entités de ces systèmes. Les véhicules proviennent de différents constructeurs, et les gestionnaires d'infrastructure diffèrent d'un pays (voire d'une région) à l'autre. Pour parvenir à faire communiquer tous ces acteurs, il est nécessaire que ces différents systèmes inter-opèrent entre eux à défaut d'une standardisation du protocole de communication, à la fois entre les véhicules (V2V) mais aussi avec les unités d'infrastructure (V2I/I2V). Les dernières années ont donc été consacrées à l'émergence de l'inter-opérabilité et de nouveaux standards dans le monde des communications inter-véhiculaires. L'évolution des technologies de communication mobiles avec l'avènement récent de la 4G a pendant un temps provoqué un regain d'intérêt pour leur utilisation dans les ITS. Le besoin d'un réseau dédié et intégrant davantage la problématique liée à la sécurité a cependant pris l'ascendant lors de l'effort de standardisation notamment opéré par le consortium *Car2Car Communication*. Une nouvelle norme baptisée WiFi 802.11p et aussi connue sous le nom de WAVE (*Wireless Access in Vehicular Environments*) a été préférée. Ce standard assure des communications à courte portée (moins de 300 mètres) entre des entités fixes ou mobiles sur des canaux dédiés. Le schéma d'échange retenu est la diffusion (*broadcast*). Des tests grandeur nature ont montré la robustesse du protocole

retenu avec des valeurs de latences observées [336] compatibles avec des applications dédiées à la sécurité (de l'ordre de 100 ms).

Les C-ITS préfigurent une nouvelle façon de gérer les flux de trafic grâce à l'interconnexion de tous les acteurs, de l'utilisateur jusqu'au gestionnaire. Les données reçues peuvent être traitées en temps réel et renvoyées aux véhicules sur la section considérée, sous forme individuelle ou groupée. En étendant les possibilités offertes par les systèmes autonomes au-delà des simples risques de collision, les systèmes coopératifs permettent au conducteur d'interagir avec son environnement et d'avoir une meilleure perception des conditions de circulation. Par exemple, une information sur une perte d'adhérence par conditions pluvieuses prendra encore plus d'importance si elle est couplée à des indications sur la circulation en aval. Ces nouvelles technologies ouvrent la voie à une régulation de vitesse adaptative et coopérative (*Cooperative and Adaptive Cruise Control – CACC* [433]). Au niveau de la sécurité routière, l'information en temps réel sur le niveau de risque ne profite pas seulement à l'utilisateur mais aussi au gestionnaire à travers des délais d'intervention réduits. De plus, le fait d'avoir accès à une gestion facilitée de la vitesse et des interdistances par une programmation aisée des limiteurs de vitesses ou régulateur de vitesse pourrait, en plus de transmettre en temps réel une information liée au risque de collision en cas de conduite à risque, être également un facteur d'incitation au développement de véhicules électriques. Les gestionnaires de trafic auront un rôle important à jouer dans l'assimilation des différentes sources de données et le contrôle résultant sur les véhicules. À travers la communication infrastructure-véhicule, les actions de régulation de vitesse pourraient être ciblées par catégories de véhicules, des véhicules plus modérés en termes d'émissions de Co2 (covoiturage, auto-partage, véhicules électriques) étant informés différemment que d'autres. La notion de contrôle d'accès (*ramp metering*) pourrait également être revisitée à la lumière des possibilités offertes par les systèmes coopératifs. Les véhicules souhaitant s'insérer recevraient une information pertinente de la part des gestionnaires sur la vitesse à adopter ou pourquoi pas une redirection en cas de congestion. D'autre part, en milieu urbain, il semble prometteur de repenser le concept d'ondes vertes, en portant le travail non seulement sur la synchronisation optimale des feux mais aussi et surtout sur les informations à fournir aux véhicules pour fluidifier le trafic et limiter la formation de longues files de véhicules aux intersections. Ces applications sont des cas concrets d'utilisation des systèmes coopératifs, qui s'inscrivent dans les objectifs des premiers projets visant à déployer les C-ITS comme *Connected Vehicle Safety Pilot*² ; *ITS Spot Service*, le Corridor C-ITS reliant Rotterdam à Vienne, et l'ambitieux projet pilote français nommé SCOOP@F³.

1.3 Les évolutions des systèmes d'information multimodale aux usagers

Une application directe des C-ITS réside dans la diffusion aux usagers d'une information fiable, temps réel ou prédictive, sur leurs conditions de circulation. Ce levier de gestion de trafic repose sur le postulat qu'une meilleure information des usagers permettrait de dissiper la congestion aux niveaux temporel et spatial. Par exemple, l'information de temps de parcours aide à choisir les itinéraires appropriés. Communiquée avant le trajet, elle autorise une flexibilité quant à l'heure de départ. Le temps de parcours est une variable facile à mesurer en milieu interurbain mais beaucoup moins à

2. Site internet du projet Safety Pilot : <http://safetypilot.umtri.umich.edu/>

3. Site internet du projet SCOOP : <http://www.scoop.developpement-durable.gouv.fr>

prédire. En milieu urbain, tout se complique : fortes dynamiques court terme, non conservation du flot (ex : parking), topographie fluctuante (ex : travaux). Dans le contexte des C-ITS, les véhicules connectés décrits précédemment s'ajoutent progressivement au parc des véhicules fournisseurs de données. Une autre source de données prometteuse est le Bluetooth, présent dans la plupart des appareils électroniques (téléphones, systèmes de navigation, lecteurs musicaux). L'intérêt d'une telle technologie pour la caractérisation du trafic sera présenté dans le chapitre 3. Ainsi, les C-ITS doivent permettre des estimations de temps de parcours plus fiables. Cette information doit ensuite être retranscrite à l'échelle des différents modes de transports et de l'ensemble du réseau de transports urbain. L'objectif est de favoriser un transfert modal, principalement de l'automobile vers les transports en commun, et cette quête du Graal monopolise tout un ensemble d'acteurs responsables de la gestion de la mobilité et des enjeux énergétiques ou environnementaux. La part des trajets en voiture reste trop importante alors que le secteur des transports reste à l'origine d'une proportion significative d'émissions globales de gaz à effets de serres. À Paris, depuis 2012, on attribue ainsi au trafic routier plus de 50% des émissions de polluants [10]. Il est donc crucial de favoriser le recours des usagers à des modes de transport alternatifs, surtout en milieu urbain où la grande majorité des déplacements concerne des petites distances. Ce report ne s'effectue pas par la seule volonté politique ; il est fonction de comportements individuels et collectifs et de la situation relative de l'offre et de la demande (concurrence/complémentarité des modes) au niveau des territoires. Cette complémentarité s'articule autour du temps relatif de déplacement entre modes, du coût et de la valeur du temps mais aussi de bien d'autres composantes psychologiques ou socio-économiques. Le choix modal et plus encore le report modal reposent aussi sur les technologies de l'information. Ces technologies mises au service de la collecte, de l'organisation et de la diffusion collective ou individualisée d'informations sur l'état du trafic routier, sur l'offre de transport collectif, sur l'offre de disponibilité de parking, participent aussi à une offre de marché de plus en plus compétitive et mondialisée. Aujourd'hui l'aspect multimodal est incontournable. Des Systèmes d'Information Multimodale (SIM) sont apparus sur nos territoires sous l'autorité de communautés d'agglomération ou urbaine, des départements et des régions. Un SIM à destination des usagers comprend⁴ :

- la mise à disposition de plans du réseau (interactifs ou en téléchargement),
- la consultation des lignes, horaires,
- un service de recherche d'itinéraires,
- des informations sur l'état du trafic et les éventuelles perturbations.

D'un point de vue recherche, la dernière brique constituante d'un SIM, après le traitement de données et l'estimation des temps de parcours, réside dans la proposition d'algorithmes de calculs d'itinéraires (problème du plus court chemin, optimisation dans les grands graphes routiers). Ces méthodes visent à intégrer et offrir des optimums sous contrainte (temps, accessibilité, coût, confort) de trajets en intégrant des critères de choix modaux, qui balayent l'ensemble des itinéraires possibles par mode ou intermodal entre deux points, ou même entre plusieurs points de chaînage complets de déplacements au cours d'une journée. La disponibilité d'une information totale et en temps réel reste, à notre avis, la seule à même de favoriser un transfert modal raisonné et effectif.

4. À retrouver sur <http://www.transport-intelligent.net/>

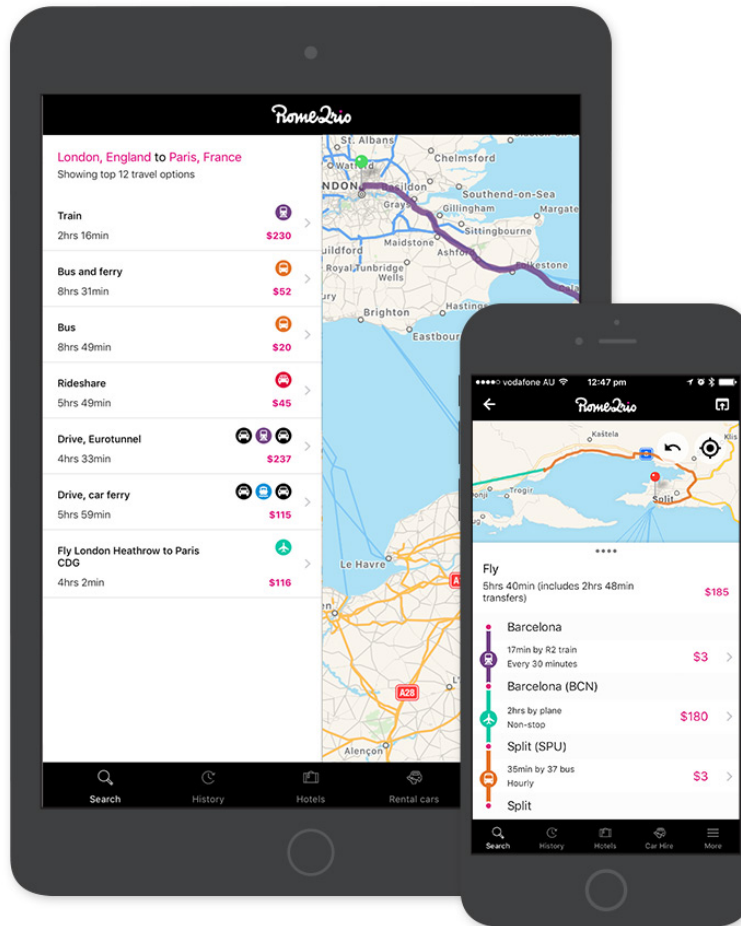


Fig. 1.2.: Exemple d'un calculateur d'itinéraires : Rome2Rio (source : <https://www.rome2rio.com/fr/>)

1.4 Défis et motivations scientifiques

Quelle est la place du chercheur dans le monde de la mobilité intelligente ? Après tout, *Google* ou *Microsoft* ne s'occupent-ils pas de récolter toutes les mégadonnées et de nous fournir les meilleures solutions sur nos smartphones pour prévoir nos trajets ? Et *BMW*, *Mercedes*, *Peugeot* ou *Renault* de concevoir, tester, déployer les véhicules connectés ? La question mérite d'être posée tant les solutions livrées "clé en main" semblent convaincantes. Néanmoins, évacuons-là d'emblée car la recherche est omniprésente dans tous les sujets abordés et les grands groupes mentionnés renforcent d'années en années leurs équipes R&D de chercheurs souvent à la pointe de la recherche scientifique, comme en témoignent les listes d'articles acceptés dans de grandes conférences en intelligence artificielle (au sens large). En réalité, de nombreux défis scientifiques émergent de ces nouvelles technologies, la plupart étant loin d'être résolus malgré le défilement des années. Par la suite, nous avons choisi de mettre en avant trois défis autour desquels gravitent un certain nombre d'enjeux et de verrous scientifiques. Ces verrous nous permettent de formuler à chaque fois des questions de recherche et contributions plus précises. L'articulation logique entre ces trois questions de recherche sera présentée dans la section suivante.

1.4.1 Défi 1 : anticiper le déploiement de nouvelles technologies

Notre participation à des projets industriels comme Co-Drive (avec Valéo) ou Scoop@F nous a permis de mettre en lumière des besoins en recherche sur les véhicules connectés et les C-ITS en général. L'écueil majeur, et évident, dès le début de nos activités sur ce thème en 2011, concernait l'absence de véhicules connectés ou autonomes. Aujourd'hui encore, les projets de déploiement sont localisés sur des sites pilotes. Le projet Scoop@F en a identifié 5, éparpillés aux quatre coins de la France. Les conclusions des expérimentations ciblées, la plupart du temps limitées à des sections d'autoroutes, seront donc toutes relatives et des études en modélisation et simulation sont indispensables pour étendre les conclusions à des échelles raisonnables. Dans ce cadre restreint, les efforts des constructeurs se limitent trop souvent à des cas d'usage critiques liés à des enjeux de **sécurité**. Les impacts des C-ITS sur la **gestion du trafic** et l'**environnement** sont occultés alors qu'un certain nombre de questions se posent indubitablement :

1. Les véhicules connectés vont-ils vraiment améliorer le trafic en terme d'indicateurs d'efficacité, de sécurité, ou d'émissions environnementales ?
2. Que se passera-t-il en cas de trafic mixte, entre véhicules non équipés et véhicules équipés, lors des phases successives de déploiement des C-ITS ?
3. À partir de quelle fraction de véhicules équipés (*i.e.* le taux de pénétration du marché) les effets sur le trafic sont-ils significatifs et fiables ?
4. Quelles informations doivent-être envoyées aux véhicules connectés pour réduire l'hétérogénéité du trafic ou le nombre de situations critiques ?
5. Quelles sont les stratégies de contrôle optimales d'une flotte de véhicules équipés ?
6. Comment prendre en compte les sources d'incertitude (ex : fiabilité de l'information) ?

Ces questions peuvent s'articuler autour d'un premier grand défi : **comment anticiper le déploiement de nouvelles technologies, en particulier de nouveaux types de véhicules ?**

La première interrogation peut sembler surprenante ou faussement naïve mais il est regrettable qu'elle soit trop rarement abordée du point de vue de la gestion de trafic. Les constructeurs mettent en avant un sur-équipement et une hyperconnectivité des véhicules, illustrés par quelques cas d'usage censés prouver les bienfaits des nouveaux véhicules en terme de sécurité. Les expérimentations, limitées à quelques véhicules, valident des spécifications fonctionnelles sans illustrer de potentiels bénéfiques sur l'écoulement du trafic, la congestion, les situations à risque ou les émissions environnementales. Cependant, une fraction partielle de véhicules équipés au sein d'un trafic hétérogène ne pourrait-elle pas créer des « interférences » au niveau des comportements de conducteurs, informés ou non, menant à des effets contre-productifs ? Ce besoin de travaux de **modélisation et simulation des C-ITS** a motivé une première direction de nos recherches. Ces recherches ont dû aborder la question d'un trafic mixte entre véhicules classiques et véhicules équipés/autonomes. Il est attendu que les phases successives de déploiement des systèmes coopératifs engendreront de nouveaux comportements, soit du conducteur au sein de son véhicule doté de nouvelles technologies, soit de la part des autres usagers face aux nouveaux types de véhicules rencontrés. Une troisième question est relative au taux de pénétration du marché des véhicules équipés. À partir de quel pourcentage les bénéfiques (ou désagréments) se feraient-il sentir ? Seules des études en simulation réalistes et reproductibles permettent de répondre à cette question. Enfin, la communication entre véhicules et avec l'infrastructure amène logiquement à revisiter la gestion dynamique du trafic, et les leviers que sont l'offre et la demande. Face à toutes les possibilités entreouvertes, il convient de déterminer quelles informations envoyer aux véhicules et plus généralement quelles nouvelles

stratégies de contrôle inventer dans un flux de trafic qui va irrémédiablement se métamorphoser à moyen terme. Une gestion coopérative du trafic doit s'adapter à la composition du flux et à l'état du trafic. La nature des informations à envoyer, leur réception et utilisation par les conducteurs, et les effets sous-jacents sur le flot de véhicules, constituent assurément un point critique devant susciter un effort inter-disciplinaire. De plus, la fiabilité de l'information, la variation des rayons de communication, les temps de latence, sont autant de sources d'incertitude devant être prises en compte dans les modèles. L'ensemble de ces défis se répercutent en autant de verrous scientifiques que nous détaillerons à la section 1.5. Ces défis et verrous ont motivé notre premier objectif de recherche sur le thème de la modélisation et simulation des C-ITS.

1.4.2 Défi 2 : extraire des connaissances pour la caractérisation temps réel et la prévision court terme du trafic routier

L'exemple introductif l'a montré : la majorité des individus (surtout les plus urbains d'entre eux), produisent des traces numériques de leurs activités et la mobilité n'échappe pas au phénomène *big data*. De nouvelles bases de données sont disponibles et les applications en matière de gestion de trafic se développent. Les gestionnaires d'infrastructures se doivent d'être présents sur les fronts du recueil et de la diffusion d'une information riche aux usagers. Toutefois, un long chemin sépare une information brute, de surcroît fournie par une nouvelle source de données mal maîtrisée, d'une connaissance exploitable pour la supervision et la prévision de l'état d'un réseau routier. Ce processus constitue l'essence même de toute démarche de fouille de données ou plus généralement d'extraction de connaissances. Ce deuxième défi comporte différents enjeux pour le gestionnaire d'infrastructures :

1. Evaluer le potentiel de nouvelles sources de données (ex : Bluetooth, FCD) pour la gestion de trafic,
2. Fournir des estimations fiables des temps de parcours sur les tronçons du réseau, qui prennent en compte l'incertitude,
3. Utiliser les données pour extraire de nouveaux indicateurs du fonctionnement d'un réseau, en particulier selon des aspects spatiotemporels, et de sa vulnérabilité,
4. Prévoir les états de trafic à court terme, à l'aide des grands historiques de données,
5. Détecter les incidents non récurrents de manière fiable.

Le premier enjeu a trait à l'évaluation d'une nouvelle technologie. Les données Bluetooth ou FCD sont de séduisantes solutions en termes de coût et de couverture spatiale du réseau (un véhicule peut émettre une position de n'importe où tandis que les capteurs Bluetooth peuvent être rapidement déployés à l'échelle d'une ville). Néanmoins, cette information est très partielle et il arrive qu'une poignée de véhicules (parfois un seul) soit utilisée pour afficher une information trafic (de type trafic coloré) à l'utilisateur sur une carte routière. La question de la représentativité des résultats se pose clairement, d'autant plus que les problématiques de filtrage de données aberrantes et de reconstruction de données sont prégnantes lorsqu'il s'agit de réseaux urbains : véhicules effectuant un arrêt entre deux points, filtrage des modes de transports (a-t-on mesuré une voiture ou un vélo ?), données manquantes, non conservation des véhicules entre deux points ou autres erreurs. Une fois ces questions de qualification de données résolues, une cartographie temps réel des états de trafic peut être construite, bien souvent en termes de débits ou de vitesses sur des tronçons du réseau. La diffusion de cette cartographie aux usagers est alors possible. Ces derniers, de plus en plus connectés, confrontent les différentes sources d'information disponibles avec des exigences

croissantes. Parmi les nouveaux enjeux de la caractérisation temps réel du trafic, on retrouve la question de l'incertitude, en particulier sur les temps de parcours. L'important n'est plus la mesure d'un temps de parcours à vide mais bien l'obtention du meilleur modèle de distribution des temps de parcours sur un arc donné. Dans cette optique, la disponibilité des historiques de données suggère de considérer le temps de parcours comme une variable aléatoire puis de déterminer les meilleures distributions de probabilité pour le modéliser. Les paramètres des distributions obtenues permettraient ainsi de quantifier les incertitudes sur les informations des différentes sections d'un réseau et de mieux estimer les durées des trajets. De manière plus large, les nouveaux historiques de données nous orientent vers l'utilisation de méthodes d'apprentissage, non supervisé ou supervisé, pour caractériser et prévoir les états d'un réseau. Les méthodes non supervisées permettent d'identifier des zones homogènes et donc de "factoriser", et paralléliser, la supervision d'un réseau routier tandis que les approches supervisées visent à prédire à court terme le niveau de service, en particulier la congestion. Les motifs réguliers identifiés par les approches de reconnaissance de forme ne sont pas les seuls observables. Le fonctionnement nominal du réseau est régulièrement perturbé par l'apparition de phénomènes non-récurrents essentiellement dûs à des causes externes [401]. De tels phénomènes trouvent diverses causes : incidents, accidents, travaux, mauvaises conditions météorologiques. Ne pouvant être anticipés, ces événements conduisent à des pertes de performances de prévision et de fonctionnement du réseau non négligeables, si des congestions en résultent. L'identification automatique de ces incidents non récurrents constitue également un défi scientifique à relever.

1.4.3 Défi 3 : améliorer la planification des trajets pour les usagers des réseaux de transports

Notre troisième défi se place du point de vue d'un utilisateur qui ne souhaite pas seulement être informé en temps réel sur les conditions de circulation mais plutôt planifier en amont la nature de son déplacement : choix du mode de transport, itinéraire, voire heure de départ. L'objectif est d'améliorer les algorithmes présents au sein des calculateurs d'itinéraires. Sur le marché, les calculateurs historiques sont très souvent soit routiers (Google ou Mappy), soit dédiés aux transports en commun (RATP, TCL, Bibus). De nouvelles solutions sont apparues sur le marché ces dernières années. En France, nous pouvons citer la solution *Optygo* de *Cityway*, à l'international *Open Trip Planner* ou *Rome2Rio*. Les critères pris en compte sont souvent basiques : temps de parcours, nombre de kilomètres parcourus, coût financier (global ou détaillé), type de mode, temps de marche à pied. Les grands enjeux sont les suivants :

1. Promouvoir des calculateurs d'itinéraires multimodaux offrant la possibilité de combiner modes individuels et collectifs afin de faciliter un report modal de la voiture vers les transports en commun ou le vélo.
2. Intégrer des formes récentes de mobilité comme les véhicules électriques ou les vélos en libre service (VLS).
3. Proposer des solutions dynamiques et stochastiques : les valeurs des arcs d'un graphe changent au cours du temps et sont soumises à une incertitude.
4. Optimiser les solutions selon plusieurs critères. Exemple : temps de parcours, coût, confort, nombre de changement de modes, émissions environnementales.
5. Promouvoir l'open-source et la comparaison d'algorithmes pour la communauté scientifique.

Cette liste d'enjeux a directement motivé nos recherches mais elle est loin d'être exhaustive et beaucoup d'autres aspects auraient mérité pareille attention. Dans une présentation aux journées techniques du réseau scientifique et technique du ministère de l'écologie, Aurélie Bousquet, spécialiste de l'optimisation des itinéraires multimodaux [61], a listé des défis additionnels, rapportés ci-après :

- Prendre en compte les contraintes de circulation sur le réseau : mouvements tournants interdits, voirie autorisée pour chaque mode (ex : vélo).
- Représenter efficacement les coûts : politique tarifaire, combinaisons etc.
- Prendre en compte les situations de handicap.
- Prendre en compte les programmes d'activités en optimisant des chaînes de déplacement multimodaux.
- Interconnecter les calculateurs : assembler les résultats fournis par différents calculateurs pour optimiser un itinéraire qui dépasse les frontières d'un calculateur particulier.

Alors que certaines tâches relèvent plus de l'ingénierie ou du politique que de la recherche (ex : voirie par mode, disponibilité des données sur le handicap, *open source*), force est de constater que la plupart des objectifs opérationnels mentionnés ci-dessus se transcrivent en véritables verroux scientifiques d'un point de vue algorithmique. Ces objectifs scientifiques seront mieux détaillés dans la section suivante.

1.5 Démarche méthodologique

Cette section brosse un portrait synthétique de notre cheminement méthodologique et justifie la cohérence de nos recherches à l'aune des trois types de défis identifiés précédemment. Trois objectifs de recherche, ou sous-thématiques, sont présentés, pour lesquels les principales contributions seront plus détaillées dans les chapitres 2, 3, 4. Dans cette section, un état de l'art succinct dans chacun des 3 thèmes permet de mieux positionner et justifier la nature de nos contributions, listées de manière concise.

1.5.1 Cheminement logique : trois objectifs de recherche par montées en échelle successives

La présentation de nos travaux s'opère selon trois types de contributions, chacune visant à répondre aux trois défis identifiés en 1.4 :

1. Modélisation et simulation des C-ITS,
2. Caractérisation temps réel et prévision court terme du trafic routier,
3. Optimisation des itinéraires sous contraintes complexes.

Le cheminement méthodologique proposé suit une succession de montées en échelle (figure 1.3). Le premier pan de nos recherches est centré sur le véhicule et la façon d'intégrer ses nouvelles fonctionnalités (communication avec d'autres véhicules ou avec l'infrastructure) dans les modèles de trafic. Sans données disponibles à l'heure d'un pré-déploiement des C-ITS, nous suivons une approche type *model driven* fondée sur une modélisation théorique des véhicules connectés. Les modèles proposés sont ensuite intégrés dans un simulateur multi-agents qui se pose en outil d'aide

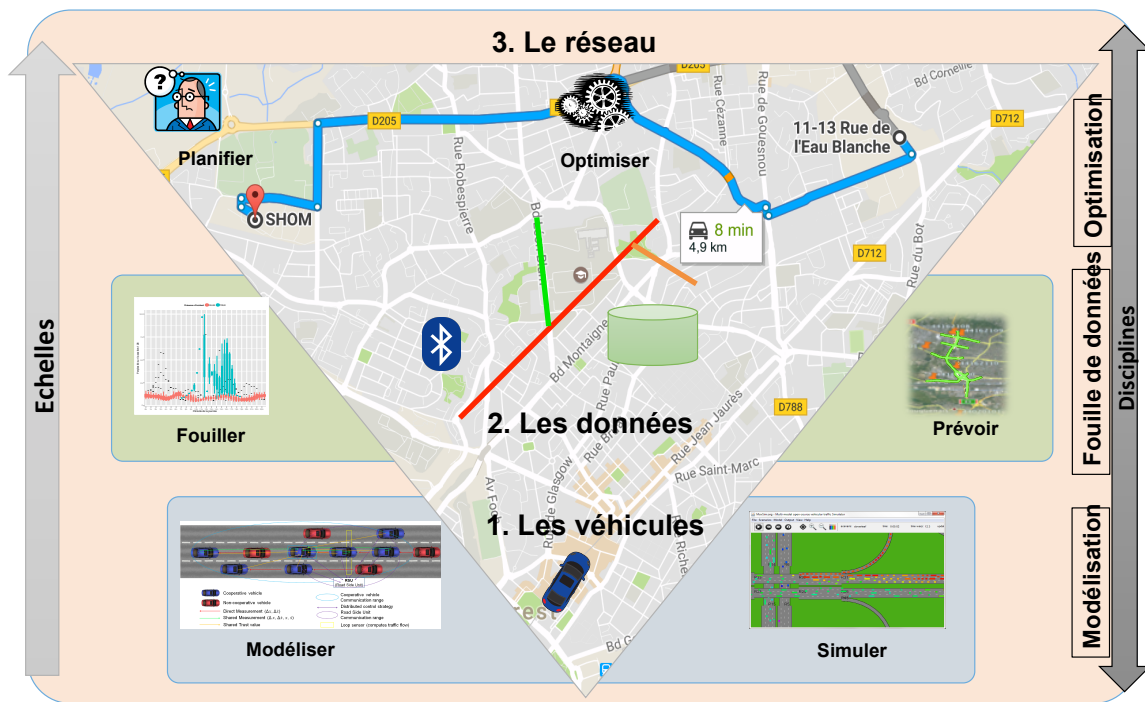


Fig. 1.3.: Complémentarité entre les échelles de travail

à la décision pour les chercheurs et gestionnaires. D'un point de vue disciplinaire, nos recherches combinent ici la modélisation théorique du trafic (mathématiques appliquées) et les systèmes multi-agents (informatique).

La deuxième échelle considérée est celle des données générées par tous ces véhicules, équipés ou non. Il s'agit d'extraire de la connaissance à partir de nouveaux dispositifs comme le Bluetooth ou les données FCD. Nous suivons des approches *data driven* et des processus classiques de fouille de données, depuis la qualification des données, la sélection de variables, jusqu'à des techniques non supervisées et supervisées validées sur des volumes de données conséquents. Les disciplines convoquées sont les statistiques (modélisation des temps de parcours par des distributions de probabilité) et l'analyse de données.

Enfin, la dernière échelle concerne logiquement le réseau. Plutôt que de nous consacrer à la simulation dynamique à l'échelle d'une ville, nos contributions ont abordé le problème de la planification d'itinéraires, en amont du trajet à effectuer. La discipline est la recherche opérationnelle, plus précisément la théorie des graphes et le problème du plus court chemin (PCC) sous contraintes complexes : aspects dynamiques et stochastiques, PCC pour véhicules électriques, techniques d'accélération pour une industrialisation des solutions proposées.

Ainsi, après avoir mieux compris la dynamique des véhicules (les modéliser, les simuler), le deuxième bloc analyse les données qu'ils génèrent. Les arcs d'un graphe routier peuvent alors être valués par des informations fiables, temps réel ou prédictives, qui sont les données d'entrées d'algorithmes du plus court chemin. Il convient à présent d'affiner le positionnement de nos travaux dans chacune des questions de recherche traitées et de le justifier par rapport à un état de l'art succinct.

1.5.2 Question de recherche Q1 : modélisation et simulation des C-ITS

1.5.2.1 Les familles historiques de modèles de trafic

La modélisation du trafic est un domaine largement étudié dans la littérature, et ce, à toutes les échelles : macroscopique, mésoscopique, et microscopique. Même si nous ne pouvons nous passer de quelques explications pour mieux situer nos travaux, nous éviterons les redondances en renvoyant le lecteur au chapitre consacré de notre thèse [52] ou à l'état de l'art référence de Dirk Helbing [193] pour de plus amples détails. L'échelle macroscopique du trafic procède par analogie avec la

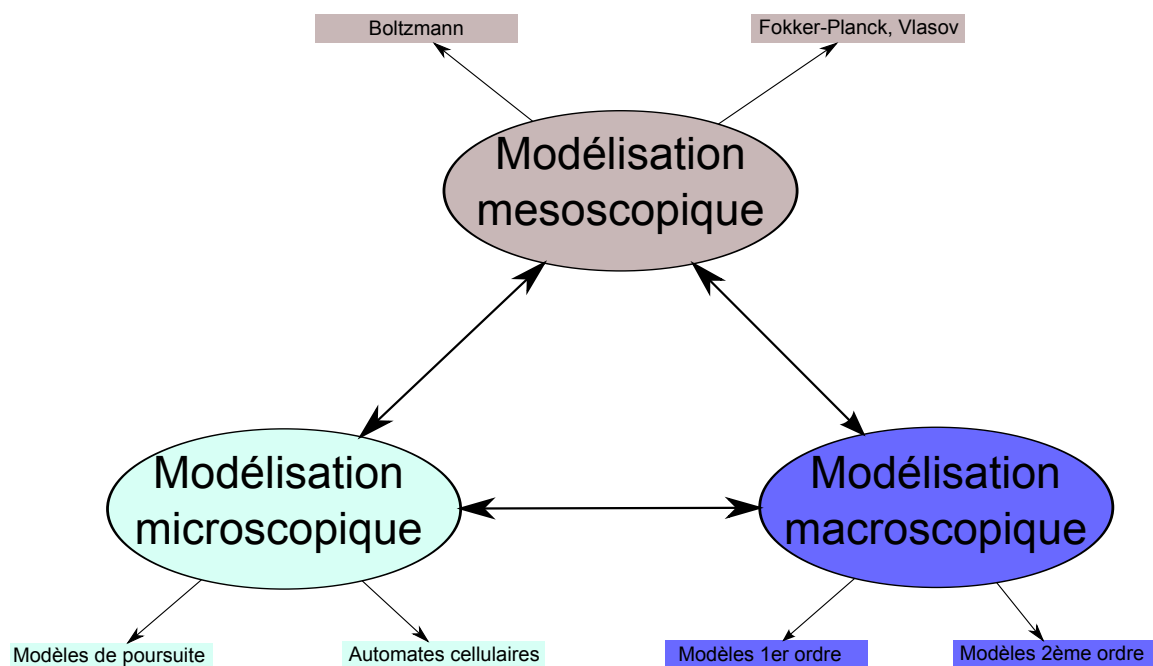


Fig. 1.4.: La modélisation du trafic routier. Une autre question d'échelles.

mécanique des fluides. Le lecteur peut s'imaginer dans un hélicoptère suivre l'écoulement du flux de véhicules comme celui d'un fluide dans un canal. Les trois variables fondamentales du trafic sont le débit, la densité et la vitesse du flux [136]. Les modèles macroscopiques sont fondés sur une loi de conservation à laquelle on ajoute une fermeture, la fameuse relation fondamentale d'abord expérimentée par Greenshields [176]. Cette relation peut prendre plusieurs formes [115] mais la forme triangulaire reste aujourd'hui largement utilisée, en particulier pour des facilités numériques [194]. Le modèle référence est un modèle du premier ordre nommé LWR, comme Lighthill et Whitham [262, 263], puis Richard [375]. C'est encore aujourd'hui une référence, pour sa simplicité, malgré quelques écueils évidents, que des modèles à plusieurs équations de type Payne peuvent combler [348, 19]. L'équation additionnelle permet de modéliser l'anticipation des conducteurs par rapport à ce qu'ils voient devant eux (en aval) et une relaxation vers une vitesse désirée qui dépend du niveau de service. Des évolutions récentes étendent la modélisation macroscopique à l'échelle d'un réseau, à partir de l'observation de diagrammes fondamentaux macroscopiques de zone (*Macroscopic or Network Fundamental Diagram*), [107, 172, 106].

Entre les niveaux microscopiques et macroscopiques se situe une représentation intermédiaire, dite mésoscopique ou cinétique, qui va manipuler des fonctions de distribution représentant la dynamique de groupes de véhicules. À ce niveau de représentation mésoscopique, les approches

dérivées de la théorie cinétique des gaz, initiées par Prigogine et Andrews dès les années 60 [359, 360], décrivent l'évolution de fonctions de densités de probabilité des principales variables du trafic. Ces approches présentent l'avantage d'être dérivables en modèles macroscopiques ou microscopiques. Des améliorations de ces modèles ont été proposées par la suite [347]. D'autre part, Illner *et al.* [207] ont montré qu'il était possible de retrouver le modèle macroscopique de Aw-Rascle [19] à partir d'équations cinétiques de type Fokker-Planck .

Le niveau microscopique considère les véhicules comme des particules dont il faut décrire la dynamique [193]. La tâche consiste à s'approcher le plus possible du comportement réel d'un véhicule en fonction de son environnement. Les modèles peuvent être soit discrets soit continus [203]. On parle alors d'automates cellulaires pour le cas discret [104, 319, 237] et de "lois de poursuite" (*car-following* ou *follow-the-leader*) pour le cas continu. Au sein des lois de poursuite, le véhicule suiveur adapte sa vitesse et son accélération en fonction de la perception de son véhicule prédécesseur (ou de ceux de son entourage dans le cas de la multi-anticipation [255]). Les références sont les modèles de Gipps [174] et Newell [324]. Actuellement, les modèles les plus utilisés fonctionnent en temps continu selon une approche "réactive". Le véhicule n est une particule qui réagit aux *stimuli* venant des véhicules prédécesseurs $n + i$. Le modèle consiste en une mise à jour de l'accélération du véhicule n en fonction de paramètres comme l'inter-distance, la vitesse courante, ou la vitesse relative. Parmi les modèles les plus prometteurs, citons le *Optimal Velocity Model* (OVM [26]) et ses évolutions : *Optimal Velocity with Relative Velocity* et *Improved Optimal Velocity Model* (OVRV [210] et IOVM [261]). Un passage en revue de ces modèles peut être trouvé dans le livre de Treiber et Kesting [427]. Les mêmes auteurs ont proposé un modèle populaire nommé *Intelligent Driver Model* (IDM), qui modélise finement l'anticipation et l'agressivité des conducteurs [424, 227]. Les modèles cités ici ont été évalués et comparés à la lumière du calibrage de données réelles et de la reproduction de comportements observés sur des trajectoires réelles [223, 308]. Comme ces modèles ne traduisent que le comportement longitudinal des véhicules, il est nécessaire de leur adjoindre des modèles de changement de voie [424, 423, 250, 224, 418]. Malgré quelques divergences, celle de Boris Kerner en l'occurrence (voir [222] pour l'essai le plus récent), c'est dans ce paradigme à trois échelles de modélisation que s'inscrivent les contributions en modélisation du trafic.

1.5.2.2 Etat de l'art des modèles pour le trafic coopératif

Si les modèles de trafic forment le cœur de la plupart des outils d'aide à la décision destinés aux gestionnaires, il faut remarquer qu'ils n'ont pas été développés pour prendre en compte toutes les interactions possibles entre des entités communicantes sur le réseau. La problématique est ambivalente puisqu'il s'agit d'un côté d'analyser l'impact des systèmes coopératifs sur les différents paramètres pour mieux calibrer les modèles (ex : vitesses ou temps inter-véhiculaire) et, d'un autre côté, de proposer dans un futur proche un nouveau cadre de modélisation pour mieux traduire la communication entre les différents éléments du réseau. Ces deux blocs de recherche interagissent forcément dans la mesure où les modèles de trafic classiques (modèles microscopiques de poursuite ou automates cellulaires, modèles macroscopiques du 1er ou 2ème ordre) pourraient être partie intégrante de la modélisation, qui consisterait à intégrer une couche d'abstraction supplémentaire pour la prise en compte des échanges d'information (communication). Une fois validée, cette modélisation coopérative du trafic permettra d'évaluer *a priori* les nouvelles solutions proposées, et de tester l'effet de déploiement des systèmes coopératifs sur le trafic en fonction de leur taux de pénétration. Mais avant une telle étape, il convient de déterminer dans quelle mesure les principes coopératifs peuvent

être introduits dans les formalismes de modélisation actuels. Comme la communication s'opère d'abord au niveau du véhicule, il est naturel de considérer le niveau de modélisation microscopique comme plus à même de reproduire les différents canaux de communication. Une revue bibliographique confirme la richesse des travaux au niveau microscopique face à une sous-représentation des deux autres échelles, malgré quelques contributions notables [259, 325, 326, 470]. Aujourd'hui, le système le plus représenté dans les véhicules est le régulateur de vitesse à contrôle de distance (*Adaptive Cruise Control – ACC* [463]). Les études de terrain ou de simulation montrent des résultats mitigés à mesure que les situations de trafic se complexifient. En termes de modélisation du trafic Kesting *et al.* [227] ont proposé une extension du modèle IDM qui intègre le comportement de ACC. Un premier pas vers la coopération est le modèle C-ACC [114, 356], extension coopérative de ACC où la communication V2V apporte des informations supplémentaires sur les véhicules prédécesseurs. Des modèles théoriques montrent les avantages des systèmes C-ACC en termes de stabilité du trafic, au sens mathématique, c'est-à-dire la propension à propager ou non de petites perturbations à partir d'un état d'équilibre [335]. Une nouvelle fois, l'inter-distance temporelle cible est choisie par le conducteur et influence directement le comportement du véhicule (progressivité de l'accélération). Les perturbations, le plus souvent induites dans la réalité par les changements de voie [9], peuvent être limitées grâce au système [433, 299, 386]. Inspirées par les technologies ACC et C-ACC, des recherches récentes en modélisation du trafic intègrent les récentes évolutions des C-ITS en tant que multi-anticipation du conducteur : la perception serait étendue à un ensemble de véhicules situés dans un rayon de communication accessible. Dans ce cadre multi-anticipatif [255], il est alors possible de considérer dans une loi de poursuite l'influence des véhicules en aval [337] ou en amont [461] du véhicule courant. La communication permet alors d'étendre les modèles microscopiques [167] afin de prendre en compte ces véhicules grâce à la multi-anticipation. Citons également une extension multi-anticipative du système ACC par nos collègues hollandais [450, 451]. Quant à la tâche de changement de voie, elle tend aussi à s'automatiser de manière coopérative [274].

1.5.2.3 Les Systèmes Multi-Agents appliqués à la modélisation et au contrôle du trafic routier

Comment modéliser la complexité des interactions entre véhicules et infrastructure ? Comment traduire la fiabilité de l'information ? Une étude bibliographique sur les véhicules connectés ou autonomes oriente indubitablement vers le thème des systèmes multi-agents (SMA). Les agents constituent des entités autonomes [153, 154] évoluant dans un système lui-même autonome, le système multi-agents [120]. L'agent est autonome car il poursuit ses propres buts [72], ses interactions avec d'autres agents ou l'environnement pouvant mener à des comportements collectifs. Au niveau du système, on souhaite mettre en valeur des comportements globaux complexes, non définis au départ, grâce à des phénomènes d'*émergence* [123] ou d'*auto-organisation* [163]. Comme pour la mécanique des fluides et la modélisation du flux, l'analogie entre SMA et modélisation microscopique du trafic est naturelle. Chaque conducteur peut être modélisé comme un agent autonome [129], les interactions au niveau microscopique se répercutant par des phénomènes à l'échelle macroscopique, comme la congestion par exemple [393]. À travers les nouvelles technologies, on souhaiterait faire émerger des phénomènes bénéfiques d'homogénéisation du flux. De plus, la nature distribuée des SMA serait parfaitement en accord avec la gestion coopérative d'une flotte de véhicules connectés ou partiellement autonomes. Il est donc logique que l'on retrouve des utilisations des concepts tirés du paradigme agent dans le domaine du trafic et des transports [81], à travers des applications très variées [66, 112], allant des approches comportementales [130, 131] à des problématiques très

classiques de gestion, régulation, contrôle du trafic [42, 41, 25, 24, 88, 239, 4, 5, 272, 435, 82, 469]. Pour la partie modélisation, les études sont plus rares et certains concepts SMA nous ont semblé prometteurs pour une transposition aux C-ITS pour la question de la fiabilité des informations échangées. On pense notamment aux systèmes fondés sur la confiance et la réputation [378, 355]. Au niveau du contrôle, la plupart des approches SMA se cantonnent à des problèmes de gestion d'intersection [134, 442] ou de régulation à l'échelle de la ville. Les concepts multi-agents reproduisent des comportements de coalition permettant de favoriser le passage de pelotons homogènes de véhicules. Le moyen d'action (*i.e.* l'application du contrôle) se limite principalement aux limitations de vitesse dynamiques (VSL) [483]. La modélisation et le contrôle d'un flux autoroutier pourraient susciter plus d'intérêt, comme le montrent les auteurs dans [391, 417]. Il s'agirait alors de revisiter des stratégies de régulation classiques comme la régulation d'accès ou la gestion dynamique des vitesses (voir [342, 229]). Enfin, des échanges scientifiques réguliers avec l'équipe SMA du LIRIS nous ont aiguillé vers les approches constructivistes telles qu'inspirées par Piaget [354]. Elles considèrent le développement d'un individu par des interactions successives avec l'environnement, vers l'acquisition de l'autonomie. Cette théorie a donné lieu à la proposition de méthodes permettant de construire une représentation sous la forme de schémas sensorimoteurs [133]. Ce concept a été largement repris par la communauté IA (voir la synthèse de [181]) avec des applications allant de l'intelligence ambiante [321, 290] à la robotique développementale [283, 292, 17]. En tout état de cause, l'amélioration de l'efficacité du trafic et de la sécurité viendront d'une combinaison entre une approche de modélisation innovante intégrant les C-ITS, et une stratégie de contrôle [260]. Dans l'attente d'une flotte suffisante de véhicules connectés, des études en simulation peuvent constituer des outils d'aide à la décision pertinents.

1.5.2.4 Les outils de simulation de trafic

La simulation microscopique du trafic a déjà fait l'objet de nombreux développements de logiciels commerciaux ou plateformes *open source*. Cependant, peu d'entre-elles traitent des communications inter-véhiculaires (voir [213] qui donne un aperçu des simulateurs de trafic routier les plus populaires). Parmi ces solutions, SUMO [242], AIMSUN [30], VISSIM [150], TRAFFSIM [22] et MovSim [428] font figure de référence. Introduire les C-ITS dans un simulateur de trafic se rapproche naturellement des simulateurs de réseaux ad hoc de véhicules (*vehicular ad hoc networks* (VANETS)) qui forment une catégorie à part [288]. En ce qui concerne les concepts agents, MATSim (*Multi Agent Transport Simulation*) et ARCHISIM [144] s'imposent comme deux outils pertinents, l'un se focalisant plus sur la demande de trafic (MATSim) et l'autre sur les aspects comportementaux (ARCHISIM). L'objectif premier de nos travaux était de comparer différents modèles de trafic et stratégies de contrôle, plutôt à l'échelle d'une section autoroutière avec entrées/sorties. La génération d'un trafic connecté à l'échelle d'une ville n'étant pas le but poursuivi, notre évaluation des simulateurs s'est orienté vers des solutions aptes à comparer des modèles microscopiques de trafic. SUMO (*Simulation of Urban MObility*) et MovSim (*Multi-model Open-source Vehicular-traffic SIMulator* [428]) ont retenu notre attention en tant qu'outils open-source, portables et déjà dotés d'un nombre appréciable de fonctionnalités. Ils sont développés respectivement en C++ et Java. SUMO est l'outil le plus utilisé et cité dans la littérature scientifique [242]. À propos de la communication V2V, une extension récente consiste à simuler les effets de la communication sur les comportements des conducteurs. D'autre part, le projet VEINS est une extension récente qui offre une suite de modèles pour la communication avec l'infrastructure [407]. MovSim est un outil plus récent qui bénéficie des concepts les plus avancés en théorie du trafic [427, 225], pensé depuis le début pour intégrer les C-ITS [391, 422, 226]. Notre choix s'est porté sur MovSim en raison de sa conception (*design pattern*) plus générique,

décentralisée et des modèles de trafic déjà implémentés. Les concepts à base d'agents ont, de plus, déjà été abordés par les auteurs [225]. En tant qu'outil *open source*, il a déjà été réutilisé pour des problèmes de prévision de trafic ou de conduites en pelotons (*platooning*, voir [471]).

1.5.2.5 Nature des contributions

Comme nous l'avons souligné dans le cadre du premier défi, la phase de déploiement de véhicules connectés sera synonyme d'un trafic mixte mêlant des véhicules connectés (coopératifs) à des véhicules non connectés. La figure 1.5 illustre nos contributions pour répondre à ce défi.

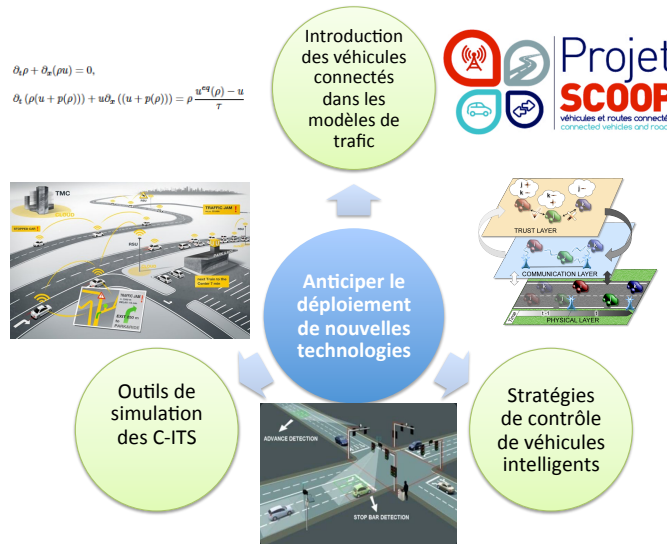


Fig. 1.5.: Thématiques de recherche en réponse au défi 1

Le besoin de mieux comprendre l'impact des systèmes coopératifs sur le trafic s'accompagne d'un manque total de données de terrain, à savoir des véhicules coopératifs en nombre suffisant pour générer du trafic. Des études de modélisation et de simulation prospectives sont nécessaires afin de déterminer à quel point et sous quelles conditions des véhicules connectés pourraient améliorer la sécurité et la mobilité. Nous restreignons notre champ d'études à des sections autoroutières ou péri-urbaines, en alignement avec les projets de déploiement actuels. L'état de l'art précédent nous a permis de décliner un certain nombre de contributions :

1. Une modélisation multi-agents multi-couches par couplage de dynamiques. Au niveau microscopique, un nouveau type de modèle est proposé pour décrire la dynamique des véhicules connectés, dont les propriétés sont validées théoriquement (études de stabilité). Le modèle est ensuite intégré dans un cadre multi-agents plus complet mieux à même d'englober les problématiques de communication et de fiabilité des données ;
2. Un simulateur multi-agents de trafic coopératif développé comme une extension de l'outil existant MovSim. La solution est suffisamment générique, robuste et réaliste pour aider à la comparaison de modèles académiques et assister le monde industriel lors des phases de déploiement des C-ITS ;
3. Une évaluation des impacts des C-ITS sur un certain nombre d'indicateurs d'efficacité, de sécurité et d'émissions environnementales, à l'aide du simulateur ;
4. Une stratégie de contrôle innovante fondée sur le paradigme de l'apprentissage constructiviste.

L'objectif final du travail consiste à enrichir le simulateur développé de nouvelles fonctionnalités venant aussi bien des chercheurs que des industriels et ainsi d'en faire un outil d'aide à la décision capable d'assister les projets de déploiement des C-ITS et d'anticiper un futur (proche, lointain ? En tout cas incertain) émaillé de nouveaux véhicules et comportements sur les routes. La présentation de nos résultats sur ce thème sera approfondie dans le chapitre 2.

1.5.3 Question de recherche Q2 : approches statistiques pour la caractérisation temps réel et la prévision court terme du trafic routier

1.5.3.1 La variabilité des temps de parcours

Le temps de parcours constitue l'indicateur-clé et connu de tous les usagers [209] et le maîtriser dans toutes les situations serait le rêve de tout un chacun. Pour le gestionnaire d'infrastructures, l'information sur le temps de parcours a une influence sur le choix des modes et des itinéraires, elle constitue un levier fort sur la demande en trafic. En termes économiques, une étude de 2005 donnait des chiffres allant jusqu'à 50% de pertes de revenus pour les opérateurs si la variabilité des conditions de trafic n'était pas prise en compte [456]. La variabilité des temps de parcours agit directement sur leur fiabilité, une faible fiabilité étant intrinsèquement liée à un faible niveau de service. Du point de vue des opérateurs, un lien sera moins fiable si l'on ne peut pas toujours y atteindre la vitesse libre, c'est-à-dire sans contraintes [141]. Ceci se répercutera sur les usagers par une plus grande variabilité des temps de parcours puisque l'heure d'arrivée sera en dehors de l'intervalle envisagé lors de la planification du trajet [140, 141]. Une revue bibliographique met en exergue de nombreuses contributions ayant trait à la notion de variabilité, selon deux tendances nettes : un premier groupe de travaux se focalise sur la maximisation d'une utilité d'un point de vue de l'utilisateur [268, 69, 68, 142, 60, 209, 403, 334, 38, 202, 158] alors qu'un deuxième ensemble s'intéresse à la maximisation du niveau de service sur un lien, notamment à l'aide d'indicateurs statistiques. C'est naturellement vers cette deuxième approche *data driven*, que les nouvelles sources de données rendent de plus en plus pertinente (voir [406, 413] pour l'utilisation de données taxis ou FCD), que nos travaux se sont tournés. Dès lors, quelles sont les causes de cette incertitude sur les liens à l'échelle du véhicule, des heures, du jour [333] ? Multiples, bien sûr. Incidents entraînant des chutes de capacité [97], comportements des conducteurs [212], travaux, conditions météorologiques [53], interruptions des transports en commun [201, 284]. Plusieurs métriques de variabilité ont déjà été proposées à l'échelle du réseau. Une contribution fondamentale de l'institut des transports Texan [273] subdivise les indicateurs proposés en trois catégories :

- Les indicateurs statistiques comme la fenêtre de temps de parcours acceptable ou autres indices de variation,
- Les indices de type *Buffer* qui vont calculer des marges de sécurité (ex : *Buffer Time Index*),
- Les indicateurs d'évaluation du retard (*Florida Reliability Method, On-Time Arrival, Misery Index*).

Depuis, des travaux plus récents se sont rapportés à des mesures plus classiques de dispersion. L'équipe du Pr. Mahmassani a mis en évidence une relation (linéaire) entre la moyenne des temps de parcours et leur écart-type [284, 230]. Ces publications s'inscrivent dans la lignée d'autres études fondées sur l'écart-type [362, 369] tandis que des alternatives capables de capturer les différentes phases de congestion ont été proposées [436, 447]. Finalement, aucune contribution

vraiment originale ne sort du lot et l'on revient rapidement à la question de la modélisation de la variable aléatoire "temps de parcours" par une distribution de probabilité adéquate [141, 218]. Dans [376], les auteurs ont procédé à des convolutions entre temps de départ et d'arrivée pour construire une distribution réaliste. À partir d'approches fines prenant en compte les comportements des conducteurs et les caractéristiques des sections, plusieurs équipes ont mis en évidence des formes bi-modales pour les distributions de temps de parcours [212, 413]. Dans [96], les auteurs calculent les premiers moments des distributions empiriques puis les ajustent à l'aide de dérivées de Gaussiennes. Partir des séries temporelles de temps de parcours pour construire des modèles est également une piste envisagée dans [406] ou [371] via des procédures Markoviennes. Dans [140], El Faouzi a étudié les propriétés de la loi log-normale pour un calage de distributions empiriques. Cette approche a été suivie par Emam *et al.* [141] qui ont élu la loi log-normale comme la plus pertinente parmi quelques autres comme la Weibull, l'exponentielle ou la Gaussienne, puis par d'autres auteurs (voir [183, 231, 413, 421]). Il est clair que la loi log-normale, tout comme la loi Gamma, est particulièrement adaptée à la modélisation d'une variable aléatoire résultant de la multiplication de plusieurs facteurs indépendants. D'autre part, les distributions empiriques de temps de parcours se prêtent bien à une forme exponentielle asymétrique accompagnée d'une longue queue de distributions pour les grandes observations [140, 413, 421]. Kim et Mahmassani utilisent une loi Gamme composée, à travers une première variable représentant la variabilité inter-véhiculaire et une deuxième relative à la variabilité inter-journalière [231]. Sur des données simulées pour le centre-ville de Chicago, les auteurs montrent une bonne adéquation avec le phénomène modélisé [231]. En Australie, l'équipe de Mike Taylor défend l'utilisation d'une distribution alternative : la loi de Burr [421, 413]. Cette distribution trouve son origine dans l'actuariat mais les recherches australiennes ont mis en évidence des performances prometteuses au regard de données de la ville d'Adélaïde, face aux lois Gamma, Weibull ou log-normal, performances confirmées par les travaux plus récents de Guessous *et al.* [183]. Finalement, l'ensemble de ces travaux suggèrent qu'il n'y a pas de juge de paix concernant la modélisation statistique des temps de parcours et que d'autres contributions pourraient améliorer la connaissance dans le domaine, notamment avec une meilleure flexibilité face aux différents types de sections modélisées.

1.5.3.2 Challenges de l'extraction de connaissance à partir de nouvelles données : le cas Bluetooth

Les véhicules connectés ou autonomes ne représentent qu'une part récente de l'évolution des systèmes avancés de gestion de trafic (SAGT) fondée sur les nouvelles sources de données [477, 366, 467] comme les véhicules traceurs (*Floating Car Data* [6], les réseaux sans fil [357] ou le Bluetooth [384, 67]. Identiquement à tout processus de traitement de données, et en particulier dans un contexte *big data* où les volumes collectés atteignent de nouvelles dimensions, les étapes de qualification et pré-traitement des données brutes revêtent une importance particulière [345, 6]. Nous avons porté de nombreux efforts sur l'analyse du Bluetooth, sujet de recherche dynamique sur les sept dernières années [357], pour lequel nous avons valorisé quelques résultats [431, 245, 247]. Le choix du Bluetooth s'est imposé à nous de par la collaboration étroite avec l'université de technologie du Queensland, à Brisbane, Australie. Le Brisbane City Council, en charge du réseau routier de cette ville australienne, a implanté, en moins de 4 ans, près de 900 capteurs. Un tel réseau de recueil de données est unique au monde. Les capteurs Bluetooth présentent peu d'emprise sur l'infrastructure routière et sont très peu coûteux, dans la mesure où ils peuvent être produits sur place (100\$/capteur à Brisbane, Australie). Le principe de la collecte de données Bluetooth est de capter tout appareil équipé (émetteur) croisant une zone de détection (récepteur) [50]. Chaque

appareil est identifié par une adresse MAC unique, rendant ainsi possible la reconstitution d'un flux partiel et la construction d'indicateurs de temps de parcours. Néanmoins, aux prémisses du Bluetooth, en 2011, les efforts étaient dirigés vers le pré-traitement, la qualification de la donnée, et son évaluation pour la gestion de trafic. Les problèmes de qualité de données sont pléthoriques. En termes d'hétérogénéité d'abord, les appareils Bluetooth ne sont pas toujours distingués entre eux et des routines de traitement de données doivent être développées pour discriminer les différents types d'utilisateurs [67, 13, 51]. D'autre part, ce type de détection induit un nombre significatif de données aberrantes (*outliers*) comme le montrent Araghi *et al.* dans [13]. L'influence des *outliers* est amplifiée par un faible taux de pénétration de la technologie sur le marché [127]. En effet, la part de véhicules équipés d'émetteurs Bluetooth est soumise à une grande variabilité, de 5 à 6% selon certains travaux de 2011-2012 [67, 13, 12, 7, 446] à 25-30% pour les plus optimistes, notamment les constructeurs. Tracer un véhicule est possible à condition que les appareils soient effectivement détectés partout où ils sont passés, ce qui est loin d'être le cas dans la réalité [12, 431]. Des algorithmes de reconstruction de trajectoires pourraient alors être développés. Pour les données manquantes, les méthodes de complétion ne manquent pas dans la littérature et les approches classiques peuvent être adaptées. Elles consistent en des régressions simples ou lissages temporels [235, 365], voire spatio-temporels : le voisinage d'une section est intégré au processus de complétion des données manquantes [44, 479]. D'autres alternatives sont possibles, allant de la réduction de dimension [367, 192] aux réseaux bayésiens [460, 44]. Une fois les difficultés précédentes surmontées, il est possible d'utiliser les détections multiples d'un même appareil au cours de son déplacement pour reconstituer un temps de parcours, global ou par liens entre deux détecteurs [13, 12, 431], ainsi que des trajectoires [298]. La majorité des publications scientifiques se concentrent sur la reconstruction de temps de parcours à partir de données Bluetooth [357, 13, 431, 363, 457]. Rappelons que l'anonymisation des adresses MAC permet un usage légal, et respectueux de la vie privée, des estimations de temps de parcours [363]. Le coût abordable de la solution la rend particulièrement attractive pour un déploiement à l'échelle urbaine. Néanmoins, les algorithmes de caractérisation ou de prévision de trafic à partir de données Bluetooth vont fortement dépendre de la qualité des données en entrées. Ainsi, comme dans tout processus de fouille, la phase de pré-traitement des données ne doit pas être négligée. En ce qui concerne le filtrage des données, des moyennes glissantes ou des seuils sont une première étape indispensable qui ne suffit pas toujours à bien discriminer les différents modes de transport ou filtrer les valeurs aberrantes [13, 431]. Dans [58], Blogg *et al.* synthétisent les difficultés méthodologiques rencontrées lors de l'analyse de données de la ville de Brisbane. Une fois ces écueils surmontés, la nouvelle source de données peut être utilisée pour estimer des matrices Origine-Destination à grande échelle [67, 7, 58, 31, 32, 161, 481, 297] ou prévoir le trafic à court-terme avec des approches d'assimilation de données [31, 32]. Nos travaux rejoignent cet ensemble de contributions, en portant l'effort sur une caractérisation spatio-temporelle des états de trafic, à partir du Bluetooth.

1.5.3.3 La prévision court terme du trafic routier

Avec les récentes technologies de recueils, la supervision temps réel du trafic est de mieux en mieux maîtrisée, ainsi en va-t-il de l'identification spatiale d'une congestion lorsqu'elle est installée. Le challenge s'est reporté sur la question de la prévision court terme des conditions de circulation, qui reste une tâche critique pour tous les gestionnaires et services d'information trafic. Le but est de développer une prévision robuste avec une complexité temporelle raisonnable pour aborder de grands réseaux urbains. S'attaquer à la prévision court terme du trafic routier, c'est investir un thème de recherche foisonnant qui a accompagné les différentes évolutions des ITS. Des revues bibliogra-

phiques plus complètes sont disponibles et nous renvoyons le lecteur à [269, 443] pour approfondir des points qui ne seront pas abordés dans ce paragraphe. Le point d'entrée classique réside dans la dichotomie entre approches paramétriques et non paramétriques. Les méthodes paramétriques sont fondées sur des hypothèses fortes quant à la forme des distributions manipulées. Les approches par séries temporelles ont été très populaires pour prévoir des chroniques de débits, vitesses, ou temps de parcours. Des techniques de lissage et les processus autorégressifs donnent de bons résultats quand le problème est simple et le calage possible. Les approches de type Box-Jenkins [62] et leur application au domaine du trafic routier ont été évaluées dans une contribution fondamentale de Smith *et al.* [405], complétée par Vlahogianni *et al.* [443]. Les auteurs confirment la sensibilité des méthodes ARMA ou ARIMA aux valeurs extrêmes, aux conditions de trafic non récurrentes (incidents, travaux), ainsi qu'une complexité temporelle non négligeable. Dans un contexte *big data* de plus en plus favorable, les méthodes non paramétriques ont gagné en popularité. Le terme non paramétrique est ici relatif à l'absence d'hypothèses sur la forme des distributions [135]. Les réseaux de neurones artificiels (ANN) ont été appliqués au trafic routier avec plus ou moins de succès, seuls [482, 444] ou combinés à des approches paramétriques [419, 434]. Le côté "boîte noire" de ce type d'approches est souvent critiqué [233] même si l'avènement du *deep learning* devrait bientôt porter ses fruits pour une application au trafic routier. La méthode de Régression par machines à Vecteurs Supports (SVR), développée au milieu des années 1990 [100, 390, 441], constitue une adaptation au problème de régression des *Support Vector Machines* (SVM), introduits par Vapnik *et al.* [440] dans le courant des années 1960. Le SVR requiert un historique moins profond que d'autres approches pour aboutir à des performances équivalentes [439]. Sa rapidité d'apprentissage et son implémentation en ligne en font un sérieux candidat pour des performances résilientes à des phénomènes non récurrents [75, 480]. Plus récemment, des approches ensemblistes comme les forêts d'arbres décisionnels (*Random Forest*) [64] et les *Gradient Tree Boosting* [478] ont mis en lumière de bonnes capacités à prévoir les états de trafic routier avec simplicité et efficacité [256, 187, 478]. Ces deux méthodes reposent sur l'apprentissage par arbres de décision. Comme beaucoup de méthodes récentes, elles nécessitent un historique de grande taille qui n'est pas toujours disponible lorsqu'une nouvelle source de données est évaluée. En prévision de trafic routier, des méthodes simples et parcimonieuses ont fait leurs preuves. On peut citer l'approche des k plus proches voisins (kNN) qui consiste en une régression non paramétrique. Intuitive et rapide d'usage, elle est souvent employée en trafic comme référence lors de comparaisons entre approches paramétriques et non paramétriques [404, 405], comme dans la contribution de Davis *et al.* qui mettent en compétition kNN et prévision par séries temporelles [113]. Elle s'avère efficace dans de nombreux cas [338, 232, 23, 449, 92]. Dans la même mouvance, la régression non paramétrique à noyaux peut être vue comme une généralisation des k plus proches voisins. El Faouzi a utilisé cette technique dès 1996 pour prédire les débits sur autoroutes [138]. Elle joue un rôle d'intermédiaire entre l'approche extrêmement simple du kNN et celle plus évoluée du SVR [192, 56, 412].

1.5.3.4 Détection et intégration des externalités dans les modèles

Par externalités du trafic, nous entendons les effets souvent négatifs de certains phénomènes non récurrents, en particulier les incidents et les conditions météorologiques. Détecter un événement non récurrent n'est pas une tâche facile car il faut le distinguer d'un phénomène de congestion dû à une trop forte demande sur une section donnée. Des méthodes de détection temps-réel des incidents ont émergé depuis les efforts pionniers de Kahn dans les années 1970 puis Payne [349], qui comparaient les conditions de circulation en amont et aval d'une section donnée pour détecter

l'anomalie. Quelques méthodes suivent cette tendance en comparant les quantités physiques du système [211] alors que d'autres se concentrent sur la comparaison entre valeurs prédites et estimées afin de capter un écart anormal, à l'aide d'un seuil. Ces dernières font appel à des outils statistiques classiques tels que la régression paramétrique ARIMA [8], les réseaux bayésiens [257] ou le lissage exponentiel [78]. La dynamique court terme très forte d'un réseau urbain a cependant empêché la popularisation de ces méthodes à l'échelle d'un réseau. En tout état de cause, si l'arrivée des données FCD a rendu le problème de la détection temps réel d'incident plus abordable [395, 476, 217, 271, 151, 122], la prévision de ces incidents et leur intégration dans les méthodes de prévision posent plus de problèmes. L'augmentation des volumes de données au cours des dernières années a permis de faire appel à des méthodes de *machine learning* qui abordent le problème sous une formulation de classification supervisée. Autant de méthodes que les réseaux de neurones, la logique floue [275, 408], la régression [452], arbres de décision et forêts aléatoires [89, 270], SVM [474, 90], ont été testées pour résoudre le problème. En dépit de bonnes performances sur certains cas d'études, les réseaux de neurones souffrent de faibles performances en généralisation malgré un temps d'apprentissage conséquent [275, 408]. La palme revient finalement aux approches de type SVM, malgré leur sensibilité au choix du noyau [474, 90]. L'intégration de l'impact de conditions météorologiques dégradées dans les modèles de prévision reste un grand défi, même si cet impact a été bien quantifié à différentes échelles [53]. Comme pour les incidents, le problème vient du peu de données d'apprentissage disponibles pour le facteur externe considéré. C'est pourquoi les méthodes *model driven* sont plus présentes dans la littérature scientifique [370, 52, 248]. On retrouve également des méthodes statistiques paramétriques [430, 281, 139, 291] mais de façon générale, l'intégration de cette externalité reste à faire.

1.5.3.5 Nature des contributions

Alors que notre premier défi suggérait naturellement des approches centrées sur les modèles *model driven* dans la mesure où les véhicules connectés ne sont pas encore (ou du moins n'étaient pas encore) sur nos routes, le deuxième défi est centré sur la génération de nouvelles données et le futur de la gestion de trafic à l'ère du *big data*. Pour y répondre, nous proposons un certain nombre d'approches statistiques qui partent toutes des données empiriques pour extraire des connaissances. Nos travaux mettent en exergue trois types de contributions illustrées par la figure 1.6.



Fig. 1.6.: Thématiques de recherches en réponse au défi 2

Dans un premier temps, nous nous sommes intéressés à la question de la modélisation statistique des temps de parcours. Nous avons proposé une nouvelle famille de distributions, les lois de Halphen, appliquées pour la première fois à la question du trafic routier, ainsi qu'un outil de sélection automatique de distribution, le diagramme de ratio des moments. La contribution a été évaluée à partir de données empiriques de type Bluetooth ou reconnaissance automatique de plaques.

La seconde contribution concerne l'extraction de connaissances à partir de données Bluetooth. Nous avons évalué le potentiel de cette nouvelle source de données pour une caractérisation temps réel d'un réseau routier urbain. Des contributions originales relatives à l'analyse spatio-temporelle du fonctionnement du réseau (approches non supervisées) ont été introduites.

Le troisième type de travail porte l'effort sur la question de la prévision de trafic. Nous avons testé une approche globale qui permet de mieux prendre en compte la structure spatio-temporelle du réseau. L'approche a été illustrée sur des données Bluetooth mais aussi sur des données FCD plus complètes. Ces dernières nous ont permis d'effectuer un *benchmark* des performances des méthodes pour la ville de Nantes. Quelques extensions sur la détection d'incidents et les phénomènes météorologiques ont également été développées. Le chapitre 3 présentera, pour le lecteur intéressé, une sélection de résultats.

1.5.4 Question de recherche Q3 : optimisation des itinéraires sous contraintes complexes

1.5.4.1 Etat de l'art sur les problèmes d'optimisation pour l'électromobilité

La question de l'électromobilité nous a intéressé dès 2012 avec la thèse de Fouad Baouche [27] qui s'est effectuée en collaboration avec le laboratoire transports environnement (LTE) de l'IFSTTAR. Outre la question de la consommation énergétique des véhicules électriques, centrale car fournisseuse des données d'entrée des algorithmes, la thèse présente des contributions sur la question de l'allocation optimale des stations de recharge et l'optimisation d'itinéraires pour véhicules électriques. Le premier n'est donc pas un problème d'optimisation des itinéraires, il se situe en amont, dans la conception et le design des stations de recharge. Sur les dix dernières années, l'essor des véhicules électriques, promus par des politiques pro-environnementales comme le bonus écologique, nous a ramené à un des tous premiers problèmes de la recherche opérationnelle, un problème de localisation-allocation⁵. En effet, sachant un ensemble de contraintes de budget, de type de stations, de demande en trafic, de séparation entre stations, quelle est l'allocation optimale des stations de recharge à l'échelle d'une ville, d'une région ? La question renvoie naturellement aux problèmes de déploiement de ressources, en particulier de véhicules ambulanciers ou services d'urgence. Le problème de déploiement consiste à déterminer les points de stationnement optimaux (les stations) à utiliser pour la localisation des véhicules d'urgence entre deux affectations. Ce problème a été formulé et résolu initialement de deux façons différentes. Premièrement, Hakimi a, dès les années 1960, formulé un problème de minimisation connu sous le nom de couverture par ensembles [184]. Il s'agissait de minimiser le nombre de stations ouvertes en fonction d'une certaine demande. Le problème est NP-complet et l'exemple le plus connu concerne la localisation des stations de pompiers. Ce cas a ensuite permis de décliner le *p-center*, qui consiste à déterminer la location de

5. En 1940, la première équipe de RO, dirigée par Patrick Blackett, a traité le problème d'implantation optimale de radars de surveillance

p centres dans un réseau de façon à minimiser la distance maximale, parmi tous les usagers, d'un usager à un centre. Un deuxième type d'approches considère un problème de maximisation. La contribution historique de Church et Revelle [95] formule une recherche de couverture maximale : la fonction objectif consiste à maximiser le service de la population et la distance couverte (maximiser la demande), selon un certain nombre de ressources disponibles. La résolution est toujours complexe et des auteurs ont proposé un certain nombre d'heuristiques comme des relaxations lagrangiennes couplées à des approches type *branch and bound* [293, 109, 374, 110]. Des variantes plus raffinées des problèmes initiaux peuvent être trouvées, au gré des années, dans la littérature scientifique, comme le modèle FLEET [387] ou d'autres extensions probabilistes [108, 372]. L'adaptation au véhicule électrique a généré de nombreuses publications au cours des dix dernières années en fonction des contraintes prises en compte [454, 453, 159, 94, 466, 11, 91].

Une fois les stations de recharge en place, la planification du trajet pour les conducteurs de véhicules électriques doit être résolue. Le problème du plus court chemin pour véhicules électriques entraîne de nouvelles contraintes liées à ce type de mobilité. La plus évidente est la possibilité, ou la nécessité selon l'état de charge de la batterie du véhicule, de s'arrêter à une station de recharge en cours de trajet. Une deuxième contrainte, qu'il ne faut pas sous-estimer, réside dans la capacité d'un véhicule électrique à récupérer de l'énergie lors des phases de descentes, impliquant des coûts négatifs sur les arcs et disqualifiant certains algorithmes d'étiquetage comme Dijkstra [124]. Une fois ces précautions prises, le problème peut être résolu par différentes techniques. À l'instar de l'allocation des stations de recharge qui s'inspire des stations de véhicules ambulanciers, beaucoup a été fait à propos des stations-service (stations essence), à l'aide de techniques de programmation dynamique [266, 265, 264, 414, 415]. Pour les véhicules électriques, il s'agit de minimiser la distance parcourue, ou l'énergie consommée, avec un ensemble de stations de recharges atteignable depuis l'origine et sachant un état de charge au début du trajet. Malgré quelques contributions notables [416, 389, 306, 411, 3], la complexité des approches empêche un passage à l'échelle d'une ville ou une intégration future dans un calculateur d'itinéraires. Ces motivations ont guidé nos travaux sur ce thème.

1.5.4.2 Le plus court chemin stochastique

La recherche d'un ou de plusieurs plus courts chemins peut intervenir dans beaucoup d'applications de la vie réelle. Cette recherche est facile dans le cas déterministe, au sens où des algorithmes peuvent la résoudre en des temps polynomiaux comme ceux de Dijkstra ou Bellmann-Ford évoqués plus hauts. Elle devient difficile si les arcs du graphe présentent un caractère aléatoire, c'est-à-dire s'ils sont soumis à une incertitude, ou s'il sont, en d'autres termes, stochastiques. À cet égard, une modélisation statistique des temps de parcours, telle que présentée dans la question de recherche Q2, revient, si elle est utilisée pour valuer chaque arc, à leur donner un caractère stochastique. Le problème du plus court chemin stochastique a été abordé très tôt [160, 305] et de façon très variée dans sa formulation et sa terminologie. La terminologie *SSPP* (ou *Stochastic Shortest Path Problem*) est la plus répandue, mais des variantes comme le *LET* (*Least Expected Time*), le *SOTA* (*Stochastic On-Time Arrival*) ou encore l'*ORP* (ou *Optimal Routing Problem*) sont possibles. Mais les problèmes associés aux différents termes peuvent varier, rendant l'état de l'art aussi foisonnant que trouble. En se plaçant du point de vue de l'utilisateur et de l'aversion au risque, il est ainsi possible de chercher à minimiser le pire temps de trajet possible, le temps moyen, rechercher le chemin le plus sûr, ou le chemin le plus fiable selon un budget-temps, chaque objectif entraînant un sous-problème différent.

Le problème *LET* consiste à rebasculer sur un problème déterministe en prenant comme poids des arcs soit la plus petite réalisation de la variable temps de parcours [301], soit son espérance [302, 300, 472, 128]. Des algorithmes d'étiquetage sont ensuite employés pour la résolution. Sous le terme *SSPP*, on retrouve une formulation plus raffinée du problème avec l'ajout dans la fonction objectif d'un terme relatif à l'incertitude, bien souvent la variance [278, 162, 392, 468, 282]. Du point de vue de l'utilisateur, cette composante peut correspondre à l'incertitude (souvent la variance) pondérée par un paramètre individuel lié à l'aversion au risque [83, 84, 85, 162, 252, 282, 332, 358, 392, 396, 438, 468, 487]. Cette bi-critérisation de la fonction objectif varie en fonction des pondérations choisies pour les deux objectifs et la façon d'intégrer la variance. Un système de pondération convexe est souvent adopté [332, 358, 85, 83, 84, 204]. Néanmoins, dans le problème *SSPP*, la non-additivité des variances sur les différents arcs rencontrés pose une difficulté, bien soulignée par Khani *et al.* [228].

Le problème *ORP* traite quant à lui de politiques de routage. Il s'agit de mettre en place une règle de décision qui va spécifier à chaque nœud dit de décision, le prochain nœud à sélectionner en fonction des temps de parcours réalisés, des coûts sur les arcs et du temps [164, 394, 125].

Dans une des premières (si ce n'est la première) contributions au problème du plus court chemin stochastique, Frank définit le chemin optimal comme celui maximisant la probabilité de réaliser un temps de parcours inférieur ou égal à un certain seuil [160]. De nombreuses années après, des chercheurs se sont inspirés de cet objectif pour décliner la formulation *SOTA* pour *Stochastic On-Time Arrival*, [146, 147], dans laquelle on maximise la probabilité d'être à l'heure, compte tenu d'un budget de temps donné. L'algorithme emploie une fonction récursive qui applique une convolution en série des distributions associées aux arcs parcourus, et dans laquelle le budget de temps de parcours diminue au gré des nœuds visités et des réalisations de temps de parcours sur les arcs. Alors que *SOTA* adapte la requête en cours de trajet, une alternative majeure, *SPOTAR*, propose une approche *a priori* en calculant les chemins localement optimaux [330]. La résolution est fondée sur la mise à jour, à chaque nœud, d'un vecteur d'étiquettes (une étiquette optimale par budget de temps), structure comparable à celle utilisée pour les calculs du plus court chemin dans les graphes dynamiques. D'autres acronymes peuvent être mentionnés dans la littérature (*PUE* [74], *RSPP* [83, 84, 85], *MRCP* [358], *MRP* [409], *RASP* [464] et *MPTTB* [487]) mais les problèmes se rapportent toujours à un des problèmes évoqués plus haut. Si la résolution du plus court chemin est claire dans le cas déterministe, grâce au principe de Bellman (tout sous-chemin d'un plus court chemin est un plus court chemin), des difficultés nouvelles se posent dans le cas stochastique. Le contournement du principe de Bellman [278] amène ainsi un certain nombre de difficultés méthodologiques. Les techniques de modélisation et de résolution convoquent la programmation mathématique, la programmation dynamique ou les algorithmes d'étiquetage. Pour ces derniers, la comparaison de vecteurs d'étiquettes implique d'introduire des notions de dominance stochastique [301, 303, 464]. D'autre part, certaines contributions prennent en compte la définition dynamique des temps de parcours [330, 340, 464].

1.5.4.3 Deux enjeux méthodologiques du point de vue des calculateurs d'itinéraires : le multimodal et le multiobjectif

Les algorithmes de calcul de plus court chemin constituent le cœur des planificateurs d'itinéraires existants. Une analyse des solutions existantes avait été proposée dans un rapport du projet FP7 *eCOMPASS* [373]. Les auteurs analysent les services web existants parmi lesquels *kvv.de*

(Karlsruhe, Allemagne), youdrive.gr (Grèce) et Google Transit et relèvent un certain nombre de lacunes à combler, en particulier un manque de fonctionnalités : intégration complète du multimodal, prise en compte des émissions environnementales et d'autres critères, véhicule à la demande, accélération du calcul avec des heuristiques non optimales, etc. Des conclusions similaires peuvent être retrouvées dans un article plus récent de notre ancien collègue Bradley Casey qui passe en revue une dizaine de planificateurs multimodaux et leur capacité à intégrer différents modes (comme l'automobile, la marche à pied, les transports en commun, bateaux, vélos, avions) ou différents objectifs (calcul du chemin le plus rapide, le moins cher, le plus respectueux de l'environnement, proposant le moins de marche à pied possible, limitant le nombre de changements de mode ?) [71]. Dans ce domaine, un fort enjeu réside dans la disponibilité et la qualité des données en entrée, et de l'usage de l'information temps réel disponible sur le réseau. Comme le confirme une étude plus récente [400], malgré de nets progrès en dix ans, l'amélioration des calculateurs d'itinéraires reste un défi pour tous les acteurs de la mobilité intelligente. Derrière la production d'une information précise pour les usagers se trouvent les données, bien sûr, mais aussi les algorithmes. Nous avons choisi de nous intéresser aux deux grands enjeux que sont les réseaux multimodaux et la résolution multi-objectifs.

Dans le cas d'un réseau multimodal, la première question est celle de la représentation en graphe pour laquelle plusieurs variantes sont possibles : multi-graphe [327, 45, 149, 280, 208], graphe multi-valué [485, 99, 244] ou encore graphe multi-niveau [34, 398]. Si le problème reste polynomial dans sa version multi-modale simple, il devient NP-difficile dès lors que l'on rajoute des contraintes telles que, par exemple, une limite sur le nombre de changements de modes : la recherche du chemin optimal dans un réseau multi-modal est alors traduite en la recherche d'une combinaison valide et optimale de modes de transport. Le fait que le problème soit NP-difficile dans ce cas implique que sa résolution se heurte rapidement à un phénomène d'explosion combinatoire. Afin de le contourner, les algorithmes sont généralement fondés sur des approches heuristiques qui ne garantissent pas de trouver la solution optimale mais, en contrepartie, ont des complexités polynomiales qui permettent un réel passage à l'échelle. Citons par exemple, Febraro et Sacone [149], qui ont fourni un des premiers modèles pour un réseau multimodal de quatre modes et proposé un algorithme heuristique, Lozano et Storchi [280], qui ont proposé un algorithme heuristique incrémental très utilisé par la suite, ou encore Pham [353], qui a récemment proposé un algorithme de recherche locale pour calculer des chemins optimaux sous diverses contraintes. Une autre difficulté inhérente aux problèmes de transport réside dans le fait que les temps de transport ne sont pas statiques mais peuvent fluctuer en fonction de l'heure de la journée ou de la météo. Ce problème peut être modélisé à l'aide de réseaux espace-temps [339, 215, 485]. Pallottino et Scutellà [339] ont développé l'algorithme Chrono-SPT pour résoudre ce problème de plus court chemin dynamique dans le cas mono-modal et montré comment l'étendre au cas multi-modal. De récents résultats peuvent être trouvés dans [15] et [234]. Comme dans la thèse d'Aurélié Bousquet [61], ces travaux font usage d'un automate fini non déterministe qui modélise les transitions, séquences d'arcs, possibles.

Les problèmes d'optimisation multi-objectifs dans les réseaux ont fait l'objet de nombreux travaux depuis les années 1980 [105, 188, 289, 455, 410, 317, 197, 102, 402, 182, 287, 420, 315, 368, 46, 383, 137]. Des synthèses bibliographiques ont été proposées dans Garroppo *et al.* [165], Di Puglia Pugliese et Guerriero [364], Jozefowicz *et al.* [216] et Banos *et al.* [21]. Dans une optique d'implémentation dans un calculateur d'itinéraires, une approche intuitive consiste à contourner une pure résolution multi-objectif en calculant d'abord les k plus courts chemins selon le critère le plus important (exemple : temps) puis choisir parmi ces k solutions selon d'autres critères. Pour

les k plus courts chemins, les algorithmes polynomiaux de Yen [473] et Eppstein [143] sont les références les plus répandues.

Pour le cas multimodal et multiobjectif, Lozano et Storchi ont proposé une version modifiée de l'algorithme Chrono-SPT de Pallotino et Scutella [339] dans laquelle le calculateur multimodal optimise un trajet *one-to-one* minimisant le temps de parcours et le nombre de changements de mode [279, 280]. Dans [101], les auteurs convertissent le problème en mono-objectif. Ils trouvent les solutions optimales au sens d'un seul critère mais en contraignant tous les autres. En conséquence, toutes les solutions se retrouvent dans un sous-ensemble sur lequel le concept de dominance (au sens de Pareto) est appliqué pour conserver celles présentant les meilleurs compromis. Cette stratégie peut néanmoins ignorer des solutions acceptables. Dans [294], les auteurs se positionnent dans une optique de développement d'un système d'information pour les usagers. L'algorithme minimise la distance parcourue, le temps de parcours et le coût du trajet, en bus et métro. Les auteurs adoptent une stratégie d'expansion du graphe initial. Le modèle prend en compte le temps de changement de mode et la marche à pied entre les stations.

1.5.4.4 Les techniques d'accélération

Ultraconnecté, exigeant, l'utilisateur du 21^{ème} siècle souhaite une information personnalisée et adaptable en temps réel. Cette exigence pose de réels problèmes algorithmiques et la modernisation des calculateurs d'itinéraires appelle à des compromis entre raffinement des fonctionnalités et rapidité des calculs. Pour répondre aux exigences industrielles de passage à l'échelle et de temps de calcul, des techniques d'accélération algorithmique (*speed-up techniques*) ont été inventées, parmi lesquelles l'heuristique A^* (*A-star*) fait figure de référence [190]. Il s'agit de guider la recherche du plus court chemin vers la direction la plus sensée en intégrant dans le coût des arcs la distance euclidienne entre chaque nœud et la destination. La méthode est exacte si l'heuristique utilisée sous-estime la distance réelle. Dans l'optique d'accélérer le populaire A^* , de nombreuses extensions ont ensuite fait leur apparition. L'algorithme ALT (*A* search, Landmarks and Triangle Inequality*) vise à améliorer l'heuristique en créant des repères (*landmarks*) qui accélèrent les calculs à l'aide d'inégalités triangulaires. Cette méthode demande un prétraitement plus long mais s'avère très utile si la topographie du réseau change peu. Plus récemment la technique du *transit node routing (TNR)* [37] insiste sur le préprocessing en pré-calculant des tables de distances sur un sous-ensemble de nœuds, les nœuds de transit. À partir d'une origine o , le nœud de transit devient un nœud d'accès de o s'il est le premier à être contenu dans un chemin le plus court depuis o . Le préprocessing va également stocker la distance entre une origine o et ses nœuds d'accès. Une autre heuristique, nommée *arc flags* [199] partitionne le graphe en K grandes zones homogènes. Chaque arc est associé à un vecteur de K bits. Le bit i est mis à 1 si l'arc appartient à un PCC vers un nœud de la cellule i . On se sert ensuite de nœuds "frontières" des cellules pour mettre à jour les *flags* dans un temps raisonnable. Dans une autre veine, les approches hiérarchiques vont chercher à gagner de l'espace mémoire en supprimant les nœuds et arcs les moins importants. Ainsi, les *highway hierarchies* [236] enlèvent les nœuds de degrés faibles en créant des raccourcis et réduisent le nombre d'arcs en supprimant ceux trop proches de l'origine ou de la destination. Si un arc est présent sur un PCC, mais situé loin de l'origine et de la destination, on le conserve (*highway edge*). La difficulté de ce genre d'approche se situe au niveau du réglage du paramètre de taille des zones "origine" et "destination". Dans la bataille des calculateurs d'itinéraires actuels, des techniques toujours plus raffinées sont proposées afin de gagner quelques millisecondes de temps de calcul pour une requête à l'échelle d'un pays (voire d'un continent). Parmi elles, les contractions

hiérarchiques consistent en une pré-classification des nœuds par ordre d'importance puis un ajout d'arcs "raccourcis" (*shortcuts*) entre certains couples de nœuds prioritaires (voir Geisberger [169] et Geisberger *et al.* [170]). La résolution de la requête s'effectue dans le nouveau graphe à l'aide d'une version bidirectionnelle de l'algorithme de Dijkstra. Cette technique d'accélération présente un pré-traitement parmi les moins longs : les expérimentations de Bast *et al.* [36] sur un graphe de 18 millions de nœuds et 42 millions d'arcs orientés rapportent un pré-traitement de l'ordre de la dizaine de minutes, pour des requêtes résolues en moyenne en 0.1 milliseconde. Par ailleurs, les contractions hiérarchiques ont été étudiées en détail par Batz [39] et Batz *et al.* [40] dans le cas d'un graphe dynamique. Dans le cas stochastique, une contribution est à mettre au crédit de Sabran *et al.* [379] dans le cadre du problème *SOTA*. En combinant deux méthodes d'accélération dont *arcs flags* les auteurs améliorent le temps de calculs par des facteurs compris entre 1.2 et 8.8 par rapport à une résolution stochastique classique (graphes allant jusqu'à 30000 nœuds et 72500 arcs environ).

1.5.4.5 Nature des contributions

Se positionner dans un état de l'art aussi riche que dispersé n'est pas chose aisée et n'importe lequel des thèmes évoqués dans notre troisième question de recherche (Q3) aurait pu concentrer l'essentiel de nos efforts. Nos contributions ont adopté un positionnement pratique, du point de vue d'un calculateur d'itinéraires et des fonctionnalités à y ajouter. Le défi est bien de proposer des nouveaux services à l'utilisateur (défi 3), comme le montre la figure 1.7.

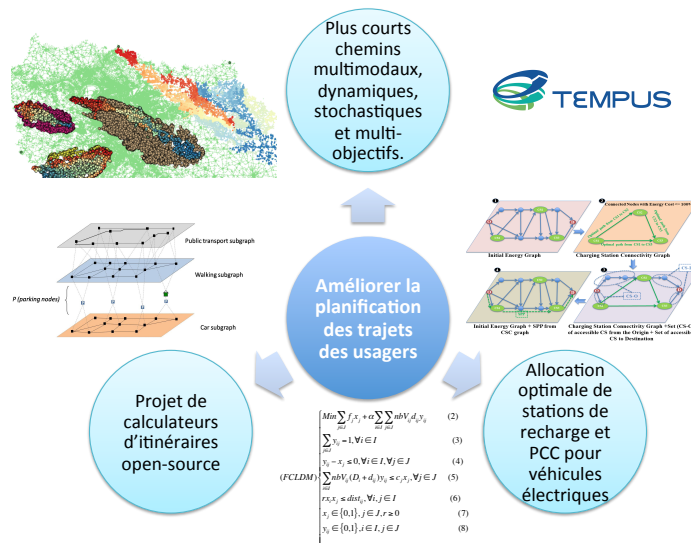


Fig. 1.7.: Thématiques de recherche en réponse au défi 3

L'utilisateur d'aujourd'hui veut personnaliser son trajet en fonction de ses propres contraintes. Une première contrainte peut être le mode de transport. Parmi les nouvelles tendances, nos travaux ont considéré l'usage d'un véhicule électrique et l'adaptation des algorithmes à cette nouvelle forme de déplacement. Dans un second temps, à la suite de nos réalisations liées à la modélisation statistique des temps de parcours, l'intégration de l'incertitude dans le calcul du plus court chemin s'est imposée comme une voie de recherche prometteuse. Tous ces travaux ont vocation à être intégrés à moyen terme dans une plateforme développée depuis 2007 au LICIT (laboratoire de l'IFSTTAR à Lyon), nommée TEMPUS, qui est un calculateur de plus court chemin. Le portage du projet en Open-Source a ouvert des discussions stimulantes avec des industriels comme OSLANDIA

et MAPPY, qui ont orienté nos travaux vers les techniques d'accélération aptes à optimiser les performances de calcul à l'échelle de grands graphes. En somme, nous choisissons de mettre en valeur dans le chapitre 4 les contributions suivantes :

1. Outils d'optimisation pour l'électromobilité : affectation des stations de recharge et plus court chemin pour véhicules électriques,
2. Contributions au plus court chemin stochastique et analyse de l'impact des modélisations statistiques employées,
3. Les contractions hiérarchiques : analyse de sensibilité et adaptation au cas stochastique,
4. Le projet TEMPUS : un calculateur d'itinéraires multimodal open-source.

D'autres travaux, notamment liés à l'optimisation multi-objectif, feront partie intégrante du projet de recherche présenté au chapitre 6.

1.5.5 Organisation de la suite du manuscrit

Ce long chapitre de contexte et positionnement de nos travaux a proposé une vision synthétique de notre démarche méthodologique et de nos contributions, pour le seul thème de la mobilité connectée. En réponse à trois grands défis, nos travaux s'inscrivent dans les disciplines des mathématiques appliquées (modélisation du trafic), de l'informatique (systèmes multi-agents et simulation), de la fouille de données et de l'optimisation. Nous avons choisi de présenter notre travail sur la santé connectée, démarré en mars 2016, dans un chapitre à part, présenté comme un chapitre d'ouverture vers un nouveau domaine d'application, qui contient néanmoins des contributions originales, publiées ou proches de l'être. Le chapitre 6 articulera les deux domaines d'application autour d'un projet de recherche en fouille de données et optimisation. Avant cela, la figure 1.8 illustre la manière dont sont organisés les quatre chapitres de contributions à suivre. Le **chapitre 2** rassemble nos travaux

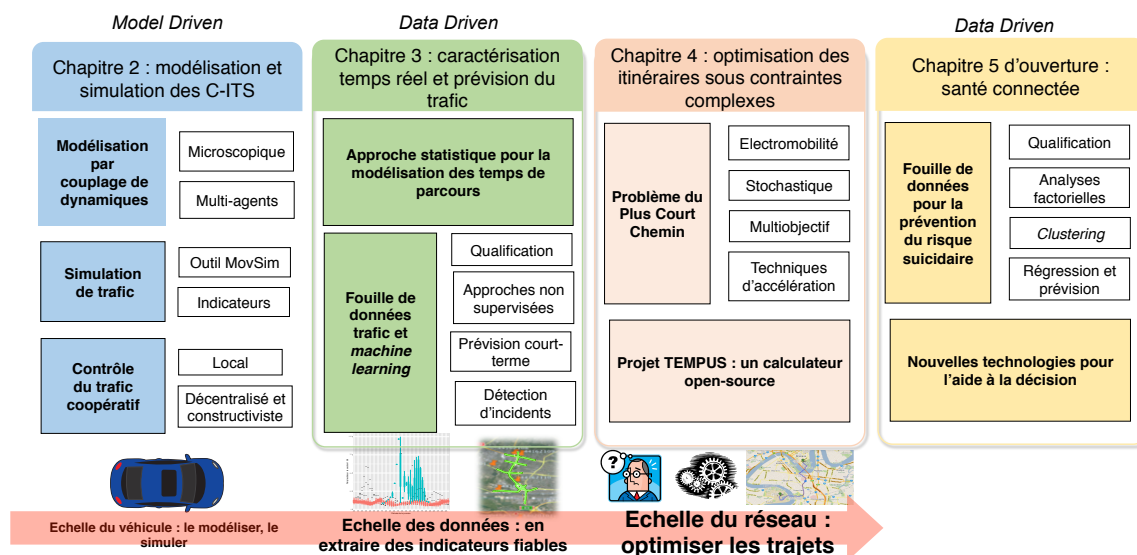


Fig. 1.8.: Organisation du manuscrit d'habilitation

sur la modélisation, la simulation et le contrôle des C-ITS. En l'absence de véhicules connectés et donc de données associées, l'approche part de modèles validés ensuite en simulation (*model driven*). Dans le **chapitre 3**, les données sont le point de départ et il s'agit de contruire des modèles statistiques ou indicateurs de caractérisation et prévision de trafic à partir des nouvelles sources de

données (approches *data driven*). Les résultats des chapitres 2 et 3 peuvent tout à fait servir de données d'entrées, historiques, temps-réel ou prédictives, d'un calculateur d'itinéraire. C'est de ce point de vue que le **chapitre 4** présentera nos résultats sur l'optimisation des itinéraires, ainsi que le projet TEMPUS.

Les chapitres suivants seront dédiés à des travaux actuels ou futurs. Dans le **chapitre 5**, nous verrons comment nos approches non supervisées et supervisées en fouille de données peuvent être déclinées en méthodes d'aide à la décision pour la santé connectée. Nous présenterons les premiers résultats d'une collaboration avec le CHU Brest, via l'encadrement de la thèse de Sofian Berrouiguet, sur le thème de la prévention du risque suicidaire. Ce chapitre d'ouverture fait la transition avec le suivant, qui présente une vision de nos futurs travaux : le **chapitre 6** montre les perspectives de l'ensemble de nos recherches et les englobe dans un projet de recherche pour les prochaines années. Le projet ne sera plus centré sur un seul domaine d'application comme la mobilité mais adoptera plutôt une approche générique autour de la fouille de données et de l'optimisation de données spatiotemporelles pouvant être adaptée à plusieurs catégories de systèmes complexes.

À la lumière de l'organisation proposée, le parcours du manuscrit peut s'adapter à différentes lectures :

- Le lecteur pressé, ou adepte des manuscrits d'habilitation très courts, peut se contenter de la lecture des chapitres 1, 5, 6. Il y trouvera une synthèse de nos travaux sur la mobilité intelligente, un chapitre d'ouverture sur la santé connectée, et le projet de recherche.
- Le lecteur souhaitant plus de détails sur les contributions sur la mobilité intelligente pourra approfondir la lecture en consultant les chapitres centraux 2, 3, 4.

Modélisation, simulation et contrôle des C-ITS

” *Les gens peuvent choisir n'importe quelle couleur pour la Ford T, du moment que c'est noir*

— **Henry Ford**
Industriel américain

Dans ce chapitre, le lecteur trouvera de plus amples détails de nos contributions sur les C-ITS, en réponse au premier défi mentionné au chapitre 1. En termes d'encadrement de la recherche et de collaboration scientifique, ces contributions ont commencé dès 2011 et notre prise de poste à l'IFSTTAR avec la thèse de Julien Monteil, accompagnée d'échanges stimulants avec le Pr. Jacques Sau, puis se sont poursuivies à travers la thèse de Maxime Guériau, co-dirigée avec Salima Hassas et Frédéric Armetta du LIRIS (université Lyon 1). Pour traiter notre premier défi et la question de recherche associée, les premiers efforts se sont concentrés sur la question de la modélisation microscopique du trafic routier et de l'intégration des C-ITS dans les modèles existants. Si des résultats mathématiques montrent que les systèmes coopératifs peuvent amener de la stabilité dans le flux (section 2.1), la complexité des interactions et l'analogie avec les systèmes complexes distribués nous ont orienté vers l'étude du paradigme multi-agent. Nous proposons ainsi une modélisation des C-ITS par couplage de dynamiques (section 2.2). Le cadre de modélisation multi-agent a ensuite été intégrée dans une plateforme de simulation construite comme un véritable outil d'aide à la décision pour le déploiement des systèmes coopératifs (section 2.3). Les premiers résultats en simulation et l'application de stratégies de contrôle locales ont montré certaines limites face à l'apparition de nouveaux comportements. C'est pourquoi nous proposons, dans une contribution plus prospective, une nouvelle forme de contrôle par apprentissage constructiviste (section 2.5).

2.1 Modélisation théorique des C-ITS : quand multi-anticipation rime avec stabilité

2.1.1 Un modèle multi-anticipatif bilatéral (BMA)

La première étape de notre travail a consisté à introduire la communication dans les modèles classiques de trafic puis à en évaluer les effets d'un point de vue théorique. Comme nous l'avons rappelé en 1.5.2.1, plusieurs approches sont possibles en modélisation du trafic. Les niveaux meso- et macroscopiques restent selon nous trop agrégés pour bien représenter les communications entre véhicules. Le niveau microscopique s'est imposé comme le plus naturel. A cette échelle, les véhicules mettent à jour leur vitesse en fonction d'une perception d'une interdistance et d'une vitesse

relative avec le véhicule précédent [427, 26, 196]. La forme générale des modèles microscopiques dits lois de poursuite est

$$\dot{x}_n = f(\dot{x}_n, \Delta x_{n+1}, \Delta \dot{x}_{n+1}) \quad (2.1)$$

où f est l'équation du modèle considéré, \dot{x}_n la vitesse du véhicule courant, Δx_{n+1} l'inter-distance entre le véhicule courant et son prédécesseur (qui nécessite de connaître leurs positions x_n et x_{n+1} ainsi que leurs longueurs l_n et l_{n+1} respectives), et $\Delta \dot{x}_{n+1}$ leur vitesse relative.

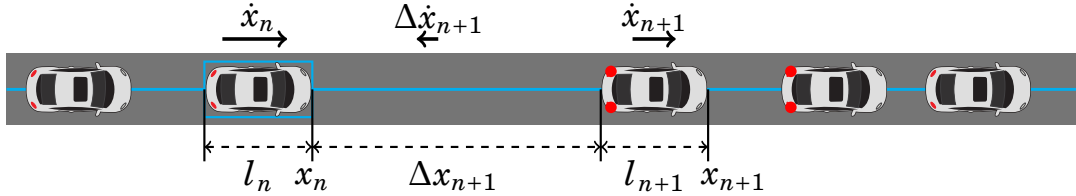


Fig. 2.1.: Interprétation graphique des paramètres utilisés dans les lois de poursuites (crédit image : Maxime Guériaux)

Une extension naturelle de ces modèles pour les C-ITS consiste à introduire la multi-anticipation. Ainsi Ge *et al.* [167] étendent le modèle *optimal velocity* de telle façon que chaque véhicule adapte sa vitesse et son accélération en fonction de plusieurs interdistances en aval. Nous avons montré que cette stratégie permet de stabiliser l'évolution des interdistances lors de conditions perturbées [310, 307]. A la suite de premiers résultats prometteurs, nous avons situé nos travaux dans le cadre d'approches multi-anticipatives dans lesquelles un véhicule adapte à chaque pas de temps son accélération selon l'information située dans un certain rayon d'interaction. Aucune source de bruit telle qu'un temps de latence dans la communication ou une éventuelle défaillance n'est, à ce stade, introduite. On note r le rayon d'interaction entre véhicules, de telle sorte que les voisins informatifs d'un véhicule i sont définis comme

$$N_i = \{j \in V : 0 < x_j - x_i \leq r\}, \quad (2.2)$$

où x_i représente la position du véhicule i et V l'ensemble des véhicules sur la section considérée. Ceci permet de définir une topologie d'un réseau de capteurs avec un certain rayon d'interaction r dont la magnitude peut varier en fonction des évolutions technologiques [277]. Les modèles microscopiques coopératifs correspondent à des modèles microscopiques multi-anticipatifs prenant en compte l'information de plusieurs véhicules devant [337] ou derrière [461] le véhicule courant, suivant l'équation générale suivante :

$$\ddot{x}_n = f(\dot{x}_n, \sum_{j=-m'}^m a_{nj} \Delta x_{n+j}, \sum_{j=-m'}^m b_{nj} \Delta \dot{x}_{n+j}), \quad (2.3)$$

où $(a_{nj})_{-m' \leq j \leq m}$ et $(b_{nj})_{-m' \leq j \leq m}$ sont les coefficients de coopération définissant l'importance de l'interaction entre le véhicule n et les véhicules environnants. Ces coefficients pondèrent l'importance de chaque véhicule dans la loi coopérative. Afin de simplifier les calculs, il est logique de considérer que $(a_{nj})_{0 \leq j \leq m} = (b_{nj})_{0 \leq j \leq m}$ et $\sum_{j=0}^m a_{nj} = 1$.

La figure 2.2 illustre le principe de l'analogie entre communication et multi-anticipation.

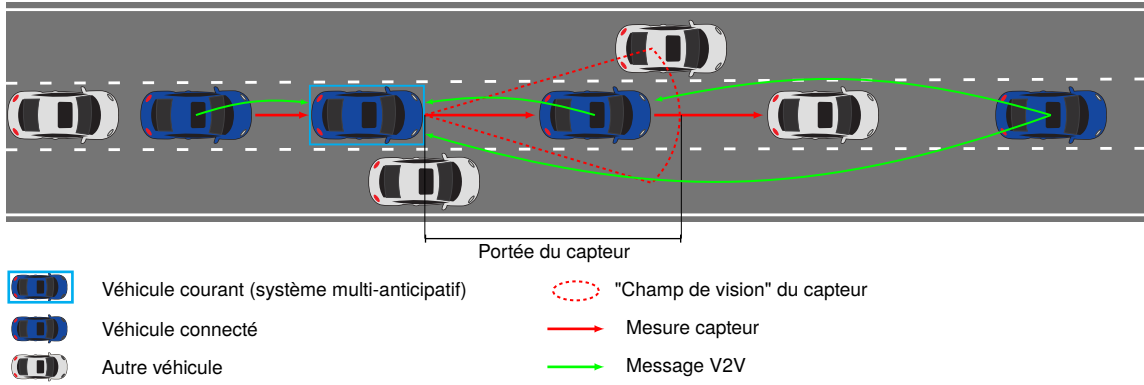


Fig. 2.2.: Représentation de la perception d'un véhicule équipé d'un système de contrôle basé sur le principe de la multi-anticipation (crédit image : Maxime Guériaux).

2.1.2 Principales conclusions des études de stabilité

A partir d'une formulation d'un modèle multi-anticipatif bilatéral, nos premiers travaux ont contribué au domaine de la théorie du trafic à travers des analyses de stabilité [309, 382], qui déterminent dans quelles conditions, à partir d'un trafic stable, l'introduction d'une petite perturbation va se propager au cours du temps. La question de la stabilité est cruciale en théorie du trafic de par l'analogie entre conditions instables (au niveau théorique) et propagation d'ondes de congestion dites *stop and go*. Alors qu'une large panoplie de méthodes est disponible [425, 426, 221], la méthode la plus classique est l'analyse de stabilité linéaire qui va déterminer les conditions pour lesquelles une perturbation autour d'un état stable du système, va grandir avec le temps [462]. Ces perturbations vont correspondre à des freinages ou changements de voie pour des conditions de trafic réelles. En plus des analyses de stabilité linéaire, les méthodes non linéaires s'attachent à décrire le profil des ondes, dans des régimes faiblement non linéaires. L'astuce mathématique consiste à retomber sur une équation de type Korteweg de Vries (KdV) qui est un cas particulier d'équations aux dérivées partielles [241, 251, 156]. Une équation kdV peut par exemple gouverner l'évolution d'une vague marine dans un régime faiblement non linéaire [121]. Les solutions sont des solitons qui maintiennent leur forme et leur vitesse dans le temps [132]. Cet outil a déjà été utilisé pour la modélisation du trafic par un groupe actif de chercheurs [316, 167, 168, 166]. Par souci de concision, nous nous limiterons à une description succincte du principe de l'analyse de stabilité linéaire et au résumé de nos principales conclusions. La première étape consiste à introduire une perturbation autour d'un état d'équilibre du trafic. Un tel état d'équilibre est atteint quand $\Delta \dot{x} = 0$ et $\ddot{x} = 0$, *i.e.*

$$f(\dot{x}_{eq}, \Delta x_{eq}, 0) = 0, \quad (2.4)$$

où \dot{x}_{eq} et Δx_{eq} sont les vitesses et interdistances d'équilibre. Des petites perturbations $y_n(t)$ autour de l'état d'équilibre sont introduites :

$$\begin{aligned} x_n(t) &= x_{neq}(t) + y_n(t), \\ \dot{x}_n(t) &= \dot{x}_{eq} + \dot{y}_n(t), \\ \Delta x_n(t) &= \Delta x_{eq}(t) + \Delta y_n(t), \\ \Delta \dot{x}_n(t) &= \Delta \dot{y}_n(t). \end{aligned}$$

L'analyse de stabilité linéaire consiste à écrire la perturbation sous forme de Fourier puis, après des développements de Taylor, à obtenir la relation de dispersion du système dynamique. La résolution de l'équation de dispersion nécessite des calculs poussés qui sont détaillés dans [309]. En regardant le signe de la partie réelle du polynôme caractéristique, des conditions de stabilité sont obtenues. Il est intéressant de constater que la coopération augmente le domaine de stabilité, c'est-à-dire augmente les jeux de paramètres possibles pour lesquels la perturbation ne se propage pas dans le diagramme espace temps. La figure 2.3 montre l'augmentation de ce domaine de stabilité dans le cas d'un trafic coopératif (modèle multi-anticipatif).

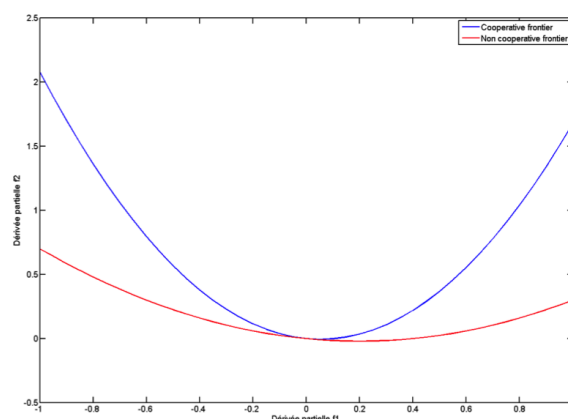


Fig. 2.3.: Domaines de stabilité (le système est stable dans la partie supérieure aux courbes).

De même, à la limite de cette zone de stabilité, il est possible de déterminer par la méthode de perturbation réductrice l'équation unique d'un soliton, qui est en fait l'équation de l'onde de perturbation près de la frontière d'instabilité. Un développement de Korteweg de Vries (KdV) donne l'amplitude et la vitesse du soliton, l'extension étant faite pour le cas bilatéral et multi-anticipatif. Les conclusions qui en découlent sont les suivantes : la coopération introduite diminue l'amplitude de l'équation du soliton, d'un facteur qui dépend des conditions de circulation et du modèle choisi. Des premières études en simulation ont permis d'illustrer ces résultats théoriques. Par exemple, la figure 2.4 montre l'atténuation d'une onde de congestion lorsque la moitié du flux de trafic est constitué de véhicules coopératifs.

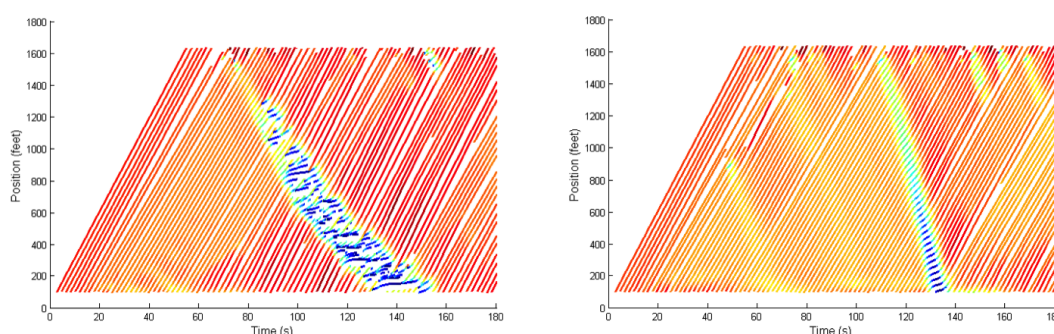


Fig. 2.4.: Développement d'ondes de congestion. A gauche 0% de véhicules coopératifs. A droite, 50% de véhicules équipés (*i.e* modèle BMA)

Notre analyse théorique des systèmes coopératifs a été complétée par la proposition d'une méthodologie graphique, nommée *root locus*, qui permet d'observer l'évolution du lieu des racines

dans le plan complexe en fonction de la variation des paramètres des modèles physiques [382]. Appliquée aux systèmes coopératifs, la méthode permet d'observer comment le réglage de certains paramètres affecte les propriétés de stabilité des modèles. La figure 2.5 illustre ainsi comment la prise en compte de plus de véhicules dans une loi multi-anticipative peut augmenter le domaine de stabilité. Un tel outil graphique se révélerait utile lors du design de stratégies de contrôle, ou la construction de pelotons de véhicules automatiques.

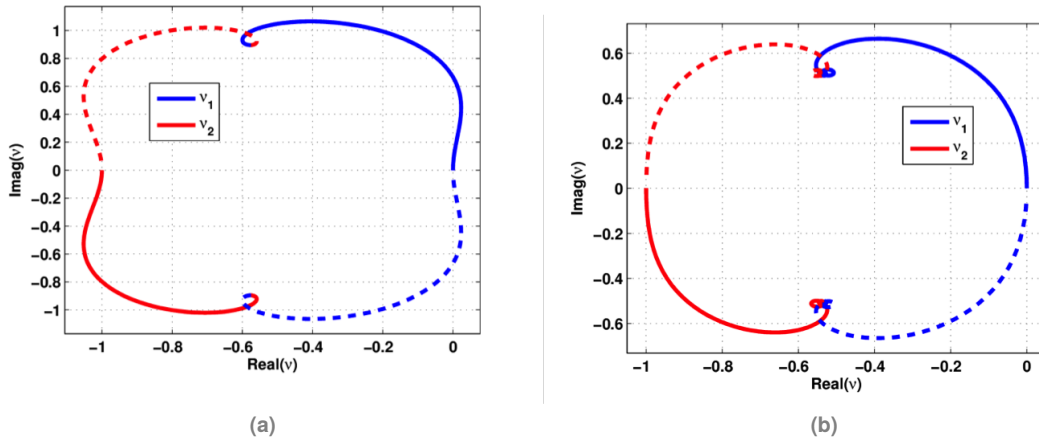


Fig. 2.5.: Root Locus : trafic instable avec prise en compte de 3 véhicules leaders (a) ou 6 véhicules leaders (b)

2.2 Modèle multi-agent par couplage de dynamiques

Le texte de cette section est largement adapté d'un article paru en anglais dans Transportation Research Part C [180].

2.2.1 Motivation et présentation générale du modèle

Malgré la pertinence des résultats théoriques précédents, qui montrent qu'un trafic coopératif tend à stabiliser le trafic, le cadre de modélisation classique de la théorie du trafic nous a semblé réducteur par rapport à la richesse des interactions induites par les C-ITS. Plutôt que de produire des extensions de plus en plus complexes de modèles déterministes existants, nous avons choisi de les englober dans un cadre de modélisation plus flexible, au sein d'un paradigme multi-agent. Par leur modularité et leur souplesse, les systèmes multi-agents se prêtent parfaitement à la modélisation d'un système complexe distribué comme celui des C-ITS [195]. Nous avons proposé un modèle générique fondé sur trois couches complémentaires qui couplent les dynamiques de contrôle et d'information. Ce couplage vise à prendre en compte la fiabilité de l'information tout en permettant une homogénéisation du flux de trafic grâce aux systèmes coopératifs. La figure 2.6 présente les interactions entre les différentes couches. La couche physique a trait à la dynamique des véhicules, elle s'appuie sur un modèle microscopique tel qu'introduit précédemment. La couche de communication gouverne les échanges d'information par des règles de proximité et de fiabilité. Cette fiabilité de l'information est elle-même calculée grâce à une couche de confiance qui permet à un agent d'évaluer la confiance qu'il peut avoir envers d'autres [388, 328]. Les couches de communication et de confiance influencent la couche physique à travers les paramètres du modèle multi-anticipatif bilatéral.

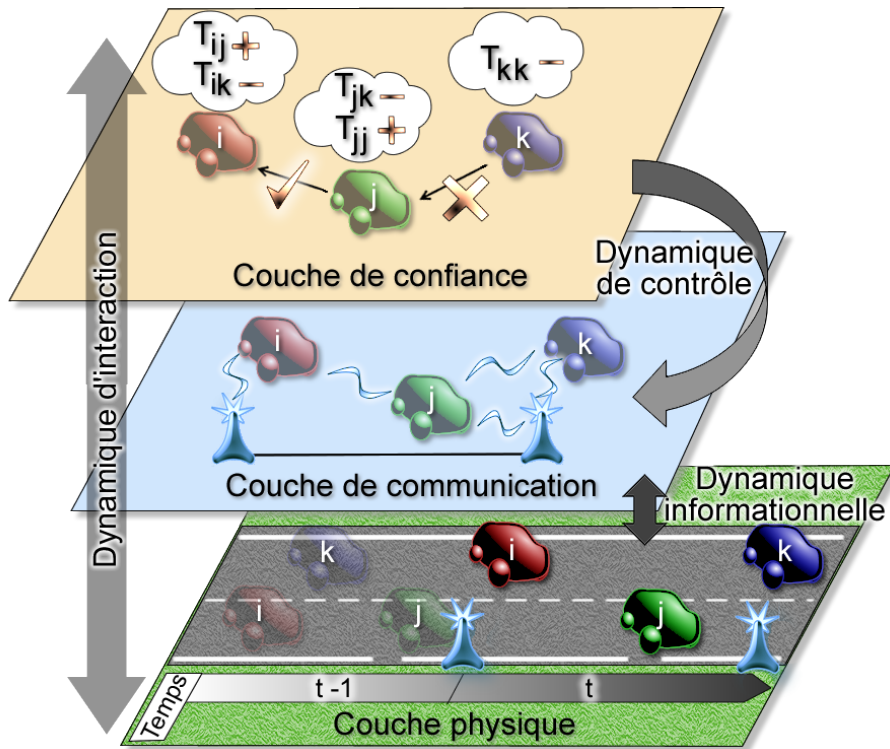


Fig. 2.6.: Modélisation conceptuelle par couplage de dynamiques

2.2.2 Couche physique

La couche physique décrit la dynamique des agents et leur capacités d'émission et de réception d'information. Les véhicules représentent la première catégorie d'agents. Il faut faire la différence entre les véhicules classiques, que nous nommons non coopératifs, et les véhicules connectés, que nous nommons véhicules coopératifs. Ces véhicules coopératifs échangent de l'information à propos des conditions de trafic ou de paramètres dépendant du temps (ex : vitesses, positions) par les canaux de communication *Vehicle to Vehicle* (V2V), *Infrastructure to Vehicle* (I2V) et *Vehicle to Infrastructure* (V2I) comme le décrit la figure 2.7 qui met en miroir les similitudes et différences entre véhicules non coopératifs et véhicules coopératifs.

Les unités de bord de route (UBR), ou *Road Side Units* (RSU), récoltent l'information issue des véhicules et peuvent estimer les conditions de trafic sur la section couverte. L'échange d'information n'est pas limité à des paramètres cinématiques mais il inclut également des mesures stockées par les capteurs intra-véhiculaires ou des valeurs de confiance que les véhicules évaluent entre eux. En somme, le concept de véhicule coopératif peut être ici compris comme un véhicule partiellement autonome capable de percevoir son environnement grâce aux capteurs et à la communication, agissant de manière collective. Si le contrôle longitudinal est délégué au système coopératif automatisé, l'humain n'en reste pas moins responsable de son véhicule, en particulier en ce qui concerne les manoeuvres de changement de voies. Dans un premier temps, les comportements individuels des agents, en particulier le caractère collectif ou égoïste de chaque conducteur, ne sont pas modélisés. De même, l'échelle spatiale se limite à une section de route multi-voies de type autoroutier ou péri-urbain. Les comportements coopératifs aux intersections, tels que ceux introduits par [488] à partir du *Cooperative Adaptive Cruise Control* (C-ACC), n'ont pas été pris en considération par nos études. Ainsi, le comportement longitudinal d'un véhicule coopératif est déterminé par l'application

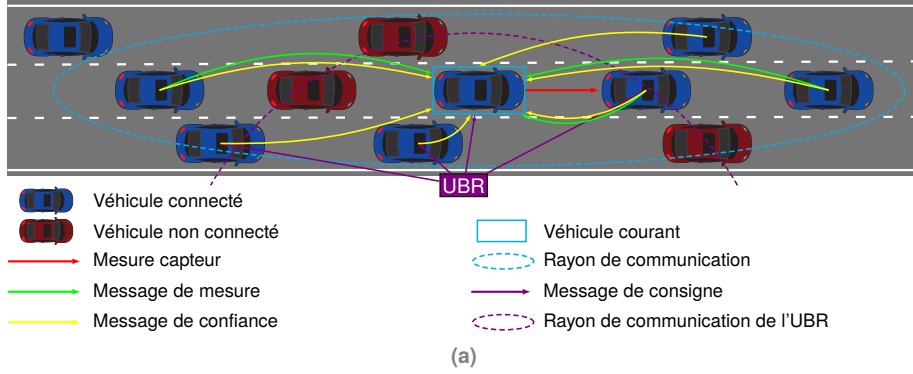


Fig. 2.7.: (a) Schéma général des systèmes coopératifs sur une section de route (b) véhicule non coopératif vs. véhicule coopératif.

d'un modèle multi-anticipatif bilatéral (BMA) tandis que la perception des véhicules non connectés se limite au seul véhicule prédécesseur, sans capacité de communication. Précisons à présent les deux modèles physiques pour véhicules non coopératifs et véhicules coopératifs.

Véhicules non coopératifs : à partir de l'équation générale de la loi de poursuite 2.1, la forme de la fonction f oriente le choix d'un modèle précis. Parmi les modèles de la littérature, le modèle IDM, *Intelligent Driver Model*, proposé par Treiber et Kesting [424, 227] a retenu notre attention. Ce modèle est bien connu pour reproduire de manière réaliste et sans collision la dynamique du trafic [427]. Le modèle est construit en utilisant des termes d'accélération et de freinage couplés à un certain nombre de paramètres physiques. La fonction f_{IDM} s'écrit

$$f_{IDM}(\dot{x}_i, \Delta x_{i+1}, \Delta \dot{x}_{i+1}) = \ddot{x}_i = a \left[1 - \left(\frac{\dot{x}_i}{v_0} \right)^{\gamma} - \left(\frac{s^*(\dot{x}_i, \Delta \dot{x}_{i+1})}{\Delta x_{i+1}} \right)^2 \right] \quad (2.5)$$

où

- $s^*(\dot{x}_i, \Delta \dot{x}_i) = s_0 + \max\left(0, \dot{x}_i T + \frac{\dot{x}_i \Delta \dot{x}_i}{2\sqrt{ab}}\right)$ exprime l'inter-distance temporelle désirée ;
- a et b sont respectivement l'accélération maximum et l'accélération désirée ;
- v_0 est la vitesse désirée ;
- s_0 est la distance inter-véhiculaire à l'arrêt ;
- T est l'inter-distance de sécurité (temporelle) ;

- γ est un coefficient d'agressivité de l'accélération.

En plus de l'équation microscopique, les véhicules peuvent effectuer un changement de voie avant de mettre à jour leur accélération. Le modèle de changement de voie appliqué est le modèle MOBIL, qui estime si un espace suffisant et sécurisé est disponible sur une voie adjacente (changement de voie dit opportuniste) [224]. La tendance du conducteur à effectuer ou non ce changement de voie est modélisée par un paramètre de politesse.

Véhicules coopératifs : pour les véhicules connectés, le modèle bilatéral multi-anticipatif (BMA) de l'équation 2.6 est repris dans la forme particulière IDM, avec une pondération des véhicules situés dans le rayon de communication :

$$\ddot{x}_i = f_{IDM}(\dot{x}_i, \sum_j a_{ij} \Delta x_{i+j}, \sum_j a_{ij} \Delta \dot{x}_{i+j}) \quad (2.6)$$

où a_{ij} est le coefficient d'interaction entre l'agent i et l'agent perçu j tandis que f_{IDM} a été écrit en 2.5. Ce coefficient d'interaction découle des couches communication et confiance, par la formule suivante :

$$a_{ij} = p_{ij} \cdot T_{ij}, \quad (2.7)$$

où p_{ij} représente une règle de proximité définie par la suite en 2.8, et T_{ij} traduit la confiance qu'un agent i a en un autre agent j , également précisée par la suite dans l'équation 2.9.

2.2.3 Couche de communication

Dans la couche de communication, deux types de message sont échangés par les véhicules : des mesures de capteurs et des mesures de confiance. Afin de pondérer les échanges de messages, nous définissons une règle de proximité p_{ij} comme suit :

$$P_{ij} = \frac{|\Delta \dot{x}_{i+j}|}{(\Delta x_{i+j})^\delta}, \text{ puis } p_{ij} = \frac{P_{ij}}{\sum_{k \in \mathcal{P}_i} P_{ik}}. \quad (2.8)$$

où \mathcal{P}_i est l'ensemble des agents perçus par i et δ un terme contrôlant les interactions en aval et en amont. Les valeurs retenues sont respectivement 2 et 0.5 afin de respecter la physique du trafic routier et assurer une stabilité du flux. L'idée est de donner plus d'importance aux véhicules proches et à ceux présentant un fort différentiel de vitesse. Alors que les véhicules connectés présentent, de par leur modèle, un comportement multi-anticipatif, l'homogénéité du flot peut être maintenue par une telle règle de proximité, comme l'illustre la figure 2.8.

En cas de problème de communication, ou tout simplement en l'absence de véhicules connectés, le modèle BMA 2.6 correspond au modèle IDM 2.5 des véhicules non coopératifs, assurant une continuité quelles que soient la densité et la qualité des zones traversées par un véhicule.

2.2.4 Couche de confiance

A l'égard de la notion de confiance, notre travail est influencé par le concept de délégation tel que décrit par Castelfranchi dans un célèbre article [73]. Un véhicule coopératif peut être affilié à un processus de délégation en ce sens qu'il a besoin des actions des autres agents pour atteindre

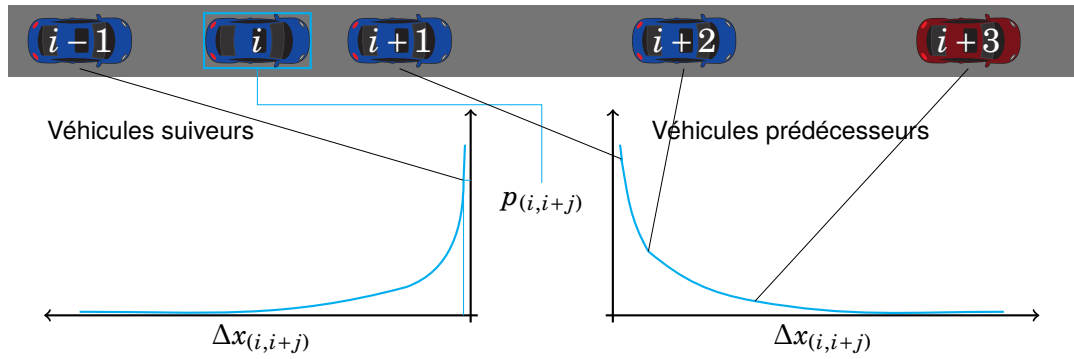


Fig. 2.8.: Illustration de l'application de la règle de proximité. Le véhicule courant n pondère l'influence des véhicules suiveurs et prédécesseurs, connectés (en bleu) et non-connectés (rouge).

son but et qu'il doit de surcroît les inclure dans ses plans. La notion de confiance est intrinsèque au concept de délégation. En fonction de la confiance en un autre agent, un agent de notre système coopératif échange ou non des informations sur ses mesures physiques ou à propos d'autres agents. Dans le cas des C-ITS, le concept de confiance permet de modéliser la fiabilité des capteurs et donc de l'information échangée. De nombreux modèles de confiance ou de réputation sont présents dans la littérature associée et nous renvoyons le lecteur à la revue complète de Pinyol et Sabater-Mir [355] pour une étude détaillée. Nos travaux ont étendu le modèle de confiance de NGuyen *et al.* pour l'adapter au trafic routier. Ce modèle repose initialement sur une structure de données de type *TrustNet* [388]. La représentation se fonde sur les expériences non agrégées des agents et repose sur la majorité. Si ce dernier point pourrait être considéré comme le grand défaut de l'approche, l'on peut espérer qu'une application aussi contrôlée que le trafic routier et les C-ITS ne mènera pas à des situations où la majorité a tort ! D'autre part, notre adaptation des *TrustNets* ne prend pas en compte les comportements égoïstes ou tricheurs : seule la fiabilité de l'information basée sur les capteurs embarqués et leurs défaillances influence les confiances que les agents ont les uns envers les autres. Pendant le processus de décision, les agents comparent l'information disponible et chaque agent i stocke et met à jour une opinion locale sur chacun des agents j précédemment perçus. Cette opinion locale est la valeur de confiance T_{ij} calculée comme suit :

$$T_{ij} = \frac{T_{ii}DT_{ij} + IT_{ij}}{T_{ii} + 1}. \quad (2.9)$$

La valeur de confiance est calculée à chaque pas de simulation t . Formellement, c'est une somme pondérée de trois composants $\in [0,1]$, que nous détaillons ci-après :

1. **Confiance directe** DT_{ij} : elle représente l'erreur de mesure. Le choix effectué consiste à mesurer un pourcentage d'erreur sur la mesure de l'interdistance par le véhicule j :

$$DT_{ij} = \frac{|\Delta x_i + x_i - x_j|}{\Delta x_i + x_i}, \quad (2.10)$$

où Δx_i et x_i correspondent aux mesures du véhicule courant i (à l'aide d'un télémètre laser et d'un appareil de localisation), et x_j est la position dans le repère global partagée par l'agent j (donné par son propre système de localisation).

2. **Confiance indirecte** IT_{ij} : cette valeur de confiance fait référence à l'information envoyée par les autres agents à propos de j . Elle se calcule comme suit :

$$IT_{ij} = \frac{\sum_{k \in \mathcal{W}_i} T_{ik} T_{kj}}{\sum_{k \in \mathcal{W}_i} T_{ik}}, \quad (2.11)$$

où \mathcal{W}_i est l'ensemble des agents k qui partagent une valeur de confiance sur j avec l'agent i . Ces agents k doivent se situer dans le rayon d'interaction de j et également avoir déjà calculé une valeur de confiance à propos de l'agent j . Ces deux contraintes réduisent significativement le risque d'explosion combinatoire.

3. **Confiance en soi** T_{ii} : il s'agit d'une confiance moyenne des autres agents envers l'agent i au temps t , et donc en connaissance de tous les échanges passés :

$$T_{ii} = \frac{\sum_{k \in \mathcal{C}_i} T_{kk} T_{ki}}{\sum_{k \in \mathcal{C}_i} T_{kk}}, \quad (2.12)$$

où \mathcal{C}_i représente l'ensemble des agents k qui partagent leurs valeurs de confiance sur i . De manière identique, les agents k doivent être accessibles depuis i au moment de l'échange d'information.

En ce qui concerne l'infrastructure des C-ITS, les unités de bord de route sont une deuxième catégorie d'agent visant à assurer une fonction de supervision et de contrôle des sections sous leur autorité. Le rôle des UBR, entités fixes installées au bord des routes par les gestionnaires, se limite à cette tâche de supervision. Ainsi, le cadre défini correspond à un système multi-agent où les agents mobiles sont les véhicules qui évoluent dans un environnement partagé avec des agents statiques, les UBR, en charge de tâches de contrôle. Cette approche se veut réaliste quant aux futurs cas de déploiement de C-ITS pour lesquels la localisation des UBR est une contrainte imposée.

2.3 D'un simulateur de trafic microscopique à un simulateur multi-agent des C-ITS

Après un effort de modélisation des C-ITS par couplage de dynamiques, notre équipe a développé un simulateur multi-agent de trafic coopératif, dans l'optique d'obtenir des résultats de simulation robustes et reproductibles pour l'évaluation des impacts des C-ITS sur le trafic routier et des stratégies de contrôle. Nous avons réalisé une extension complète d'un simulateur existant pour s'accorder aux concepts SMA et intégrer notre architecture de modélisation du trafic coopératif. L'objectif est de garantir aux véhicules une perception locale et partielle de leur environnement (donnée par leurs capteurs respectifs), de modéliser leurs interactions complexes (communication V2V/V2I) et de leur assurer un processus de décision autonome.

2.3.1 Architecture

Comme nous l'avons vu en 1.5.2.4, après une étude comparative des simulateurs existants, nous avons sélectionné l'outil MovSim (*Multi-model Open-source Vehicular-traffic SIMulator* [428]) comme base de travail à partir de laquelle une extension multi-agent a été proposée. De nombreux objets ont été créés, parfaitement intégrés à la base existante puisque la plupart des classes originales héritent de ces nouvelles classes (voir figure 2.9b).

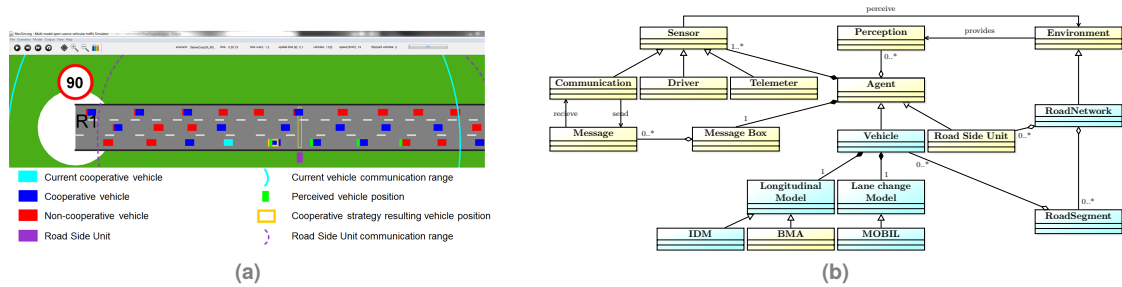


Fig. 2.9.: (a) Capture d'écran de l'interface de MovSim. (b) Diagramme de classes de l'extension multi-agent de MovSim. Les classes bleues sont les classes du simulateur d'origine

La version originale de MovSim inclut une interface graphique qui affiche une vue de la simulation en cours avec plusieurs paramètres microscopiques (du véhicule) ou macroscopiques (ex : débits en entrée et sortie) ainsi que la représentation du réseau et des véhicules y circulant (figure 2.9a). Un paramétrage chromatique est offert afin d'attirer l'attention sur des aspects ciblés : vitesses des véhicules, accélérations/freinages, changements de voie etc. Notre extension implémente le modèle multi-agent de trafic coopératif présenté plus haut et permet de simuler un véritable trafic coopératif. La plupart des changements sont ainsi relatifs à des mécanismes de perception et de communication. Pour accompagner ces changements, de nouvelles visualisations permettent de distinguer les véhicules connectés des véhicules non connectés, d'afficher les rayons de communication et de couverture des véhicules connectés et des UBR. De plus, un outil dynamique de représentation des diagrammes espace-temps offre une visualisation de l'évolution des trajectoires individuelles de véhicules au cours du temps. Ces diagrammes illustrent la propagation des ondes de choc au cours du temps et les effets des C-ITS sur l'homogénéisation du flux. Une vidéo de ces développements a été présentée lors de la conférence AAAI 2015 [179].

2.3.2 Perception-Décision-Action

L'environnement était déjà au centre de l'outil original et il reste le point d'entrée de la simulation. Nous nous plaçons dans le cadre de systèmes multi-agent situés qui doit garantir l'autonomie et l'indépendance des agents dans le processus de décision. En particulier, l'ordre d'exécution des agents ne doit pas affecter le résultat de la simulation, qui repose sur une exécution ordonnancée à temps discret. La continuité temporelle entre les différents pas de temps de simulation est assurée par une boucle perception-décision-action d'un agent autonome [154]. L'idée consiste à partir d'une situation figée de l'environnement à un temps t puis de calculer la perception correspondante pour chaque véhicule. Ensuite, chaque véhicule décide de l'action à effectuer en fonction de son environnement, choix fondé sur des mesures des capteurs embarqués et un processus autonome de décision. Enfin, l'environnement est mis à jour en appliquant les actions des véhicules. Le triptyque perception-décision-action est développé dans la figure 1.8.

Perception : grâce aux capteurs embarqués (télémetre laser, récepteur GPS) et à la communication, les véhicules connectés construisent une perception étendue de leur environnement. Il s'agit de passer d'une représentation globale du système multi-agent à une représentation locale et partielle. Les mesures directes concernent des positions, interdistances, vitesses relatives tandis que les mesures indirectes sont relatives à des messages de communication partagés par les autres

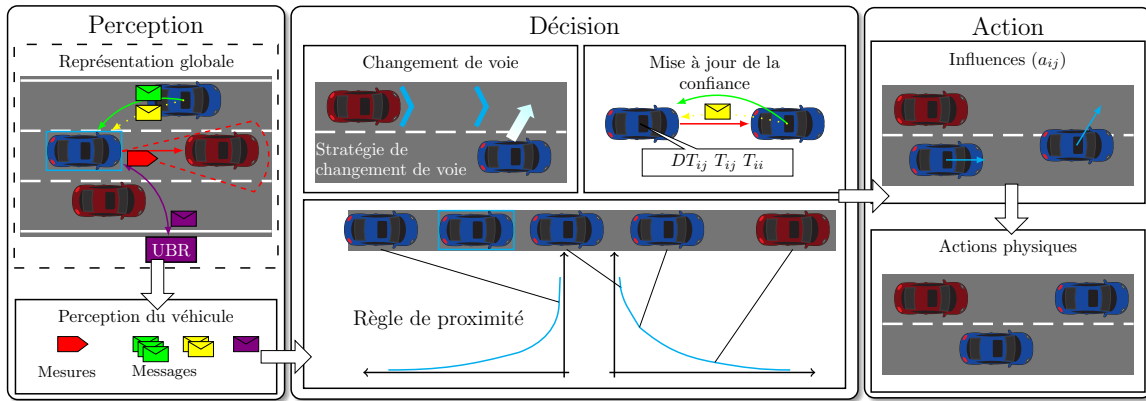


Fig. 2.10.: Schéma d'opérations de la simulation

véhicules connectés. La perception des UBR peut intégrer de l'information provenant d'un ou plusieurs capteurs et les UBR ont la capacité d'échanger de l'information via la communication.

Décision : chaque véhicule connecté peut être vu comme un agent réactif avec son propre processus de décision. Après l'étape de perception, chaque véhicule traite les messages reçus pour étendre sa perception. Trois types de message sont possibles :

- Les **messages de mesure** (V2V et V2I) : informations des capteurs, partagées par un véhicule à ses voisins (diffusion) ;
- Les **messages de confiance** (V2V) : portent le triplet $(i, j, T_{i,j})$, *i.e.* la valeur de confiance qu'un agent i a en un autre agent j ;
- Les **messages de consigne** (I2V) : permettent aux UBR de transmettre des recommandations afin de contrôler un flux de véhicules connectés.

Conformément aux spécifications du standard Wi-Fi pour les ITS, nous ne considérons pas les échanges d'informations partiels, les messages modifiés, les pertes ou la duplication d'informations. Tous les messages qui sont envoyés sont reçus par les véhicules équipés qui sont à portée de l'émetteur, après une latence de 100 ms. Les messages n'ont donc pas de destinataires connus, ils sont diffusés en *broadcast* dans la zone d'émission. Une fois la perception des véhicules connectés étendue grâce à la communication, chaque véhicule peut pondérer l'importance des véhicules présents dans sa perception en suivant la stratégie coopérative, c'est-à-dire en appliquant les règles de proximité et de confiance. Les agents mémorisent les coefficients calculés sous la forme d'un vecteur d'influence. Cette structure de données encapsule l'information concernant les véhicules perçus (devant et derrière) jusqu'à la dernière étape d'action de la boucle de simulation.

Action : la dernière étape du triplet perception-décision-action correspond à la couche physique de l'architecture trois couches. Les positions et vitesses des agents véhicules sont modifiées après application des décisions calculées à l'étape précédente. Les véhicules non connectés appliquent un modèle microscopique IDM classique (équation 2.5) tandis que les véhicules connectés mettent en oeuvre la loi bilatérale multi-anticipative BMA (2.6). En ce qui concerne le modèle de changement de voie, une seule stratégie est appliquée pour les deux type de véhicules. Il s'agit du modèle MOBIL (*Minimizing Overall Braking decelerations Induced by Lane changes*) [224].

2.4 Résultats en simulation

2.4.1 Données et plan expérimental

La variabilité des comportements de conduite est une cause majeure du déclenchement de perturbations et ainsi de la formation d'ondes de choc. L'objectif opérationnel de nos travaux, qui a motivé le développement du simulateur, est de fournir un outil d'aide à la décision capable d'évaluer l'impact des C-ITS et en particulier leur capacité à lisser la variabilité du trafic. Le but premier d'un outil de simulation est de reproduire cette variabilité, si possible à partir de données réelles. Dans un travail non détaillé dans ce manuscrit d'HDR, nous avons analysé un jeu de données réelles du programme américain NGSIM [308]. Le travail a mené à l'élaboration d'une chaîne complète d'analyse de points méthodologiques fondamentaux allant du filtrage des données à l'échantillonnage de jeux de paramètres calibrés pour une simulation de trafic réaliste. Le cas d'études correspond à un trafic d'autoroute urbaine américaine dense (40 veh/km/voie) évoluant sur une section de 3 voies de dix kilomètres de long. Le modèle IDM a été calibré pour reproduire fidèlement cette situation. Le trafic est initialisé aléatoirement avec un débit d'entrée de 2000 véhicules par heure et une vitesse de 90 km/h qui correspond à la vitesse limite sur ce type de sections. Les capteurs des véhicules coopératifs simulés présentent un écart type de mesure de 5%, qui va traduire l'incertitude. Les résultats présentés par la suite répondent à deux objectifs :

1. Evaluer l'impact du pourcentage de véhicules connectés ainsi que celui de la fiabilité de l'information dans le cas d'un trafic mixte et dense,
2. proposer et évaluer des stratégies locales de gestion de trafic à partir des deux types d'agents en présence (UBR et véhicules connectés).

2.4.2 Impact des C-ITS sur la stabilité du trafic et prise en compte de la fiabilité de l'information

Homogénéisation du flux : dans le premier scénario, nous montrons des trajectoires de véhicules au cours de 30 minutes de simulation pour différents taux de véhicules équipés (vs. véhicules non équipés). Le diagramme de la figure 2.11a nous montre ainsi les trajectoires de la situation de référence avec 0% de véhicules équipés.

La figure nous montre que de petites perturbations se propagent au cours du temps. Ces perturbations sont des ondes de congestion, de type *stop and go*, qui résultent des fluctuations intrinsèques du flux de trafic, c'est-à-dire de l'hétérogénéité naturelle des véhicules et des manoeuvres de changement de voie. En augmentant le pourcentage de véhicules équipés, ou taux de pénétration (de 20% - figure 2.11.b - à 100% - figure 2.11.f) il apparaît qu'un trafic coopératif est plus résilient face à la propagation de telles perturbations. Mieux, au delà d'un certain seuil, qui s'établit à 40% de véhicules équipés dans le présent scénario, l'anticipation induite par la communication V2V permet de lisser les perturbations et d'obtenir un trafic homogène. En effet, des forts pourcentages de véhicules équipés homogénéisent le trafic autour d'un état d'équilibre (vitesses et interdistances similaires), réduisant *de facto* le nombre de changements de voie.

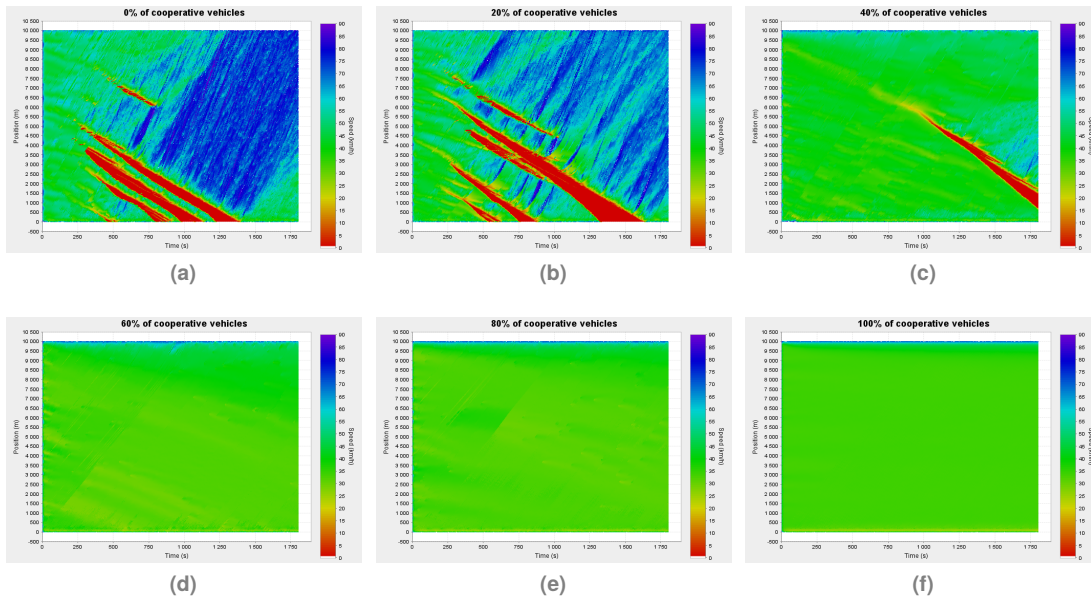


Fig. 2.11.: Trajectoires de véhicules pour différents taux de pénétration. (a) 0% (b) 20% (c) 40% (d) 60% (e) 80% (f) 100%

Comparaison avec le système ACC (*Adaptive Cruise Control*) : le système ACC automatise la fonction de conduite longitudinale à travers des capteurs de distance et de vitesse relative par rapport au prédécesseur direct. Un véhicule coopératif va plus loin en prenant en compte plus d'information dans son système de décision. L'objectif du deuxième scénario est de comparer le comportement d'un flux partiellement composé de véhicules coopératifs vs. un flux composé de véhicules équipés d'ACC. Les paramètres d'initialisation sont les mêmes que dans le cas précédent. Afin d'évaluer l'homogénéité du flux de façon plus rigoureuse, l'indicateur de **désaccord de groupe ou Group Disagreement Value**, noté G_d , est calculé : cet indicateur est lié à l'objectif d'obtenir un trafic fluide grâce à la coopération entre les véhicules. Adapté des travaux de Choi *et al.* [93], il permet d'évaluer l'homogénéité d'un groupe d'agents qui s'exprime :

$$G_d = \sum_{(i,j) \in N^2} b_{ij} (\dot{x}_j - \dot{x}_i)^2, \quad (2.13)$$

où b_{ij} est une valeur binaire d'adjacence (0 si les véhicules ne sont pas dans leur rayon de communication, 1 sinon), et N désigne le nombre de véhicules sur la section. L'indicateur G_d permet d'évaluer l'homogénéité d'un flux sur une section lorsque le nombre de véhicules est comparable. Un G_d nul représente un trafic parfaitement homogène.

La figure 2.12 montre les valeurs de l'indicateur G_d de désaccord de groupe pour différentes fractions de véhicules équipés. On remarque que pour de faibles taux de pénétration, un trafic coopératif apporte un bénéfice plus fort sur le flux en terme d'homogénéité des vitesses. D'autre part, les transitions entre régimes congestionnés et fluides sont plus souples avec les véhicules connectés qu'avec un système ACC. Ces résultats peuvent évidemment dépendre des contextes d'initialisation et des systèmes testés. L'approche devra être comparée avec d'autres technologies coopératives comme le *Cooperative Adaptive Cruise Control* (C-ACC) [433] ou d'autres contributions récentes sur le thème [450].

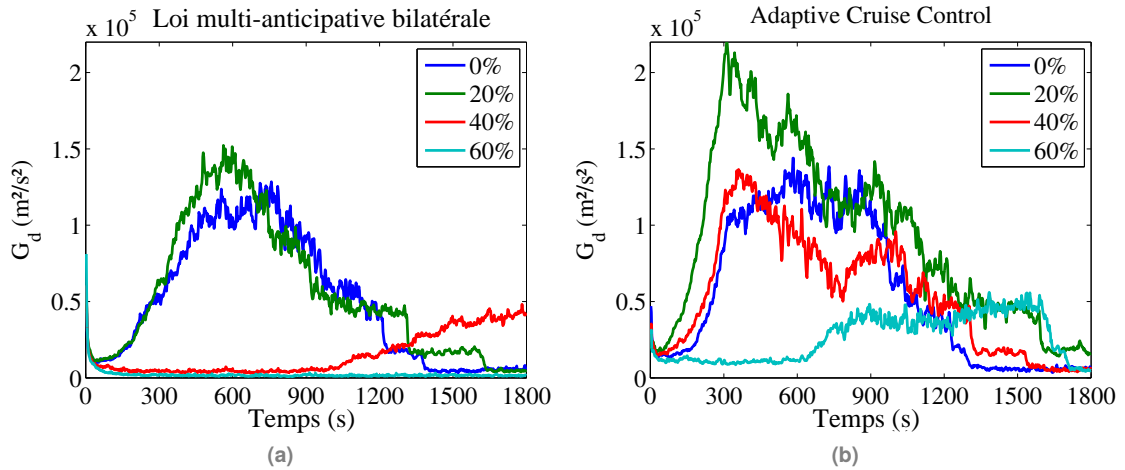


Fig. 2.12.: Valeurs de désaccord de groupe pour différents taux de véhicules équipés : modèle multi-anticipatif bilatéral (BMA) vs. ACC.

Confiance et fiabilité de l'information : le but est de montrer comment la notion de confiance peut modéliser les dérives de certains capteurs. Ainsi, dans ce scénario, certains véhicules connectés sont équipés de capteurs perturbés, qui produisent des erreurs (+/- 50%) sur leurs mesures d'interdistances et de vitesses relatives. L'indicateur choisi dans ce scénario est le **Temps intervéhiculaire – Post-Encroachment Time PET** : il est défini comme la différence temporelle entre deux véhicules passant par une même zone spatiale. Le *PET* représente ainsi un risque direct de collision (qui survient lorsqu'il prend la valeur zéro). Il s'écrit [304, 173] :

$$PET_i(t) = \frac{\Delta x_{i+1}}{\dot{x}_i} \quad (2.14)$$

La figure 2.13 montre les distributions de *PET* pour différents modèles : un modèle sans couche de confiance, un modèle comparatif nommé FIRE [206], et notre modèle de confiance. La mémoire stocke une limite de 6000 résultats d'interaction qui correspond à 10 minutes de simulation avec un pas de temps de 0.1s. Nous renvoyons le lecteur à [206] pour plus de détails sur le réglage sur le paramétrage du modèle FIRE. Dans la figure 2.13, le pourcentage de capteurs bruités ne concerne que les véhicules connectés. Par exemple, si cette valeur est fixée à 25% et que le taux de pénétration des véhicules connectés est de 25%, la totalité des véhicules connectés ont des capteurs défectueux. Les résultats montrent que la couche de confiance, quel que soit le modèle utilisé, apporte un effet positif sur les valeurs de *PET*, en termes d'homogénéisation et de valeurs numériques. D'autre part, notre modèle de confiance inspiré des *TrustNets* se comporte de manière satisfaisante face à un modèle concurrent, en particulier lorsque le taux de pénétration augmente.

2.4.3 Stratégies de contrôle locales

La suite de nos expérimentations numériques s'est attachée à illustrer le rôle potentiel des unités de bord de route comme agents de supervision capables de collecter et disséminer une information de type *broadcast* sur l'état de trafic courant ou un évènement particulier sur la section. Sur ce dernier point, le message envoyé par l'UBR aux véhicules connectés inclut la localisation précise du message et sa nature. Les deux scénarios présentés dans ce manuscrit concernent une fermeture de voie et une insertion de véhicules. L'UBR, située en amont des zones concernées, envoie un

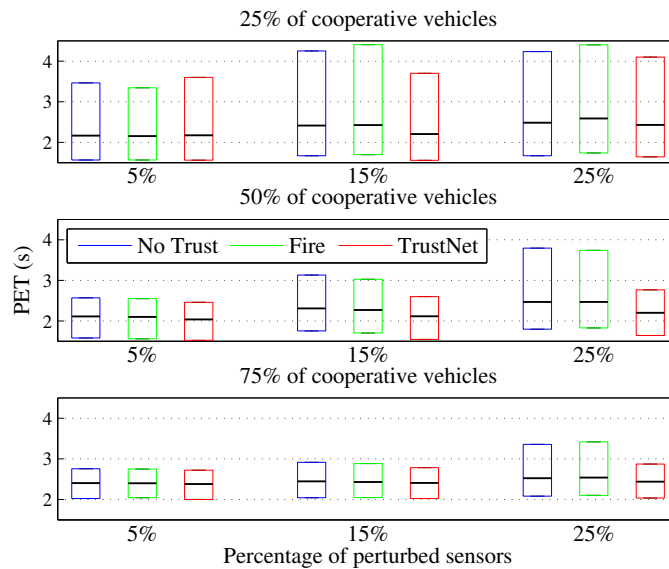


Fig. 2.13.: Distributions des *PET* (médiane, 1^{er} et 9^{ème} déciles) pour trois variantes de modèles de confiance. Du haut vers le bas : augmentation du % de véhicules équipés. De gauche à droite : augmentation du % de capteurs défaillants.

message de consigne de changements de voie qui sera reçu par 100% des véhicules situés dans le rayon d'interaction et rediffusé en amont par chaque véhicule. Nous faisons ici l'hypothèse que chaque véhicule recevant le message essaie de changer de voie, si les conditions le permettent.

Scénario d'insertion : une section convergente est représentée et l'expérimentation consiste à simuler 1000 veh/h sur toutes les voies pendant 20 minutes. La figure 2.14 montre l'état du trafic après 17 minutes et 3 secondes.

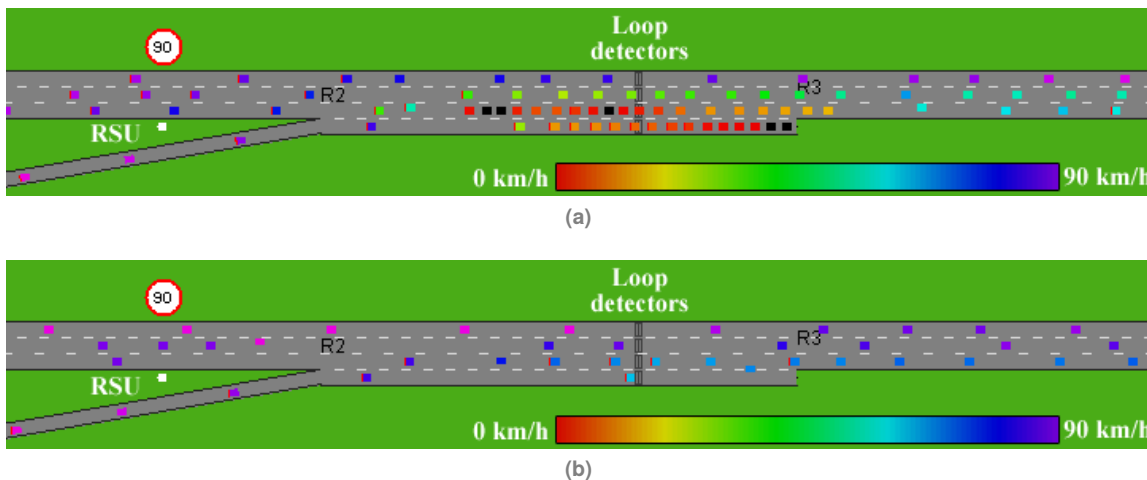


Fig. 2.14.: Scénario d'insertion (a) instabilité du trafic avec 0% de véhicules connectés et (b) 40% de véhicules connectés.

La figure 2.14a présente un cas typique d'insertion dans lequel, au bout d'un certain temps, la densité de trafic ne permet plus aux véhicules candidats de s'insérer dans la section principale, créant une onde de congestion se propageant en amont. Sur la figure 2.14b, le carré blanc représente la

position d'une UBR située en amont du convergent. Cette UBR demande aux véhicules d'appliquer un changement de voie "obligatoire" au lieu "d'opportuniste" dans la stratégie MOBIL de changement de voie. Ainsi, les véhicules connectés vont dans la mesure du possible se décaler vers les deux voies adjacentes jusqu'à ce que la perturbation s'estompe. La figure montre que, là encore, 40% de véhicules connectés suffisent à garantir un régime fluide au même temps de simulation, dans les mêmes conditions initiales. Des capteurs de type boucles électromagnétiques, associés à l'UBR, permettent de reporter sur la figure 2.15 les diagrammes fondamentaux débit/concentration sur les différentes voies. La figure confirme les bénéfices d'une stratégie de contrôle par les UBR. La

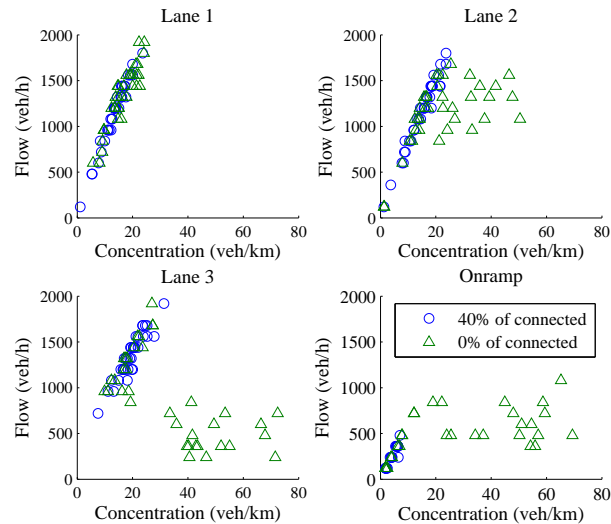


Fig. 2.15.: Scénario d'insertion. Diagrammes fondamentaux de données agrégées par pas de 1 min. (a) Voie 1 (b) Voie 2 (c) Voie 3 (d)

partie congestionnée du diagramme fondamental n'est jamais atteinte quand la consigne ciblée est envoyée à 40% des véhicules.

Scénario de fermeture de voie : le deuxième scénario est très courant puisqu'il concerne une fermeture temporaire de voie avec un passage de 3 à 2 voies qui implique un certain nombre de manoeuvres des véhicules tentant de se rabattre depuis la voie concernée. Ce cas d'école mène à des perturbations, que la figure 2.16 illustre au temps $t = 15 : 13$ minutes.

La stratégie de contrôle de l'UBR reste la même. Les véhicules de la voie 2 sont invités à se reporter sur la voie 3 afin de laisser de l'espace pour l'insertion des véhicules de la voie 1. Dans l'expérimentation présentée, cette consigne simple, suivie par 100% des 50% de véhicules connectés, suffit à résoudre le problème. Sur la figure 2.17, le diagramme fondamental confirme l'impression visuelle de la copie d'écran de l'interface du simulateur, avec une amélioration substantielle sur les voies 2 et 3, tandis que la voie 1 n'est pas affectée dans les deux cas.

Discussion sur les stratégies de contrôle locales : les résultats de ces stratégies de contrôle locales doivent être considérées comme des preuves de concepts, construites à partir d'hypothèses fortes qui amènent à nuancer la portée des résultats :

- les stratégies n'ont pas été testées dans toutes les situations de trafic et un algorithme de contrôle adaptatif doit être développé pour s'accomoder de différentes densités de trafic, composition du flux de véhicule, etc.

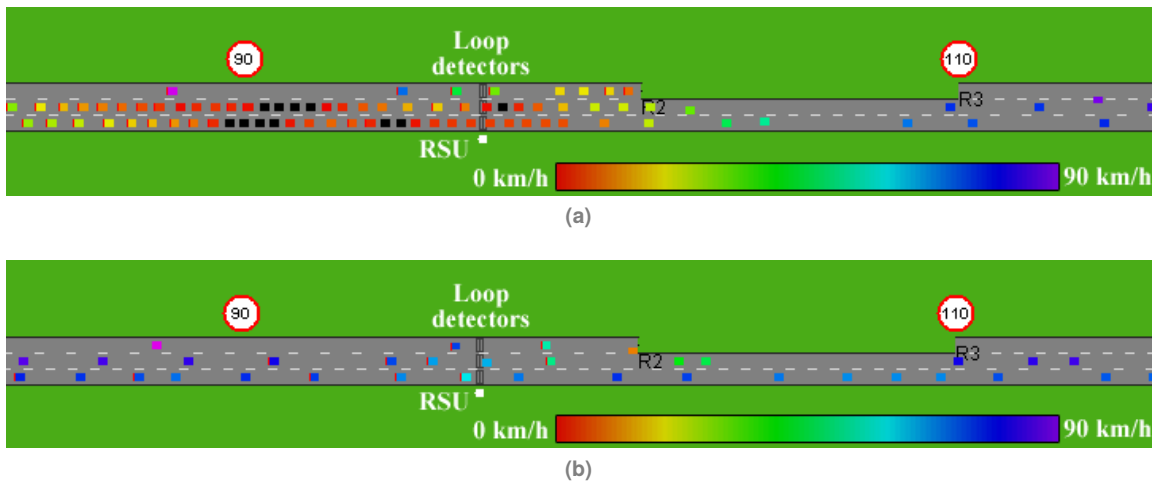


Fig. 2.16.: Scénario de fermeture de voie (a) instabilité du trafic avec 0% de véhicules connectés et (b) 50% de véhicules connectés.

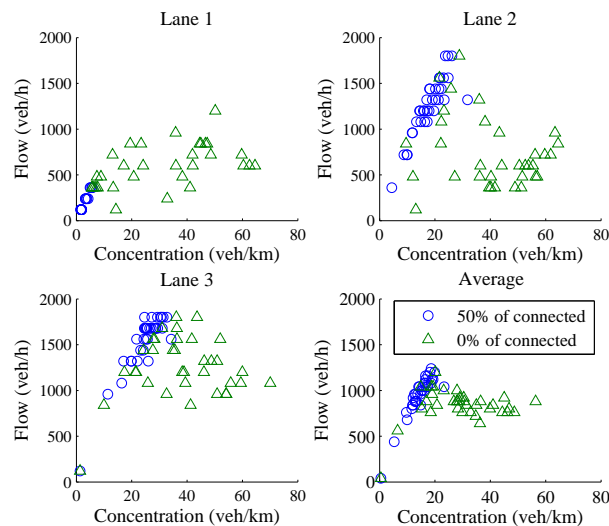


Fig. 2.17.: Scénario de fermeture de voie. Diagrammes fondamentaux de données agrégées par pas de 1 min. (a) Voie 1 (b) Voie 2 (c) Voie 3 (d) moyenne des 3 voies

- L'UBR répète un même message à tous les véhicules, sans adaptation dynamique.
- Le taux de suivi de consigne est de 100%.

Ainsi, des analyses de sensibilité complémentaires sont nécessaires pour généraliser, voire implémenter de tels mécanismes en situations réelles. Cependant, la flexibilité de l'architecture du simulateur permet de tester tout un éventail de stratégies de régulation et l'outil pourrait tout à fait assister la mise en oeuvre d'expérimentations réelles sur des sites ciblés.

2.4.4 Discussion

L'effet local d'un trafic coopératif reste, à ce stade de nos investigations, trop réducteur : la coopération entraîne une baisse des vitesses en parallèle d'une homogénéisation des vitesses et interdistances. La stratégie de contrôle local doit être raffinée en permettant des accélérations des véhicules dans

des situations bien choisies. D'autre part, ce travail a mis en lumière le rôle crucial que pourraient avoir les UBR dans la régulation du trafic. En tant qu'agents mésoscopiques, les UBR permettent un contrôle décentralisé mieux adapté à la diversité des situations rencontrées. En anticipant sur un monde composé de véhicules connectés, autonomes, ou même des formes hybrides de véhicules, il est logique d'envisager de laisser plus d'autonomie à l'agent UBR, qui pourrait alors apprendre les meilleures stratégies de contrôle à adopter en fonction d'un contexte et de sa mémoire des expériences passées. C'est ce à quoi s'attache la contribution suivante sur l'apprentissage constructive.

2.5 Contrôle par apprentissage constructiviste

Le texte suivant est repris de deux publications récentes dans les conférences ICTAI 2016 et JFSMA 2017 [177]. Nous avons eu l'honneur d'obtenir le prix du meilleur article scientifique pour le travail présenté en juillet 2017 aux JFSMA [178].

Le premier objectif est de proposer un modèle permettant à un système autonome de construire une représentation à partir de ses interactions bas-niveau avec son environnement. Le modèle suit une perspective constructiviste et couple deux processus inter-dépendants : la construction d'hypothèses d'états "perception-action" et l'élaboration d'une "stratégie de contrôle". Ces processus sont gérés par une population d'agents en compétition pour construire une représentation précise de l'environnement (ensemble d'états perception-action) et un processus de prise de décision permettant de choisir une action pertinente. Ce dernier mécanisme est modélisé comme une forme d'apprentissage par renforcement où le retour sur expérience est fourni par l'environnement et utilisé pour renforcer les hypothèses d'état perception-action proposées par les agents. L'originalité de notre contribution est d'utiliser un ensemble de représentations concurrentes modélisées sous la forme d'agents autonomes qui s'adaptent dynamiquement grâce à des mécanismes d'intelligence collective, dans le but de produire une représentation émergente d'un contexte dynamique.

Le second objectif du travail a été de proposer une application innovante de notre modèle conceptuel au domaine des C-ITS. Nous avons choisi l'environnement du véhicule connecté où une UBR est un agent de contrôle du trafic qui implémente notre modèle.

2.5.1 État de l'art

Les approches classiques en IA reposent sur une abstraction de l'environnement proposée par le concepteur du système. Cette représentation dépend très fortement des spécificités du problème visé et de l'environnement associé (capteurs/effecteurs). Un système exploitant une telle représentation aura donc une capacité limitée d'adaptation à d'autres problèmes/environnements [65]. Comme l'autonomie d'un système dépend de sa capacité d'adaptation, les approches constructivistes proposent des mécanismes de construction itérative de la représentation de l'environnement par un agent, à partir de ses interactions. Ces approches s'inspirent des travaux précurseurs en sciences cognitives de Piaget [354] sur le développement de l'enfant, et ont convergé vers la définition de caractéristiques permettant de modéliser l'autonomie [486] : l'agent doit être incarné, situé, et exploiter un processus développemental épigénétique. La notion d'épigénétique a été proposée par Piaget, et décrit le processus de développement de l'individu à travers ses interactions avec

son environnement. L'approche la plus prometteuse en IA développementale consiste à reproduire le processus d'acquisition de connaissance de l'être humain. Seuls quelques travaux novateurs proposent de construire une représentation uniquement à partir des interactions sensorimotrices de l'agent [313]. Ces approches se heurtent au problème de l'amorçage de la représentation, qui survient dès que l'on cherche à démarrer la construction de la représentation de l'environnement sans utiliser de connaissances expertes *a priori*. Pour contourner ce problème, il est possible d'aider le système en lui donnant quelques mécanismes de bas niveau, permettant d'amorcer le processus d'apprentissage [361]. Dans [361], les auteurs utilisent le gaz neuronal croissant (GNG) - une adaptation des cartes auto-organisatrices - afin de générer une première discrétisation de l'espace des perceptions du système. Ceci permet de réduire le nombre de dimensions des données perçues (espace d'entrée) afin de simplifier l'identification des actions les plus pertinentes correspondant à cet espace. Cependant, pour fournir une telle discrétisation, il faut réaliser un partitionnement des données perçues en amont du processus d'apprentissage du système. Cet amorçage s'oppose en partie au cadre strict du constructivisme où cette discrétisation devrait être le résultat des interactions avec l'environnement. Le système doit donc être en mesure de faire évoluer cette représentation. Une stratégie possible consiste à proposer différentes discrétisations au système, qu'il devra combiner afin d'en déterminer les plus utiles à son usage (actions). Dans notre travail, nous proposons des mécanismes permettant au système d'utiliser plusieurs représentations concurrentes, qui sont mises à jour en parallèle. Ces abstractions (discrétisations) des données brutes perçues peuvent être générées en amont de la phase d'apprentissage du système, à partir de différents algorithmes de regroupement (*K-means*, GNG).

2.5.2 Construction d'une représentation par apprentissage concurrent

L'objectif à long terme est de concevoir un système de prise de décision dynamique capable de définir les frontières de chaque état perception-action, afin d'identifier précisément le périmètre de chaque action étant donné un ensemble de perceptions. En effet, un tel système serait capable de mieux comprendre les dynamiques de l'environnement et aurait une probabilité plus élevée de sélectionner l'action la plus pertinente dans un contexte donné. Plusieurs problèmes sont à traiter afin d'arriver à un tel résultat. D'abord, il est nécessaire de choisir quelles sont les connaissances minimales à fournir au système sans pour autant incorporer trop de connaissances expertes dans sa représentation. Ensuite, le périmètre exact des actions peut ne pas être atteignable selon le problème. La difficulté est de trouver la meilleure façon de modéliser cette notion de périmètre sans pour autant limiter l'ensemble des états possibles. Enfin, un défi se situe au niveau du renforcement de la représentation, qui doit mener à un contrôle efficace de l'environnement par le système. Nous proposons un modèle conceptuel générique dont les mécanismes traitent les problématiques précédentes, et qui peut être facilement adapté dans son implémentation pour traiter des problèmes spécifiques, en particulier le contrôle des C-ITS.

2.5.2.1 Aperçu

Le modèle proposé s'appuie sur des mécanismes d'intelligence collective afin d'assister le système dans sa tâche itérative de construction de sa représentation de l'environnement. Cette tâche prend la forme de deux processus couplés, perception et décision, comme illustré dans la figure 2.18. Les données d'entrée, collectées par les capteurs, représentent des observations de l'environnement et constituent la perception de bas-niveau du système. Ces données sont continues et non identifiées

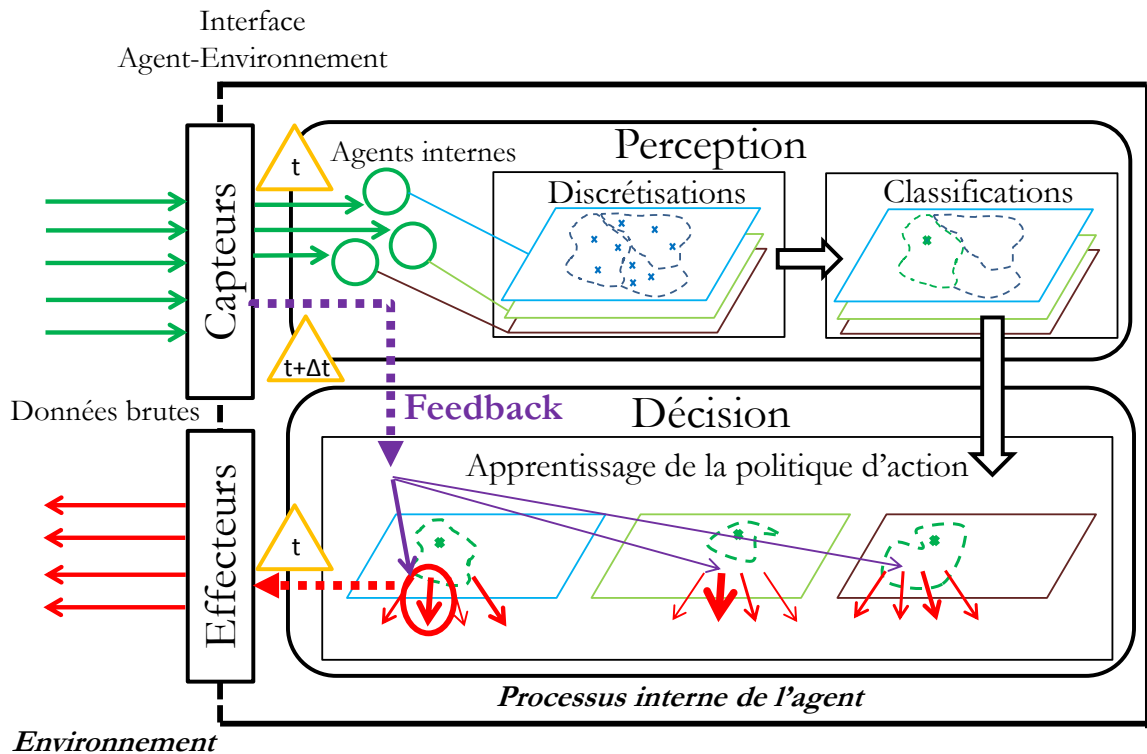


Fig. 2.18.: Aperçu du modèle.

dans la mesure où aucune information experte n'est communiquée au système au sujet de la pertinence ou de l'importance de chaque variable perçue. Cette perception est transmise aux agents "discrétiseurs", dont le rôle est de fournir des abstractions à partir de ces informations. Ces abstractions correspondent à des façons alternatives d'interpréter le signal d'entrée. Chaque agent applique une stratégie de discrétisation interne. Les discrétisations générées peuvent évoluer en temps réel pour adapter le périmètre des états (classes) que l'agent propose. Plusieurs mécanismes peuvent améliorer la précision des discrétisations proposées par les agents. Le processus multi-agent produit un ensemble de partitions concurrentes pour le système correspondant à différentes représentations possibles. La sélection de l'action repose sur un apprentissage par renforcement qui fait intervenir toutes les représentations concurrentes. Le modèle, présenté dans l'algorithme 1, vise à combiner dynamiquement les états perception-action précédemment appris afin de construire une représentation encore plus précise de l'environnement.

2.5.2.2 Perception de haut-niveau

Le système utilise des représentations individuelles proposées par ses agents discrétiseurs pour construire une représentation de plus haut-niveau. Chaque discrétiseur exploite son propre processus de classification exécuté dans son propre espace de perception, qui peut varier d'un agent à l'autre. Ceci offre la possibilité d'évaluer des représentations dédiées à des canaux spécifiques de la perception. Ainsi, les discrétiseurs peuvent percevoir des données à partir des mêmes capteurs ou utiliser des dimensions différentes à l'échelle de la perception du système. Nous proposons de décrire le modèle de façon générique, ce qui permet de repousser le choix de la méthode de classification à l'étape d'implémentation du modèle. Les algorithmes en ligne ou hors ligne peuvent être utilisés à la condition qu'ils soient compatibles avec le type des données d'entrées

Algorithme 1 Processus général du modèle

```

Instancier l'ensemble  $A$  des actions possibles
Instancier l'ensemble  $D$  des agents discrétiseurs
Instancier l'ensemble  $S$  des états sélectionnés concurrents (vide)
Boucler # boucle qui représente le processus de décision du système
  Si  $S$  n'est pas vide Alors # condition pour la première itération
     $F = \text{feedback}(P)$ 
    PourTout  $s_i \in S$  Faire # récupérer le lien  $L$  entre l'état  $s_i$  et l'action  $a^*$ 
       $L = \text{getPerceptionActionLink}(s_i, a^*)$ 
       $\text{incrementReward}(L, F)$ 
    FinPour
     $\text{reinforce}(D, a^*, F)$ 
  FinSi
   $P = \text{getPerception}()$  # début du processus de sélection de l'action
   $S.\text{removeAll}()$ 
  PourTout  $d_i \in D$  Faire
     $p_i = \text{discretizerPerception}(P)$  # les agents peuvent percevoir différentes variables
     $d_i.\text{discretize}(p_i)$ 
     $s_i = D_i.\text{classify}(p_i)$ 
     $S.\text{add}(s_i)$  # ajout de l'hypothèse d'état  $s_i$  à l'ensemble  $S$ 
  FinPour
   $a^* = \text{actionSelection}(S)$ 
   $\text{execute}(a^*)$ 
FinBoucle # fin du processus de décision avant la prochaine perception (et feedback)

```

et qu'ils permettent de générer une représentation sous la forme d'un ensemble d'états discrets. L'ensemble des variables perçues par chacun des agents peut être défini de manière experte ou initialisé aléatoirement, en fonction du problème visé. Le processus de perception du système peut se résumer en deux étapes consécutives (étapes 1 et 2), telles que représentées dans la figure 2.19. Chaque fois que le système perçoit son environnement, tous les agents discrétiseurs

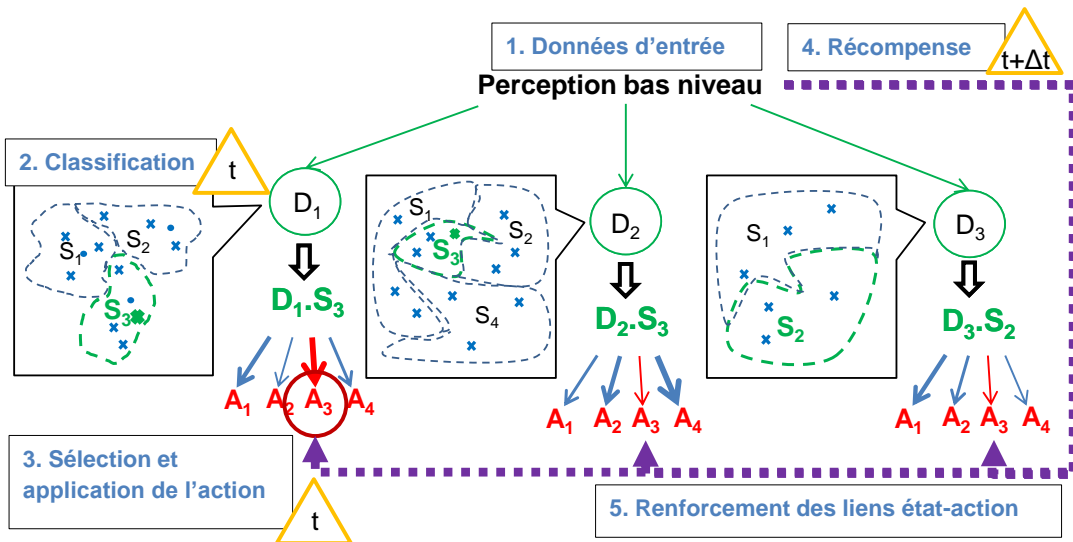


Fig. 2.19.: Processus de discrétisation, de classification et de sélection d'action.

D mettent à jour en parallèle leur propre espace d'hypothèses d'état, en fonction de leur méthode de classification choisie (K-means, GNG, etc.). Par exemple, la discrétisation de l'agent D_i est composée de l'ensemble des hypothèses d'état : $\{s_1, s_2, \dots, s_j\}$. Pour chaque pas de temps de

perception du système, chaque agent D perçoit ses variables d'entrée et classe sa perception comme une hypothèse d'état discret. A partir de l'ensemble des hypothèses d'état (une par agent), le système tente de choisir l'action la plus pertinente dans le contexte défini par les différents agents.

2.5.2.3 Contrôle de l'environnement

Les discrétiseurs mémorisent la probabilité de sélection de chaque action lorsqu'ils classent leur perception en tant qu'une hypothèse d'état donnée. Ces probabilités sont modélisées sous la forme de liens état-action qui permettent de garder une trace au niveau du système du résultat des actions consécutives, évalué à travers le retour sur expérience (*feedback*) fourni par l'environnement. A l'issue de l'étape de perception, tous les discrétiseurs ont sélectionné une de leurs hypothèses d'état en réponse à la perception de bas-niveau propagée. Le système doit désormais choisir une action qui soit la plus pertinente possible en fonction des différentes hypothèses proposées (une par agent). Il s'agit alors de trouver un compromis entre l'exploitation (appliquer l'action générant le meilleur *feedback*) et l'exploration (tester des actions non utilisées jusque là). Ce choix est illustré par les étapes 3 à 5 dans la figure 2.19. Le système fait face à un dilemme classique d'exploration-exploitation [18], puisque l'espace des actions est supposé fini et qu'il n'y a pas de corrélation évidente entre des hypothèses d'état distinctes. Cette situation fait écho au problème dit du bandit à n bras, relatif à l'élaboration d'une stratégie optimale de sélection et de test de différentes machines à sous dans un casino. Plusieurs stratégies, disponibles dans la littérature, peuvent être utilisées dans notre modèle. La plupart utilisent un paramètre objectif à minimiser fondé sur la notion de regret, exprimé comme la différence entre la récompense courante et la récompense optimale (théoriquement inconnue). Notre modèle ne nécessite pas que la stratégie soit optimale dans la mesure où la sous-exploration de l'espace des actions risque de nuire à la performance globale du système. Néanmoins, cette hypothèse dépend encore une fois du type de problème visé et pourra être étudiée en comparant différentes implémentations du modèle.

Le retour sur expérience fourni au système fait partie de sa perception de bas-niveau et doit être défini lors de la formalisation du problème (*i.e.* pour l'implémentation). Cette fonction de *feedback* permet une évaluation par le système du résultat de ses actions dans l'environnement. Cette valeur est utilisée comme une récompense pour l'algorithme d'exploration-exploitation et contribue au renforcement de la représentation construite. Le résultat de la sélection est un lien état-action dont chaque état est directement lié à un agent discrétiseur. Ainsi, le processus de renforcement contribue à l'apprentissage du système, en l'aidant à identifier les hypothèses d'état les plus précises parmi l'ensemble des discrétisations proposées par la population d'agents discrétiseurs.

2.5.3 Application au trafic coopératif

Le modèle générique a été appliqué à la régulation du trafic coopératif, dans le cadre de simulation multi-agent introduit dans la section 2.3. Nous donnons ici les détails de l'implémentation de notre modèle de contrôle : description des caractéristiques des agents discrétiseurs et actions/*feedback* utilisés. Ensuite, les résultats en simulation mettent en avant le comportement du modèle et confirment les bénéfices de notre approche qui, grâce à des représentations concurrentes, permet de réguler un flux partiellement composé de véhicules connectés.

2.5.3.1 Cas d'étude

Comme illustré par la figure 2.20, toutes les expérimentations sont menées dans un même scénario représentant une section autoroutière rectiligne à trois voies avec la présence d'un convergent (voie d'insertion) vers sa fin. L'UBR (représentée par le disque blanc) perçoit les débits et les vitesses grâce aux capteurs sur la route (rectangles gris) ainsi que les vitesses des véhicules connectés dans son rayon de communication (150 mètres).

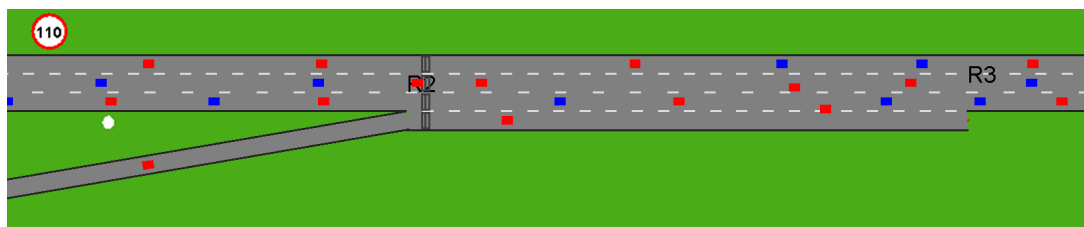


Fig. 2.20.: Capture d'écran du scénario dans le simulateur.

Afin de générer différents états distincts pour l'UBR, nous avons déterminé les flux d'entrée (section principale et voie d'insertion) à partir de données réelles issues de boucles électromagnétiques. Le jeu de données est composé d'un mois (juin 2001) de données boucles, récoltées entre 8h et 9h sur l'autoroute A6. Nous n'avons conservé que les données de deux boucles, choisies pour la correspondance entre le réseau réel et celui de notre scénario (figure 2.20). Cela a permis de produire 30 simulations de 60 minutes chacune. La diversité des situations de trafic simulées permet d'étudier le comportement du modèle dans des contextes sensiblement variables (trafic fluide, régime critique ou congestionné). L'unité de bord de route implémente le modèle afin de choisir les consignes les plus pertinentes à envoyer aux véhicules connectés pour éviter, ou au moins réduire, la congestion sur sa section. Les consignes sont propagées sous la forme de messages I2V aux véhicules connectés. La voie d'insertion, à cause de la différence de vitesse avec la section principale, déclenche, suivant les conditions de trafic dense, l'apparition de congestion qui se propage depuis la voie de droite jusqu'à l'ensemble des voies. Nous étudions l'effet de la stratégie de régulation d'une UBR sur le flux en faisant l'hypothèse d'un pourcentage fixé à 30% de véhicules connectés, mélangés de façon homogène parmi les véhicules non connectés sur la section principale.

2.5.3.2 Définition des comportements

Unité de Bord de Route : le résultat de la stratégie de contrôle de l'UBR est le choix, à chaque pas de temps de décision (fixé à 120 secondes), d'une action parmi les suivantes :

- A_1 : pas d'envoi de consigne.
- A_2 – A_3 : consigne de changement de voie (droite à gauche – gauche à droite).
- A_4 – A_5 : consigne de modification d'interdistance (1,8 s–1,2 s).
- A_6 – A_7 : consigne de limite de vitesse (110 km/h–50 km/h).

Toutes les actions mènent à la propagation d'un message de consigne (sauf A_1) contenant le type de consigne et le paramètre associé aux véhicules connectés. Ces messages définissent aussi une zone de pertinence (qui couvre l'ensemble de la section jusqu'à la voie d'insertion) ainsi qu'une date d'expiration (fixée elle-aussi à 120 secondes). La consigne contenue dans un message est appliquée par le véhicule qui l'a reçue tant qu'il se trouve dans la zone de pertinence du message

et que celui n'a pas expiré.

Véhicules : la dynamique des véhicules est régie par les modèles de la couche physique présentés en 2.2. Seuls les véhicules connectés sont équipés de dispositifs de communication sans fil leur permettant de percevoir les messages de consigne dans un rayon de 150 mètres, ce qui correspond aux technologies actuelles. Nous ne considérons pas les messages partiellement reçus ou les pertes d'information dans la mesure où le protocole réel est censé les traiter. Les messages de consigne reçus par les véhicules connectés sont retransmis aux autres véhicules (à portée) jusqu'au début de la section (*i.e.* tant qu'ils se trouvent dans la zone de pertinence du message).

Les véhicules connectés affichent dans leur interface embarquée la consigne contenue dans les messages, afin que le conducteur puisse l'intégrer à son comportement. Nous proposons donc une interprétation par le véhicule de chacune des consignes. Pour les actions A_2 et A_3 , la consigne de changement de voie modifie la stratégie opportuniste tant que le message reste pertinent (zone et durée). Le véhicule va donc essayer de rejoindre la voie la plus à gauche (respectivement à droite) en entreprenant un ou des changement(s) de voie jusqu'à atteindre cette voie cible. Le modèle de changement de voie intègre une estimation du créneau spatiotemporel disponible. Cette estimation permet au véhicule de décider si un changement de voie est suffisamment sûr (au sens de ces paramètres) pour être effectivement exécuté. Cela implique qu'une plus grande proportion de véhicules connectés changera de voie lorsque le trafic sera fluide (et inversement lorsque le trafic sera dense). Pour les actions A_4 à A_7 , la seule modification consiste à changer temporairement les paramètres correspondants dans le modèle longitudinal (respectivement, l'interdistance désirée et la vitesse désirée) pour les valeurs indiquées dans le message de consigne. Toutes ces modifications sont maintenues tant que le message reste pertinent et sont automatiquement annulées dès que le message expire ou que le véhicule quitte la zone de pertinence définie.

2.5.3.3 Implémentation du modèle

Dans l'application C-ITS, le modèle est implémenté en tant que processus de décision d'une UBR. Cette UBR est capable de percevoir par l'intermédiaire d'un capteur (double boucle électromagnétique) le débit moyen, la concentration et les vitesses sur la voie d'insertion et sur les trois voies de la section principale. En plus, les véhicules connectés situés dans son rayon de communication partagent leurs vitesses respectives. L'ensemble de ces données constitue la perception bas-niveau du système. La perception et la décision du système sont ici discrétisées et exécutées toutes les 120 secondes. L'UBR perçoit des données à partir de ses capteurs, agrégées sur la période précédente de 120 secondes, puis choisit et applique immédiatement une action (*i.e.* envoie une consigne). Après une période de 120 secondes supplémentaires, la consigne expire et l'UBR reçoit, au même moment, un *feedback* utilisé pour renforcer la représentation et une nouvelle perception de l'environnement.

Agents discrétiseurs : afin de montrer les bénéfices du modèle, nous proposons une implémentation avec deux agents discrétiseurs. Chaque agent a pour rôle de générer une représentation (discrétisation) à partir des variables qu'il perçoit. Nous avons décidé de lier chaque discrétiseur, nommés D_1 et D_2 , à un capteur spécifique. Les agents diffèrent donc principalement par leur perception. Ainsi, D_1 perçoit les données des boucles électromagnétiques (débit et densité moyens)

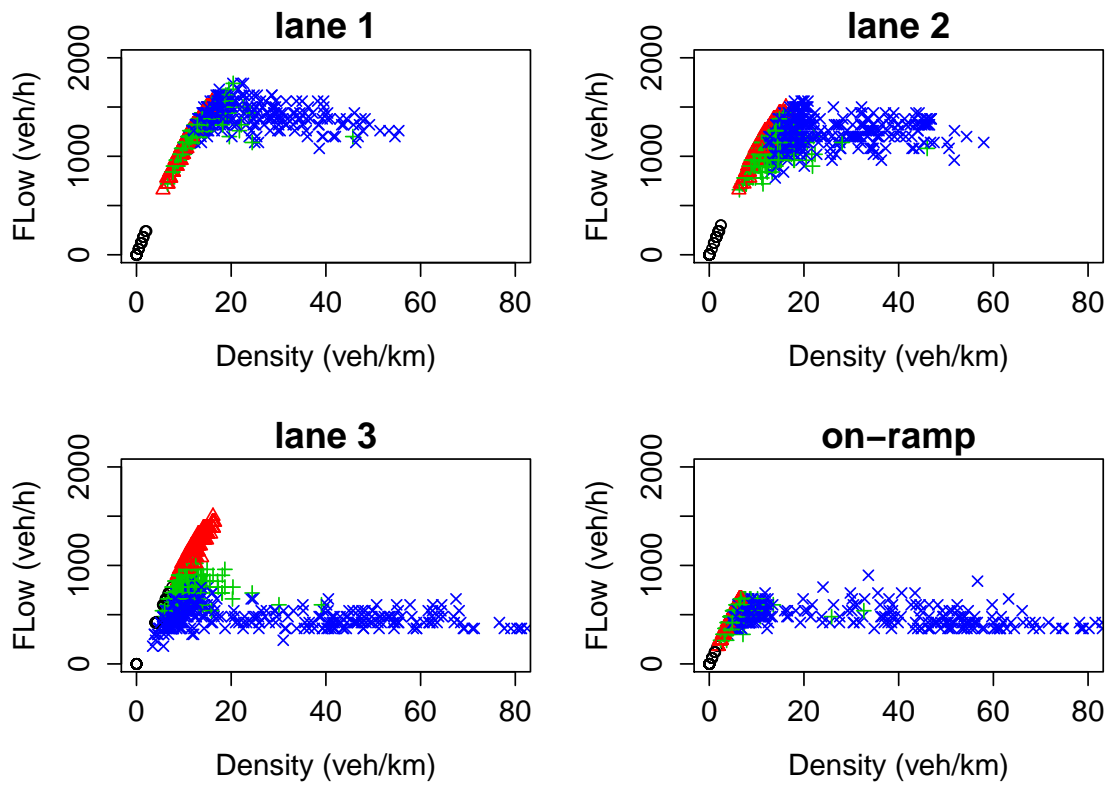
pour chacune des trois voies et la voie d'insertion, agrégées toutes les 120 secondes. D_2 exploite les données provenant du capteur de communication de l'UBR et construit un modèle (moyenne et écart-type) de la distribution estimée des vitesses à partir des messages des véhicules connectés qui sont passés dans son rayon de communication durant son dernier pas de temps de décision. Les discrétisations des agents sont générées à partir d'une classification par l'algorithme des *K-means* (stabilisé avec 10000 itérations à partir d'initialisations aléatoires) pour respectivement 4 et 3 classes (D_1 et D_2). Les résultats de la discrétisation par les deux agents sur les données issues des 15 premières simulations sont représentés sur la figure 2.21. L'apprentissage des états perception-action de chaque agent est réalisé sur les 15 mêmes simulations, en exploitant le *feedback* du système.

Renforcement de la représentation : les récompenses recueillies par les agents sous la forme de liens état-action sont calculées à partir d'une variable de l'environnement choisie. Dans le scénario proposé, nous avons retenu la vitesse moyenne la plus faible (sur les quatre boucles électromagnétiques) en tant que *feedback* du système, puisque maximiser cette valeur correspond à l'objectif global d'amélioration de l'écoulement du flux. Ce retour sur expérience externe, est utilisé sous la forme de récompenses dans l'algorithme d'exploration-exploitation pour renforcer à la fois la stratégie de contrôle et les représentations proposées. Parmi les algorithmes de bandit à n bras de la littérature, notre choix s'est arrêté sur *Upper Confidence Bound* [18] (UCB). Son initialisation consiste à essayer chaque machine une seule fois. Ensuite, l'algorithme sélectionne systématiquement la machine j qui maximise $\hat{x}_j + \sqrt{\frac{2 \ln n}{n_j}}$ où \hat{x}_j correspond à la récompense moyenne obtenue pour la machine j , n_j est le nombre de fois que la machine j a été jouée jusqu'ici et n est le nombre total d'essais. La mise à jour des récompenses concerne tous les liens perception-action (notés L dans l'algorithme 1), et non pas seulement celui qui a été effectivement sélectionné par l'algorithme. De cette façon, le bénéfice d'une action est également propagé aux liens qui ont contribué à sa sélection. Cette récompense permet d'accélérer la phase d'exploration, en s'appliquant à autant de liens que d'agents discrétiseurs pour chaque itération.

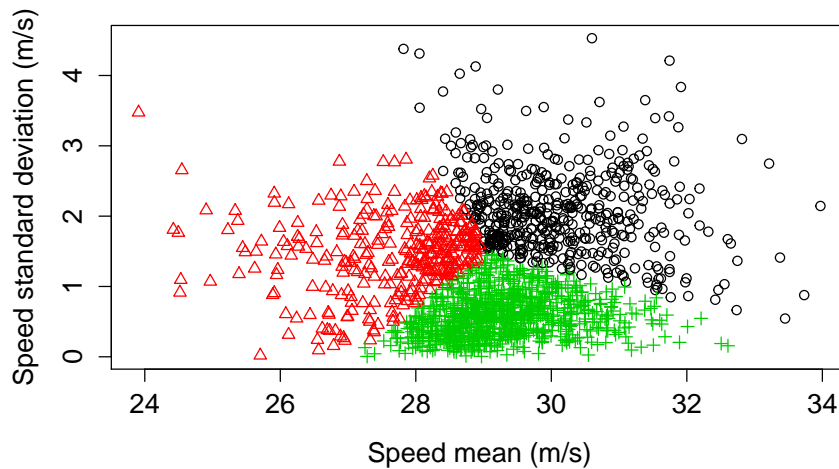
2.5.3.4 Expérimentations

Les expérimentations sont menées en deux étapes principales. Dans un premier temps, un ensemble de données d'entrée (la perception des agents D) est nécessaire afin d'appliquer les algorithmes de classification et obtenir des discrétisations stables (sans appliquer les actions). Puis, un jeu de simulations est utilisé pour l'apprentissage des liens perception-action en utilisant le processus complet d'exploration et d'application des actions (sur les 15 simulations). La seconde étape confronte le système à des scénarios alternatifs, pour vérifier que la représentation construite par le système l'aide à appliquer une stratégie de régulation efficace sur le flux. Cette étape est menée sur la deuxième partie des simulations. La situation témoin de chaque simulation est l'observation d'un ensemble d'indicateurs dans le même scénario mais sans utiliser l'infrastructure (aucune consigne n'est envoyée). Nous comparons trois implémentations du modèle : le cas avec un seul agent (D_1 et D_2 séparément), et une combinaison dynamique des représentations de ces deux agents grâce au modèle.

Indicateurs : les indicateurs sélectionnés sont le temps total passé (TTS), la vitesse moyenne sur la section et le pourcentage de congestion. Le TTS est simplement la somme des temps de trajet de tous les véhicules sur la section ; des valeurs plus faibles indiquent un meilleur écoulement. La vitesse moyenne permet d'observer, au cours du temps, l'homogénéité du flux. Le pourcentage de conges-



(a)



(b)

Fig. 2.21.: Discrétisations générées par (a) D_1 [8 dimensions] et (b) D_2 [2 dimensions] sur le jeu de données d'apprentissage.

tion est la proportion spatiale de la section où les véhicules circulent à moins de 30 km/h. On peut s'attendre à ce que l'implémentation du modèle profite des représentations individuelles proposées par les agents et mène à l'amélioration de l'écoulement du flux, observé au regard de ces indicateurs.

Résultats : parmi les quinze simulations disponibles, nous en avons sélectionné trois qui présentent des situations de trafic différentes. La figure 2.22 présente le tracé des trois indicateurs sur ces

simulations. Les indicateurs permettent d'évaluer les effets des différentes implémentations du modèle, en les comparant avec la simulation témoin (courbes rouges). Dans les simulations (a) et (c), sans régulation, les instabilités générées par les changements de voies depuis la voie d'insertion se propagent à l'ensemble du flux. Les trois simulations montrent des comportements différents du modèle. Dans la simulation (a), la stratégie de contrôle de l'agent D_1 permet d'éviter l'apparition de congestion alors que celle de l'agent D_2 donne de moins bons résultats que dans la simulation témoin. Ces résultats sont bien visibles en observant le tracé des TTS. Des valeurs faibles de temps total passé indiquent que les véhicules parcourent la section plus rapidement en moyenne. La combinaison dynamique de D_1 et D_2 produit une stratégie mixte qui donne une amélioration mitigée. La simulation (b) illustre un cas spécifique où l'un des deux discrétiseurs propose une mauvaise représentation (D_2), puisque l'on n'observe pas de congestion dans la simulation témoin. Néanmoins, le modèle est capable de converger vers les meilleures représentations individuelles (D_1) dans ce contexte, ce qui permet de ne pas empirer la situation observée dans la simulation témoin. Dans la figure 2.22 (c), l'état de trafic modélisé se rapproche du régime critique du flux, puisque l'ampleur de la congestion est limitée et la vitesse moyenne ne diminue que peu (comparée à la simulation (a)). Dans ce contexte, les deux stratégies apprises par chaque discrétiseur ont un impact positif sur le flux, et l'amélioration de la représentation par le modèle est visible. Le système est capable, grâce aux récompenses exploitant le *feedback*, de construire une représentation de plus haut niveau à partir de celles des agents, qui permet d'éviter la propagation de congestion sur la section. La figure 2.23 représente de façon simplifiée la contribution de chaque discrétiseur (en reprenant le code couleur de la figure 2.22) à la stratégie de régulation globale du système. Le schéma montre que des états perception-action des deux agents sont utilisés par le modèle. L'action A_2 , la plus utilisée, semble logique puisqu'elle permet de laisser plus de place aux véhicules entrants pour s'insérer dans le flux principal. Ainsi, les résultats confirment la capacité du modèle à utiliser différentes représentations concurrentes et à les combiner en une représentation plus précise.

2.5.3.5 Discussion

Comme dans la plupart des problèmes d'apprentissage par renforcement, le choix du *feedback* est critique et complexe. De plus, comme le système exploite différentes discrétisations concurrentes, le choix des données d'entrée (perception) est important. Cela est dû à une problématique de fusion de capteurs : certains phénomènes sont perceptibles ou améliorables en utilisant seulement certaines dimensions. Cela signifie que le système doit percevoir des variables pertinentes pour son usage pour être capable d'adapter sa stratégie de contrôle aux variations de l'environnement. Dans les expérimentations, nous avons vu que le concepteur doit fournir un premier ensemble de discrétiseurs capables d'identifier des états pertinents, avant même de les faire combiner par le modèle. Ce problème pourrait être traité par l'ajout de mécanismes permettant au système de créer de manière autonome de nouveaux agents, voire de les faire évoluer dynamiquement, en exploitant son expérience d'interaction.

Comme les données d'entrée peuvent provenir de différentes dimensions (ou variables), une possibilité est de donner une perception complète au système. Dans ce cas, lors de l'implémentation du modèle, le temps d'exploration de l'ensemble des combinaisons de toutes les variables pour chaque discrétisation risque d'être très important. C'est pourquoi nous proposons d'utiliser différents agents qui ne travaillent que sur un sous-ensemble de la perception de bas-niveau. Les résultats obtenus en simulation laissent entrevoir le potentiel de convergence du système (en termes de profils

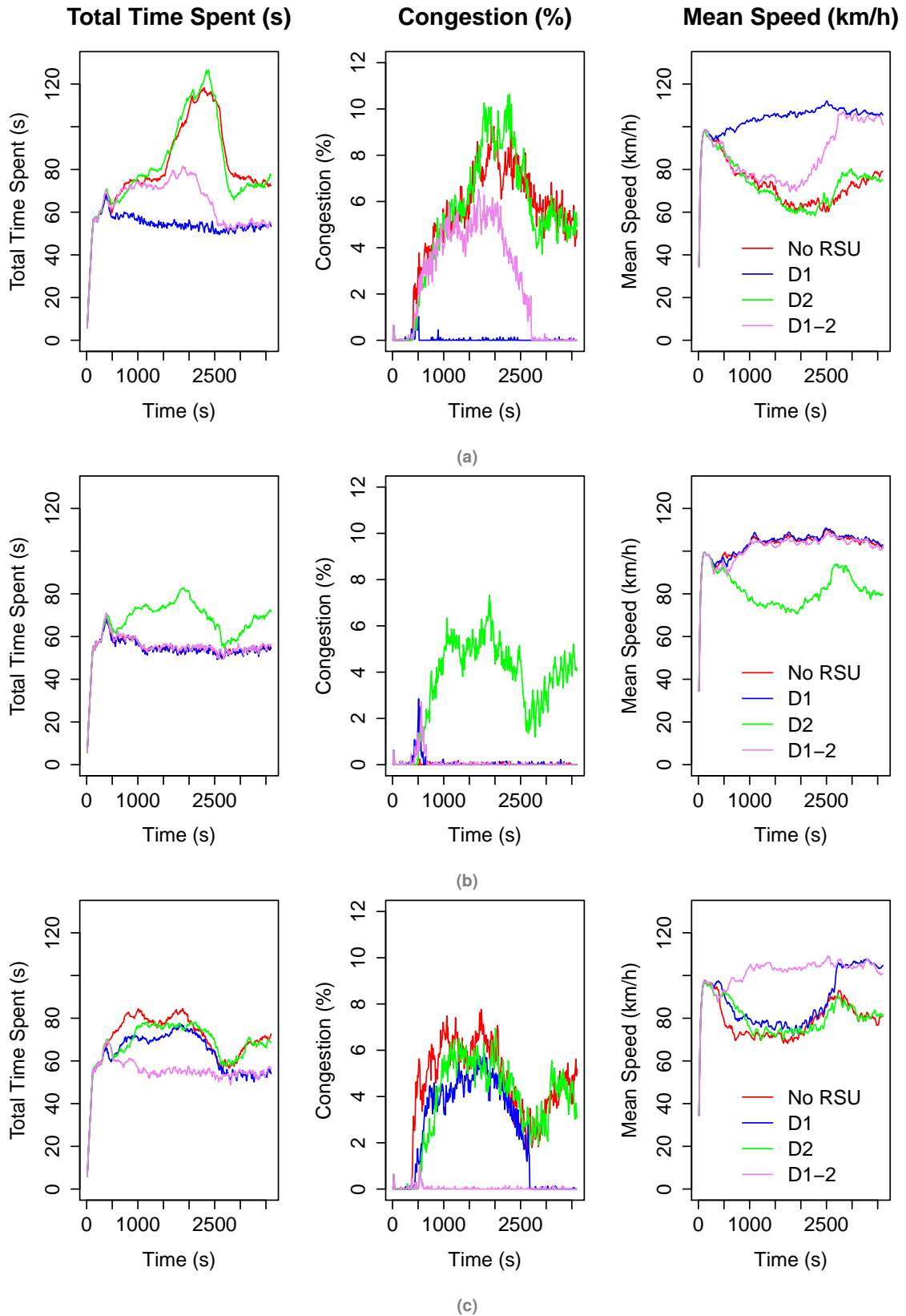


Fig. 2.22.: Résultats des 3 simulations : (a) Lundi 18, (b) Jeudi 21 et (c) Mardi 28.

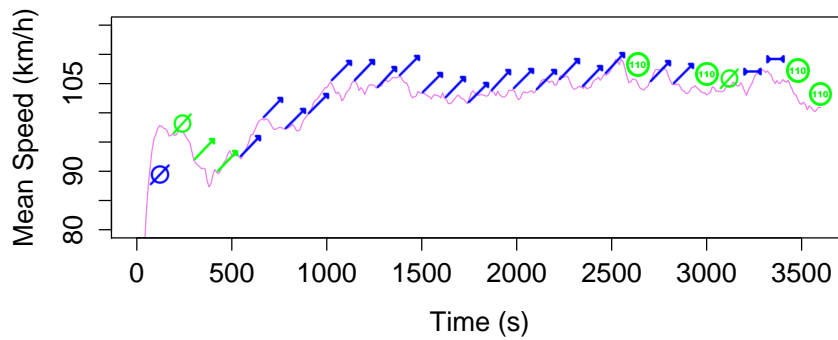


Fig. 2.23.: Répartition entre les agents des consignes choisies par le modèle dans la simulation (c).

de liens perception-action) vers les combinaisons de variables les plus pertinentes pour l'usage du système (*i.e.* au regard de la stratégie de contrôle produite). Ceci pourrait être particulièrement utile pour des problèmes où la combinaison des variables d'entrée n'est pas connue *a priori*. Parmi les suites évidentes du travail, une comparaison avec des méthodes concurrentes permettrait de valider et généraliser les résultats préliminaires introduits dans ce chapitre.

2.6 Conclusion et perspectives

Ce chapitre a rassemblé une sélection de contributions sur le thème des C-ITS. Après des résultats théoriques sur la stabilité apportée par les véhicules connectés, nous avons proposé un modèle multi-agent par couplage de dynamiques qui propose de traduire de façon flexible les interactions complexes entre véhicules connectés et avec l'infrastructure. Le contrôle de la qualité de l'information par la notion de confiance permet de capturer la fiabilité de l'information. Notre approche a été fidèlement implémentée dans un simulateur de trafic multi-agent, qui se pose comme un outil de virtualisation du trafic coopératif. Les premiers résultats confirment les bénéfices attendus des C-ITS en terme de mobilité et de sécurité. Le simulateur, quant à lui, se pose comme un outil d'aide à la décision pour le design des stratégies de contrôle associées. Dans une vision plus futuriste de contrôle de systèmes complexes par intelligence artificielle, nous avons ensuite proposé un modèle générique permettant à un système de construire une représentation de son environnement à partir de ses interactions. Par un apprentissage constructiviste visant à construire itérativement le processus de prise de décision d'un système de contrôle, le modèle utilise une population interne d'agents discrétiseurs pour amorcer la construction de la représentation. La représentation qui en est le fruit se compose d'états perception-action qui évoluent dans un apprentissage par renforcement exploitant un retour de l'environnement afin d'explorer l'espace de recherche. Après avoir présenté le modèle d'un point de vue théorique, nous avons proposé une application dans le cas d'étude de la régulation de trafic coopératif. Les résultats obtenus en simulation montrent que la combinaison dynamique des discrétisations individuelles permet au système d'adopter une stratégie de régulation plus efficace. La prochaine étape de notre travail se concentrera sur les évolutions du modèle. La construction du modèle conceptuel est aussi itérative, afin de pouvoir tester et valider chacun des mécanismes par l'implémentation à chaque étape. Nous cherchons désormais à concevoir un nouveau type d'agent capable d'exploiter les hypothèses d'état proposées par les discrétiseurs en proposant des associations entre ces états. Dans un premier temps, deux types d'associations

seront considérées : la spécialisation ou l'agrégation. La question du compromis entre le temps nécessaire à l'exploration et le nombre d'associations reste à étudier.

Approches statistiques *data driven* pour la caractérisation et la prévision du trafic routier

” La photographie, c'est la vérité. Et le cinéma, c'est 24 fois la vérité par seconde.

— Jean-Luc Godard
Cinéaste

Alors que notre premier chapitre est logiquement parti des modèles théoriques, dans une approche *model driven* justifiée par l'absence de données sur les systèmes coopératifs, une grande part de nos recherches s'est consacrée à l'analyse de nouvelles sources de données disponibles comme le Bluetooth, dans une approche *data driven*. Le monde des C-ITS, en pleine transition vers le *big data*, doit s'appuyer sur des processus de fouille de données innovants capables de générer des indicateurs de caractérisation temps réel de trafic et de prévision à court terme. C'est donc par la présentation des différentes données utilisées que nous commencerons ce chapitre (section 3.1), avant de détailler trois types de contribution relatives à la modélisation statistique des temps de parcours (section 3.2), l'analyse spatiotemporelle d'un réseau de capteurs Bluetooth (section 3.3) et la prévision du trafic par apprentissage statistique (section 3.4). Les travaux présentés sont les fruits de l'encadrement de travaux de Master (Justine Dinh, Pierre-Antoine Laharotte, Fabien Baradel), de deux thèses (Pierre-Antoine Laharotte et Raphaël Delhome), en collaboration avec les milieux académiques (Université de Technologie du Queensland, Austrian Institute of Technology), industriels (Mediamobile) et ministériels (CEREMA) pour l'obtention des données.

3.1 Présentation et qualification des données

3.1.1 Le Bluetooth

3.1.1.1 Principe général

Un scanner Bluetooth est un instrument capable de percevoir les ondes émises par tout émetteur Bluetooth, identifié par une adresse physique (*Media Access Control (MAC)*) unique. Ces capteurs sont caractérisés par leur position physique (longitude et latitude) et leur aire de détection, généralement de l'ordre de 100m. Les appareils émetteurs, comme les téléphones mobiles, radios, GPS embarqués, ont leur propre rayon d'émission, généralement inférieur (de l'ordre de 40m). Tout l'enjeu repose sur le fait de capter l'appareil émetteur lorsqu'il traverse la zone de détection, avec un réglage adéquat d'une fréquence de balayage du scanner. En principe, le placement des détecteurs aux intersections, où la vitesse est localement ralentie, permet de réduire les pertes de

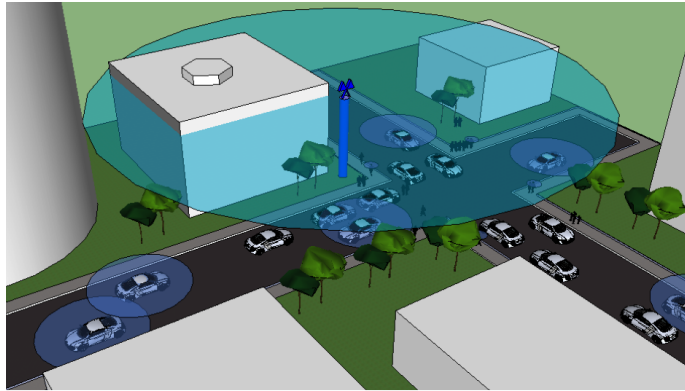


Fig. 3.1.: Principe de détection de la technologie Bluetooth.

données [50]. À chaque traversée de la zone de détection associée à un capteur Bluetooth, les données suivantes sont récoltées : l'identifiant MAC unique et propre à l'appareil détecté, l'identifiant propre au détecteur permettant la géolocalisation de l'individu, la date d'entrée dans la zone de détection et la durée passée au sein de cette zone de détection ou *duration*. À l'issue de premiers tests sur les capteurs Bluetooth, nous avons conclu que la variable *duration* ne pouvait être significative du fait de sa forte dépendance à la fréquence de balayage des détecteurs Bluetooth [431, 322]. Les horodatages d'entrée dans les zones de détection successives constituent ainsi la matière première pour le calcul de temps de parcours. De telles données, disponibles à l'échelle d'un réseau, fournissent suffisamment d'information pour adopter une approche individualisée en suivant les itinéraires de tout usager au cours d'un voyage, mais également globale en construisant des matrices Origine–Destination à condition de considérer chaque détecteur comme une Origine et une Destination potentielle. **Dans la suite de ce chapitre, seul le trafic routier de voitures particulières sera analysé.** Deux jeux de données ont été analysés. Le premier a permis de traiter la donnée brute de A à Z, à l'échelle d'un réseau, pour caractériser les états de trafic et évaluer le Bluetooth dans une optique de supervision et prévision du trafic. Le second (ville de Portland) se limite à une seule section de données temps de parcours, le but poursuivi étant la modélisation statistique de distributions à l'échelle d'un tronçon.

Premier cas d'étude : Brisbane (Australie)

Le premier cas d'étude concerne les données Bluetooth de la ville de Brisbane, obtenues dans le cadre de notre collaboration avec l'université de technologie du Queensland. Les données recueillies sont des données brutes issues de 79 détecteurs de la ville de Brisbane, couvrant de façon homogène le centre-ville et sa proche périphérie (figure 3.2 (a)). Comme évoqué ci-dessus et en section 1.5.3.2, le Bluetooth nécessite de nombreux pré-traitements avant d'être exploitable pour des analyses plus poussées. Le premier d'entre eux est un filtrage temporel qui détecte les *outliers*, vérifie la qualité des données brutes et discrimine les différents modes de transports. Les temps de parcours, calculés à partir des différentiels de temps de passage d'un même appareil sur deux détecteurs consécutifs, nécessitent un filtrage, comme l'ont pratiqué des études comparables [457, 363, 1, 285, 286]. Tout d'abord, comme pour les véhicules traçeurs, il est nécessaire de fixer un seuil maximum afin de filtrer les véhicules faisant de longues pauses (ex : *shopping*), qui pourraient biaiser les résultats [31]. D'autre part, un problème critique concerne la discrimination entre les différents modes de transports. Dans la lignée de Araghi *et al.* [12], un filtrage temporel a été effectué sur la base d'une analyse de répartition des vitesses (figure 3.2 (b)).

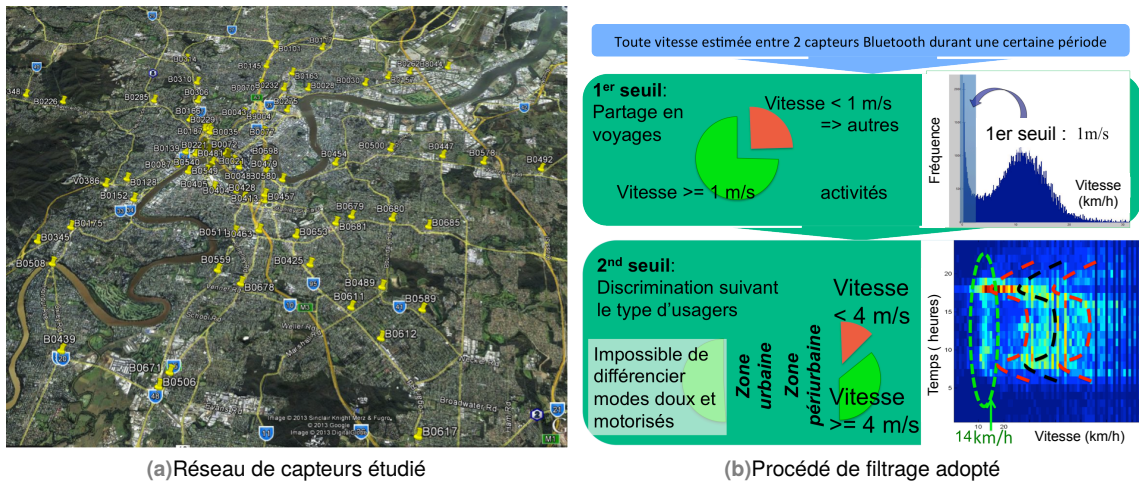


Fig. 3.2.: Cas d'études australien.

En définissant un lien comme la section comprise entre deux capteurs Bluetooth, un seuil minimum de vitesse de 1m/s est fixé. Pour la discrimination des différents modes de transports, la tâche ne s'avère pas si simple et après de nombreux essais, parfois raffinés (variations autour du *clustering*), une stratégie triviale a été retenue, qui consiste à appliquer un seuil de 4m/s (15 km/h) pour distinguer les véhicules motorisés en zones péri-urbaines uniquement. En milieu urbain, la discrimination n'est malheureusement pas possible et un simple seuil minimum de 1m/s a été conservé.

Cas d'étude secondaire : Portland (USA)

Un deuxième jeu de données Bluetooth est utilisé dans les travaux du manuscrit. Il est proposé en *open access* par l'université de Portland¹ et maintenu par l'administration américaine des autoroutes (FHWA). La zone sélectionnée consiste en une simple section urbaine, la 82ème avenue, en particulier un corridor équipé de 2.1km de long, dans la direction Nord-Sud (figure). La vitesse limite sur la section est de 56 km/h et les temps de parcours analysés concernent la période du 15 septembre 2011 au 15 novembre 2011.

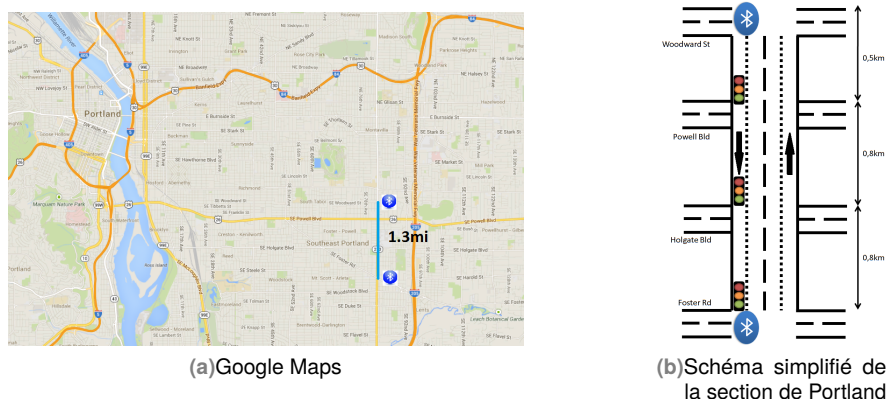


Fig. 3.3.: Description du site test de Portland et localisation des capteurs Bluetooth.

1. <http://portal.its.pdx.edu/Portal/index.php/fhwa>

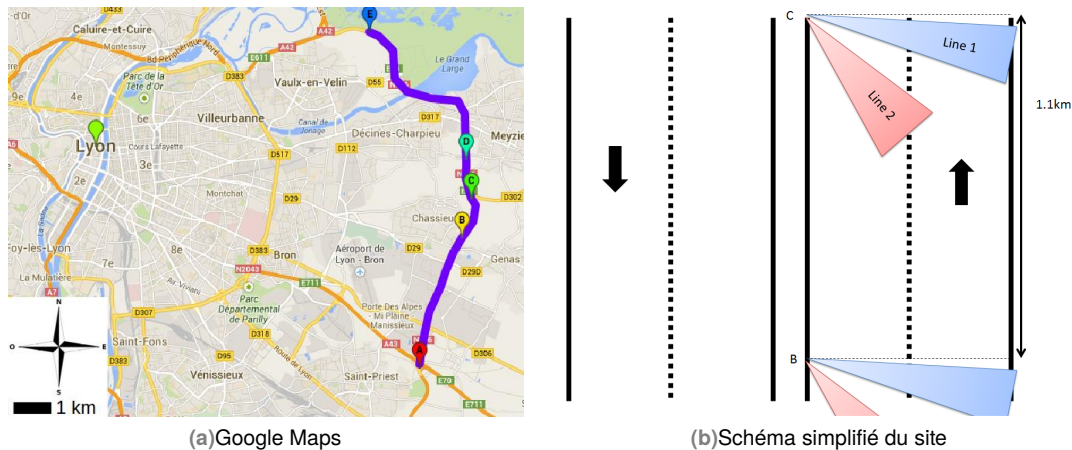


Fig. 3.4.: Cas d'étude de la rocade Est de Lyon.

Face aux difficultés inhérentes au Bluetooth, deux procédures ont été mises en place pour diminuer le biais venant des modes de transports alternatifs comme la marche à pied, les bus, les vélos. Les temps de parcours sont filtrés à vingt minutes et les données aberrantes supprimées. Ce processus est similaire au cas de Brisbane. Après cette phase, 400 observations valides par jour sont retenues pour la direction choisie, sur un total de 29 jours de mesure entre le 18 octobre et le 15 novembre 2011. Les données sont ensuite agrégées sur des périodes temporelles d'une heure afin de comparer des distributions entre elles. En conservant un minimum de 30 observations par tranche, 140 périodes ont été finalement retenues pour l'analyse.

3.1.2 Lecture Automatique de Plaques d'Immatriculation (LAPI)

Dans l'optique d'une validation expérimentale d'un travail de modélisation statistique des temps de parcours, une deuxième source de données a été envisagée à la suite du Bluetooth. Les données de Lecture Automatique de Plaques d'Immatriculation (LAPI) constituent, lorsqu'elles sont disponibles, une référence pour le calcul de temps de parcours, avec une grande disponibilité et une très bonne précision. Une section périphérique lyonnaise, la RN346, a été équipée par cinq dispositifs LAPI sur 12 kilomètres de long. La vitesse limite est fixée à 90km/h sur cette rocade de l'Est lyonnais. Dans le travail présenté, un seul tronçon de 1100 mètres a été pris en compte (voir section BC, figure 3.4), sans entrées ni sorties, ce qui permet d'assurer une quasi-exhaustivité des données.

Les données de temps de parcours disponibles ont été mesurées entre le 1er Mars 2014 et le 7 Août 2014. Si quelques périodes de maintenance entraînent des absences ponctuelles de données, l'échantillon n'en reste pas moins très pertinent avec 143 journées complètes d'enregistrements comportant une moyenne de 840 mesures valides par heure. De ces données brutes, 13503 sous-périodes de 15 minutes ont été construites. Finalement, une attention a été portée à 13 journées comparables, des jeudis de semaines régulières, résultant en 1026 périodes après une dernière phase de qualification.

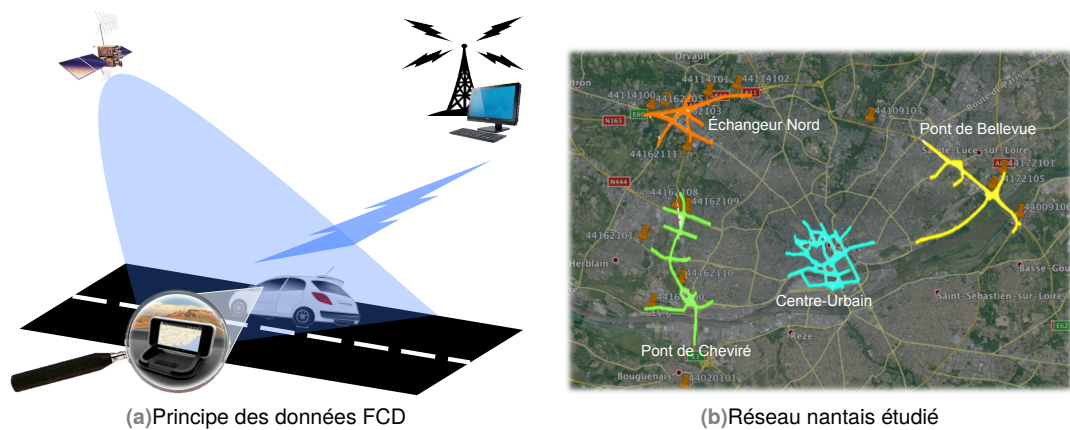


Fig. 3.5.: Suivi de véhicules traceurs.

3.1.3 Véhicules traceurs ou *Floating Car Data (FCD)*

Une collaboration avec l'industriel MEDIAMOBILE a ouvert la voie à l'usage de données issues de véhicules traceurs, ou *floating car data*. Ce type de données est une mine d'or pour un bon nombre d'application d'information trafic indépendantes des gestionnaires d'infrastructure (comme V-TRAFFIC ou GOOGLE MAPS). MEDIAMOBILE possède une flotte de véhicules traceurs qui sont géolocalisés environ toutes les minutes. L'horodatage, la latitude, longitude, ainsi qu'un identifiant unique et anonyme permettent de reconstituer des trajectoires de véhicules (figure 3.5). La qualification de données FCD brutes a fait l'objet du travail de Master de Justine Dinh et a mené par la suite à la proposition d'un algorithme de détection de queue de bouchon, en collaboration avec l'autoroutier AREA. Ces travaux, non présentés dans le manuscrit d'HDR, ont été publiés en conférence et revue [126]. À la suite de cette première étude, un deuxième jeu de données a été demandé afin de valider les capacités de généralisation de nos algorithmes de prévision de trafic à court terme, dans le cadre de la thèse de Pierre-Antoine Laharotte. MEDIAMOBILE a gracieusement mis à disposition trois mois de recueil de données (septembre à novembre 2013) sur quatre zones sensibles de la ville de Nantes et sa périphérie (figure 3.5). Les quatre zones retenues couvrent des portions bien connues des Nantais pour les difficultés aux heures de pointe : le centre-ville près de la rue de Strasbourg, les ponts de Cheviré et de Bellevue et une zone d'échangeurs autoroutiers reliant la Bretagne au reste de la France. D'un point de vue du graphe routier correspondant, on décompte 796 arcs ou sections de routes pour le centre-ville, 177 pour la zone du Pont de Cheviré, 183 pour la zone de Bellevue et 127 pour l'échangeur Nord. Pour cette étude, MÉDIAMOBILE a fourni une donnée pré-traitée par ses soins. Une vitesse harmonique est fournie toutes les 3 minutes sur chacune des sections pré-définies. La représentativité de ces données reste inconnue et un processus de filtrage a également été appliqué. Dans le cadre de nos travaux sur la prévision de trafic, un pas de temps de 30 minutes a été retenu, bon compromis entre représentativité des données et précision.

3.2 Modélisation statistique des temps de parcours

La majorité du texte de cette section a été traduit et retravaillé à partir d'un article paru récemment dans la revue *Journal of Intelligent Transportation Systems* [119]. Cette première contribution du chapitre se consacre à la modélisation statistique des temps de parcours du point de vue de la fiabilité. Les deux jeux de données considérés sont les données Bluetooth de Portland et les données LAPI de Lyon, décrits dans la section précédente. La contribution principale réside dans l'usage de la famille de distributions de Halphen, appliquée à la modélisation de jeux de données empiriques de temps de parcours. Ce système trouve son origine dans les travaux de Étienne Halphen [185], un hydrologue français qui cherchait à modéliser les débits mensuels par la meilleure distribution statistique possible. Complétée par la contribution majeure de Morlat [312], l'approche a répondu à un besoin de flexibilité pour modéliser une grande variété de phénomènes. C'est ainsi que trois distributions de Halphen ont émergé au sein d'un système éponyme : la distribution de type A , de type B et de type B^{-1} ou Inverse B (notée par la suite IB). Les propriétés mathématiques des trois distributions ont ensuite été analysées finement par Perreault *et al.* [350]. Dans la lignée de ces travaux, nos contributions sont les suivantes :

- La première application du système Halphen au domaine du trafic routier.
- Une comparaison empirique complète avec un ensemble de distributions concurrentes, sur deux jeux de données différents et un volume conséquent d'échantillons (plus de 1000 distributions).
- Un outil graphique de sélection de distributions, le diagramme de ratio des moments (δ *Moment Ratio Diagram* noté par la suite δ -MRD). Ce diagramme permet de sélectionner la meilleure distribution à partir d'un jeu de données empirique et d'en estimer les paramètres. De plus, l'outil décrit bien la dynamique des temps de parcours et leur fiabilité [118, 117].

3.2.1 Présentation du système de Halphen

Les éléments suivants rappellent brièvement les définitions mathématiques des lois de Halphen, que l'on peut retrouver dans l'article de Perreault *et al.* [350].

3.2.1.1 La loi Halphen de type A (H_A)

Dans [185], la loi Halphen de type A est définie comme suit :

$$f_A(x) = \frac{1}{2m^\nu K_\nu(2\alpha)} x^{\nu-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right], \quad (3.1)$$

où $x > 0$, m est un paramètre d'échelle strictement positif, α est un paramètre de forme strictement positif tout comme $\nu \in \mathbb{R}$. $K_\nu(x)$ est la fonction de Bessel modifiée de second ordre. Sous un paramétrage bien choisi, cette distribution est équivalente à une Gaussienne Inverse Généralisée.

3.2.1.2 La loi Halphen de type B (H_B)

Halphen a également développé une loi de type B , qui apporte de la souplesse de par les différentes formes possibles en fonction de la gamme de paramètres. La fonction de densité est donnée par

$$f_B(x) = \frac{2}{m^{2\nu} e f_\nu(\alpha)} x^{2\nu-1} \exp \left[-\left(\frac{x}{m}\right)^2 + \alpha \left(\frac{x}{m}\right) \right], \quad (3.2)$$

où $x > 0$, m et ν sont deux paramètres d'échelle et de forme strictement positifs et α un paramètre de forme tel que $\alpha \in \mathbb{R}$, tandis que $e f_\nu(x)$ est la fonction exponentielle factorielle définie par [186] de la façon suivante :

$$e f_\nu(\alpha) = 2 \int_0^\infty x^{2\nu-1} \exp[-x^2 + \alpha x] dx \quad (3.3)$$

3.2.1.3 La loi Halphen de type IB (H_{IB})

La loi Halphen de type IB a été ajoutée au système par Morlat [312], afin d'en étendre les possibilités à travers une relation de symétrie la liant à la loi de type B , comme suit :

$$f_{IB}(x) = \frac{2}{m^{-2\nu} e f_\nu(\alpha)} x^{-2\nu-1} \exp \left[-\left(\frac{m}{x}\right)^2 + \alpha \left(\frac{m}{x}\right) \right], \quad (3.4)$$

où les paramètres sont les mêmes que pour la loi de type B .

3.2.1.4 Temps de parcours et système de Halphen

Les distributions de Halphen, et donc les données empiriques d'hydrologie, partagent de nombreuses propriétés avec les données empiriques de temps de parcours :

- elle sont définies seulement pour des variables réelles positives, *i.e.* $D_f \in \mathbb{R}^+$,
- elles tendent vers 0 pour de petites valeurs, *i.e.* $\lim_{x \rightarrow 0^+} f(x) = 0$,
- elle possèdent une asymétrie à droite, positive *i.e.* $\gamma_1 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] > 0$,
- elles présentent une longue queue de distribution, *i.e.* $C_k = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] > 3$.

De surcroît, quelques propriétés mathématiques en font un système pratique et manipulable :

- Si X suit une loi $H_B(\alpha, \nu, m)$, alors $Y = 1/X$ suit une loi $H_{IB}(\alpha, \nu, 1/m)$.
- Si X suit une loi $H_A(\alpha, \nu, m)$, alors $Y = 1/X$ suit une loi $H_A(\alpha, -\nu, 1/m)$.
- La loi Gamma est un cas limite entre la loi H_A et H_B .
- La loi Gamma Inverse est un cas limite entre la loi H_A et H_{IB} .

L'attractivité du système Halphen réside également dans le puissant outil graphique qui l'accompagne, comme nous allons le montrer par la suite.

3.2.2 Le diagramme de ratio des moments δ -MRD

Le δ -MRD est un outil visuel développé par Morlat [312] afin de représenter l'ensemble des paramètres dans un espace de ratio des moments. Il ne doit pas être confondu avec le β *Moment Ratio*

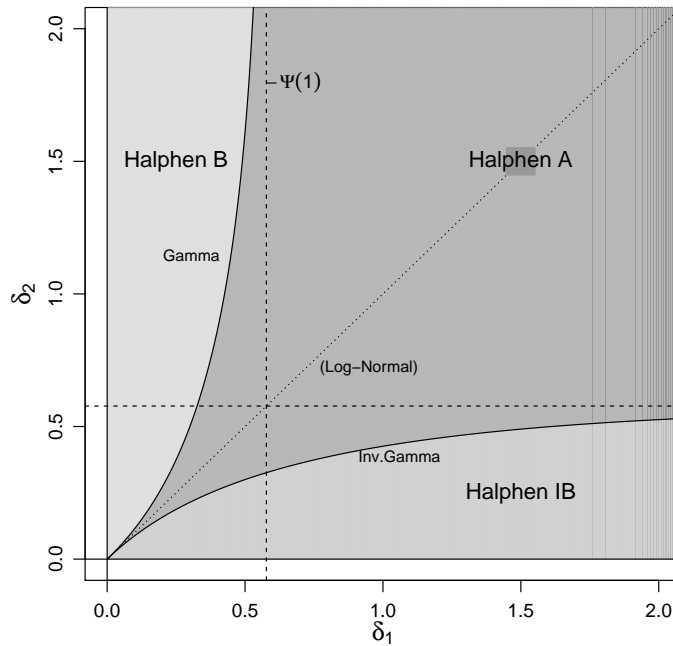


Fig. 3.6.: δ -MRD théorique [350].

Diagramme qui représente un espace liant les coefficients de dissymétrie et d'aplatissement (*skewness* et *kurtosis*). Les deux diagrammes sont décrits et distingués dans [59].

Le δ -MRD lie les moments d'ordre r ($|r| \leq 1$) par les coordonnées suivantes :

1. $\delta_1 = \ln(A/G)$, *i.e.* le logarithme naturel du ratio de la moyenne arithmétique A et de la moyenne géométrique G .
2. $\delta_2 = \ln(G/H)$, *i.e.* le logarithme naturel du ratio de la moyenne géométrique G et de la moyenne harmonique H .

La figure 3.6 représente le δ -MRD du système Halphen dans l'espace (δ_1, δ_2) .

À partir de ce diagramme, il a été possible de dériver une technique innovante d'estimation des paramètres des lois à partir de la localisation d'un échantillon sur le δ -MRD, c'est-à-dire à partir de ses seules coordonnées δ_1 et δ_2 . Cette procédure permet d'établir une passerelle entre une représentation graphique et un calibrage statistique.

3.2.2.1 Sélection de distribution et estimation des paramètres

La première étape de la procédure consiste à choisir la meilleure loi de Halphen correspondante à un échantillon empirique. Pour ce faire, la comparaison de la localisation de l'échantillon avec les courbes limites de la figure 3.6 suffit. D'après la contribution initiale de Bobee *et al.* [59], développée dans nos travaux [118], les ratios des moments pour les lois Gamma et Gamma Inverse sont liés de la façon suivante :

$$\delta_2 = g(\delta_1) \quad (3.5)$$

$$\delta_1 = g(\delta_2) \quad (3.6)$$

où

$$g(x) = x + \left(2\Psi(f^{-1}(x)) + \ln \frac{\Gamma(f^{-1}(x)-1)}{\Gamma(f^{-1}(x)+1)} \right) \quad (3.7)$$

$\Gamma(\cdot)$ et $\Psi(\cdot)$ sont respectivement les fonctions de densité des lois Gamma et Digamma, $f(x) = \ln(x) - \Psi(x)$ et $g(x)$ est définie $\forall x \in]0; -\Psi(1)[$. L'équation 3.5 traduit le cas limite Gamma tandis que l'équation 3.6 exprime le cas limite Gamma Inverse. Ainsi, trois situations sont possibles :

- si $\delta_2 > g(\delta_1)$, la distribution de type *B* est sélectionnée ;
- sinon, si $\delta_1 > g(\delta_2)$, alors la distribution de type *IB* est choisie ;
- dans les autres cas la distribution de type *A* est considérée comme le meilleur modèle.

Ensuite, il est possible de calculer analytiquement les coordonnées du diagramme pour chaque loi tandis que la procédure inverse, qui vise à estimer les paramètres d'une loi Halphen à partir de la position d'un échantillon sur le diagramme, demande plus d'efforts. Une procédure d'optimisation couplée à une méthode hybride entre maximum de vraisemblance et méthode des moments permet d'estimer ces paramètres [119]. Cette approche s'inspire d'autres techniques d'estimation proposées par ailleurs [351, 80, 79] en les adaptant au δ -MRD.

Cet arsenal théorique impose le δ -MRD comme un puissant outil graphique de description du système Halphen. Il est possible de sélectionner la meilleure distribution à partir des seules moyennes arithmétique, géométrique et harmonique d'un échantillon donné mais aussi de situer une distribution théorique et ses paramètres associés sur ce même diagramme, ce qui faciliterait les démarches d'échantillonnage.

3.2.3 Méthodologie d'évaluation des distributions concurrentes

Notre méthodologie repose sur l'usage du δ -MRD pour caractériser et comparer différentes fonctions de densité possibles pour la modélisation d'échantillons de temps de parcours. Dans un second temps, nous montrons comment cet outil apporte une pierre à l'édifice de la fiabilité des temps de parcours en analysant dans le temps les échantillons d'une même section. Même si nous partons d'une inspiration théorique, le système Halphen, la démarche est résolument guidée par les données (*data driven*) et le point de départ consiste en un ensemble de mesures de temps de parcours sur un segment de route simple. Dans chaque cas, les temps de parcours sont normalisés par la distance parcourue de façon à pouvoir comparer plusieurs sections de route. Un point critique a trait à la variable d'intérêt : le temps de parcours est une variable aléatoire composée d'une partie fixe, qui correspondrait à la vitesse maximum (vitesse libre) autorisée qu'un véhicule pourrait adopter tout au long du trajet, et d'une partie aléatoire relative au retard accumulé par rapport à la situation idéale. Dès lors, il est pertinent d'ajouter un décalage correspondant au temps de parcours minimal enregistré (*i.e.* véhicule le plus rapide) durant la sous-période d'intérêt. Ce décalage adaptatif permet de se focaliser sur la partie réellement aléatoire tout en améliorant les performances de calage (en particulier à gauche pour des densités nulles).

Le δ -MRD détermine dans un premier temps, pour un échantillon donné, le meilleur modèle au sein de la famille Halphen, selon les conditions énoncées plus haut. Cette étape indique également les performances d'autres distributions représentées sur le δ -MRD comme la **Gamma**, la **Gamma Inverse**, la **lognormale** ou la **Weibull** [59]. De manière intuitive, plus la distance entre l'échantillon et le modèle théorique est grande, plus la qualité d'estimation diminue. D'autres lois vont compléter le *benchmark*, à deux ou trois paramètres :

- La **loi normale**, qui fera office de référence.
- La **loi lognormale à trois paramètres**. Le troisième paramètre vise à mieux prendre en compte le décalage par rapport à l'origine. Dans la loi lognormale à deux paramètres, les estimateurs du maximum de vraisemblance sont directement obtenus à partir de la moyenne et de l'écart-type empiriques du logarithme naturel des observations. Dans le cas d'une loi à 3 paramètres, une procédure plus raffinée doit être employée, telle que définie dans [86].
- La **loi de Burr à trois paramètres**. Popularisée par Mike Taylor pour la modélisation des temps de parcours [421], des estimateurs des paramètres ont été proposés dans divers articles [448, 458, 397]. À partir de ces contributions, des procédures d'optimisation de type Newton-Raphson peuvent être utilisées. À noter que la loi de Weibull est un cas limite de la loi de Burr à trois paramètres [141].

En ce qui concerne les techniques d'estimation, les modèles de Halphen estimés par la seule méthode du maximum de vraisemblance sont inclus dans l'étude : on distinguera donc le modèle de Halphen simple (théoriquement optimal par la méthode du maximum de vraisemblance, voir [351]) du modèle Halphen MRD et l'on pourra évaluer la pertinence de l'approche δ -MRD pour l'estimation des paramètres. Tous les autres modèles sont estimés par des procédures classiques de log-vraisemblance. Trois critères de performances sont mobilisés :

1. la valeur de la log-vraisemblance,
2. l'erreur quadratique moyenne (EQM) ou *Root Mean Square Error (RMSE)*,
3. la distance de Kolmogorov-Smirnov. Cette dernière se fonde sur la comparaison de fonctions de répartition empiriques. Pour les distributions de Halphen, l'absence d'expressions analytiques nous a amené à discrétiser la variable et fournir des estimations ponctuelles.

3.2.4 Résultats empiriques

Les deux jeux de données, Bluetooth de Portland et LAPI de la rocade est lyonnaise, correspondent respectivement à des sections urbaines et périurbaines. Ces données, agrégées, ne permettent pas d'analyser la variabilité individuelle à l'échelle des véhicules, mais plutôt de comparer des échantillons de temps de parcours sur des périodes temporelles variées. Les résultats présentés comparent neuf modèles : Burr, Halphen, Halphen paramétrée depuis le δ -MRD (Halphen MRD), Gamma, Gamma Inverse, lognormale à deux et trois paramètres, normale et Weibull. Ils ont été estimés pour 140 sous-échantillons dans le cas urbain de Portland, et 1026 sous-échantillons dans le cas périurbain de Lyon. Pour chacun des trois indicateurs d'évaluation choisis, les modèles sont classés du meilleur (numéro 1) au moins bon (numéro 9). Les meilleurs modèles seront donc en moyenne plus proches de 1 avec un faible écart inter-quartile.

La figure 3.7 résume les résultats pour la ville de Portland. En ce qui concerne le critère du maximum de vraisemblance, trois groupes de niveau peuvent être mis en avant. En haut de l'échelle de performance, les lois de Burr et de Halphen dominent le classement tandis que les lois Gamma et

Weibull présentent de bons résultats. Rappelons que les deux derniers modèles sont des cas limites des lois de Halphen et Burr. Les autres modèles candidats n'atteignent pas de bonnes performances sur ce cas.

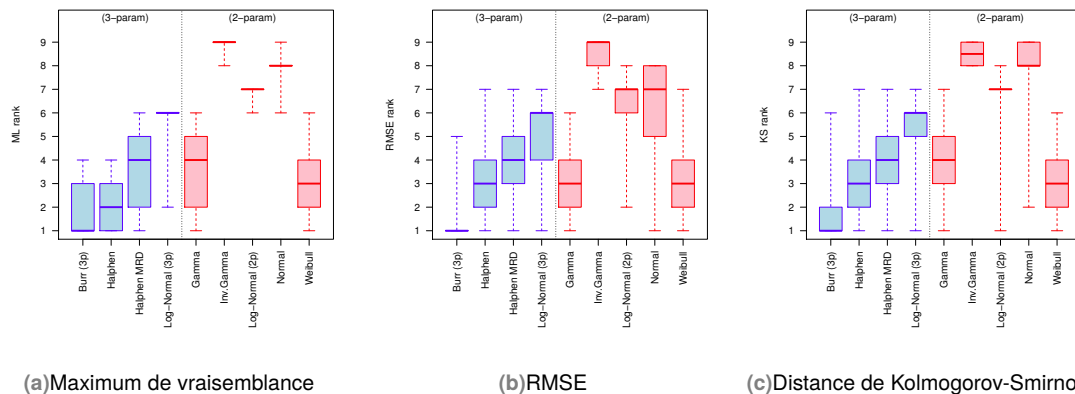


Fig. 3.7.: Boîtes à moustache des distributions de mesures de performance sur des échantillons d'une heure de données à Portland.

Des conclusions similaires sont retrouvées pour le deuxième critère, *i.e.* le RMSE, à la différence près que le modèle de Burr prend l'avantage sur Halphen. L'observation de la distance de Kolmogorov-Smirnov confirme les premières tendances. Quant à la loi lognormale, qui est peut-être la plus utilisée dans l'état de l'art, ces performances sont limitées pour chacun des trois critères d'évaluation. En observant particulièrement les performances du modèle Halphen, on note une chute de performance pour la méthode d'estimation fondée sur le δ -MRD, même si cette baisse n'entraîne pas un déclassement significatif.

Les résultats sur les échantillons lyonnais sont résumés sur la figure 3.8. Comme pour le cas d'études américain, un classement a été effectué pour chaque sous-période et chaque indicateur. Ainsi, les figures illustrent les répartitions de ces classements (à travers des boîtes à moustaches) pour les trois indicateurs de performance choisis.

Dans le cas lyonnais, les différents critères se comportent de manière différente et le maximum de vraisemblance apporte un meilleur pouvoir discriminant. Le classement général reste néanmoins très proche des données de Portland. Les modèles de Burr et Halphen se détachent et la loi de Halphen prend même l'avantage en sélectionnant les trois critères. D'autre part, les lois de Gamma et Weibull dominent dans la catégorie des modèles à deux paramètres. Le deuxième jeu de données apporte des confirmations sur les mauvaises performances de certaines lois très utilisées dans la littérature : la loi lognormale n'est pas assez performante sur ces échantillons de temps de parcours tandis que les performances de la loi normale sont trop instables, comme l'illustrent les boîtes à moustache des RMSE et des distances de Kolmogorov, de manière flagrante dans le cas lyonnais. Eu égard au système Halphen, la méthode d'estimation à partir du δ -MRD propose un modèle Halphen MRD dans le trio de tête, ce qui valide sa potentielle utilisation en conditions opérationnelles.

3.2.5 Mesurer la fiabilité des temps de parcours avec le δ -MRD

Outre un outil de sélection du meilleur modèle statistique, les coordonnées du δ -MRD peuvent être considérées comme des métriques de fiabilité : de faibles valeurs de δ_1 et δ_2 correspondent à une

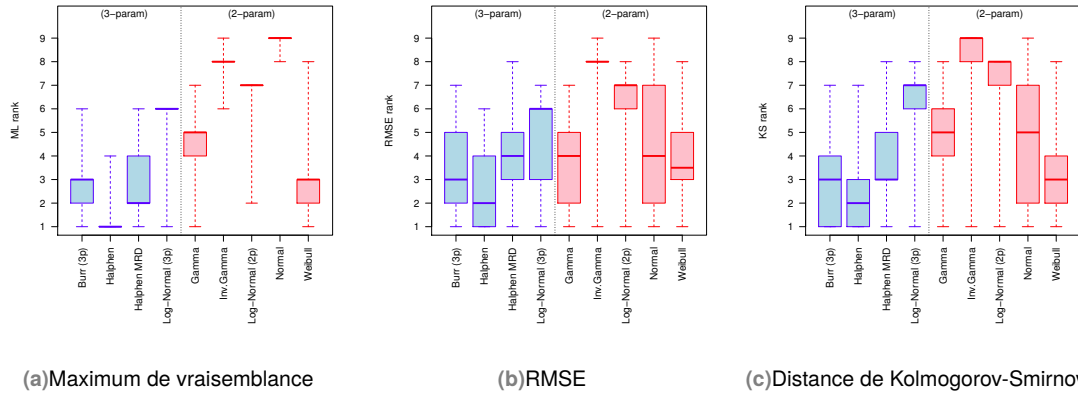


Fig. 3.8.: Boîtes à moustache des distributions de mesures de performance sur des échantillons de 15 minutes de données à Lyon.

faible dispersion des valeurs de l'échantillon, et ainsi un haut niveau de fiabilité. Cette hypothèse a été validée par une analyse de corrélation des deux coordonnées avec les indicateurs de fiabilité présents dans l'état de l'art, à savoir : le *Buffer Time (BT)*, le *Buffer Time Index (BTI)*, le *Planning Time Index (PTI)*, le coefficient de variation (CV_T), le *Misery Index (MI)*, l'indicateur floridien (IF), l'étendue de la distribution λ_{var} , l'asymétrie de la distribution λ_{skew} , la fréquence de congestion (FC), et tout simplement l'écart type σ_T . Le détail des indicateurs peut être trouvé dans la revue bibliographique de Lomax [273]. La plupart se fondent sur des considérations sur les fonctions de répartition censées exprimer la variabilité. Par exemple, les index BT et BTI réfèrent aux différences et différences relatives entre les 90ème et 50ème centiles, exprimant un budget-temps qu'un automobiliste doit prévoir en plus d'un temps de parcours nominal tandis que l'indicateur floridien mesure la proportion de véhicules présentant un temps de parcours inférieur à 110% du temps médian. La figure 3.9 présente les valeurs des coefficients de corrélation des deux paramètres δ_1 et δ_2 du δ -MRD avec ces indicateurs de l'état de l'art. La comparaison est faite pour les deux cas d'études et sur trois type de modèles :

- Sans décalage (TT), c'est-à-dire sur données brutes ;
- Avec décalage constant (TDF) : les distributions sont toutes décalées à droite d'un même temps de parcours minimal ;
- Avec décalage adaptatif (TD) : c'est le modèle adopté dans les résultats précédents, le décalage est propre à chaque sous-période.

Les résultats mettent en exergue des coefficients de corrélation élevés entre les métriques de l'état de l'art et les coordonnées du δ -MRD, surtout pour le modèle sans décalage (TT). Ce résultat est logique car les métriques de l'état de l'art sont proposées pour des distributions de temps de parcours et non de délais. Ainsi, le δ -MRD peut, dans certaines conditions, se substituer à des métriques de fiabilité existantes. D'autre part, un dernier résultat concerne l'analyse temporelle des distributions de temps de parcours. La variabilité intra-journalière peut par exemple être observée en traçant sur un δ -MRD plusieurs échantillons relatifs à plusieurs sous-périodes homogènes d'une même journée. Les positions des points sur le δ -MRD renseignent sur la variabilité des périodes étudiées. De même, plusieurs jours peuvent être comparés entre eux afin d'observer une variabilité inter-journalière. A titre d'exemple, la figure 3.10 compare quatre échantillons de temps de parcours relevés à la même heure, pendant quatre jours différents sur la section de Portland.

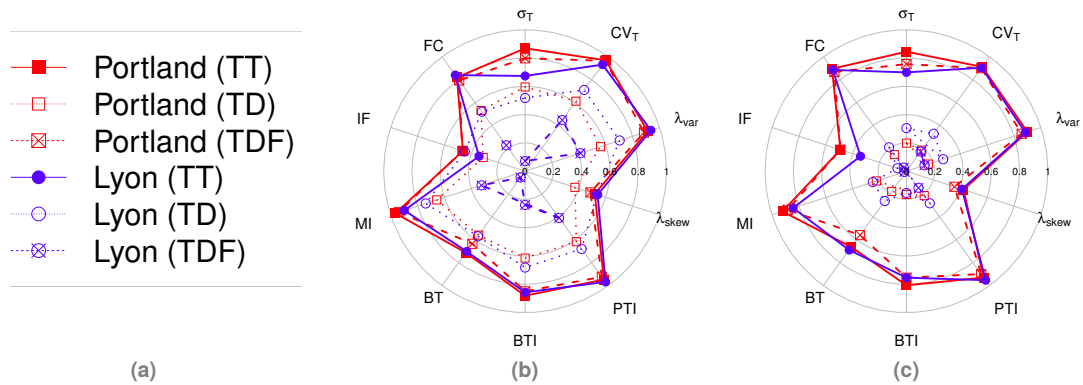


Fig. 3.9.: Similarité entre les indicateurs de fiabilité classiques et les coordonnées du δ -MRD : (a) légende (b) coefficient de corrélation avec δ_1 (c) coefficient de corrélation avec δ_2 .

On remarque que les périodes numérotées 42 à 64, qui concernent des périodes à faible trafic, s'opposent clairement à deux journées plus chargées, numérotées 2 et 55. La variabilité des conditions de circulation sur ces deux dernières journées mène en effet à des histogrammes plus dispersés. Le δ -MRD présente une vision synthétique de ces dégradations de fiabilité.

3.2.6 Discussion

Le système Halphen se compose de trois types de lois dont la souplesse s'accorde particulièrement bien à la modélisation du trafic routier. Ce système, appliqué pour la première fois au domaine de la mobilité intelligente, est accompagné d'un outil puissant, le δ -MRD, qui synthétise de manière graphique et synthétique un échantillon de données au sein de la famille Halphen. Après une présentation théorique des propriétés des lois de Halphen et du δ -MRD, une méthodologie d'évaluation de modèles statistiques concurrents a été proposée pour deux jeux de données complémentaires, à Portland et Lyon. Les principales conclusions de cette étude comparative sont les suivantes :

- Sur la base des trois critères d'évaluation choisis, les meilleurs modèles parmi les neuf testés sont les lois de Burr et de la famille Halphen. Elles présentent une variété de forme leur permettant de s'adapter à la diversité des situations rencontrées. La loi de Burr, bien connue pour les problèmes de fiabilité [448], complète la distribution plus connue de Weibull. Elle est par ailleurs utilisée dans la littérature récente sur les temps de parcours [413]. Le système Halphen, quant à lui, se distingue par ces trois formes complémentaires. Selon le critère choisi et l'échantillon concerné, le meilleur modèle des trois peut varier.
- L'étude empirique souligne également que les lois Gamma et Weibull sont les plus performantes parmi les distributions à deux paramètres. En tant que cas limites des systèmes Halphen et Burr, elles présentent un bon compromis entre simplicité et puissance de modélisation. La loi Gamma a d'ailleurs été appliquée récemment au trafic routier dans une version composée [231].
- Les lois lognormales à deux ou trois paramètres ne sont pas propices à la modélisation des temps de parcours.
- La technique d'estimation des paramètres d'une loi de Halphen à partir du δ -MRD est globalement satisfaisante même si l'erreur d'estimation fait baisser les performances générales

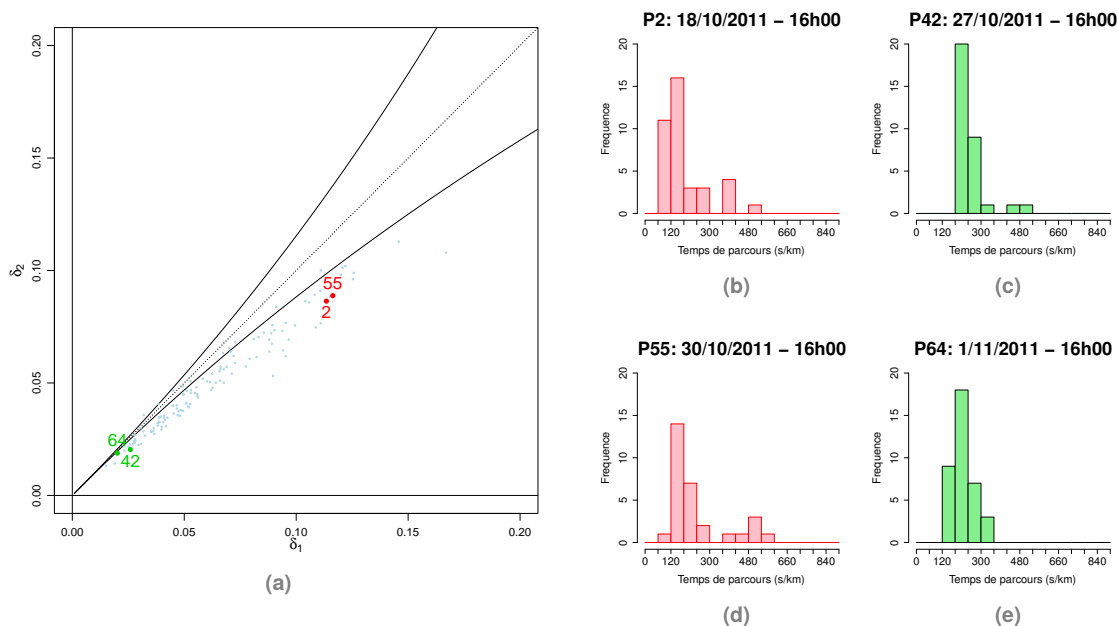


Fig. 3.10.: Comparaison des distributions de temps de parcours mesurés entre 16h et 17h à Portland, sur différents jours; (a) Position des échantillons de temps de parcours sur le δ -MRD; (b) et (d) Exemples d’histogramme de temps de parcours, distributions à forte variabilité; (c) et (e) Exemples d’histogramme de temps de parcours, distributions à faible variabilité.

et ne permet pas de conclure à une supériorité nette du modèle Halphen MRD par rapport à la loi Gamma à deux paramètres. La méthode d’estimation, en particulier la procédure d’optimisation, doit être améliorée dans de futurs travaux.

Afin d’illustrer la flexibilité des trois types de lois Halphen, la figure 3.11 montre un diagramme de dispersion des différents échantillons de Portland et Lyon sur le δ -MRD.

À chaque échantillon correspond un point du diagramme, dont la localisation oriente vers une des trois distributions. Le graphique souligne la prépondérance des distributions de type *B* pour le jeu de données lyonnais tandis que l’échantillon Bluetooth de Portland correspond, selon les conditions, à des lois de type *A* ou *B*. On remarque que les points de la rocade est lyonnaise sont concentrés dans la partie inférieure gauche du δ -MRD. Ceci correspond à un fonctionnement stable et sans perturbation majeure, tel qu’il peut être attendu d’une section péri-urbaine. De leur côté, les échantillons de Portland sont plus dispersés et situés plus en haut à droite par rapport à leurs homologues français. Cette différence s’explique par la plus grande variabilité du trafic sur une section urbaine forcément sujette à des dynamiques court-terme plus fortes.

Les perspectives de ce travail sont nombreuses et certaines ont déjà été abordées récemment dans des travaux annexes. Outre la question de la fiabilité, abordée dans ce manuscrit, il semble important d’établir une correspondance claire entre le δ -MRD et les états de trafic, tel que le diagramme fondamental peut les représenter à l’équilibre à travers les grandeurs macroscopiques que sont le débit, la densité et la vitesse du flot. Ce travail a été esquissé dans notre récente publication [117] grâce à une simulation mesoscopique du trafic. D’autre part, des résultats récents montrent une

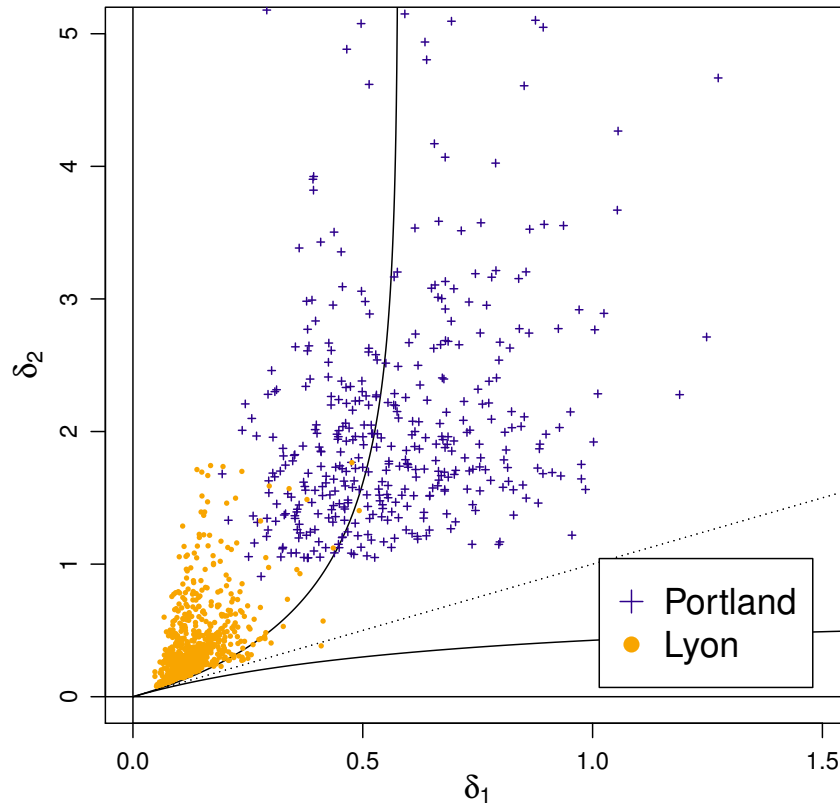


Fig. 3.11.: Représentation des échantillons de Portland et Lyon sur le δ -MRD.

adéquation entre des clusters d'états de trafic sur le diagramme fondamental et des zones du δ -MRD. Cette passerelle entre les deux outils théoriques pourrait être mieux mise en valeur, en s'intéressant par exemple aux régimes de vitesse libre, pré-congestion, congestion, ou post-congestion, tels que décrits dans [436].

La principale faiblesse du travail présenté concerne l'étude d'un seul segment de route, pour les deux cas d'études. Même si cette simplification a permis d'obtenir des premiers résultats solides et facilement interprétables, le lecteur pourra se référer à la thèse de Raphaël Delhome pour des considérations sur l'additivité des lois de Halphen. En particulier, l'auteur y développe une approche empirique pour constituer des sommes de deux distributions présentes sur le δ -MRD. L'étape suivante consisterait à évaluer les performances du système Halphen à l'échelle d'un itinéraire. Les conclusions sur les performances des différents modèles y seraient-elles les mêmes ? Le δ -MRD conserverait-il ses capacités d'analyse intra- et inter-journalière à l'échelle du réseau ? Autant de questions qui mériteraient d'être traitées dans les années à venir, en couplage avec des outils plus récents de la théorie du trafic comme le diagramme fondamental macroscopique ou *macroscopic fundamental diagram* [171].

3.3 Analyse spatiotemporelle d'un réseau de capteurs Bluetooth

Après une première contribution en statistiques appliquées sur la modélisation des temps de parcours, nous proposons ici une méthodologie de fouille de données pour la technologie Bluetooth, partant de la **donnée brute** du réseau de Brisbane. L'accent est porté sur la supervision et la prévision du trafic routier à l'échelle d'un réseau urbain. À l'issue d'une phase de préparation des données qui filtre les imperfections et construit les structures aptes à caractériser le fonctionnement d'un réseau de capteurs urbain, une analyse spatiotemporelle du réseau est introduite à l'aide d'approches non supervisées comme l'allocation latente de Dirichlet ou les *K-means*. Cette deuxième étape valide la pertinence de la donnée Bluetooth pour la caractérisation temps-réel des états de trafic. Ensuite, nos travaux les plus récents ont été consacrés à la prévision court-terme du trafic par apprentissage statistique.

3.3.1 Préparation des données

La motivation opérationnelle du travail est l'évaluation du potentiel des données Bluetooth pour la gestion du trafic, après le déploiement de détecteurs aux intersections du réseau de la ville de Brisbane, Australie (voir section 3.1.1). L'opérateur en charge du réseau, le *Brisbane City Council*, a ainsi fourni un échantillon de deux mois de données pour 80 détecteurs, bien répartis sur le réseau. Comme dans tout processus de fouille de données, la phase de qualification est très chronophage et nécessite un effort pointilleux afin que la qualité des données en entrées des algorithmes soit satisfaisante. Notre démarche repose sur 3 étapes illustrées sur la figure 3.12 :

1. Un pré-filtrage temporel qui détecte les *outliers*, vérifie la qualité des données et discrimine les différents modes de transports. Cette étape a été présentée en section 3.1.1.
2. La construction d'une matrice *Bluetooth-Origin-Destination* ou *B-OD* qui stocke le nombre de détections et les temps de parcours entre chaque couple de capteurs.
3. Un post-filtrage qui traite les cas de détections manquées.

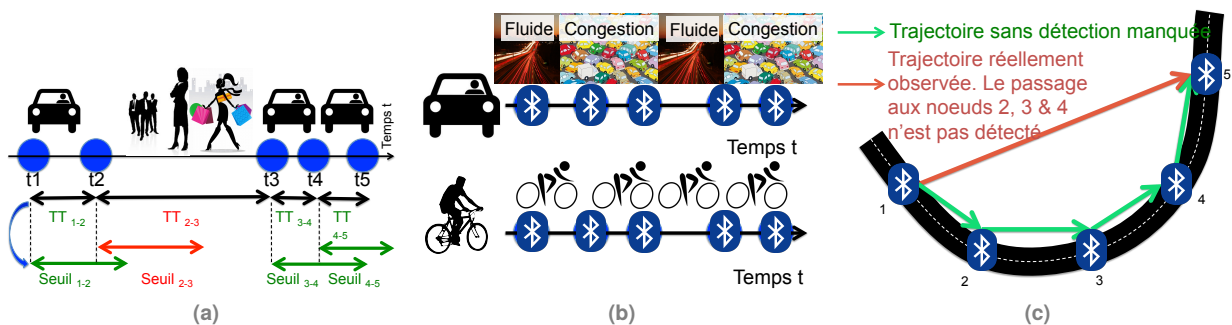


Fig. 3.12.: Illustration des difficultés rencontrées à partir des données Bluetooth : a) Scission en itinéraires du set de détecteurs franchi par un usager b) Distinction des modes de transports c) La problématique des détections manquées.

3.3.1.1 Le concept de la matrice Bluetooth Origine-Destination (B-OD)

Dans l'optique d'agréger les données et de caractériser les conditions de circulation à l'échelle du réseau, nous avons introduit le concept de matrice Bluetooth Origine-Destination (B-OD). Ce concept diffère des matrices OD classiques manipulées pour prévoir la demande de déplacement sur un réseau. Les matrices B-OD contiennent le nombre de détections entre chaque couple de capteurs Bluetooth, associées aux relations vitesse/débit sur les sections traversées entre ces détecteurs. Chaque capteur Bluetooth, nœud du réseau, est associé à une origine et une destination dans la matrice B-OD et chaque couple de détecteurs correspond à un élément dans la matrice B-OD. Pour des conditions fixées, deux types de B-OD peuvent être créés :

- Une matrice B-OD de volume, qui stocke le nombre de détections sur chaque couple,
- Une matrice B-OD temporelle qui stocke les temps de parcours.

3.3.1.2 Post-filtrage et matrice B-OD finale

Comme l'illustre la figure 3.12 (c), une des difficultés rencontrées avec les données Bluetooth réside dans les détections manquées, qui s'apparentent à des trajets pour lesquels l'ensemble des détecteurs croisés par le véhicule est incomplet. Ce phénomène dégrade la qualité des matrices B-OD et peut faire rater la structuration réelle d'un réseau, en saturant certaines zones et en sous-estimant d'autres. Le processus de post-filtrage intervient pour pallier cette inconsistance et améliorer la qualité de la matrice B-OD. Il s'agit de réaffecter un certain nombre de trajets suspects au vu de la première matrice B-OD. Pour ce faire, des hypothèses de plus courts chemins entre deux capteurs permettent de reconstituer les trajets les plus plausibles entre deux capteurs. Cette réallocation fait passer d'une matrice B-OD complète à un graphe d'incidence connexe : chaque nœud est connecté à chacun des autres nœuds par un chemin composé de nœuds intermédiaires.

3.3.2 Représentation et représentativité

La visualisation des matrices B-OD donne une première idée du potentiel du Bluetooth pour la supervision du trafic. Nous introduisons également le concept de diagramme fondamental Bluetooth (B-FD) pour observer les relations entre vitesse/nombre de détections et ainsi la qualité de l'écoulement. Enfin, les représentations des états de trafic n'ont un sens que si elles sont représentatives de l'ensemble des véhicules du réseau. C'est un problème majeur quand on sait que les applications disponibles aujourd'hui sur les *smartphones* produisent parfois de jolies cartes construites avec très peu de véhicules. Grâce à des données issues de boucles électromagnétiques sur le réseau de Brisbane, une estimation du taux de pénétration de la technologie Bluetooth peut être donnée.

3.3.2.1 Première représentation spatiale

La matrice B-OD finale peut être directement représentée et scrutée afin d'évaluer dans quelle mesure elle reconstruit de façon fidèle les conditions de trafic de la ville de Brisbane. La figure 3.13 montre une visualisation sur un fond de carte de la ville de Brisbane de la matrice B-OD volume pour des heures creuses et heures pleines. Le même type de résultats est possible pour la matrice B-OD temporelle en affichant une vitesse médiane pour chaque lien. Cette première visualisation est prometteuse tant l'on retrouve les dynamiques propres à la ville de Brisbane. Par exemple, l'accès

à l'aéroport et la zone du port au nord-est de la ville expérimentent souvent de forts volumes aux heures de pointes tandis que des autoroutes urbaines comme la *Pacific Highway* liant le centre de la ville aux banlieues sud sont toujours très fréquentées.

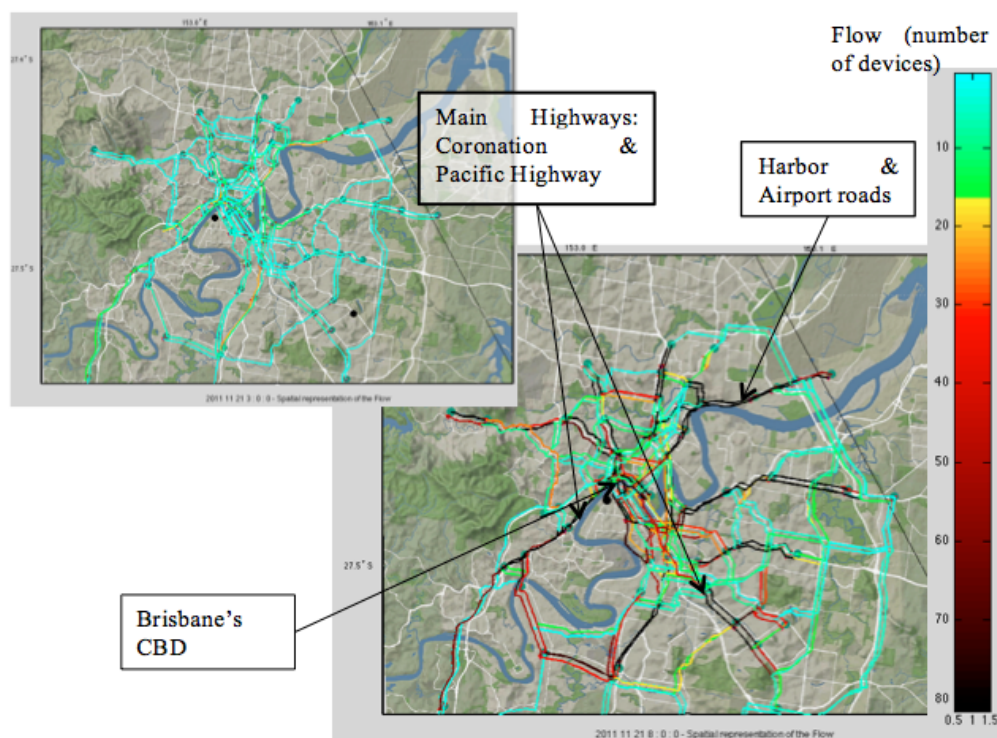


Fig. 3.13.: Représentation spatiale de matrices B-OD de volume à différentes heures pour la journée du 7 novembre 2011. À gauche : entre 3h et 4h du matin (période creuse). À droite : entre 8h et 9h du matin (pic de trafic).

Cette première visualisation, bien que séduisante à l'oeil nu, reste très superficielle, et des outils adaptés de la théorie du trafic peuvent améliorer l'observation des états du trafic et du niveau de service du réseau.

3.3.2.2 Le diagramme fondamental Bluetooth (BFD)

La figure 3.14 ((a) & (b)) représente l'évolution sur un mois de données, des volumes de trafic et des vitesses pour lien Bluetooth. Les périodes de pointe du matin et du soir sont bien identifiables sur ces séries temporelles.

Nous proposons sur la figure 3.14 (c) un nouvel indicateur, le diagramme fondamental Bluetooth (B-FD), qui représente pour chaque couple B-OD (lien) la relation vitesse/débit. Cet outil est adapté du concept de diagramme fondamental macroscopique en théorie du trafic [171]. Comme le montre la figure 3.14, le B-FD reproduit les états de trafic et constitue un outil pertinent pour l'évaluation des performances du réseau.

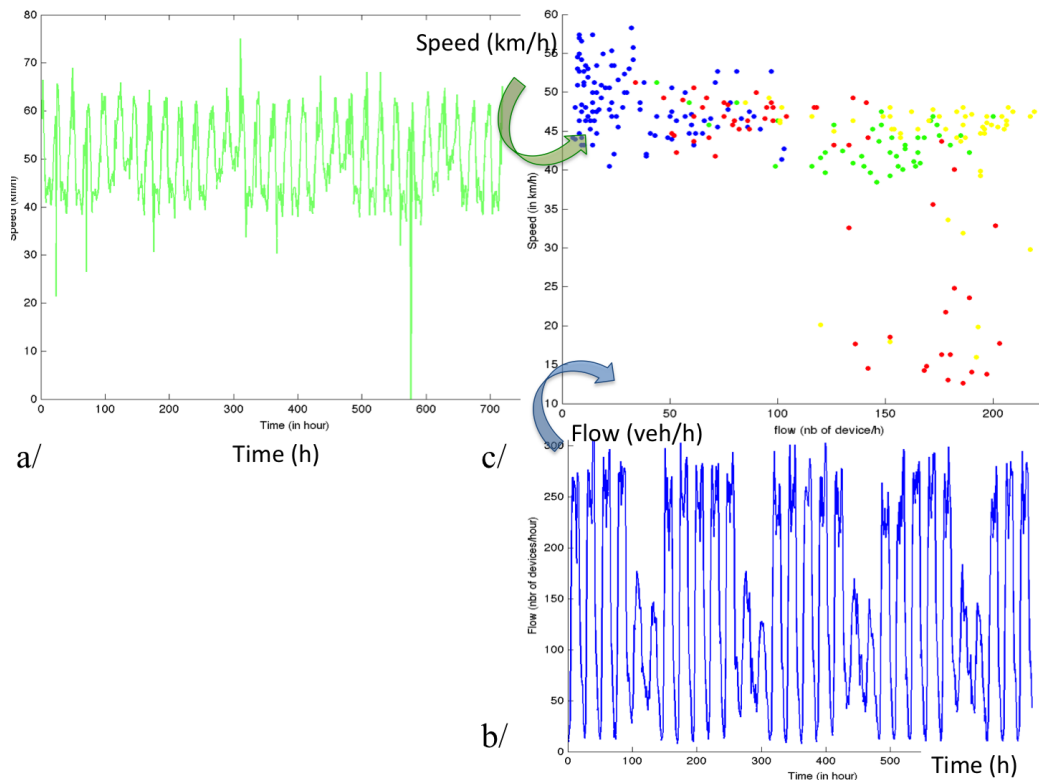


Fig. 3.14.: Illustration des conditions de trafic pour un couple de capteurs Bluetooth : (a) Série temporelle de la vitesse horaire sur un mois. (b) Série temporelle du débit moyen sur un mois. (c) Diagramme fondamental Bluetooth (B-FD) débit/vitesse colorié par périodes temporelles ([bleu : soirée et nuit]; [rouge : matin]; [vert : midi]; [jaune : après-midi]).

3.3.2.3 Fluctuations du taux de pénétration

Le taux de pénétration représente, à l'instar des véhicules connectés du chapitre précédent, la proportion de véhicules équipés d'émetteurs Bluetooth au sein de l'ensemble des véhicules circulant sur le réseau. Dans l'état de l'art, le chiffre de 20% circule quant à cette représentativité du trafic Bluetooth [323, 32]. Nous avons souhaité confirmer ces estimations à travers une étude des données de boucles électromagnétiques disponibles pour la ville de Brisbane. En sélectionnant un sous-ensemble d'intersections et en comparant les volumes collectés pour différentes journées, les analyses rapportent un taux de pénétration fluctuant entre 2 et 45%, comme le montre la figure 3.15.

Une telle variation temporelle s'explique par différentes hypothèses : problèmes de détection, comportements particuliers des usagers Bluetooth, problèmes également liés aux boucles. Nous retiendrons la première hypothèse liée à une fluctuation de la qualité de détection des capteurs Bluetooth. Toutefois, en moyenne, la qualité de détection Bluetooth reste satisfaisante et permet de conclure quant à la représentativité de cette nouvelle source de données. Il est alors possible d'aller plus loin dans l'analyse spatiotemporelle des états de trafic.

Representativeness of Bluetooth data in comparison to loops data on one day (22 October 2012)

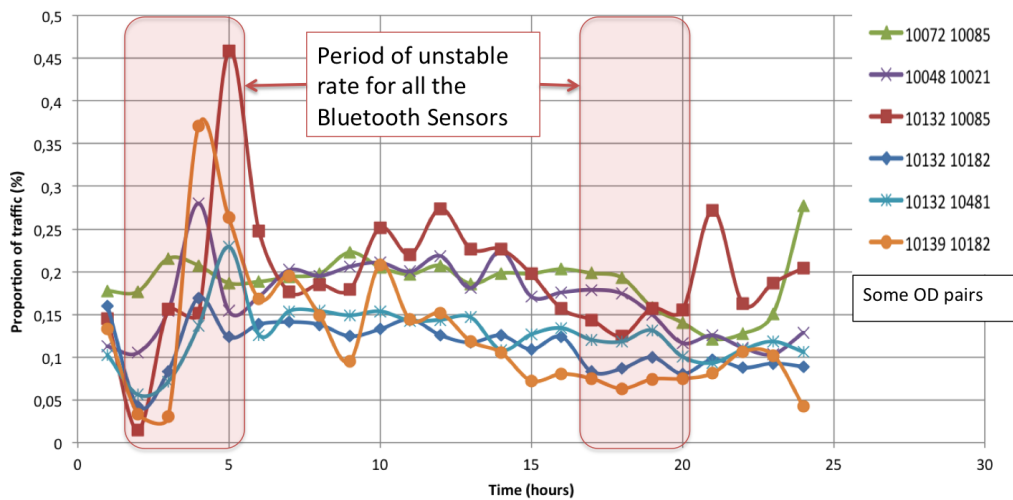


Fig. 3.15.: Observation du taux de pénétration sur un jour pour différents couples de B-OD.

3.3.3 Clustering spatiotemporel

Nous présentons ici des résultats originaux sur l'analyse spatiotemporelle des états de trafic. Nos approches non supervisées vont manipuler les deux outils introduits précédemment :

- Une analyse temporelle globale des matrices B-OD *via* une allocation latente de Dirichlet ;
- Une analyse spatiale des zones similaires à travers un clustering des B-FD.

3.3.3.1 Analyse temporelle globale du réseau : approche par allocation latente de Dirichlet

Un premier résultat concerne l'analyse du fonctionnement temporel de matrices B-OD successives, comme autant d'images instantanées de l'état d'un réseau. Ce type de point de vue se rapproche du travail effectué par Delafontaine *et al.* sur les mouvements humains enregistrés par la technologie Bluetooth, dans lequel les auteurs utilisent des techniques d'alignement séquentiel [116]. Ici, l'analogie est différente puisque nos travaux s'inspirent de l'allocation latente de Dirichlet, ou *Latent Dirichlet Allocation (LDA)*, un modèle de catégorisation textuelle introduit par David Blei en 2003 [57]. L'idée consiste à considérer que les documents peuvent être caractérisés par un ensemble de thèmes (*topics*) abordés, ces mêmes thèmes étant générés par une distributions de mots. Le modèle LDA appartient à la famille des modèles probabilistes génératifs qui vont retrouver la structure latente du document. Cette approche a déjà été appliquée à d'autres domaines que la fouille de données textuelles. Farrahi et Gatica-Perez [148] l'ont par exemple adapté pour l'analyse de données de téléphonie mobile, Huynh *et al.* [205] s'en sont servis pour extraire des routines de comportement humain à partir de capteurs portables tandis que Niebles *et al.* [331] l'utilisent pour

la classification sémantique d'images. Dans le domaine du transport, le modèle a été utilisé pour fouiller des données générées par les systèmes de vélos en libre-service (VLS) [311, 98]. L'idée sous-jacente repose sur l'hypothèse que l'usage d'un mode de transport pourrait être résumé par un ensemble fini de profils de demandes, représentés par des *pattern* typiques d'OD. Ensuite, la représentation spatiale de ces *pattern* peut amener à une meilleure compréhension des dynamiques urbaines.

L'adaptation du LDA à l'analyse spatiotemporelle du trafic Bluetooth procède, une fois encore par analogie entre *text mining* et fouille de données trafic : les documents (ou sacs de mots) sont remplacés par des sacs de trajets, les matrices B-OD, et les mots sont les couples de capteurs, éléments de la matrice B-OD. Dans ce travail, seule la matrice B-OD volume est considérée. En d'autres termes, un mot sera apparenté à la valeur d'une case de la matrice B-OD, c'est-à-dire le nombre de détections pour une période de temps donnée entre deux capteurs Bluetooth (la période de temps est fixée à 1h). Ainsi un document représente l'ensemble des éléments de la matrice B-OD, c'est-à-dire une cartographie des états de trafic du réseau à une heure donnée. Le corpus de documents comprend les différentes images du réseau au cours du temps. La méthode fait apparaître des thèmes latents, les *topics*, qui ne sont pas directement observables. Dans le cas du trafic routier, ces thèmes peuvent être interprétés comme des *pattern* de B-OD, à partir desquels les documents pourront être définis par des vecteurs de proportions. Ainsi, le vocabulaire est composé de couples B-OD qui sont les éléments du produit cartésien des deux ensembles finis des capteurs Bluetooth. Les sacs B-OD sont décrits comme des matrices d'occurrence, où chaque élément (i, j) correspond au nombre de détections au temps t . À partir de cette analogie, le processus génératif LDA peut être réécrit dans la forme suivante : tout d'abord, K profils latents de demande, ou *pattern* B-OD, sont échantillonnés depuis une loi de Dirichlet de paramètre β :

$$\Lambda^{(k)} \sim \mathcal{D}(\beta), \forall k \in \{1, \dots, K\}. \quad (3.8)$$

Ensuite, on considère que chaque matrice B-OD, c'est-à-dire chaque document, ou sac de trajets i dans l'ensemble $\{1, \dots, M\}$ est généré selon deux étapes :

1. Génération des proportions de *pattern* dans le document :

$$\pi^{(i)} \sim \mathcal{D}(\alpha).$$

2. Pour chaque trajet dans le sac i :

- a) Génération du *pattern* T :

$$T \sim \mathcal{M}(1, \pi^{(i)}).$$

- b) Génération d'un couple B-OD W à partir du *pattern* T :

$$W \sim \mathcal{M}(1, \Lambda^{(T)}).$$

Où \mathcal{M} représente la loi multinomiale et α , $\pi^{(i)}$, $\Lambda^{(T)}$ des paramètres du modèle, à estimer.

La réinterprétation du LDA pour le problème du trafic routier ne change pas l'essence de la méthode. Le modèle classique du LDA est adapté avec les mêmes hypothèses : le système générant les B-OD est considéré comme stationnaire pour de courtes périodes de temps (1 heure dans notre cas), les distributions des documents sont fixés et les mêmes pour tous les mots. Le processus génératif aléatoire est fixé pendant un laps de temps t avant de basculer sur un autre régime. Les

deux principales sorties du modèle sont les variables $\Lambda^{(k)}$ et $\pi^{(i)}$ qui doivent être réinterprétées dans notre contexte de matrices B-OD dynamiques :

- $\Lambda^{(k)}$: la distribution discrète sur les couples B-OD. Ces *pattern* constituent des profils de demande qui permettent de décrire géographiquement les trajets par rapport à des profils typiques, les *pattern*.
- $\pi^{(i)}$: les proportions qui résument chaque matrice B-OD par un vecteur de proportion des *pattern*.

Une question fondamentale réside dans le choix du nombre de *pattern* K utilisés pour décomposer les données. Ce paramètre peut être estimé à l'aide d'une mesure de perplexité du modèle [57], évaluée sur une base de données test pour différentes valeurs de K . La valeur de K associée au modèle minimisant la valeur de perplexité sur de nouvelles données est ensuite sélectionnée. Dans notre cas, cette procédure mène à la sélection de 7 *pattern*.

Comme les matrices B-OD sont triées de façon temporelle, les *pattern* de B-OD ont le potentiel de donner une représentation compacte du comportement temporel du système. De plus, les régularités cycliques du système doivent être mises en valeur par une telle approche. Ainsi le comportement observé à l'heure de pointe du lundi matin, par exemple entre 8h et 9h, pourrait être similaire à celui d'autres journées à la même heure. En d'autres termes, des activités régulières comme les déplacements pendulaires, les loisirs durant les soirées ou les week-ends, pourraient être retrouvées par le modèle, selon leur significativité. Ces *pattern* d'activités latents, non directement observables, pourraient confirmer des hypothèses empiriques que les gestionnaires formulent, par expérience, sur le fonctionnement de leur réseau.

Les premiers résultats concernent les proportions de trajets associées à chaque *pattern* au cours du temps. La figure 3.16 (a) donne le nombre de détections susceptible d'être générée par heure par chaque *pattern*. Ce chiffre est obtenu en multipliant les proportions de chaque *pattern* dans chaque intervalle de temps par le nombre de détections observées durant cet intervalle.

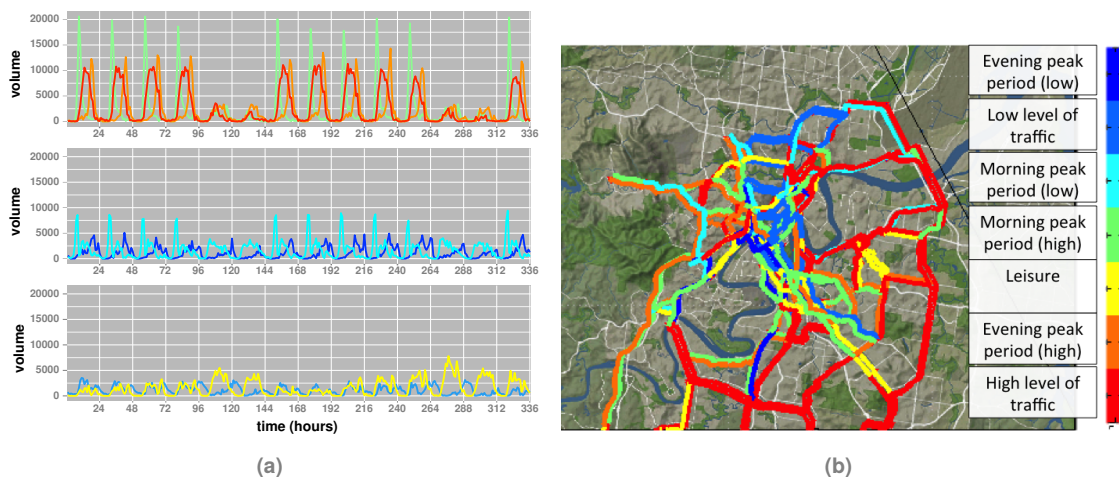


Fig. 3.16.: (a) Deux semaines de données Bluetooth. Nombre de détections générées par heure par chaque *pattern*. (b) Représentation spatiale des couples B-OD en fonction de leurs *pattern* associés (Novembre 2011).

Les régularités cycliques attendues sont clairement visibles sur ces figures. La première différence claire est classique, entre semaine et week-end, 5 jours comparables étant suivis par deux autres

jours moins fréquentés. Sur le haut de la figure 3.16 (a), 3 *pattern* sur 7 ont été représentés, facilement identifiables : on reconnaît les pics du soir hebdomadaires (en orange), les pics du matin (en vert) et du midi (en rouge). Ces trois *pattern* sont responsables de la majorité des trajets, si on compare les volumes aux quatre autres. Ces derniers présentent des caractères moins tranchés même si quelques hypothèses peuvent être faites comme par exemple des activités de loisirs du soir (couleur jaune), des sorties du week-end et du soir (bleu). En considérant à présent chaque couple B-OD indépendamment, il est possible d'affecter un couple d'OD au *pattern* depuis lequel il a la plus grande probabilité d'être généré. À partir de cette opération, la figure 3.16 (b) produit une cartographie du réseau où chaque couple B-OD est associé à un *pattern*, montrant spatialement différents comportements temporels. Si une analyse *pattern* par *pattern* est laissée à la discrétion des spécialistes de la ville de Brisbane, la visualisation proposée valide néanmoins la consistance de la méthode : les couples B-OD les plus influencés par les *pattern* vert et orange, *i.e.* pics hebdomadaires du matin et du soir, sont cohérents avec ce que les usagers, en particulier les salariés, expérimentent chaque semaine. De plus, les sections dominées par le *pattern* jaune correspondent bien à des destinations de loisirs (ex : plages, parcs). Ainsi cet aperçu spatial guidé par les profils latents issus du LDA donne une vision cohérente des dynamiques temporelles du trafic urbain.

3.3.3.2 Clustering des diagrammes fondamentaux Bluetooth (B-FD)

La méthode LDA se concentrant uniquement sur des comptages de trafic, une deuxième méthode a été mise en place afin de mieux caractériser les états de trafic grâce au B-FD introduit précédemment. En effet, une observation préliminaire des différents B-FD souligne quelques formes typiques qu'un algorithme de classification non supervisée pourrait automatiquement retrouver. Les techniques de clustering ont déjà été adoptées pour la caractérisation des conditions de trafic sur les réseaux routiers [467, 20]. Le but est de mettre en valeur des structures spatiales ou temporelles représentatives et similaires puis de se concentrer conjointement sur ces groupes à des fins de supervision temps-réel ou prévision court-terme. Ce type d'approches va gagner du terrain dans les prochains mois avec l'arrivée des données massives pour la gestion du trafic par les opérateurs routiers. Pour le cas du Bluetooth, nous proposons d'adapter une méthode simple, les *K-means*, pour le clustering des B-FD [135]. La procédure repose sur trois étapes résumées dans la figure 3.17 (a).

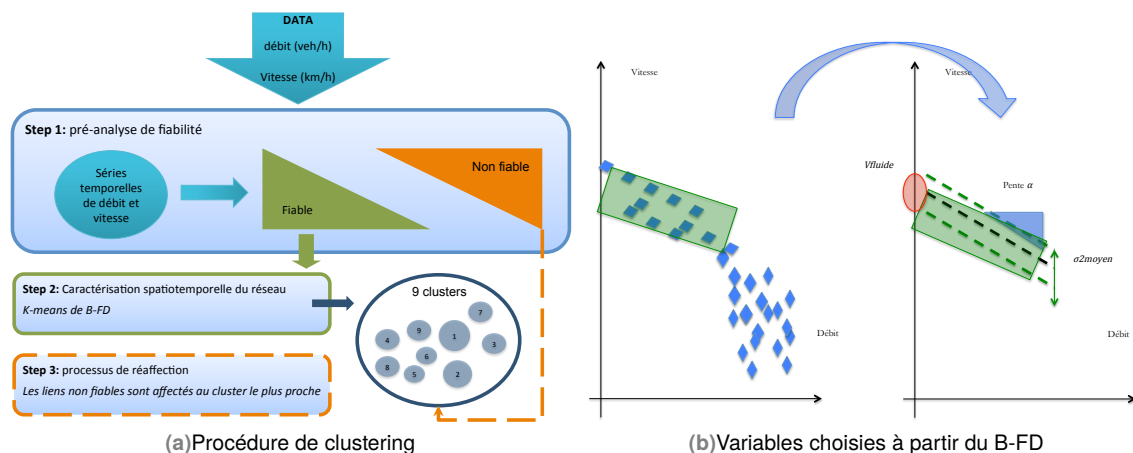


Fig. 3.17.: Méthodologie de clustering de B-FD.

Sur un réseau aussi vaste que celui de Brisbane, une analyse de fiabilité des liens est nécessaire afin de détecter les couples B-OD présentant de trop nombreux problèmes de détection ou de nombreuses données manquantes. La notion de fiabilité se fonde ici sur trois critères appliqués à l'observation des séries temporelles de vitesse et débit pour chaque couple B-OD :

- Le pourcentage de valeurs manquantes sur la période considérée,
- La longueur maximale d'un vecteur de valeurs manquantes consécutives,
- La longueur moyenne d'un vecteur de mesures non interrompues.

Cette méthode permet de classer un sous-ensemble de couples B-OD dans une catégorie dite "non fiable" tandis que le clustering n'est effectué que sur le sous-ensemble de liens fiables (figure 3.17 (a)). L'algorithme des *K-means* est appliqué à trois variables relatives à la forme du B-FD (vitesse libre, pente de l'état fluide, et écart-type moyen de la vitesse), comme illustré sur la figure 3.17 (b). Ces trois paramètres reflètent le fonctionnement de l'infrastructure : la vitesse libre est liée à la limitation de vitesse et donc le type de section, la pente du régime fluide représente une contrainte du trafic sur le lien tandis que l'écart-type des vitesses traduit une homogénéité ou non des conditions de trafic sur le couple B-OD. Une fois que la meilleure partition en K clusters est formée, les couples B-OD non fiables sont réaffectés au cluster le plus proche en fonction de leur proximité aux centres des classes. Comme pour le LDA, la question classique du nombre optimal K de classes doit être traitée avec soin [220]. L'indice de qualité choisi est la silhouette [377], qui mène dans cette étude à un nombre optimal K fixé à 9. clusters. En plus de subtilités qui échappent aux non-spécialistes de la ville de Brisbane, ce nombre de clusters mène finalement à trois grandes catégories de fonctionnement du réseau urbain : des zones très urbaines (clusters 3 et 8), plutôt péri-urbaines (clusters 2, 4, 7), et une dernière catégorie dite tampon (*buffer*) avec quelques spécificités dues à la présence de feux de circulation (clusters 1, 5, 6, 9). Cette interprétation se retrouve sur la figure 3.18, qui permet d'affiner l'hypothèse d'un groupe tampon : ces zones *buffer* sont affectées par deux types de trafic : l'un, urbain et l'autre venant d'autoroutes urbaines ou péri-urbaines comme la Coronation drive ou la Pacific Highway.

3.3.4 Discussion

Un processus complet d'extraction de connaissance a été proposé à partir des données Bluetooth de la ville de Brisbane. Après une phase de qualification de la données, la préparation des variables d'intérêt (*feature engineering*) s'est effectuée à travers l'introduction de deux nouveaux outils inspirés de la théorie du trafic. La matrice Bluetooth Origine Destination (B-OD) permet de caractériser les trajets entre deux détecteurs tandis que le diagramme fondamental Bluetooth (B-OD) est un moyen efficace d'évaluer la performance d'un lien. Ensuite, deux méthodes de clustering ont été explorées dans l'optique de construire une cartographie spatiotemporelle originale du réseau de Brisbane. Les deux typologies peuvent effectivement aider un gestionnaire à mieux comprendre les dynamiques spatiales et temporelles du trafic urbain ou, à défaut, peuvent confirmer de manière automatique, sans connaissance *a priori* les particularités de fonctionnement bien connues des experts. C'est bien dans la mise en exergue de nouvelles structures, sans hypothèse *a priori*, offerte à l'interprétation des gestionnaires d'infrastructures, que de telles approches se destinent. En plus d'outils d'aide à la décision, les résultats permettraient de factoriser les zones d'un réseau et d'agir de façon groupée pour des stratégies de régulation du trafic. Parmi les perspectives du travail, la question de la discrimination des différents modes de transports reste un challenge même si des détecteurs Bluetooth récents semblent capables de donner la classe de l'appareil capté et de déduire le type de véhicule (ex : autoradio = voiture, tout simplement). D'autre part, il apparaît

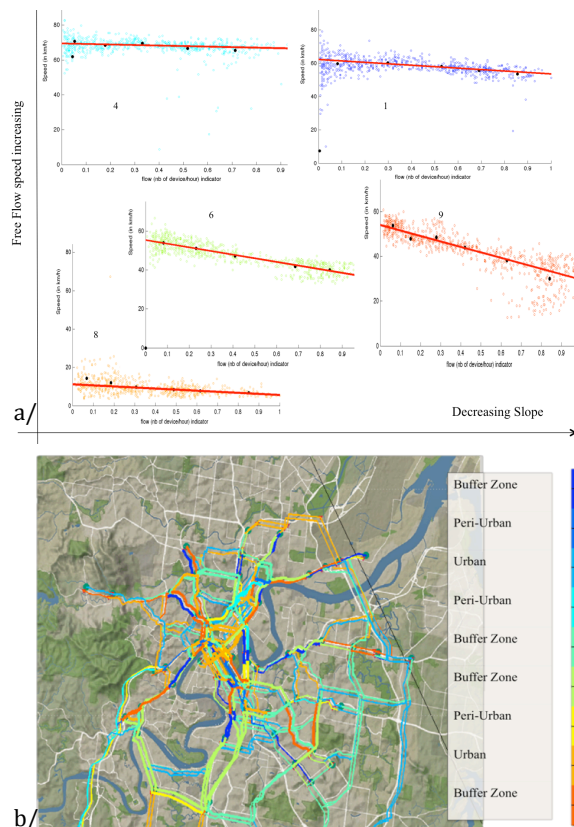


Fig. 3.18.: Illustration des résultats d'un clustering des diagrammes fondamentaux Bluetooth B-FD (Novembre 2011) : (a) formes typiques des B-FD quelques clusters. (b) Représentation spatiale du réseau à partir des clusters de B-FD.

naturel d'utiliser le LDA pour ses capacités prédictives, et notamment dans une optique de détection d'anormalités. C'est par exemple le sujet de notre récent article présenté à la conférence ESANN 2017 [246]. De manière générale, le défi de la prévision court-terme du trafic routier a occupé la suite des travaux en fouille de données. Nos efforts se focalisent sur l'adaptation de méthodes d'apprentissage supervisé pour la prévision des états de trafic à l'échelle d'un réseau [198, 135]. La section suivante résumera quelques unes de nos contributions sur ce thème.

3.4 Prévision court terme du trafic par apprentissage statistique

3.4.1 Utiliser les données pour prévoir le trafic à l'échelle du réseau : verrous et synthèse des contributions

Même si le *big data* n'est pas encore une réalité pour le monde des systèmes de transports intelligents, la disponibilité de nouvelles bases de données doit permettre d'aller plus loin que la simple tâche de supervision et ainsi d'être capable de prévoir à court terme la dynamique du réseau. Les approches centrées sur la modélisation du trafic et la propagation dans le temps de

modèles d'écoulement peuvent être remplacées par des techniques d'apprentissage supervisé qui apprennent des règles de décision sur des historiques de données. Comme le rappelait un article récent sur le sujet [445], le corpus méthodologique relatif à la prévision court-terme du trafic routier souffre de quelques lacunes parmi lesquelles l'intégration des relations spatiotemporelles entre les sections de route dans le modèle prédictif, la prévision unifiée à l'échelle de tout le réseau et la résilience face aux évènements non-récurrents. D'autre part, la liste des méthodes déjà appliquées ou potentiellement applicables est vaste, et il reste à déterminer les conditions dans lesquelles une classe de méthodes (ex : SVM vs. réseaux de neurones) serait plus adéquate qu'une autre. Ces différents verrous ont guidé nos recherches, qui ont cherché à proposer des **modèles parcimonieux de prévision de trafic fondés sur l'apprentissage statistique des historiques de données**. Au cours des dernières années, nous avons proposé plusieurs contributions résumées comme suit :

1. L'élaboration d'une approche globale pour la prévision à l'échelle d'un réseau urbain. Au lieu d'apprendre et prévoir section par section, un historique d'images du réseau sert à prévoir une image future du réseau.
2. L'adaptation de méthodes parcimonieuses (ex : plus proches voisins) à ce nouveau cadre prédictif.
3. L'évaluation des performances de la nouvelle approche par rapport à un *benchmark* de méthodes de l'état de l'art, fondée sur deux jeux de données différents : les données Bluetooth de Brisbane et les données FCD de la ville de Nantes.
4. Une méthode d'identification des sections critiques, sous-ensemble de sections suffisant pour obtenir de bonnes performances en généralisation.
5. Des *patches* additionnels au cadre de base pour la détection d'incidents et l'intégration des conditions météorologiques.

L'approche globale proposée s'inscrit véritablement dans un cadre *Multi-Input Multi-Output (MIMO)* et va un cran plus loin que les approches fondées sur les séries spatiotemporelles (voir [191]). Nous désignerons par le terme "approches locales" les différentes alternatives. La figure 3.19 illustre la différence entre une approche par séries spatiotemporelles et notre approche globale, en termes d'entrées et de sorties du modèle.

Il est évident que des questions satellites sous-tendent le genre d'approches abordé. En premier lieu, la prise en compte des données manquantes a mobilisé de longs efforts, comme dans tout processus de fouille de données. À partir de l'historique d'apprentissage, des profils nominaux de caractéristiques de trafic ont été construits, pour chaque type de journée (ex : profil mardi, profil week-end, profil jour de congés). Ensuite, ces profils sont intégrés aux séries temporelles par lissage. Nous renvoyons le lecteur aux travaux de Haworth pour des contributions originales sur ce thème [191, 192]. D'autre part, la question de la délimitation optimale du réseau n'a pas été traitée et les cas d'études ont été construits de façon empirique. De même, d'autres travaux abordent le problème de façon approfondie [200, 344].

3.4.2 Présentation de l'approche globale par l'exemple des k plus proches voisins

Nous illustrons ici l'approche globale à travers l'exemple de la méthode des k plus proches voisins (ou kNN pour k *Nearest Neighbours*), qui s'est révélée séduisante de par sa simplicité, et tout

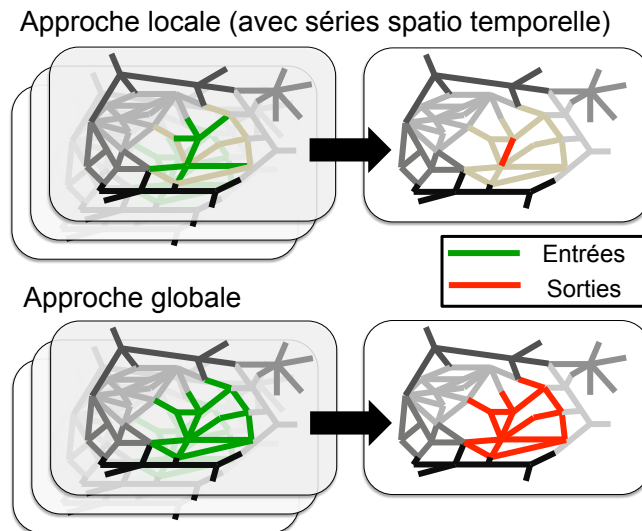


Fig. 3.19.: Différence entre une prévision fondée sur des séries spatio-temporelles (approche locale) et une approche globale.

à fait efficace en terme de performances. Deux autres méthodes ont été adaptées à l'approche globale : les méthodes à noyaux et la régression par machines à vecteurs supports (*Support Vector Regression (SVR)*). Les performances des trois méthodes par rapport à des méthodes de l'état de l'art seront résumées à la section suivante. Le premier cas d'études est le réseau Bluetooth de Brisbane, dans la lignée de la section précédente. Nous adaptons la méthode des k plus proches voisins à une version multidimensionnelle capable de prédire une image du réseau de Brisbane. Cette image peut être vue comme une photographie instantanée de la ville à un instant t , c'est-à-dire les valeurs des états de trafic sur toutes les section Bluetooth considérées. Le vecteur d'état à prédire s'écrit

$$\mathbf{X}_t = (X_{1,t}, \dots, X_{n,t})^T$$

où n est le nombre de segments de route Bluetooth considérés. Pour le cas de Brisbane, le vecteur \mathbf{X}_t est composé de $n = 466$ valeurs de débits sur les 466 segments, plus précisément le nombre de détections pendant une heure t pour ces 466 segments. En d'autres termes, l'image à prédire à un temps donné t correspond à tous les états de trafic (nombre de détections) sur tous les segments pendant l'heure passée. Le problème de prévision de trafic consiste à prévoir l'état du réseau au temps $t+h$, noté $\widehat{\mathbf{X}}_{t+h}$, où h est l'horizon de prévision, soit

$$\widehat{\mathbf{X}}_{t+h} = f(\mathbf{X}_t)$$

où f est la fonction de prévision relative à la méthode choisie, ici les k plus proches voisins. L'accent est mis sur la prévision court-terme du trafic routier et l'horizon h n'excèdera pas les trois heures. **La fonction f est dans notre cas une fonction de régression** car il s'agit de prédire des valeurs numériques continues, le nombre de détections sur tous les segments à l'horizon $t+h$. La technique des k plus proches voisins constitue une méthode de reconnaissance des formes qui va aller chercher des profils similaires dans l'historique de données afin de prédire des états futurs [135]. Plus précisément, la méthode prédit une cible numérique à partir d'une métrique entre le prototype

considéré (l'image du réseau) et ses k plus proches voisins dans la base historique, comme le montre la figure 3.20.

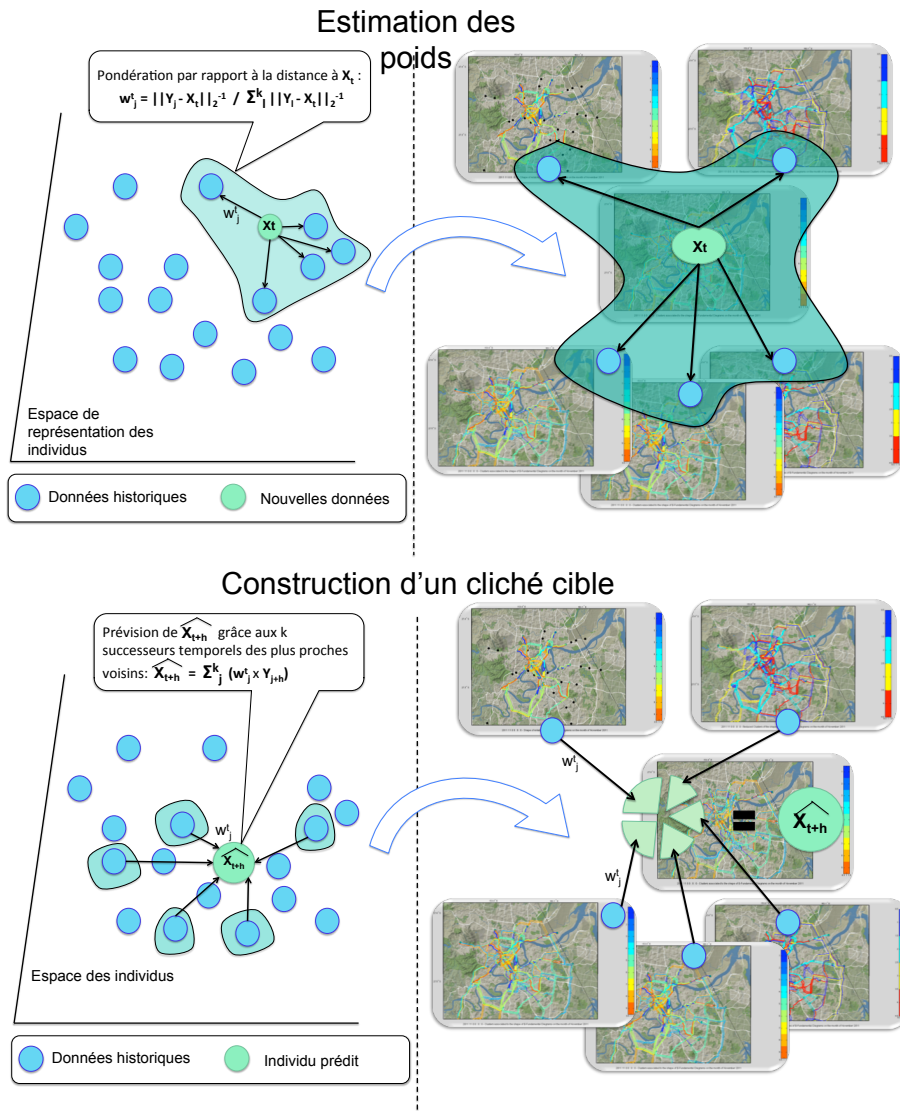


Fig. 3.20.: La méthode des k plus proches voisins adaptée à l'approche globale.

Cette figure montre l'adaptation de l'algorithme original pour le problème de régression à l'échelle du réseau. Une moyenne pondérée des états futurs des k plus proches voisins d'un individu statistique au temps t permet de générer un individu prédit au temps $t+h$. Les différentes étapes du processus sont les suivantes :

1. **Données :** les entrées sont constituées d'une image, prototype X_t à prédire, et d'un ensemble d'apprentissage constitué d'images-prototypes Y_l . Chaque prototype historique Y_l représente une cartographie des états de trafic à un temps l . Aucune relation n'est imposée entre t et l .
2. **Estimation des poids :** elle est fondée sur la distance euclidienne $\|Y_l - X_t\|_2$ et les poids des k plus proches prototypes sont calculés comme suit :

$$w_j^t = \frac{\|Y_j - X_t\|_2^{-1}}{\sum_{l=1}^k \|Y_l - X_t\|_2^{-1}}, j \in 1, \dots, k.$$

3. **Profils cibles** : les successeurs temporels des k plus proches voisins sont utilisés pour construire une moyenne pondérée. Ainsi, l'image prédite, c'est-à-dire l'individu statistique estimé, noté $\widehat{\mathbf{X}}_{t+h}$ est écrit

$$\widehat{\mathbf{X}}_{t+h} = \sum_{j=1}^k (w_j^t \times \mathbf{Y}_{j+h}).$$

La méthode a été déclinée dans une version "sans mémoire" : étant données les valeurs de débits sur tous les segments à un temps t , qui suffisent à définir le vecteur d'état \mathbf{X}_t à prédire, des images similaires sont recherchées dans l'ensemble d'apprentissage, et ce sans considération des images passées, c'est-à-dire des états $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}$ etc. Ajouter une propriété de mémoire est une extension possible qui a été implémentée et incluse dans l'évaluation des performances. Néanmoins, cette extension entraîne l'ajout d'un paramètre s relatif à la taille de la mémoire, en plus du paramètre k correspondant au nombre de voisins pris en compte.

3.4.3 Processus d'apprentissage

Un processus d'apprentissage rigoureux a été mis en place afin de garantir des performances satisfaisantes en généralisation, c'est-à-dire sur de nouvelles situations de trafic à prédire. Pour le réseau Bluetooth de Brisbane, le jeu de données final est limité à 31 jours. En utilisant un pas de temps d'une heure, chaque jour de données est constitué de 24 images du réseau. Ainsi, 744 images des états de trafic sur les 466 segments du réseau sont disponibles. L'ensemble des données a été divisé en deux sous-ensembles : un ensemble d'apprentissage pour déterminer les paramètres du modèle et un ensemble de tests pour évaluer les performances. Dans le cas d'études australien, 24 jours ont été utilisés pour la phase d'apprentissage tandis que les 7 derniers ont été utilisés pour la phase de test. Les paramètres à estimer sont la valeur k du nombre de voisins et, éventuellement, un paramètre de mémoire s . Comme pour toute méthode d'apprentissage supervisé, le processus vise à obtenir un processus satisfaisant entre biais et variance, comme l'illustre la figure 3.21. Il s'agit d'éviter les situations de sous-apprentissage ou sur-apprentissage en utilisant des modèles parcimonieux et en structurant intelligemment les données disponibles.

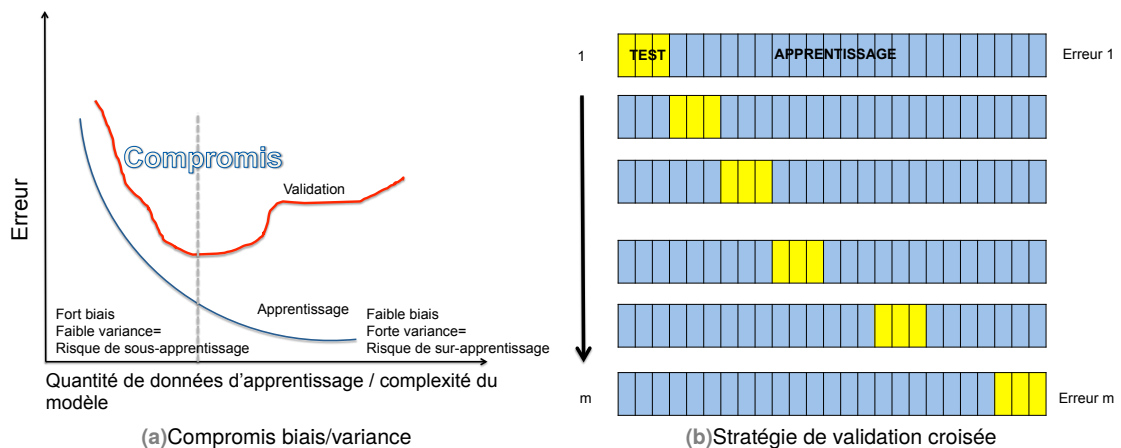


Fig. 3.21.: Processus d'apprentissage.

En ce qui concerne l'ajustement des paramètres, une méthode de validation croisée a été employée (voir figure 3.21). L'ensemble d'apprentissage est divisé en m sous-ensembles et la méthode apprend ensuite m fois sur les m sous-ensembles, en gardant à chaque fois un sous-ensemble différent

pour la validation [135]. Cette méthodologie de type *leave-one-out* mène aux valeurs optimales des paramètres k et s .

3.5 Application au cas d'étude de Brisbane

3.5.1 Performances générales

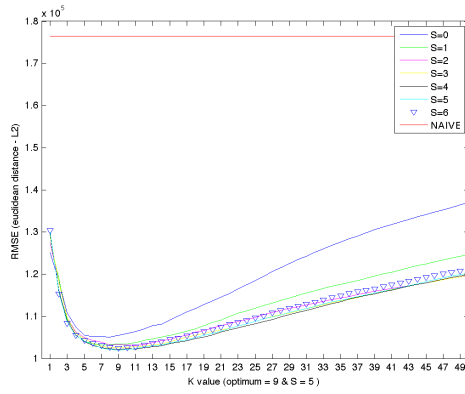
Dans un premier temps, nous évaluons classiquement les performances de l'approche globale à l'aide d'erreurs de prévision. Deux types d'erreurs sont introduites :

- **L'erreur globale** : c'est la racine carrée de l'erreur quadratique moyenne (RMSE) entre l'image prédite et l'image réelle, notée $\|\widehat{\mathbf{X}}_{t+h} - \mathbf{X}_{t+h}\|_2$.
- **L'erreur locale** : elle est évaluée au niveau de chaque segment et est notée $|\widehat{X}_{k,t+h} - X_{k,t+h}|_1, k = 1, \dots, n$.

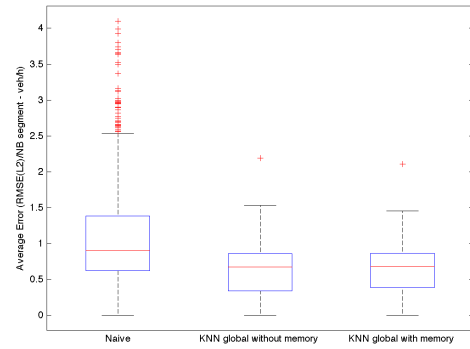
Pour étayer les conclusions, deux méthodes de référence sont introduites, avec lesquelles les versions globales des k plus proches voisins (kNN), sans mémoire et avec mémoire s , seront comparées. Un ensemble de quatre méthodes est donc évalué sur les données test de Brisbane :

1. **Méthode naïve** : pour chaque segment, la valeur observée au temps t est recopiée au temps $t+h$,
2. **kNN local avec mémoire** : pour chaque segment, une règle des k plus proches voisins est appliquée avec l'ajustement de deux paramètres : le nombre k et la taille de la mémoire temporelle s ;
3. **kNN globale** : à l'échelle du réseau, une image du réseau \mathbf{X}_t est utilisée pour prévoir l'état suivant au temps $t+h$, sans aucune information sur le passé. Un seul paramètre, k , doit être optimisé ;
4. **kNN global avec mémoire** : à l'échelle du réseau, une image du réseau \mathbf{X}_t est utilisée pour prévoir l'état suivant au temps $t+h$, avec une taille de mémoire s relative au nombre de pas de temps précédents devant être pris en compte. k et s doivent être estimés une fois.

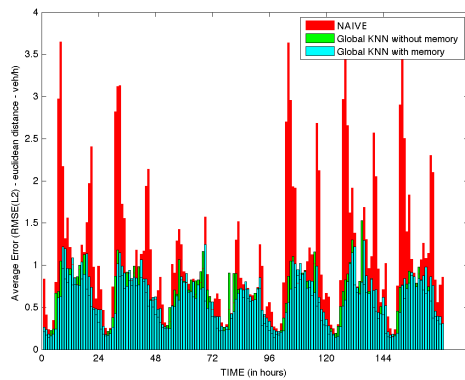
Tout d'abord, la figure 3.22 (a) présente la détermination des valeurs de k et s pour l'approche kNN globale. Une stratégie d'exploration par grille fait varier les valeurs des deux paramètres selon plusieurs plages de valeurs possibles. En appliquant la validation croisée précédente, pour un horizon de prévision d'une heure, l'approche globale sans mémoire nécessite la prise en compte de $k = 5$ plus proches voisins tandis que l'introduction d'une mémoire est optimisée à travers la prise en compte de 5 pas de temps passés ($s = 5$) et 9 plus proches images ($k = 9$). Ces valeurs doivent être réajustées quand l'horizon de prévision varie. La figure 3.22 (b) présente les distributions générales du RMSE. À ce stade, l'apport des deux approches kNN globales sur une prévision naïve est évident. Cependant, l'introduction d'une mémoire temporelle offre une distribution d'erreur comparable, pour un paramètre de plus à estimer. Cette tendance est confirmée par la troisième sous-figure qui montre l'évolution de l'erreur moyenne pour les sept jours de tests. Ce résultat est important quand on sait que l'approche kNN sans mémoire est la plus simple et puissante en termes de complexité temporelle. Cette erreur varie forcément en fonction du niveau de service. La figure 3.22.d montre la répartition des erreurs pour 6 classes différentes de niveau de service. À cette fin, six groupes de segments B-OD ont été définis en fonction des niveaux de débits. Par exemple, le



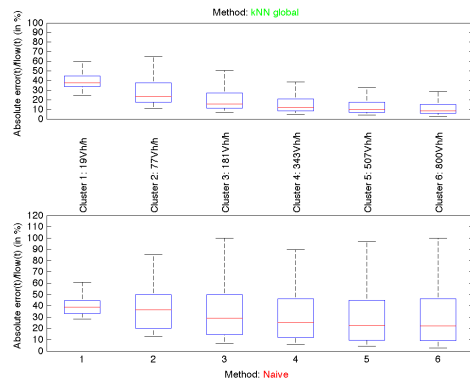
(a) Sélection des paramètres k et s



(b) Distribution de l'erreur RMSE par lien



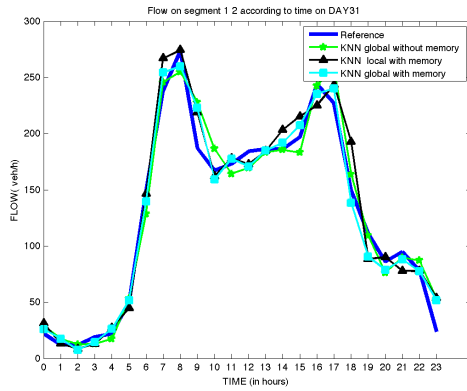
(c) Diagrammes en bâtons de l'erreur RMSE par périodes de la journée



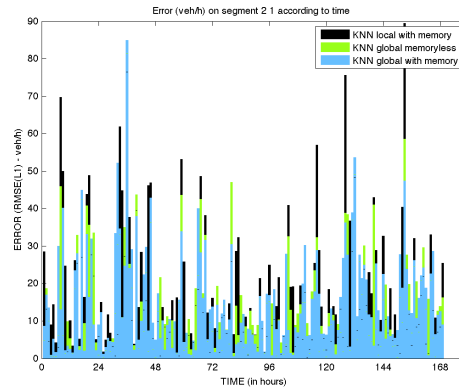
(d) Impact du niveau de service sur l'erreur de prévision

Fig. 3.22.: Evaluation de l'approche kNN globale.

premier groupe contient les segments avec un maximum de 19 détections par heure tandis que le 6ème contient des segments recensant, de façon médiane, 800 véhicules par heure. On remarque en premier lieu que l'approche globale (ici sans mémoire) se distingue particulièrement dans des situations à fort débit, se comportant ainsi efficacement pour les forts niveaux de service. Une fois ces premiers résultats introduits, il convient de comparer notre contribution en terme d'échelle, globale (*network-wide*), *versus* locale (*link-based*). Le premier avantage est la complexité temporelle. Rien que pour la phase d'optimisation des paramètres, un facteur 7 d'accélération a été mesurée en faveur de l'approche globale. Pour la phase de prévision, une expérimentation ultérieure sur 1 mois de données Bluetooth a permis de conclure à un temps de calcul de 14.9 secondes pour l'approche globale contre 2854 secondes pour une méthode kNN locale. En termes de performances, les résultats montrent que l'approche kNN globale apporte une plus-value dans de nombreux cas, en particulier les heures de pointes, comme le souligne la figure 3.23. La faculté de l'approche à performer dans des situations denses ou congestionnées est critique quand on sait que les périodes de pointe concentrent l'attention des gestionnaires, et l'intérêt des usagers.



(a) Performances journalières

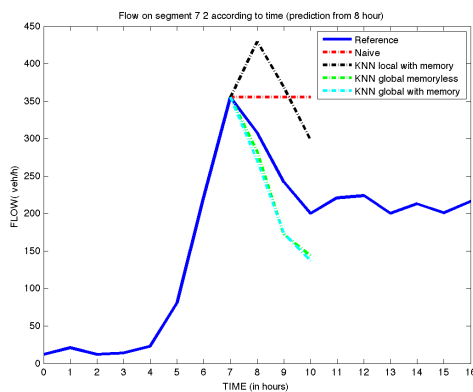


(b) Diagrammes en bâtons des erreurs locales

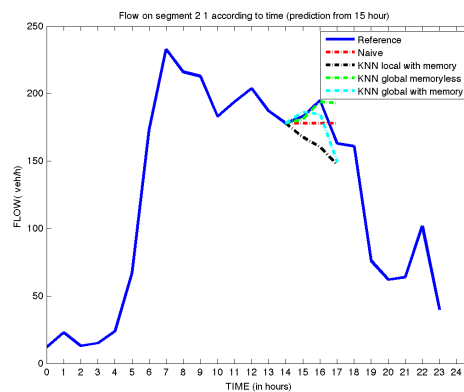
Fig. 3.23.: Approches locale vs. globale : comparaisons sur un segment-témoin.

3.5.2 Résilience à l'horizon de prévision et aux incidents

En plus de la complexité temporelle, nous avons analysé la résilience de l'approche face à différents horizons de prévision ainsi que face à des évènements non récurrents (incidents). La figure 3.24 compare les prévisions pour différentes valeurs de h . Le temps est stoppé à deux moments précis (8 heures du matin à gauche et 15h à droite) à partir desquels les valeurs futures sont prédites pour différents horizons. Une dégradation de la performance est attendue au fur et à mesure que l'horizon de prévision croît.



(a) Prédiction à partir de 8 heures du matin



(b) Prédiction à partir de 15h

Fig. 3.24.: Prédiction des états de trafic à partir d'une heure choisie à l'avance, pour les horizons successifs fixés respectivement à $t + 1$, $t + 2$, $t + 3$.

La figure montre de meilleures performances pour les deux variantes à l'échelle du réseau : pour le segment B-OD choisi, les prévisions à trois heures sont plutôt proches de la réalité tandis que les autres méthodes échouent quand les valeurs de h augmentent. Ceci pourrait s'expliquer par la structure multidimensionnelle de notre méthode qui capture la corrélation spatiale entre les segments du réseau. C'est pourquoi l'ambition est également d'atteindre une meilleure résilience face à des évènements non-récurrents, hypothèse confirmée par les résultats de la figure 3.25. Une situation particulière a été identifiée dans la base de données : à 22h, un incident survient sur un segment

B-OD, résultant en une période de congestion inhabituelle. Cet évènement se propage sur les liens adjacents (vers l'amont) au pas de temps suivant.

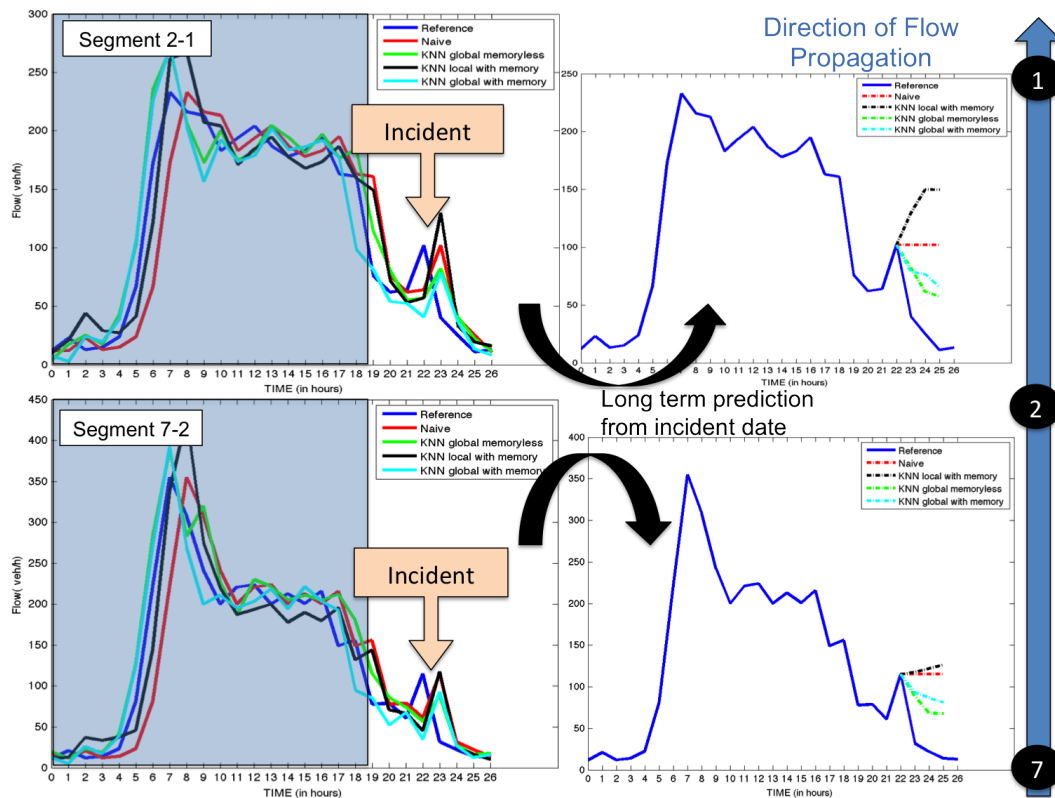


Fig. 3.25.: Résilience de la méthode kNN globale aux incidents.

En lançant les algorithmes de prévision de trafic à partir de l'heure de l'incident, 22h, on remarque que les méthodes locales produisent des estimations défailtantes, la congestion et sa propagation n'étant pas prévues par le modèle. Les méthodes usuelles prévoient un profil de débits classique, ce qui mène à une importante erreur qui peut avoir des conséquences opérationnelles non négligables. En incluant la corrélation spatiale entre les segments, la propagation est mieux captée et propagée par l'approche globale. Ce type de résultats milite pour un raffinement des méthodes de prévision court-terme à des visées opérationnelles. Néanmoins, les données Bluetooth sont souvent affectées par des variations de qualité de détection, des erreurs menant à des taux de pénétration instables. Elle n'offrent pas un cadre de comparaison assez stable pour généraliser plus en avant les performances des méthodes introduites dans cette section. C'est pourquoi un autre cas d'étude a été envisagé via les données FCD de la ville de Nantes.

3.6 Résumé des autres résultats

3.6.1 Comparaison à d'autres méthodes sur le cas nantais, données FCD. Vers une approche portfolio ?

L'approche globale a également été mise en place sur le réseau nantais, en particulier la section "à problème" du périphérique nantais à l'approche du Pont de Cheviré. Le réseau est composé de

177 sections. Trois mois de données ont été mis à disposition par notre partenaire MEDIAMOBILE : septembre, octobre et novembre 2013. **La variable d'intérêt change car des vitesses harmoniques des véhicules sont recueillies avec un pas de temps de 30 minutes.** Les horizons de prévision s'étendent de 30 minutes à 3 heures. En plus de nouvelles données, un effort a été porté sur l'adaptation de nouvelles méthodes de prévision au cas global : les méthodes à noyaux et la régression par machines à vecteurs supports (SVR). Ces deux dernières méthodes sont supposées plus "raffinées" que l'approche par k plus proches voisins. La figure 3.26 synthétise des résultats pour le cas nantais.

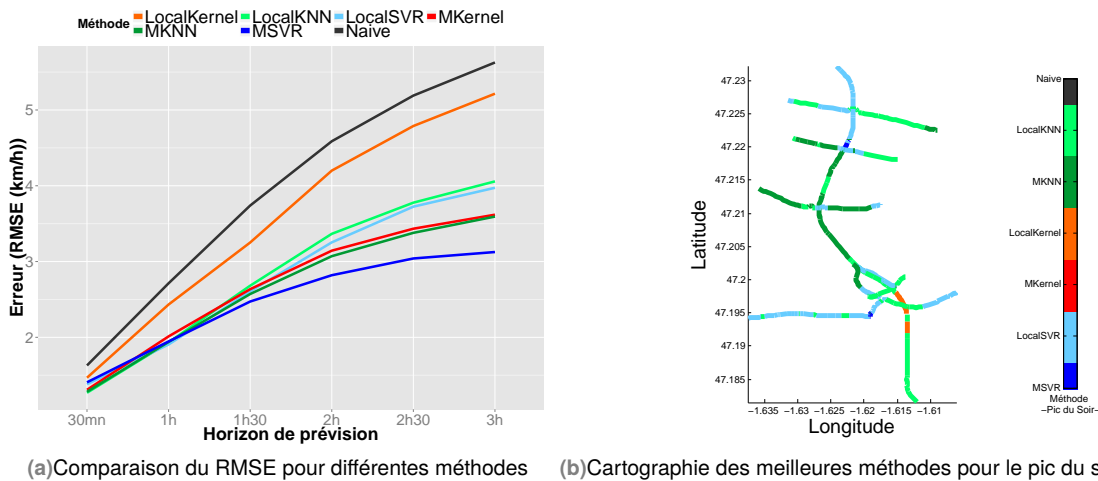


Fig. 3.26.: Performances des méthodes sur le cas de Chevirié, Nantes.

Les conclusions sont sensiblement les mêmes quant aux performances des méthodes globales *versus* locales. Lorsque l'horizon de prévision grandit, la méthode SVR produit de meilleurs résultats que les autres. Toutefois, ces résultats doivent être nuancés en fonction des zones et des conditions de circulation. La figure 3.26 montre une cartographie des meilleures méthodes par segment, pour le pic du trafic du soir. À cette heure de la journée, l'approche kNN globale truste les meilleures performances au côté de simples régressions locales type kNN ou SVR. En analysant plus finement d'autres zones du réseau nantais, les conclusions suivantes peuvent être dressées :

- Les sections les plus difficiles sont mieux traitées par les approches globales kNN et SVR ;
- Les approches locales et à noyaux prévoient mieux les zones mal représentées (données manquantes, relations spatio-temporelles partielles),
- Les approches globales produisent de meilleures performances, avec un avantage pour le SVR global lorsque l'horizon h grandit.

Plus récemment, des analyses ont été effectuées pour comparer les performances de nos approches face à des méthodes populaires dans les concours de *machine learning*, comme les *random forest* ou le *gradient boosting machine*. Ces deux dernières produisent de bons résultats, comparables aux kNN et SVR globaux, qui varient néanmoins en fonction de l'heure de la journée. Ces disparités entre performances motivent des perspectives de recherche sur une approche portfolio, qui permettrait de sélectionner la meilleure méthode d'apprentissage en fonction de critères particuliers comme l'heure de la journée, le type de section, l'horizon, etc. C'est notamment l'objectif poursuivi par Julien Salotti (LIRIS, INSA Lyon) dans ses travaux de thèse [380].

3.6.2 Identification de sections critiques

Parmi les développements liés au cadre prédictif global, un processus de sélection de variables original a été mis en place afin de sélectionner un sous-ensemble de sections B-OD dites "critiques", c'est-à-dire un sous-ensemble de liens B-OD nécessaires pour obtenir de bonnes performances en généralisation. La méthode propose de classer les liens selon leur criticité, qui est vue ici comme leur propension ou non à propager, ou catalyser, une perturbation en amont ou en aval. Cette hypothèse nous a amené à proposer une définition de la configuration locale d'une section puis à identifier, à partir d'un processus génératif fondée sur le LDA, une configuration critique nominale (dominante). Les sections sont ensuite classées par rapport à leur contribution à cette configuration nominale critique. Comme le montre la figure 3.27, il est remarquable de voir qu'une telle méthode de sélection de variables permet de passer des 177 sections de départ à une trentaine de sections critiques. La figure 3.27 montre un aperçu de la notion de configuration locale sans aller plus en avant dans les détails par souci de concision.

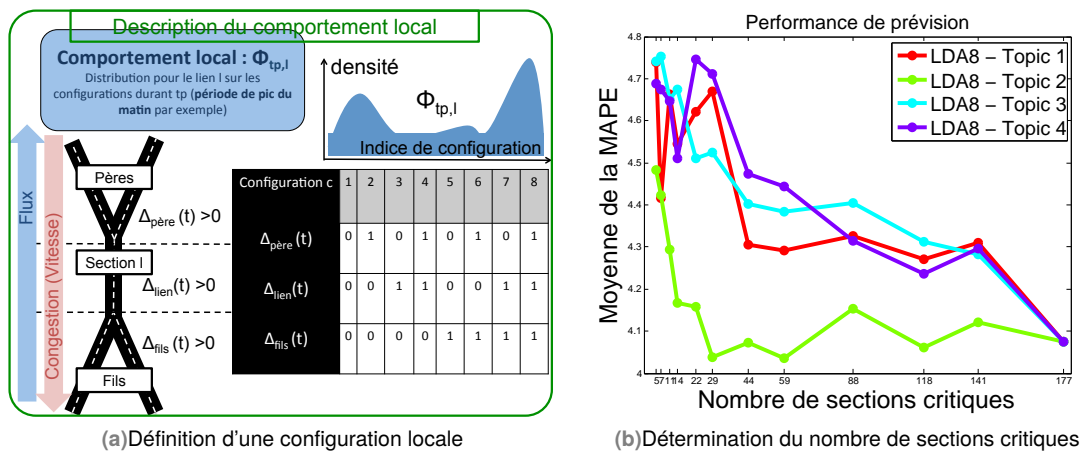


Fig. 3.27.: Identification des sections critiques. Les Δ représentent les variations du signal de vitesse sur les liens adjacents en tenant compte de la physique spatiotemporelle du trafic routier.

3.6.3 Traitement des externalités

Parmi les autres contributions non détaillées dans ce manuscrit d'HDR, nous aimerions mentionner deux facteurs exogènes du trafic qui représentent de forts enjeux pour la gestion temps-réel du réseau. Les incidents et les conditions météorologiques dégradées. Une première contribution a porté sur la détection temps-réel d'incidents. Nous avons proposé une classification non supervisée des états de trafic qui permet de bien discriminer les évènements non récurrents des autres perturbations. À partir de la détection des évènements non récurrent, la base d'apprentissage peut ensuite être enrichie afin d'intégrer véritablement cette information dans l'historique d'apprentissage. Cette contribution a été valorisée dans la conférence ESANN 2017 [246]. Ce type de stratégies peut être suivi pour l'intégration des conditions météorologiques dans les modèles de prévision. Des données locales obtenues par Météo France pour la ville de Nantes nous ont permis d'envisager un moteur de prévision météo-sensible. Néanmoins, les différentes agrégations, discrétisations, des informations trafic et météorologiques ne permettent pas de quantifier assez finement l'effet météorologique sur le réseau nantais. À moins que le manque de données pluvieuses sur la ville n'en soit la cause, ce qui serait, pour le choix de la ville de Nantes, un comble ! Quelques preuves de

concepts ont toutefois permis de montrer une accentuation de l'erreur de prévision lors de conditions météorologiques pluvieuses.

3.7 Conclusion et perspectives

La massification des données trafic ouvre la voie à des méthodes statistiques robustes qui ont vocation à intégrer les outils d'aide à la décision pour la gestion du trafic. Loin de plaider pour le remplacement de la théorie du trafic, plutôt fondée sur les modèles (*model driven*) par le *big data* et les techniques *data driven*, ce chapitre soulève le besoin de passerelles entre les deux communautés, et appelle au développement de méthodes hybrides. Dans le domaine de la modélisation statistique des temps de parcours, c'est l'intuition d'un nouveau modèle, le système Halphen, qui a pu être validée grâce à des données suffisamment nombreuses pour en attester la pertinence. La modélisation et la simulation de trafic pourraient intervenir pour générer des données synthétiques sur un itinéraire et évaluer l'additivité des lois ou la propagation de l'incertitude sur une suite de segments. En ce qui concerne la supervision temps-réel du trafic, le taux de pénétration fluctuant (et donc la représentativité partielle des données) ne permet pas d'obtenir une exhaustivité à propos de la couverture du réseau. Là encore, la simulation de trafic à partir du *macroscopic fundamental diagram* pourrait compléter les informations manquantes sur le réseau. Il en est de même en prévision court terme du trafic routier où les facteurs exogènes comme les incidents et conditions météorologiques pourraient être modélisés par des modèles de propagation afin d'inclure leurs impacts aux sections voisines au temps $t + h$.

En revisitant l'ensemble du chapitre qui s'achève, on peut distinguer deux angles de vue différents : tout d'abord, la modélisation des temps de parcours par le système Halphen propose une approche véritablement statistique, où la donnée est supposée suffisamment pertinente pour pouvoir se concentrer sur la recherche du meilleur modèle possible. Ensuite, l'exemple "fil rouge" des données Bluetooth de la ville de Brisbane suit un processus typique de reconnaissance de forme ou *pattern recognition*, tel que présenté par Duda dans son célèbre ouvrage [135]. Après une phase de capture (*sensing*) des données à l'aide d'instruments Bluetooth, où la donnée est qualifiée et corrigée, la phase de **segmentation** prépare les données en créant des structures B-OD et B-FD. Les prototypes considérés sont les segments Bluetooth sur lesquels les variables d'intérêt (*features*) ont été **extraites**. La variable la plus présente a été le débit (nombre de détections par heure) mais il aurait été tout à fait possible de travailler sur les vitesses ou temps de parcours, comme le montre d'ailleurs le cas d'étude nantais avec les données FCD. La méthode de **prévision** a ensuite été ajustée sur un ensemble d'apprentissage et à l'aide d'un procédé de validation croisée pour l'estimation des paramètres optimaux. L'approche globale proposée constitue finalement une méthode simple et efficace de prévision à l'échelle d'un réseau, qui permet de capturer toute la structure spatiotemporelle des sections à prédire. Cette classe d'approche, déclinée pour les régressions par k plus proches voisins, par noyaux, ou par machines à vecteurs supports (SVR), propose de très bonnes performances et des temps de calcul beaucoup plus intéressants que les méthodes locales qui raisonnent segment par segment. Néanmoins, la confrontation à un ensemble de méthodes de l'état de l'art suggère d'approfondir la question d'une approche port-folio, qui serait capable de sélectionner la méthode la plus performante en fonction d'un ensemble de paramètres à définir.

Optimisation des itinéraires sous contraintes complexes

” *L'humour est le plus court chemin d'un homme à un autre*

— **Georges Wolinski**
Dessinateur

Ce chapitre présente un autre pan de nos recherches dédié à l'optimisation dans les réseaux de transports sous contraintes complexes, où l'on voit bien que les résultats des recherches précédentes, sur la simulation des C-ITS, la construction d'indicateurs de supervision et prévision de trafic, ont vocation à nourrir en entrée les algorithmes. Les vitesses ou temps de parcours, obtenus à partir de simulations ou plus probablement à partir de nouvelles sources de données, estimés ou prédits, constituent les valuations des arcs d'un graphe où les nœuds sont les intersections et les arcs les sections de route entre intersections. Nous déplaçons ici le problème vers la recherche du plus court chemin entre deux points, dans un tel graphe. Nous choisissons de commencer ce chapitre par la présentation d'un projet, TEMPUS, qui a rythmé et orienté quelques unes de nos recherches au cours des dernières années (section 4.1). Ce calculateur d'itinéraires, aujourd'hui *open source*, est disponible pour la communauté scientifique et les industriels, pour tester et comparer des nouveaux algorithmes. Les contributions présentées dans la suite du chapitre, portant sur l'électromobilité (section 4.2), l'optimisation dans les graphes stochastiques et son accélération (section 4.3), pourront être intégrées dans une telle plateforme, une fois atteint un niveau de maturité scientifique suffisant. En termes d'encadrement de la recherche, les thèses de doctorat de Fouad Baouche et Raphaël Delhome ont structuré les activités entre 2012 et 2016.

4.1 La plateforme TEMPUS : quand les besoins industriels inspirent les travaux de recherche.

4.1.1 Présentation de TEMPUS, calculateur d'itinéraires multimodaux

TEMPUS est une plateforme *open source* de calculs de plus courts chemins qui a été développée et enrichie depuis 2007, année marquant le début de la thèse d'Aurélié Bousquet sur l'optimisation des itinéraires dans les réseaux multimodaux. Orientée recherche, son but est de servir comme plateforme de comparaison de performances d'algorithmes de PCC, dans un contexte multimodal voire multi-objectif. Développée dans une logique de modularité et de généricité du code, elle est conçue comme une extension de la bibliothèque C++ Boost et est interfacée avec le logiciel d'information géographique QGIS ainsi que le système de gestion de bases de données PostgreSQL, tous deux sous licence *open source*. La plate-forme s'accompagne également d'un module de chargement

automatisé des données de réseaux de transport aux formats les plus courants permettant une mise en œuvre très aisée sur tout territoire qui dispose de ce type de données. L'outil a été conçu dans une logique orientée « *plugin* », chaque utilisateur pouvant intégrer un nouvel algorithme de calcul de plus court chemin écrit en C++. L'interface graphique a elle-même été développée comme une extension python de QGIS. La comparaison des performances et des résultats des algorithmes sur différents territoires est ainsi facilitée. La figure 4.1 présente le modèle conceptuel de l'outil.

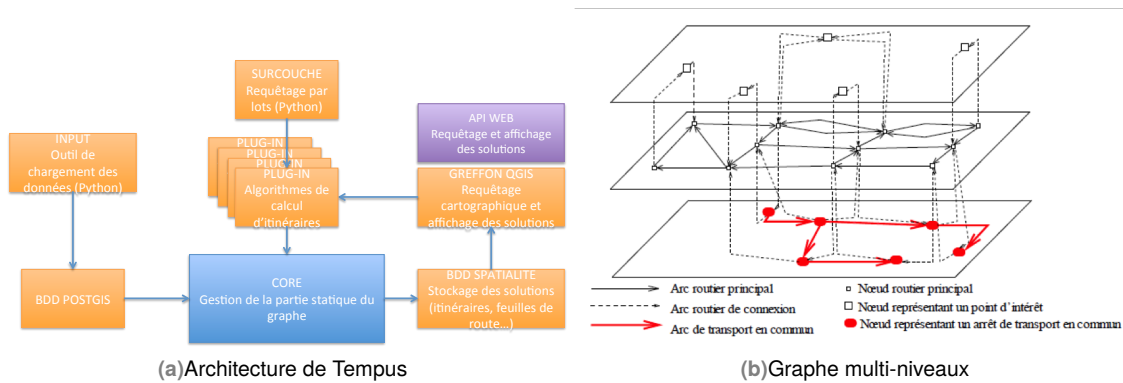


Fig. 4.1.: TEMPUS : modèle conceptuel.

La plateforme TEMPUS est un calculateur d'itinéraires porte-à-porte **entièrement multimodal**. Contrairement à beaucoup de calculateurs existants, le mot multimodal englobe ici également l'usage du véhicule particulier, qui peut être sélectionné ou garé au cours du trajet. La structure de données principale de TEMPUS est un graphe multimodal de type multi-niveaux (figure 4.1), composé du réseau routier, des graphes de transports en commun et de différents points d'intérêt (vélo en libre-service, auto-partage, etc.). Un loader de données fournit un ensemble de scripts permettant d'importer ces données à des formats divers (NavTeq, Tom Tom MultiNet, OSM, GTFS). L'implémentation des graphes revêt ensuite un haut niveau de généricité grâce à l'utilisation des potentialités de la BGL (*Boost Graph Library*). Les premiers algorithmes implémentés sont un Dijkstra et un A* multimodaux, avec prise en compte des mouvements tournants interdits sur le réseau. À travers l'interface QGIS, l'utilisateur peut sélectionner graphiquement une origine et une destination, régler les options de l'algorithme, puis apprécier la feuille de route ainsi que certains critères, comme l'altimétrie (figure 4.2).

Une surcouche de requêtage par lots permet de multiplier les requêtes et de mieux tester le passage à l'échelle des algorithmes ainsi que la performance globale de l'outil. Des premiers résultats ont été proposés à l'échelle de la ville de Nantes et, plus récemment, sur la ville de Bordeaux. Les données ouvertes des différentes agglomérations peuvent être intégrées à l'outil. La figure 4.3 montre l'exemple d'une solution proposée à l'utilisateur dans laquelle le trajet commence par l'usage d'un vélo en libre-service pour se terminer par une séquence en tramway.

L'incorporation de la voiture comme mode de transport à part entière permet d'intégrer des fonctionnalités que peu de calculateurs au monde peuvent offrir : par exemple, il est possible de préciser l'endroit où la voiture est garée, si un garage est disponible ou non à destination etc. TEMPUS se veut avant tout une boîte à outils de comparaison d'algorithmes. À titre d'exemple, la figure 4.4 illustre l'accélération permise par l'heuristique A* par rapport à l'algorithme de Dijkstra classique, en terme de nombre de nœuds explorés. Ce type de scénarios, qui possède des vertus pédago-

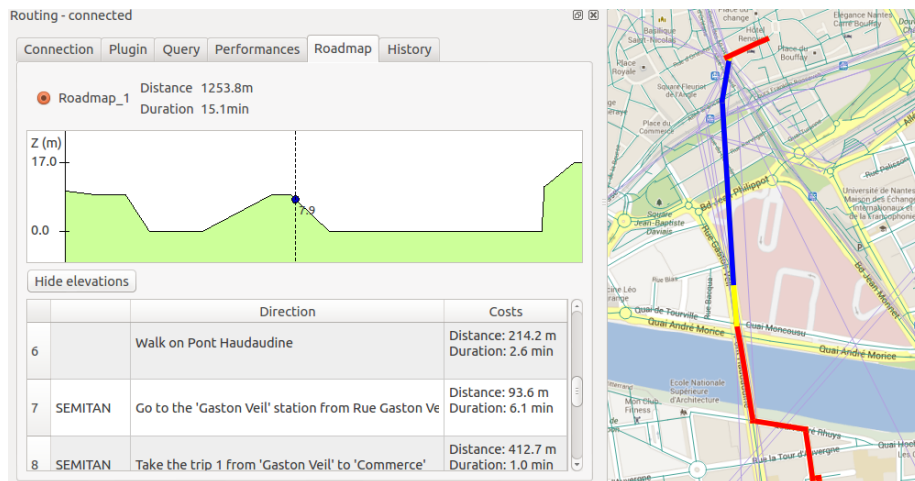


Fig. 4.2.: Interface de TEMPUS.

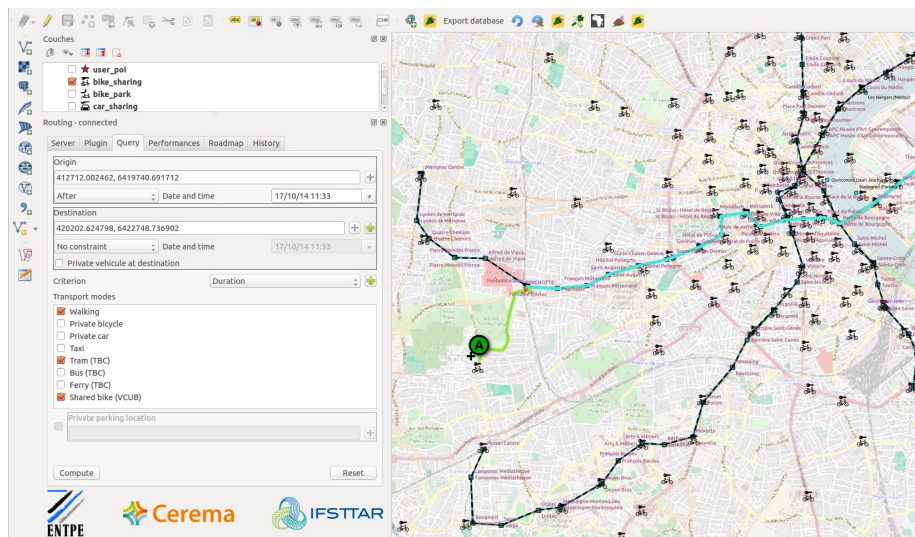


Fig. 4.3.: TEMPUS : exemple de requête multimodale.

giques certaines, est actuellement utilisé dans certains cours d'optimisation dispensés en écoles d'ingénieurs.

4.1.2 Passage en *open source* et rencontre avec Mappy. Retour d'expérience.

À mon arrivée au LICIT en 2011, l'outil consistait en un démonstrateur des travaux du laboratoire, encore immature au niveau logiciel, et trop confidentiel pour pouvoir éveiller l'attention de partenaires académiques et industriels. Pendant cinq ans, nos efforts ont porté sur l'industrialisation du produit et le portage du logiciel en licence *open source*, type de migration très rare à l'échelle des instituts publics de recherche en France, qui doit passer sous les fourches caudines des services de valorisation, qui y voient de prime abord un manque à gagner pour les différentes structures. Le passage en *open source* de TEMPUS, opéré en 2015, a pourtant vite été couronné de succès. En

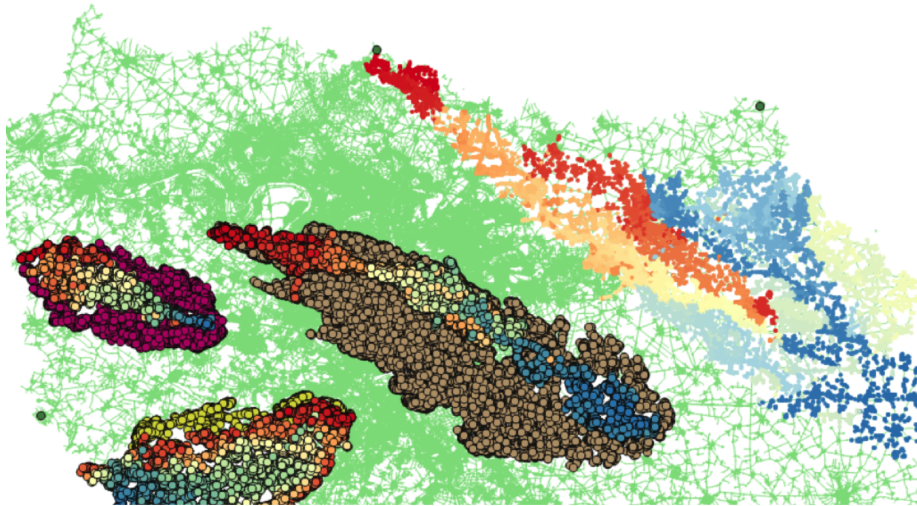


Fig. 4.4.: TEMPUS : A* vs. Dijkstra. Sur le sous-graphe de droite, les nœuds de couleur marron sont ceux explorés en plus par l'algorithme de Dijkstra.

partenariat avec OSLANDIA, entreprise de services spécialisée dans le géospatial *open source*, le nouveau statut de TEMPUS a permis de lever les verrous institutionnels habituels qui condamnent trop d'outils de qualité à vieillir dans l'anonymat des laboratoires de recherche. Grâce aux différentes opérations de communication dans des conférences comme Roadef, Foss4G, OGRS, TEMPUS a été remarquée et intégrée au paysage des calculateurs d'itinéraires *open source*. Un prototype d'une application TEMPUS mobile a ainsi été proposé aux conférenciers de l'ITS World Congress de Bordeaux, en 2015 (voir figure 4.5).

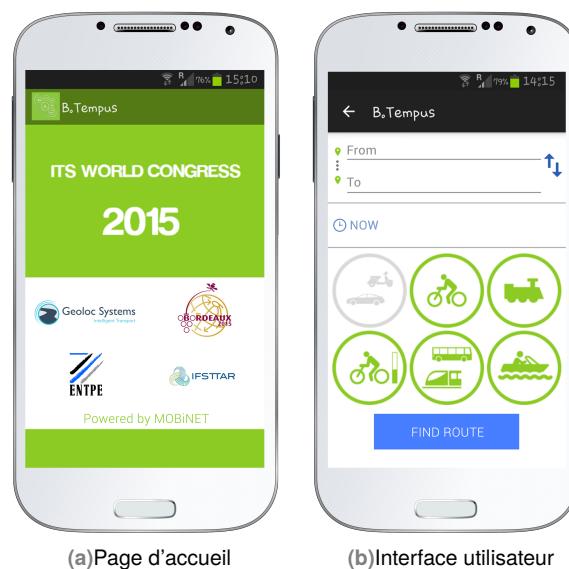


Fig. 4.5.: Application TEMPUS proposée lors de la conférence *ITS World Congress*, à Bordeaux en 2015.

Mais la consécration du passage à *l'open source* est venue de l'industriel MAPPY, en pleine refonte de son calculateur d'itinéraires, qui nous a contacté par rapport à la possibilité d'intégrer TEMPUS comme candidat potentiel pour son nouveau moteur de calcul. TEMPUS a été sélectionnée grâce à son architecture modulaire et les différentes possibilités d'extension offertes. Néanmoins, le

moteur de recherche de MAPPY n'étant pas encore multimodal, des discussions scientifiques ont écloé autour des besoins actuels de l'industriel et en particulier les techniques d'accélération qui permettraient de rendre le moteur actuel de MAPPY compétitif face à la concurrence. L'objectif ne réside plus dans un raffinement multimodal mais dans un gain de quelques centaines de millisecondes de calcul pour une simple requête Lille-Marseille dans le graphe routier de la France. Ces échanges scientifiques ont ainsi mené à l'ajout des techniques d'accélération à notre agenda de recherche, d'abord par des analyses de sensibilité qui ont aidé MAPPY à optimiser leurs paramétrages, puis, dans des perspectives moyen et long terme, dans l'adaptation de techniques d'accélération au cas de graphes stochastiques. Quelques exemples de recherche sur ce thème seront présentés dans ce chapitre. D'autre part, des échanges scientifiques avec les équipes de Microsoft à la conférence EURO 2015 ont permis de dégager de fortes tendances :

- La plupart des développements algorithmiques sont encore concentrés sur les graphes routiers pour lesquels les différentes équipes de recherche rivalisent en termes d'heuristiques d'accélération ;
- La gestion de l'information dynamique et temps réel concentre beaucoup d'énergie au détriment des raffinements méthodologiques ;
- Les prochaines étapes seront l'adaptation des techniques à des graphes multimodaux et stochastiques ;
- L'intégration de nouveaux types de mobilité comme l'électromobilité n'est pas encore d'actualité ;
- Le multiobjectif intéresse parfois plus les chercheurs que les industriels : l'utilisateur sait ce qu'il veut lorsqu'il lance une requête !

Ces interactions avec des équipes R&D d'acteurs importants du monde de l'information voyageurs ont clairement inspiré quelques unes de nos recherches, bien sûr plus en amont que les contraintes industrielles soulevées. Dans ce manuscrit d'HDR, nous nous attarderons sur des travaux en électromobilité, optimisation dans les graphes stochastiques et techniques d'accélération, qui pourront, à moyen terme, intégrer des boîtes à outils de l'information des usagers, telle que TEMPUS peut l'être.

4.2 Outils pour l'électromobilité

Comme nous l'avons souligné dans le chapitre introductif 1, les systèmes de navigation existants ne prennent pas en compte les particularités des véhicules électriques comme l'autonomie restreinte, l'état initial de la batterie et la capacité de récupération au freinage. Quand une des deux premières contraintes ne permet pas d'effectuer sans arrêt le trajet demandé, le système doit être capable de proposer un détour par une station de recharge, en assurant un compromis entre consommation énergétique et temps de parcours. Nos travaux sur les véhicules électriques ont choisi de traiter toute la chaîne méthodologique, depuis la question de la génération d'un graphe énergétique au problème du plus court chemin, en passant par l'affectation optimale des stations de recharge sur le réseau.

4.2.1 Génération d'un graphe énergétique

Sans des données d'entrée correctes, tout effort de validation des algorithmes serait vain et la construction du graphe énergétique représente une première difficulté de taille. Le point de départ est un graphe routier $G(V, E)$ où V est l'ensemble des sommets et E l'ensemble des arcs. Les arcs sont les sections de route et les sommets représentent les intersections entre les sections de route. Une vitesse moyenne (notée AS pour *average speed*) est affectée à chaque arc. Après cette première affectation naïve, la deuxième étape consiste à créer un profil de vitesse pour chaque arc à partir de quatre cycles de vitesse possibles, montrés sur la figure 4.6.

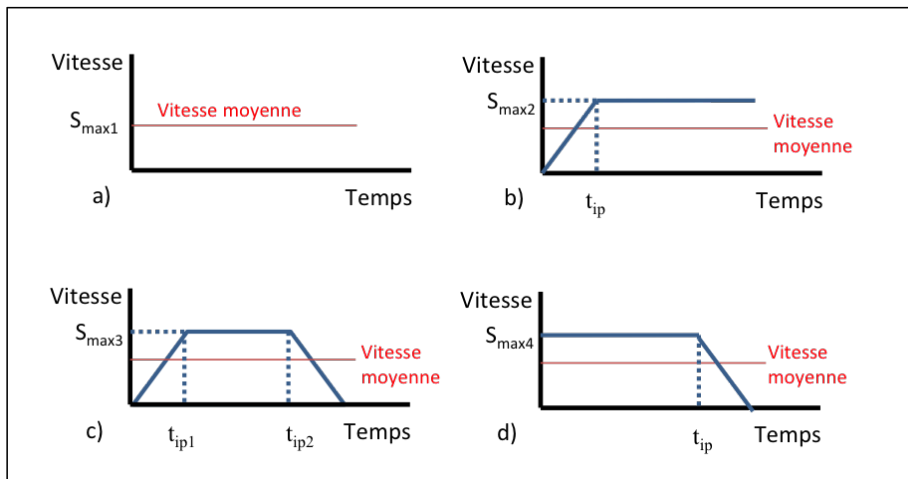


Fig. 4.6.: Cycles de vitesse.

Le premier cycle correspond au cas le plus simple de vitesse constante, il s'agit de la vitesse moyenne pour traverser la section de route. Le deuxième cycle est un cas d'accélération dans lequel le conducteur atteint la vitesse de croisière, qui reste constante jusqu'à la fin de la section. Le troisième cycle ajoute une phase de décélération à la fin de l'arc tandis que le dernier est symétrique du second (d'une vitesse libre à la décélération jusqu'à l'arrêt total). Ces quatre schémas peuvent être ensuite intelligemment affectés aux arcs, y compris en fonction des requêtes de plus courts chemins. La troisième étape de construction du graphe énergétique consiste à définir le coût énergétique nécessaire à la traversée des différents arcs du réseau avec un véhicule électrique. Notre principale contribution réside dans l'utilisation de la bibliothèque VEHLIB, utilisée dans le cadre de la co-direction de thèse de Fouad Baouche par le Laboratoire Transports Environnement (LTE) de l'IFSTTAR. VEHLIB est un modèle dynamique d'estimation de la consommation pour véhicule électrique, conçu comme une librairie modulaire combinant plusieurs modèles de véhicules [429]. La démarche de modélisation vise à considérer le véhicule étudié comme une association de différents sous-systèmes, chaque sous-système représentant un composant du véhicule (figure 4.7).

Les paramètres des composants sont identifiés par expérimentation ou directement spécifiés par le constructeur. Les entrées du modèle sont les profils de vitesse présentés ci-dessus ainsi que l'information sur l'altitude de chaque point intermédiaire. Le modèle reproduit ensuite les forces agissant sur le véhicule en fonction de ces données puis fournit une estimation précise de la consommation. Le modèle utilisé est un modèle indirect calé pour une Peugeot I-ON (Citroën C-zéro), puis estimant la consommation à partir des différents profils de vitesse traversés, donnés

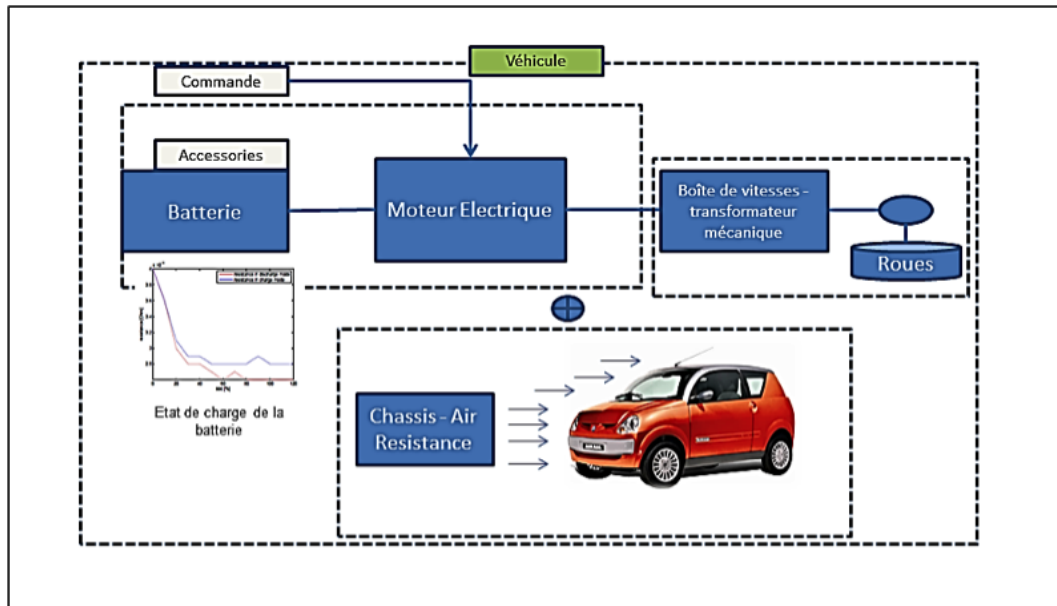


Fig. 4.7.: Approche systémique du modèle de consommation VEHLIB.

par les cycles. Des simulations attestent d'une erreur d'environ 5% par rapport à des mesures sur des bancs de test [29, 429]. Notre terrain d'études a été la zone urbaine de Lyon, comme l'illustre la figure 4.8.

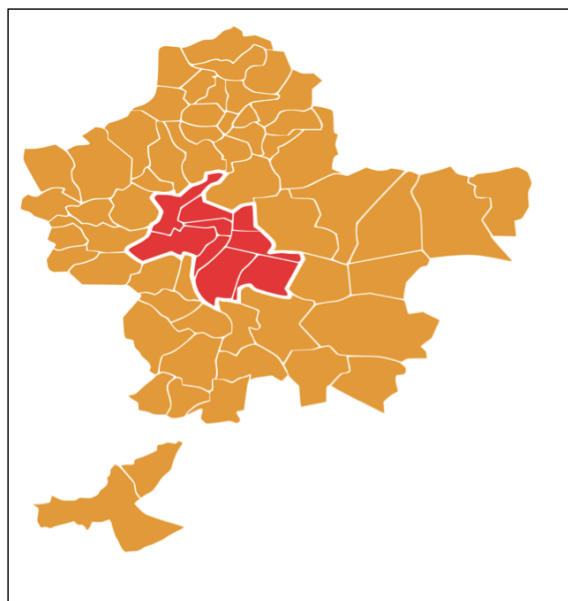


Fig. 4.8.: Carte du Grand Lyon. Lyon et ses 9 arrondissements (en rouge) et la métropole de Lyon (en orange)

Le graphe de l'aire urbaine de Lyon comporte 6946 sommets et 9492 arcs tandis que le graphe de la métropole de Lyon se compose de 44389 sommets et 53032 arcs. Ces graphes énergétiques ont été construits à partir de *Open Street Map* pour les données routières et de l'Institut Géographique National (IGN) pour les données d'altitude.

4.2.2 Affectation des stations de recharge

La génération d'un graphe énergétique grâce à une bibliothèque aussi précise que VEHLIB a apporté à nos travaux une certaine légitimité par rapport à de nombreuses études de l'état de l'art, qui se fondent sur des approximations simplistes des consommations. La deuxième phase du travail a consisté à proposer un modèle d'affectation des stations de recharge sur un réseau. Il s'agit de sélectionner, d'affecter, des stations de recharge par rapport à une liste de stations potentielles. Cette thématique était prégnante en 2013, quand le développement d'infrastructures pour véhicules électriques en était à ses prémices. Par rapport à l'état de l'art présenté à la section 1.5.4 du chapitre 1, notre approche combine plusieurs variantes existantes tout en s'appuyant sur des données réelles pour construire une demande énergétique réaliste (figure 4.9).

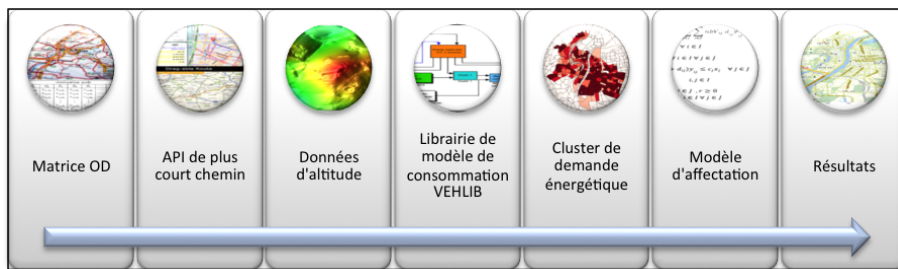


Fig. 4.9.: Démarche méthodologique pour l'affectation des stations de recharge.

Un point saillant de notre démarche concerne la création de clusters de demande à partir d'une analyse des matrices Origine-Destination découlant des Enquêtes Ménages Déplacements (EMD). En particulier, les destinations des trajets domicile-travail ou domicile-loisirs servent à créer ces zones de demande. Pour chaque Origine-Destination, un trajet type est défini à partir d'une hypothèse de plus court chemin. Et à partir de chaque destination, la consommation vers chacune des stations de recharge potentielles est également estimée, afin de la prendre en compte dans le modèle d'affectation. Le modèle d'affectation proposé choisit un ensemble de stations de recharge parmi une liste de sites candidats, dans un contexte d'optimisation des coûts. Ces coûts traduisent les frais d'installation et de mise en service des stations de recharge ainsi que les frais énergétiques de déplacements des véhicules électriques vers les stations. Le problème est formulé sous la forme d'un programme linéaire qui doit respecter des contraintes de couverture des clusters de demande, de capacité des stations de recharge ou de rayon minimal entre chaque couple de stations de recharge. Les variables de décision du problème sont

$$x_j = \begin{cases} 1 & \text{si une station de recharge est en service au site } j, \\ 0 & \text{sinon.} \end{cases} \quad (4.1)$$

et

$$y_{i,j} = \begin{cases} 1 & \text{si le cluster de demande énergétique } i \text{ est couvert par une station de recharge } j, \\ 0 & \text{sinon.} \end{cases} \quad (4.2)$$

Et le programme d'optimisation s'écrit

$$\min_{y,x} \sum_{j \in J} (f_i x_j) + \alpha \sum_{j \in J} \sum_{i \in I} n_i^{ev} d_{i,j} y_{i,j} \quad (4.3)$$

$$\begin{cases} \sum_{j \in J} y_{i,j} = 1 & \forall i \in I \\ (y_{i,j} - x_j) \leq 0, & \forall i \in I, \forall j \in J \\ \sum_{i \in I} n_i^{ev} (D_i + d_{i,j}) y_{i,j} \leq c_j x_j, & \forall j \in J \\ r x_i x_j \leq dist_{i,j}, & \forall i \in I, \forall j \in J \\ x_j \in \{0, 1\}, y_{i,j} \in \{0, 1\}, r \geq 0 \end{cases} \quad (4.4)$$

où les différentes variables sont les suivantes :

- I : ensemble des clusters de demande ;
- J : ensemble des stations de recharge potentielles ;
- f_i : coût d'affectation de la station de recharge i ;
- D_i : demande énergétique au cluster i ;
- $d_{i,j}$: énergie nécessaire pour qu'un véhicule atteigne une station de recharge j depuis un cluster de demande i ;
- r : rayon minimal entre deux stations de recharge ;
- $dist_{i,j}$: distance entre les stations de recharge i et j ;
- c_j : capacité de la station de recharge j ;
- α : coût d'un kilowatt heure ;
- n_i^{ev} : nombre de véhicules électriques se déplaçant vers le cluster i .

La fonction de coût 4.3 contient un coût total d'investissement et un coût énergétique engendré par le déplacement des usagers vers les stations de recharge. Les contraintes expriment ensuite des exigences de couverture des demandes émanant des clusters par les stations de recharge, tout en respectant les capacités de ces dernières. Alors que D_i représente la somme de toutes les demandes de déplacement vers un point i , à partir des enquêtes ménages déplacements, les $d_{i,j}$ sont ajoutés pour couvrir toute déviation d'un véhicule vers une station de recharge. L'avant-dernière contrainte s'inspire du problème p -dispersion en imposant une distance minimale de séparation entre chaque couple de stations. Le modèle a été testé sur la ville de Lyon et les résultats en simulation sont cohérents au vu des différentes analyses de sensibilité effectuées. Le détail des expérimentations a été publié dans la revue *IEEE Intelligent Transportation Systems* [28].

4.2.3 Algorithmes de plus court chemin énergétique

Le problème du plus court chemin (PCC) pour véhicules électriques peut être considéré comme un PCC avec contraintes de capacité de batterie et d'arcs négatifs (récupération d'énergie au freinage). Dans un article de 2010, Artmeier *et al.* ont adapté l'algorithme de Bellman-Ford pour intégrer ses contraintes tout en garantissant une résolution polynomiale [16]. Nous avons souhaité contribuer à cette question de recherche en ajoutant la possibilité d'une recharge pour pouvoir finir le trajet (cas le plus probable au vu des technologies et pratiques actuelles) puis en généralisant à n recharges. En plus d'un graphe routier et d'un graphe énergétique, tel que construit précédemment, il s'agit de disposer d'une liste des stations de recharge avec leur disponibilité et leur capacité. La résolution proposée est opérationnelle et destinée à intégrer à terme un calculateur d'itinéraires. Elle utilise en

son coeur l'algorithme de Artmeier *et al.* [16] tout en l'entourant d'un certain nombre d'opérations permettant les possibles recharges et donc la faisabilité de la requête origine-destination. L'état de charge initial est noté SOC_0 et une première liste, nommé RCS , contient l'ensemble des plus courts chemins entre l'origine s et les stations de recharge atteignables avec SOC_0 . Le deuxième ensemble, noté RD se compose des chemins reliant les stations de recharge vers le nœud de destination. Une hypothèse forte est de considérer que le véhicule va recharger complètement sa batterie à chaque station de recharge : $SOC = 100\%$. Une fois les deux listes prêtes, une première étape balaye les trajets origine-destination réalisables avec une seule recharge. Il s'agit de croiser les ensembles RCS et RD et de conserver les couples de sous-chemins qui partagent une station de recharge en commun. L'ensemble des trajets réalisables est noté SFP , pour *Set of Feasible Paths*. La figure 4.10 illustre la construction des différents ensembles. La solution finale est ensuite sélectionnée comme celle minimisant la consommation énergétique sur l'ensemble du trajet.

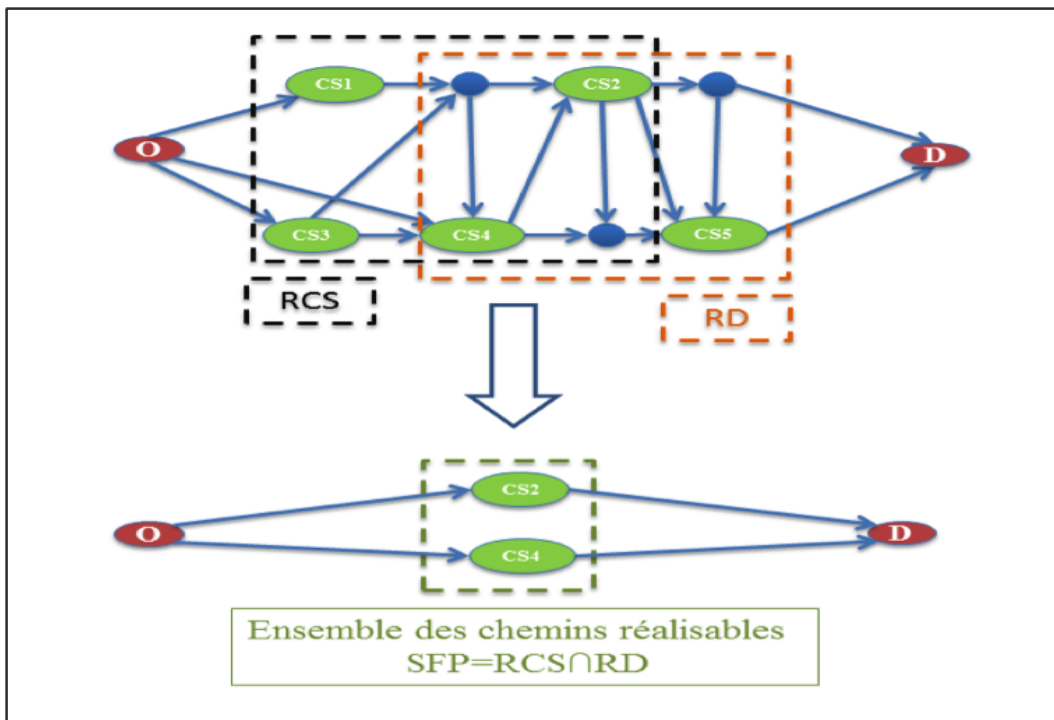


Fig. 4.10.: Construction des sous-ensembles RCS , RD et SFP .

Une amélioration possible consiste à introduire le temps de parcours total comme critère de sélection, tout en estimant plus finement le temps de recharge nécessaire pour arriver à destination à partir de la liste SFP . Ce dernier peut être déduit de l'état de charge initial, de la consommation sur les sous-chemins de RCS et de RD , et des caractéristiques des stations de recharge. Par la suite, nous avons cherché à généraliser notre contribution pour le problème à n recharges, tout en conservant une forte visée opérationnelle. Pour ce faire, un pré-traitement crée une nouvelle structure de données, dénommée *reachability matrix* (RM), qui est une matrice carrée constituée des valeurs énergétiques nécessaires pour voyager entre tout couple (i, j) de stations de recharge. La taille de la matrice dépend donc du nombre de stations de recharge sur le réseau et une valeur $RM(i, j) = 0$ s'il n'est pas possible de relier deux stations de recharge avec une batterie pleine. Cette phase d'initialisation va permettre de gagner en temps de calcul, les stations de recharge étant localisées *a priori*. La résolution d'une requête procède ensuite de la façon suivante. Lorsqu'un chemin sans recharge n'est pas possible pour relier la destination, la procédure de déviation vers une ou plusieurs

stations de recharge est enclenchée. La matrice RM est associée à l'origine et la destination à travers un graphe de connectivité (CG) et la matrice RM augmentée contient naturellement les ensembles RCS et RD introduits pour le problème à une recharge. Une fois la création du graphe effectué, un appel d'un algorithme classique (Dijkstra ou Bellman) permet de trouver la solution minimisant la consommation énergétique. Il est trivial de montrer que le plus court chemin dans le graphe de connectivité est le plus court chemin dans le graphe initial, puisque le principe de Bellman est respecté par construction. La figure 4.11 montre deux exemples de graphe de connectivité, avec une ou deux recharges nécessaires.

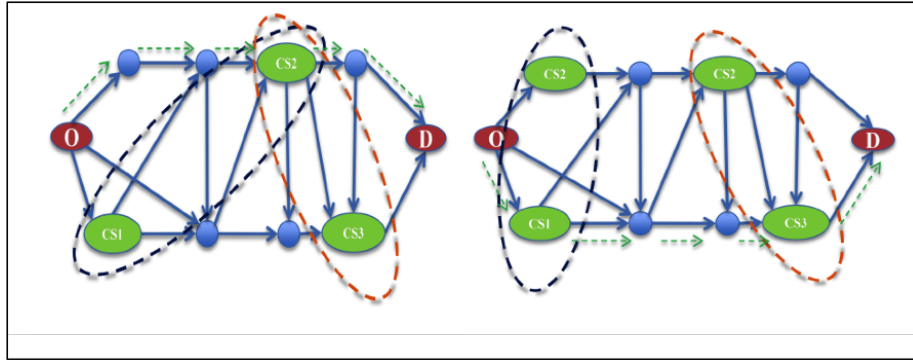


Fig. 4.11.: Graphe de connectivité

Des extensions de la méthode de base ont été proposées, non détaillées dans ce manuscrit d'HDR. Elles étendent le problème à une optimisation multi-critères en intégrant le temps de parcours total, le nombre d'arrêts à des stations de recharge et la quantité de recharge à chaque station. Une méthode simple consiste à générer les k solutions énergétiques optimales grâce à l'algorithme de Yen [473] (et la méthode des potentiels de Johnson dans le cas d'arcs négatifs [214]) puis à classer les solutions selon d'autres critères, éventuellement à l'aide de programmes d'optimisation.

4.2.4 Quelques résultats

La figure 4.12 montre un exemple de résultats pour le réseau de la ville de Lyon. Le trajet origine-destination proposé comprend des passages par des zones à fort trafic ou forte dénivellation. Le chemin en rouge s'effectue avec une batterie pleine et ne nécessite pas de recharge en route tandis que le chemin en vert imposait un état initial de batterie faible ($SOC_0 = 13\%$).

La quantité rechargée à chaque station a été limitée afin d'illustrer un trajet avec plusieurs recharges. Pour un ensemble de requêtes effectuées sur le graphe de la ville de Lyon, la valeur moyenne des temps de calcul est inférieure à la seconde. Pour le grand graphe de la métropole de Lyon, les temps de calcul vont varier en fonction du nombre de stations de recharge, en particulier pour la construction du graphe de connectivité, pré-traitement qui monopolise la majorité du temps de calcul (figure 4.1).

| Nb de stations de recharge (SR) | 5 SR | 10 SR | 20 SR | 30 SR | 40 SR | 50 SR | 60 SR | 70 SR |
|---------------------------------|-------|-------|--------|---------|---------|---------|---------|---------|
| Graphe de connectivité | 9s | 30s | 2m | 4m | 6m | 10m | 14m | 19m |
| PCC énergétique | 3.4s | 3.4s | 3.4s | 3.4s | 3.4s | 3.4s | 3.4s | 3.4s |
| Temps total | 12.4s | 33.4s | 2m3.4s | 4m3.4 s | 6m3.4 s | 10m3.4s | 14m3.4s | 19m3.4s |

Tab. 4.1.: Répartition des temps de calcul lors de la résolution du PCC énergétique

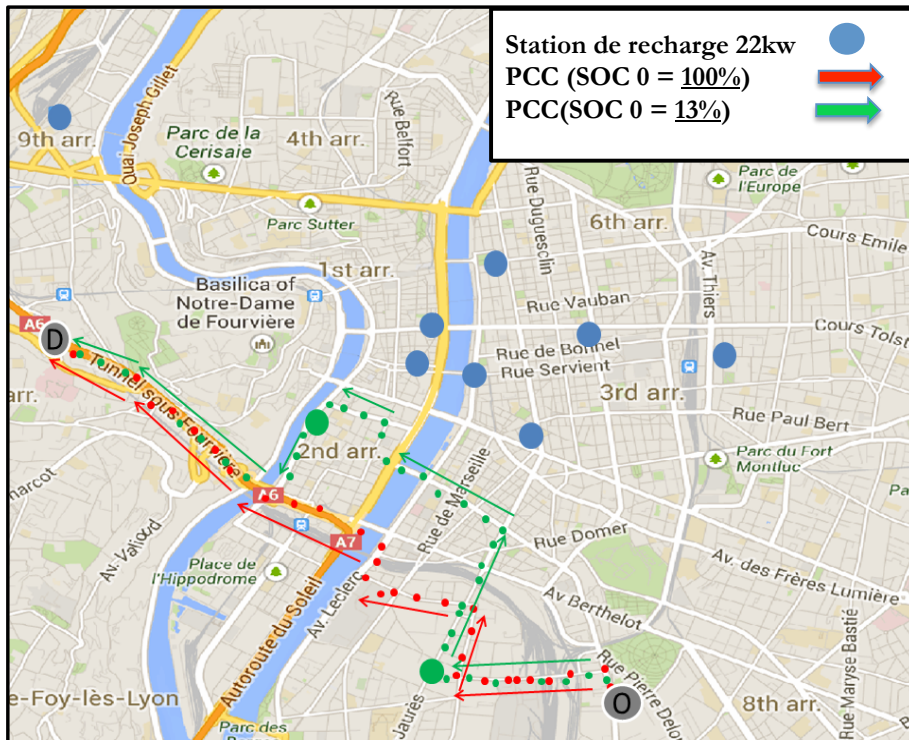


Fig. 4.12.: Exemple de solution du problème de PCC énergétique.

Un certain nombre d'analyses de sensibilité ont été conduites pour le graphe de la métropole de Lyon. Plusieurs milliers de requêtes ont été simulées, en particulier pour les nœuds les plus éloignés du graphe, parfois en simulant des instances qui ne se reproduiront jamais dans la réalité : le nombre maximum de recharges en route atteint 1000 arrêts à des stations de recharge. Les résultats confirment que notre solution est capable de passer à l'échelle avec des temps de calcul qui ne dépassent jamais les dix secondes. D'autre part, dans le cadre de la thèse de Fouad Baouche, une expérimentation grandeur nature a été menée à Lyon en équipant un véhicule électrique avec des capteurs de position et des mesures de consommation. L'expérimentation a permis d'évaluer les solutions proposées par notre calculateur d'itinéraires énergétiques, pour chacun des quatre cycles de vitesse. Si les solutions du calculateur sont pertinentes et particulièrement bien adaptées à un véhicule électrique, des erreurs de consommations sont à déplorer, en particulier à cause de la qualité des données OpenStreetMap et de la simplification du problème par des cycles de vitesse. Cependant, lorsque les schémas de modélisation des cycles de vitesse du graphe lyonnais sont corrigés et plus réalistes, les erreurs d'estimation deviennent tout à fait acceptables. La figure 4.13 montre un exemple d'un trajet proposé par le calculateur de PCC énergétique, entre le centre-ville et la banlieue lyonnaise Est (IFSTTAR à Bron), où la solution diffère de celle renvoyée par le calculateur *Google Maps*.

C'est l'impact du profil altimétrique du plus court chemin habituel (à droite sur la figure 4.13) qui mène notre calculateur à choisir un itinéraire mieux adapté à un véhicule électrique.

En conclusion, nous pouvons dire que nos travaux sur le thème de l'électromobilité ont fortement profité d'une collaboration interdisciplinaire avec l'équipe de Rochdi Trigui du laboratoire transports environnement de l'IFSTTAR. L'équipe a apporté une modélisation fine du comportement des véhicules électriques et nous a ensuite permis de valider nos algorithmes en conditions réelles,

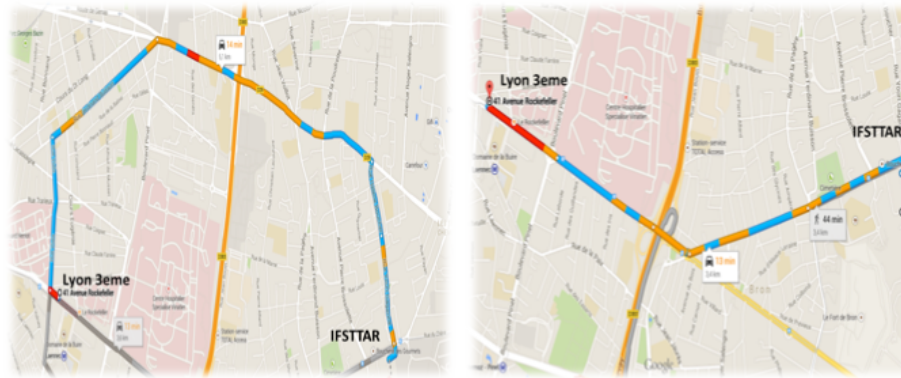


Fig. 4.13.: Exemple de deux solutions à une même requête. PCC énergétique (gauche) vs. *Google Maps* (droite).

chose rare en théorie des graphes. Nos contributions ont porté sur le déploiement optimal des infrastructures de recharge à l'échelle d'une agglomération urbaine, puis l'optimisation des itinéraires avec possibilité de recharge en cours de route. Cette dernière solution pourrait être intégrée à court terme dans un calculateur d'itinéraires de type *web service*. L'utilisateur aurait simplement à cocher une case "véhicule électrique" pour obtenir des itinéraires personnalisés en terme de consommation énergétique et d'arrêts possibles pour la recharge, voire d'une optimisation multi-critères de son trajet.

4.3 Plus court chemin stochastique : sensibilité aux lois de probabilité et techniques d'accélération

Les modèles statistiques de temps de parcours, tels que ceux introduits au chapitre 3, peuvent être logiquement pris en compte dans l'optimisation des itinéraires afin d'en raffiner leur résolution. Cette question de recherche, loin d'être évidente, a mobilisé notre attention dans le cadre de la direction de thèse de Raphaël Delhome (2013-2016).

4.3.1 Impact des distributions de probabilité sur le plus court chemin stochastique

Dans le domaine de l'optimisation des itinéraires, une question déjà ancienne mais toujours d'actualité est le plus court chemin stochastique (PCCS), c'est-à-dire un problème de graphe dans lequel les valuations des arcs sont sujettes à l'incertitude. Nous revenons ici à un graphe simple $G = (V, E)$, par exemple routier, où V est l'ensemble des nœuds et E l'ensemble des arcs. Les arcs sont valués par des temps de parcours, définis comme des variables aléatoires distribuées. Plus formellement, on note T_{ij} la variable aléatoire à laquelle est associé le temps de parcours entre les nœuds i et j , c'est à dire sur l'arc $(i, j) \in E$. Cet arc est valué par la réalisation τ_{ij} de la variable aléatoire. La fonction de densité théorique de T_{ij} est notée classiquement $f_{ij}^*(x)$ et la fonction de répartition théorique $F_{ij}^*(x)$. À partir de ce cadre théorique classique, il s'agit de construire la distribution des temps de parcours sur un itinéraire donné. Deux hypothèses fortes ont été assumées dans nos travaux :

1. Les temps de parcours peuvent être discrétisés. En particulier, $f_{ij}^*(x)$ est discrétisée en une fonction de masse $p_{ij}(x)$ à l'aide d'un intervalle de temps Δ_t sur les T_{max} premiers entiers (non nuls) consécutifs. Le temps de parcours maximal considéré est donc $T_{max}\Delta_t$.
2. Les temps de parcours sur les différents segments sont indépendants les uns des autres.

Nos efforts se sont tout d'abord concentrés sur le **choix des distributions de probabilité et leurs impacts dans la résolution du PCC stochastique**. Il s'agissait, dans la lignée de nos travaux sur le système Halphen présentés au chapitre 3, de comparer une loi de probabilité plus raffinée, en l'occurrence la loi Gamma, avec les deux lois couramment utilisées dans les modèles de PCC, les lois normale et lognormale. La loi Gamma, qui est un cas limite du système Halphen, respecte, comme les deux autres, la propriété de fermeture par addition. Les trois distributions ont été comparées sur une base d'échantillonnage commune, au sein d'un même algorithme, SPOTAR [330]. Avant de détailler l'algorithme, quelques notions doivent être expliquées, en particulier les opérations associées aux graphes stochastiques puis les concepts de dominance stochastique et politiques de routage. Tout d'abord, trois opérations fondamentales peuvent être effectuées dans un graphe stochastique :

- L'évaluation : association d'un budget-temps à une probabilité d'arriver à l'heure. L'évaluation peut se faire en évaluant la fonction de répartition en t , pour calculer la probabilité d'accomplir le trajet en une durée inférieure à un budget temps donné. D'autre part, à partir de l'inverse de la fonction de répartition, la valeur du quantile pour une probabilité ϕ donnée traduit le temps de trajet nécessaire au parcours du chemin dans ϕ % des trajets.
- La concaténation : constitution du modèle de temps de parcours sur une succession d'arcs, pour un itinéraire entier. De manière classique, cette opération appelle des convolutions de lois de probabilité.
- L'agrégation : fusion des arcs reliant le même couple origine-destination. En effet, lorsque plusieurs itinéraires permettent de rejoindre l'origine à la destination, une synthèse de la distribution de temps de parcours entre les nœuds doit être produite.

Pour les graphes stochastiques, le principe de Bellman n'opère pas toujours et de nouveaux mécanismes de parcours et de comparaison d'étiquettes sont mis en place.

La figure 4.14 propose un exemple classique pour illustrer cette difficulté. En posant $\Delta_t = 1$, la question posée est la suivante : quelle est la probabilité d'atteindre u_4 à partir de u_1 en seulement deux unités de temps ? Par des relations de convolution, il est assez rapide d'obtenir la valeur de $F_4(2)$, qui correspond à la probabilité recherchée. Deux prédécesseurs sont possibles, pour chaque budget de temps : u_2 ou u_3 , la meilleure solution étant obtenue grâce à la comparaison de deux distributions :

- la première est issue de la convolution entre la fonction de répartition associée au nœud u_2 et la fonction de masse associée à l'arc e_4 , reliant u_2 à u_4 ;
- la seconde est le produit de convolution entre la fonction de répartition associée au nœud u_3 et la fonction de masse de l'arc e_5 , issu de u_3 et dirigé vers u_4 .

En considérant la seconde distribution comme la meilleure pour $t = 2$, le meilleur prédécesseur est u_3 . La question se pose ensuite pour ce nœud u_3 . À chaque budget de temps, le prédécesseur de u_3 est donné par la distribution optimale au nœud, *i.e.* par le produit de convolution maximisant la probabilité cumulée d'arriver à u_3 . Or, la date de passage au nœud u_3 n'est pas garantie : atteindre le nœud u_4 en $t = 2$ est possible dans 3 situations (atteindre u_3 respectivement en $t = 0$, $t = 1$ ou $t = 2$ et traverser e_5 respectivement en 2, 1 ou 0 unité(s) de temps). Quelle que soit la réalisation

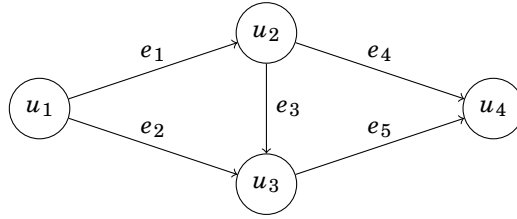


Fig. 4.14.: Exemple de reconstitution d'un itinéraire optimal entre u_1 et u_4

du temps de parcours sur e_5 , le résidu de budget devra être alloué selon le même procédé sur les arcs en amont, jusqu'à atteindre le nœud source. Comme le montre le raisonnement précédent, la reconstitution du chemin optimal dans le PCC stochastique implique de stocker les détails de chaque scénario avant la reconstruction de l'itinéraire optimal, ce qui équivaut finalement à une énumération de solutions optimales [330]. D'autre part, le caractère fortement combinatoire du graphe des prédécesseurs a mené Nie *et al.* à proposer une nouvelle structure de données, la *politique de routage*. Cette structure est constituée par l'ensemble des chemins non-dominés entre une origine et une destination données. Tout d'abord, la notion de dominance permet de comparer les vecteurs d'étiquette entre eux. Elle consiste à comparer deux distributions statistiques entre elles à travers les fonctions de répartition. Un concept-clé est la dominance stochastique de 1er ordre. On dira que la distribution 1 domine la distribution 2 si et seulement si [301, 303]

$$F_1(t) \geq F_2(t) , \forall t \in [0; T_{max}] \quad (4.5)$$

$$\exists t \in [0; T_{max}] \text{ tel que } F_1(t) > F_2(t) \quad (4.6)$$

En complément de cette définition, Miller-Hooks a introduit la notion de dominance de groupe : on dit qu'une distribution caractérisée par la fonction de répartition F est non dominée par rapport à un ensemble de distributions D si et seulement si il existe une réalisation $t \in [0; T_{max}]$ pour laquelle :

$$F(t) > F_d(t) , \forall d \in D \quad (4.7)$$

où F_d représente la fonction de répartition associée à une distribution de l'ensemble D . La définition opposée est ainsi la suivante : une distribution est dominée lorsque $\forall t \in [0; T_{max}], \exists d \in D$ telle que :

$$F(t) \leq F_d(t) \quad (4.8)$$

Comme le montre la figure 4.15, la dominance de groupe est plus stricte que la dominance individuelle car moins de distributions peuvent prétendre au statut de distribution non-dominée. La politique de routage retenue dans notre travail se fonde sur cette notion de dominance de groupe et il s'agit alors de ne conserver que les chemins contribuant directement à la distribution enveloppe. Ce choix accélère le temps de résolution au prix d'une réduction limitée de la précision de l'algorithme. Des définitions plus formelles peuvent ainsi être posées, comme suit :

Définition 1 On appelle politique de routage d'un nœud source s à un nœud destination d le sous-ensemble C^* de chemins reliant les deux nœuds et dont les distributions vérifient l'équation 4.7.

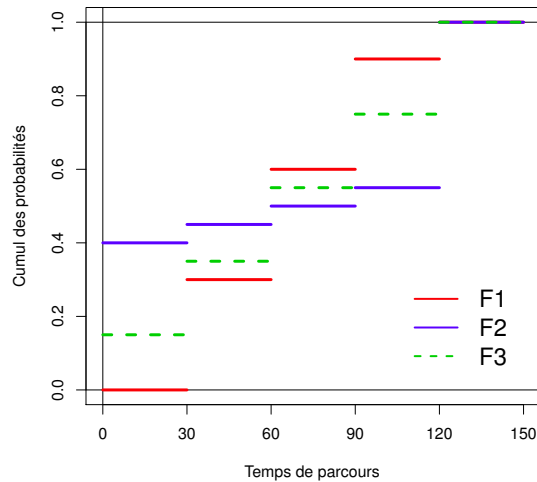


Fig. 4.15.: Illustration de la dominance stochastique de premier ordre : $F3$ est dominé par le groupe $\{F1;F2\}$, mais pas par $F1$ et $F2$ pris individuellement.

Définition 2 L'agrégation des distributions relatives aux chemins $c \in C^*$, i.e. leur enveloppe supérieure, forme la frontière de Pareto de la politique de routage, une fonction de répartition optimale que l'on note ici F^* . Cette frontière marque la probabilité maximale de relier deux nœuds en un budget de temps donné.

Définition 3 On note c_t^* le chemin optimal entre s et d pour un budget équivalent à t . Cette information répond à la requête simple de recherche de chemin compte tenu d'un budget de temps disponible.

Les définitions précédentes introduisent le problème du plus court chemin stochastique dans sa formulation SPOTAR. L'algorithme 2 propose un pseudo-code simplifié de la résolution du problème, adapté de [330]. Il s'agit d'une stratégie d'étiquetage fondée sur la création et la mise à jour de politiques de routage $[P_v^*]_{v \in V}$, depuis chaque nœud du graphe jusqu'à une destination donnée, résolution de type *all-to-one*.

Une file d'attente Q permet de stocker les chemins optimaux, son initialisation est réalisée à la ligne 2, les chemins en étant extraits à la ligne 4. À chaque itération de la boucle principale, le chemin courant c est étendu en un chemin c' à l'aide de chacun des nœuds prédécesseurs du nœud j , le plus en amont de c (ligne 5). À la condition de n'être pas constitué d'un cycle, le nouveau chemin est susceptible d'être intégré à la politique du nœud prédécesseur, soit en tant qu'unique chemin d'une nouvelle politique (ligne 9), soit en tant qu' i ème chemin d'une politique déjà existante (ligne 13). Dans le second cas, le choix de l'intégrer ou non est décidé à travers la comparaison entre la fonction de répartition candidate et la frontière de Pareto, au moyen de la relation de dominance (ligne 12). En effet, si la fonction de répartition candidate n'est pas dominée par la frontière de Pareto au sens de la dominance stochastique du premier ordre, il est alors nécessaire d'intégrer le nouveau chemin à la politique. En complément, il est nécessaire de vérifier si les anciens chemins optimaux (en particulier ceux tout juste surpassés par le chemin candidat) sont toujours non-dominés par la frontière de Pareto : si ce n'est plus le cas, ils sont à exclusion de la politique de routage.

Algorithme 2 (SPOTAR) Recherche du plus court chemin stochastique (d'après Nieet al. [330], version pseudo-code)

ENTRÉES: Graphe $G = (V, E)$, nœud destination $d \in V$

SORTIES: Politiques de routage $[P_v^*]_{v \in V}$ pour rejoindre d depuis les nœuds $v \in V$

```

1:  $P_v^* \leftarrow \{\emptyset\}$ 
2:  $Q \leftarrow \{d\}$            # Initialiser la file d'attente avec le chemin menant de  $d$  à lui-même
3: TantQue  $Q \neq \emptyset$  Faire
4:   Extraire le chemin courant  $c$  depuis  $Q$ , soit  $j$  le nœud le plus en amont de  $c$ 
5:   Pour  $e = (i, j) \in E$  Faire
6:     Vérifier que le chemin  $c' = \{i, c\}$  est acyclique
7:     Calculer la fonction de répartition  $F_{c'}$ 
8:     Si  $c'$  est le seul chemin reliant  $i$  à  $d$  Alors
9:       Initialiser  $P_i^*$  avec  $c'$            # Le chemin optimal est  $c'$  quel que soit  $t$ 
10:       $Q \leftarrow Q \cup \{c'\}$ 
11:     Sinon
12:       Si  $\exists t \in \{0, \Delta_t, \dots, \Delta_t T_{max}\}$  tel que  $c'$  améliore la frontière de Pareto Alors
13:         Mettre à jour  $P_i^*$            # La frontière est modifiée, le chemin  $c'$  ajouté à
           la politique de routage, et les anciens chemins optimaux pour les budgets concernés
           sont susceptibles d'en être retirés
14:          $Q \leftarrow Q \cup \{c'\}$ 
15:       FinSi
16:     FinPour
17:   FinTantQue
18: Retourner  $P_v^*$ 

```

Une fois l'algorithme présenté, il convient de rappeler la question de recherche visée : **les sorties de l'algorithme SPOTAR vont-elles différer en fonction des lois utilisées pour la pondération des arcs ?** Et peut-être, une deuxième question, sous-jacente : quels sont les autres paramètres d'influence sur les solutions proposées ? Un plan d'expérience complet a été mené afin d'apporter des éléments quantitatifs et solides à ces deux questions. Cinq facteurs ont été croisés pour construire un total de 324 configurations :

- Le type de **graphe** : quatre graphes ont été utilisés, typiques de la littérature sur le PPCC. Le premier, *Sioux Falls* (SF), contient 24 nœuds et 76 arcs. Le second, *ChicagoSketch* (CS) contient 546 nœuds et 2176 arcs. Les deux autres graphes sont des grilles synthétiques, régulières de respectivement 10 (G10) et 20 (G20) nœuds de côté pour un total de 100 nœuds, 360 arcs, et 400 nœuds et 1520 arcs ;
- Le **scénario de variabilité** : il s'agit d'un échantillonnage du retard par rapport à un temps de parcours à vide, à partir de l'espérance et l'écart-type propres à chaque arc ;
- La **précision des distributions de temps de parcours** : plus ou moins grande selon le nombre de points supports et la largeur des classes de l'histogramme. Les trois quantités de points supports retenues sont 60, 120 et 600 ;
- La **distance** de la requête OD : en fonction, des premier, cinquième et neuvième déciles de la distribution des distances dans le graphe, trois échelles sont créées : distance courte, moyenne ou grande ;
- La **probabilité d'être à l'heure** : en considérant cette fois les déciles de la frontière de Pareto associée à la politique de routage, trois valeurs de probabilité sont retenues (0.1, 0.5 et 0.9). A chaque probabilité donnée correspond un chemin optimal minimisant le budget temps. Cette valeur peut également être vue comme une aversion au risque de l'utilisateur.

Les expérimentations ont été réalisées par Raphaël Delhome, dans le cadre de sa thèse, sur une machine Windows dotée de 32Go de mémoire vive, d'un processeur *Intel Xeon E3-1225 3.2GHz* équipé de 4 cœurs. L'algorithme a été codé en C++11 et compilé. Les conclusions des expérimentations sont les suivantes :

- Dans la majorité des cas, les trois lois (Gamma, lognormale, normale) fournissent des solutions identiques quelles que soient la nature du graphe, la probabilité d'arriver à l'heure fixée, la précision et le niveau de variance du scénario ;
- Le graphe le plus proche d'un réseau routier urbain, CS, exhibe quant à lui des différences notables notamment pour des grandes distances entre origine et destination ;
- La précision de la discrétisation est au moins aussi importante que la loi de probabilité utilisée ;
- Les divergences entre modèles déterministes et stochastiques sont importantes ;
- Les temps de calcul restent trop importants pour entrevoir en l'état un passage à l'échelle.

La figure 4.16 montre la proportion moyenne de cinq situations de résultats au sein de configurations possibles, pour les graphes *Sioux Falls* (SF) et Chicago (CS). Les situations sont les suivantes :

- S1 : les solutions optimales des trois modèles (Gamma, lognormale, normale) sont toutes différentes ;
- S2 : seule la loi Gamma propose une solution différente ;
- S3 : seule la loi lognormale propose une solution différente ;
- S4 : seule la loi normale propose une solution différente ;
- S5 : les solutions sont les mêmes quel que soit le modèle.

Sur la figure 4.16, des requêtes de distances plutôt longues entre origine et destination ont été testées. Dans la majorité des cas, les trois modèles mettent en valeur des solutions optimales identiques, en particulier pour le petit graphe SF. Néanmoins, une analyse plus fine du graphe CS montre que le choix du modèle statistique peut avoir un impact sur les solutions. Lorsque la précision de la discrétisation est faible, les modèles divergent dans 80% des cas. Un dernier résultat majeur nous permettra d'opérer une transition avec la séquence suivante : le temps de calcul. Le problème est bien connu pour les algorithmes stochastiques et l'algorithme SPOTAR ne déroge pas à la règle. Le tableau 4.2 montre les temps d'exécution de l'algorithme en fonction de différents paramètres.

Tab. 4.2.: Temps d'exécution de l'algorithme SPOTAR (en secondes), par groupe de requêtes

| Graphe | Variance | gam | low log | nor | gam | med log | nor | gam | hig log | nor |
|--------|----------|--------|------------|--------|--------|------------|--------|--------|------------|--------|
| | Modèle | | | | | | | | | |
| SF | 60 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | 120 | 0.03 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| | 600 | 0.40 | 0.40 | 0.44 | 0.40 | 0.41 | 0.42 | 0.44 | 0.47 | 0.43 |
| G10 | 60 | 0.30 | 0.24 | 0.26 | 0.27 | 0.27 | 0.29 | 0.28 | 0.29 | 0.29 |
| | 120 | 0.56 | 0.37 | 0.41 | 0.44 | 0.43 | 0.48 | 0.49 | 0.53 | 0.55 |
| | 600 | 9.27 | 5.71 | 6.51 | 6.37 | 6.28 | 7.21 | 7.65 | 8.21 | 8.55 |
| G20 | 60 | 4.07 | 3.50 | 3.81 | 3.98 | 4.00 | 4.10 | 3.57 | 4.07 | 3.47 |
| | 120 | 6.45 | 5.28 | 5.76 | 5.94 | 6.15 | 6.28 | 6.77 | 7.56 | 6.53 |
| | 600 | 78.57 | 60.76 | 70.67 | 79.61 | 78.33 | 92.74 | 98.98 | 108.05 | 98.42 |
| CS | 60 | 13.35 | 12.15 | 10.88 | 13.17 | 12.68 | 12.81 | 13.12 | 16.53 | 15.14 |
| | 120 | 17.55 | 13.48 | 14.46 | 15.88 | 16.62 | 15.36 | 19.18 | 22.06 | 18.19 |
| | 600 | 144.23 | 98.73 | 106.37 | 140.65 | 146.26 | 145.18 | 137.47 | 159.63 | 137.32 |

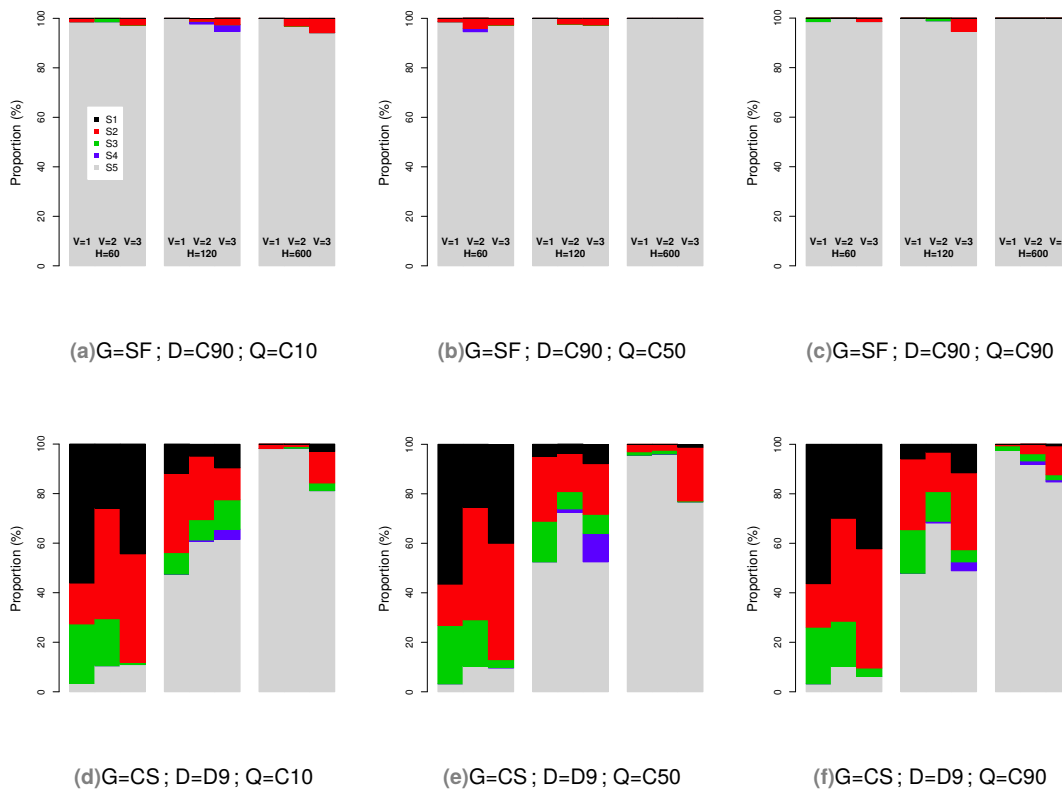


Fig. 4.16.: Proportions des 5 situations constituant le premier indicateur pour un sous-ensemble de configurations. H : nombre de points support (discrétisation) ; V : niveau de variance de 1 (faible) à 3 (élevé) ; G : type du graphe ; D : niveau de distance entre origine et destination (ex : D9 = 9ème décile) ; Q : probabilité d'être à l'heure (ex : C90 = 90ème centile).

Une lecture horizontale du tableau 4.2 montre l'effet des lois choisies et du niveau de variance du modèle statistique sur le temps de calcul. L'augmentation du niveau de variance entraîne un temps de calcul sensiblement plus long, en particulier pour les plus grands graphes. Une lecture verticale montre l'effet de la taille du graphe et de la précision de la discrétisation. On voit que ces deux paramètres revêtent une importance majeure dans les temps de calcul. Les plus hauts temps de calcul sont ainsi obtenus pour le graphe de Chicago (CS), avec une variance moyenne à élevée, et une discrétisation très précise. Les valeurs, qui atteignent les 150 secondes, sont prohibitives pour tout calculateur d'itinéraires du monde réel. L'accélération du traitement algorithmique s'est ainsi naturellement imposée comme la suite possible des travaux sur le plus court chemin stochastique.

4.3.2 Vers une accélération du PCC stochastique à l'aide des contractions hiérarchiques

Les contractions hiérarchiques constituent une des techniques d'accélération les plus performantes de l'état de l'art. Pour les graphes déterministes, des expérimentations récentes menées par Hannah Bast [36] rapportent un pré-traitement d'une dizaine de minutes et des requêtes résolues en quelques millisecondes pour des graphes routiers comportant des millions de nœuds et arcs. Des discussions parallèles avec l'industriel *Mappy* ont par ailleurs renforcé nos convictions quant à l'attractivité de la méthode. Plutôt que la course à la milliseconde, qui dépend également du

talent des informaticiens et des ressources disponibles, nous avons essayé d'analyser dans quelle mesure les contractions hiérarchiques pourraient être adaptées au cas stochastique. La technique des contractions hiérarchiques repose sur quelques principes simples, puis beaucoup de réglages. Les étapes de base sont les suivantes :

1. Une hiérarchisation des nœuds en fonction d'un coût de contraction,
2. Une contraction du graphe, qui suit l'ordre hiérarchique,
3. La résolution des requêtes dans le graphe hiérarchisé, à l'aide de recherches bidirectionnelles.

La phase de hiérarchisation des nœuds classe les différents sommets en fonction de leur importance dans le graphe. Ensuite, les nœuds seront visités par ordre d'importance, les nœuds les moins importants ayant le moins de chances d'être visités. Pendant cette phase de hiérarchisation, les nœuds sont contractés, c'est-à-dire virtuellement supprimés, y compris leurs arcs incidents, entraînant l'apparition de raccourcis. Le niveau de priorité d'un nœud est donc évalué à travers une **simulation de contraction** : il s'agit de mettre en valeur l'impact de la suppression du nœud sur des critères bien choisis. Le premier d'entre eux concerne la variation d'arcs au sein du graphe. La contraction d'un nœud peut entraîner la suppression d'arcs mais aussi l'ajout de nouveaux raccourcis. Un nœud sera jugé important si sa contraction entraîne un plus grand ajout de raccourcis que d'arcs supprimés. Un autre critère concerne ainsi le nombre d'arcs originaux que représente un nouveau raccourci généré. Cet indicateur renseigne sur la complexité de chaque arc du graphe résiduel. D'autre part, la notion de profondeur rentre également en jeu : plus l'indicateur est élevé, et plus le nœud aura été contracté tardivement par rapport à ses voisins. Finalement, le coût de contraction d'un nœud est évalué à travers une somme pondérée des différents critères. Plus ce coût est faible, plus le nœud sera susceptible d'être contracté de manière prématurée. Cette phase permet ainsi d'associer à chaque nœud un niveau de priorité. Sans rentrer dans les détails, cette brève présentation laisse apparaître les degrés de liberté et les stratégies d'optimisation possibles, à propos du choix des critères et de leur pondération.

La deuxième phase concerne la véritable contraction du graphe et l'ajout de raccourcis. Afin d'identifier les motifs possibles de contraction, un scénario simple est proposé : soit l_i le niveau de tout nœud i dans la liste de priorité, soient un nœud source s et un nœud cible d , nous posons u un nœud intermédiaire adjacent à s et d . La contraction du nœud u et l'ajout d'un raccourci pour relier directement s et d ne seront effectués qu'à deux conditions :

- Il possède un coût de contraction plus faible que les deux autres soit $l_u < \min\{l_s; l_d\}$;
- Il n'existe pas de meilleur chemin alternatif entre s et d .

La figure 4.17 illustre les différents cas possibles de notre scénario de contraction.

Le processus d'ajout de raccourcis consiste ensuite à boucler sur tous les nœuds dans l'ordre hiérarchique puis à étudier, à chaque itération, les nœuds adjacents du nœud courant qui lui sont supérieurs dans la hiérarchie. Si une configuration telle que celle présentée ci-dessus est présente (figure 4.17), un raccourci est ajouté. Ainsi est obtenu un graphe de recherche augmenté des raccourcis découlant de la contraction. Ce graphe enrichi va permettre une recherche bidirectionnelle dans deux sous-graphes :

1. Depuis la source, une recherche de PCC empruntant des séquences de nœuds de niveaux croissants (recherche *forward*),
2. À partir de la destination, une recherche *backward* empruntant des séquences de nœuds de niveaux décroissants (recherche *backward*).

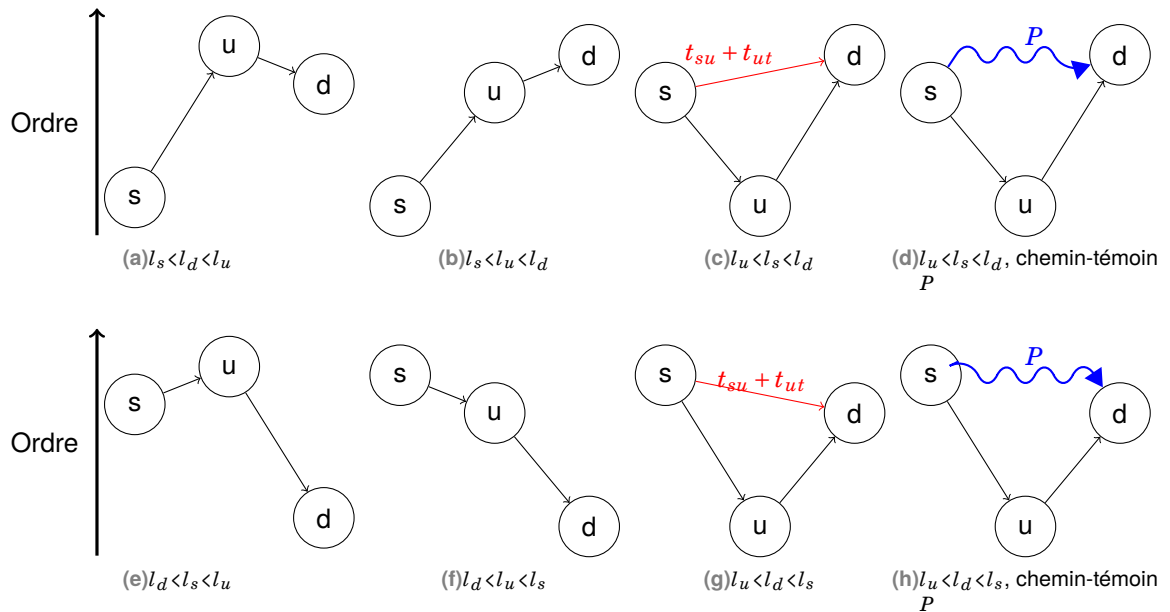


Fig. 4.17.: Configurations de hiérarchisation possibles. L'ajout de raccourcis n'est nécessaire que dans les cas (c) et (g)

La figure 4.18 illustre les deux composantes de la recherche bidirectionnelle. Le nœud u_m , qui a été contracté, permet, grâce à l'ajout d'un arc raccourci (u_l, u_n) , d'accélérer la recherche. Les deux recherches convergent en haut de la hiérarchie, aux nœuds les plus importants du chemin optimal et il s'agit alors de minimiser la somme de temps associés aux deux chemins partiels obtenus.

La méthode a été adaptée au cas dynamique, dans lequel les temps de parcours sont dépendants de la date de départ [40]. Les modifications consistent à agréger, concaténer des fonctions de temps de parcours, jusqu'à obtenir le profil optimal de temps d'arrivée à la destination. Nos contributions ont essentiellement porté sur l'adaptation des contractions hiérarchiques au cas stochastique : à la fin d'un algorithme comme SPOTAR, des politiques de routage sont disponibles, qui se caractérisent par un ensemble de sous-chemins non dominés associés à une distribution frontière et un vecteur identifiant les chemins optimaux pour chaque budget-temps. Adapter les contractions hiérarchiques au cas stochastique implique une manipulation des politiques de routage en particulier en cas de recherche bidirectionnelle dans un graphe contracté : fusion entre deux politiques de routage associées au même nœud, concaténation de politiques venant de recherches *backward* ou *forward*. La phase de hiérarchisation des nœuds n'induit pas de modification majeure : à partir d'une somme pondérée de critères, comme précédemment, un coût de contraction est associé à chaque nœud. En revanche, pour la contraction en tant que telle, la recherche d'un chemin-témoin doit s'effectuer dans un graphe stochastique dont les étiquettes sont des distributions de temps de parcours, évaluées entre elles grâce aux politiques de routage. Cette adaptation partage nombre d'aspects avec l'adaptation dynamique proposée par Batz *et al.* [40]. En particulier, l'espérance du temps de parcours sur la séquence entre le nœud source et le nœud cible est comparée avec le plus petit temps de parcours possible espéré sur un itinéraire alternatif. Si cette comparaison ne permet pas de mettre en exergue un chemin alternatif, l'ensemble des distributions doivent être comparées afin de mettre en lumière ou non un chemin-témoin dont la distribution de temps de parcours dominerait la séquence en question. La technique des contractions hiérarchiques stochastiques a été testée dans des conditions expérimentales comparables aux précédentes, en se limitant à des graphes de taille

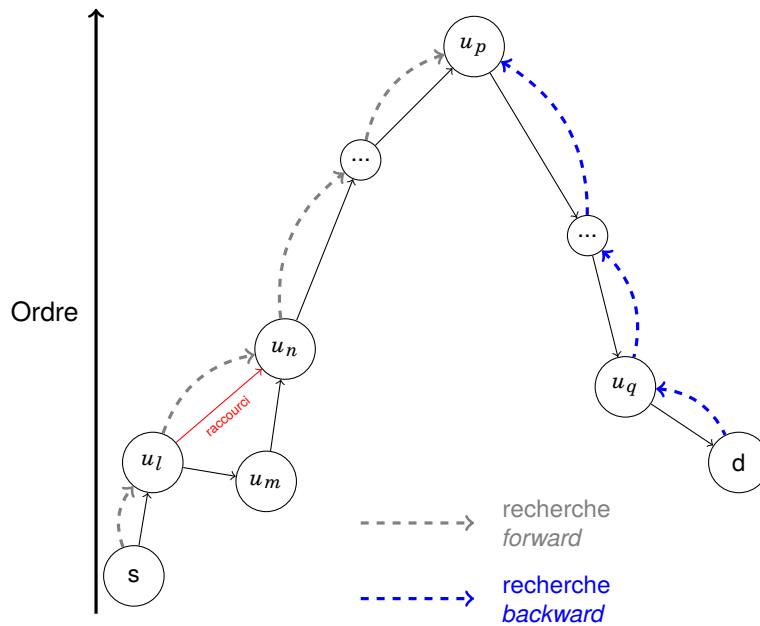


Fig. 4.18.: Type de chemin mis en valeur par la procédure des contractions hiérarchiques

modeste : le réseau de Sioux Falls (*SF*), les grilles *G10* et *G20* ainsi qu'une grille *G5* comparable à *SF*. Les distributions ont été discrétisées à l'aide de fonctions de masse en considérant des niveaux de variance et de précision graduels. En ce qui concerne la phase de pré-traitement, opérée hors ligne une fois pour toute, le temps maximum s'établit à 7h pour le plus grand graphe *G20*, avec une haute précision de discrétisation (600 points) et une forte variance sur les arcs. La taille du graphe est, plus que les autres, le critère le plus impactant en termes de pré-traitement. Pour le temps de calcul des requêtes, l'influence des différents facteurs est plus répartie : le temps de calcul est proportionnel à la taille du graphe mais la précision des distributions de temps de parcours a une importance encore plus forte sur les résultats. Le résultat principal concerne évidemment le facteur d'accélération par rapport à la méthode originale. La figure 4.19 rapporte les coefficients d'accélération de temps de calcul pour un lot de 1000 requêtes sélectionnées aléatoirement.

Les résultats sont très satisfaisants puisque l'accélération est présente dans 99% des cas. La technique des contractions hiérarchiques est deux fois plus rapide dans 89% des instances et dix fois plus rapide dans 35% des cas. Un facteur 50 peut rapidement être atteint pour de grands graphes. L'amplitude de l'accélération dépend de la taille du graphe et les performances pour le graphe *G20* laissent entrevoir un fort potentiel pour des graphes d'échelle urbaine. Toutefois, le temps de calcul de la phase de pré-traitement ne doit pas être sous-estimé : nos interactions avec l'industriel MAPPY nous ont permis d'apprécier les contraintes d'un portage industriel d'une solution : la logique temps-réel peut amener à mettre à jour la hiérarchie à chaque pas de temps (ex : 6 min). L'optimisation des implémentations, les approches de parallélisation des calculs, mieux maîtrisées par les industriels, devraient pouvoir permettre le passage à l'échelle de telles techniques d'accélération dans un certain nombre de calculateurs d'itinéraires, au cours des prochaines années.

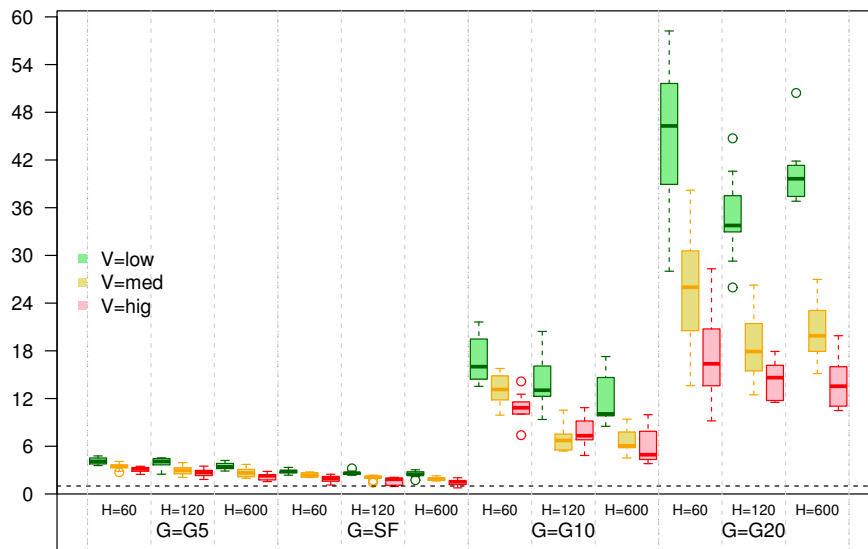


Fig. 4.19.: Baisse du temps de calcul de l'algorithme SPOTAR grâce aux contractions hiérarchiques. Facteur d'accélération d'un lot de 1000 requêtes sélectionnées aléatoirement.

4.4 Conclusion et perspectives

Les travaux de recherche présentés dans ce chapitre sont susceptibles d'être intégrés à court terme dans la plateforme TEMPUS de calculs de plus court chemins. En ce qui concerne l'électromobilité, le travail consisterait à intégrer le graphe énergétique dans les bases de données puis à coder l'algorithme de résolution du plus court chemin énergétique avec recharge en tant que nouveau *plugin*. Ainsi l'utilisateur pourrait sélectionner une option spécifique au véhicule électrique qui délivrerait une solution minimisant la consommation énergétique et éventuellement d'autres critères spécifiés. L'optimisation multi-critères fait clairement partie des perspectives de nos travaux. Pour le cas des réseaux multimodaux, nous avons déjà abordé la question au cours d'un travail commun avec Sergey Kovalev [240]. Dans cette contribution, les objectifs sont classés par priorité et nous proposons une résolution à l'aide d'un ordre lexicographique. Un solveur est équipé de plusieurs programmes linéaires booléens qui résolvent successivement des problèmes mono-objectif avec une déviation par rapport à l'optimum du problème précédent, de priorité plus importante. Les résultats sont prometteurs et permettraient d'implémenter, au sein de TEMPUS, des curseurs réglant les préférences de l'utilisateur quant à la personnalisation de la solution calculée. Néanmoins, comme les équipes de MAPPY l'ont suggéré, de telles approches, certes intéressantes d'un point de vue recherche, n'auraient peut-être pas un grand intérêt pour des usagers qui préfèrent autant choisir parmi une liste de solutions mono-objectif différentes que d'avoir une solution trop complexe clé en main. D'ailleurs les discussions avec les partenaires industriels soulignent le décalage entre les objectifs des chercheurs, souvent ambitieux et long terme, et les impératifs du monde industriel. Nos travaux sur les contractions hiérarchiques stochastiques ont cherché à concilier les deux visions même si le pré-traitement est encore trop coûteux pour une opérationnalisation de la solution. Gageons que les travaux futurs sur les stratégies de hiérarchisation devraient pallier ce manque d'efficacité. En testant les algorithmes sur des graphes de grande taille (échelle d'un pays, voire d'un continent), l'impact du choix des distributions statistiques choisies sur les sorties du modèle devrait apparaître de façon plus claire et justifier l'usage de distributions raffinées, comme les lois

Gamma ou le système Halphen, au lieu de l'éternelle loi normale. Surtout, l'usage d'algorithmes stochastiques devrait mener à l'élaboration d'indicateurs de confiance quant aux solutions fournies à l'utilisateur, en indiquant les solutions à temps de parcours garanti, par exemple.

Chapitre d'ouverture à la santé connectée : fouille de données et aide à la décision pour la prévention du risque suicidaire.

” *Le plus terrible dans ce monde c'est que chacun a ses raisons*

— **Jean Renoir (sous les traits d'Octave dans le film *La règle du jeu*)**

Cinéaste

Depuis 2016 et mon arrivée à l'IMT Atlantique, j'ai eu l'opportunité de m'ouvrir à un nouveau domaine d'application, la santé, et plus précisément le domaine de la santé mentale. Ces recherches s'effectuent dans le cadre de la co-direction de thèse de Sofian Berrouiguet, psychiatre au CHU Brest, et spécialiste de l'usage des nouvelles technologies pour la prévention et l'intervention en santé mentale. Nous travaillons plus précisément sur la question de la prévention du risque suicidaire avec l'ambition de développer un système d'aide à la décision dynamique, qui s'appuie aussi bien sur la collecte de données temps réel grâce aux nouvelles technologies que sur leur analyse et la fouille de grandes bases de données existantes, pour extraire de nouvelles connaissances sur un enjeu majeur de santé publique. Nous présentons dans ce manuscrit les principaux objectifs du projet ainsi qu'une première contribution au domaine à travers la fouille d'une base de données de patients reçus dans des hôpitaux européens après une tentative de suicide (TS).

5.1 Actualités de la prévention du risque suicidaire

5.1.1 Nouvelles technologies et prévention

On recense plus de 800 000 décès par suicide par an dans le monde et l'on estime qu'à chaque décès équivaut, de manière comptable, environ 20 autres tentatives. En France, le suicide est l'une des premières causes de mort évitable. Les chiffres nationaux sont de 15.7 suicides pour 100 000 habitants (25.3 pour les hommes et 80 pour les femmes) soit plus de 10 000 morts par an. Les décès par suicide ne représentent que la partie immergée de l'iceberg et le nombre annuel de tentatives de suicide (TS) en France est estimé entre 180 000 et 200 000. Depuis les années 2000, des progrès significatifs ont été réalisés dans le champ de la prévention du suicide. Ainsi, une stabilisation voire une légère décroissance de la mortalité par suicide est observée dans tous les pays occidentaux [49]. Néanmoins, les conduites suicidaires sont des phénomènes éminemment polyfactoriels et il est donc toujours difficile de définir des stratégies univoques à partir d'une tentative de suicide.

En France, comme ailleurs dans le monde, les tentatives de propositions de soins spécifiquement destinés aux suicidants ont toutes peine à montrer leur efficacité sur la diminution des récurrences suicidaires, voire sur la mortalité par suicide. Si une première tentative représente un très grand facteur de risque [155], d'autres critères doivent être pris en compte dans le processus de prise de décision thérapeutique et de prévention. Les facteurs de risque suicidaire ont déjà été étudiés parmi des populations de suicidants, qui sont une cible privilégiée pour l'élaboration de stratégies de prévention et d'intervention. Si les échelles d'évaluation du risque permettent d'identifier des facteurs clés [77], les limites des procédures actuelles sont bien connues et la confusion entre évaluation du risque et prédiction ne facilite pas l'émergence de dispositifs efficaces d'aide à la décision pour les cliniciens [314]. Les procédures actuelles recommandent que chaque patient reçu dans les services hospitaliers, présentant un cas ou une suspicion d'auto-mutilation, doit passer une évaluation psychologique avant sa sortie afin d'évaluer le risque de récurrence [77]. Cette évaluation doit également permettre au clinicien de choisir le meilleur traitement possible en fonction des facteurs de risque identifiés et des protocoles en vigueur. Néanmoins, de nombreuses études mettent en exergue la difficulté de telles évaluations, menées dans un contexte d'urgence qui rend peu homogène le remplissage des dossiers électroniques du patient. Dans le cadre de la thèse de doctorat de Sofian Berrouiguet, psychiatre au CHU de Brest, nous proposons de développer un système dynamique d'aide à la décision qui va combiner la fouille de grandes bases de données patients avec une auto-évaluation dynamique et individuelle qui repose sur les nouvelles technologies. Bien que des travaux attestent de l'efficacité de l'auto-évaluation par les patients, cette dernière est rarement mise en oeuvre [267]. Aujourd'hui, la démocratisation de l'internet mobile offre pourtant un parfait terrain d'application à cette auto-évaluation, en particulier après une prise en charge.

On regroupe sous le concept d'évaluation écologique instantanée (ou *Ecological Momentary Assessment* (EMA)) l'ensemble des protocoles qui permettent à un participant d'envoyer au cours du temps des informations, évaluations, sur ses symptômes, son comportement, son état mental ou autres activités, élargissant ainsi les possibilités de recueils de données réelles et temps-réel [399]. Les études utilisant l'EMA sont capables de recueillir des symptômes au moment précis où ils apparaissent, ou peu après. Internet et les technologies mobiles sont les instruments idéaux d'une telle auto-évaluation : un *smartphone* est généralement gardé sur soi toute la journée et emmené partout. Depuis sa création, l'EMA a été utilisée pour investiguer une large variété de conditions psychiatriques comme les troubles affectifs [329], la psychose [318], l'anxiété et la dépression [320], les troubles alimentaires [238], ou les abus de substances [267]. Dans un article récent, nous avons également montré son efficacité pour la supervision des prescriptions [48] tandis que l'acceptabilité patient d'une application mobile pour l'auto-évaluation de la santé mentale a été positivement évaluée par nos collègues espagnols [35]. La figure 5.1 montre un exemple d'application mobile, MEmind, qui permet entre autres de suivre l'état de santé de patients ayant été reçus en hôpital psychiatrique pour des troubles de la santé mentale.

D'autre part, Husky *et al.* ont conduit une étude qui illustre les apports de l'EMA pour l'identification de l'idéation suicidaire chez une population de patients à risque. Bien sûr, on pourrait penser que la répétition des auto-évaluations par un patient suicidaire, chez lui, peut poser question, mais une étude récente a montré que des évaluations répétées de pensées suicidaires n'avaient pas d'effets sur leur intensité ou leur fréquence [111]. Les données provenant des EMA sont utiles aussi bien pour les médecins que pour les patients, qui pourraient ainsi prendre part au processus de décision [437]. Finalement, l'ensemble des études s'accordent sur le potentiel de l'EMA et suggèrent que les technologies mobiles sont encore sous-utilisées pour la détection d'états cognitifs à haut risque dans



Fig. 5.1.: Exemple d'application mobile développée à Madrid : MEmind.

la vie quotidienne des patients [111]. Pourtant, alors que la plupart des praticiens hospitaliers ont un usage quotidien des dossiers électroniques de patients dans les services d'urgences ou les unités psychiatriques, trop peu d'institutions tirent réellement partie des avancées technologiques pour la prise de décision. Les nouvelles tendances de la recherche en prévention du risque suicidaire, dans lesquelles notre équipe projet s'inscrit, combinent la santé connectée et la fouille de données : les nouvelles technologies permettent de recueillir une information temps-réel tandis que les traitements des bases de données patients et des données temps-réel concourent à l'élaboration d'approches prédictives pour la prise en charge des patients. Ce nouveau paradigme, à l'instar de la mobilité connectée souvent dénommée mobilité intelligente, peut être défini selon le concept de santé intelligente (*i-health*). Il doit permettre de développer des stratégies d'intervention personnalisées mais également d'identifier des nouveaux facteurs de risques, ou mécanismes temporels, à intégrer aux stratégies de prévention. Les analyses multi-dimensionnelles de grandes bases de données, aujourd'hui disponibles, ont quant à elles le rôle de faire émerger des conclusions robustes qui viendront confirmer, infirmer des avis empiriques de cliniciens, ou peut-être révéler des déterminants cachés du risque suicidaire.

5.1.2 Vers un système d'aide à la décision dynamique

Les systèmes d'aide à la décision assistent les praticiens au cours du processus de prise en charge du patient. En s'appuyant sur des études cliniques publiées, ces outils orientent le décideur vers la meilleure suite à donner pour le patient reçu dans ses services. Un système d'aide à la décision clinique, ou *Clinical Decision Support Systems (CDSS)*, est un système d'information couplé au dossier électronique du patient et donc directement accessible au praticien. L'ambition de nos recherches sur ce thème concerne le *design* et l'implémentation d'un CDSS dynamique qui inclut les données individuelles du patient, un moteur de calcul de règles de décision et une base de

connaissances qui permettent l'élaboration d'un plan de prise en charge personnalisé [76]. La figure 5.2 illustre notre démarche méthodologique dans laquelle la fouille de données occupe une place centrale.

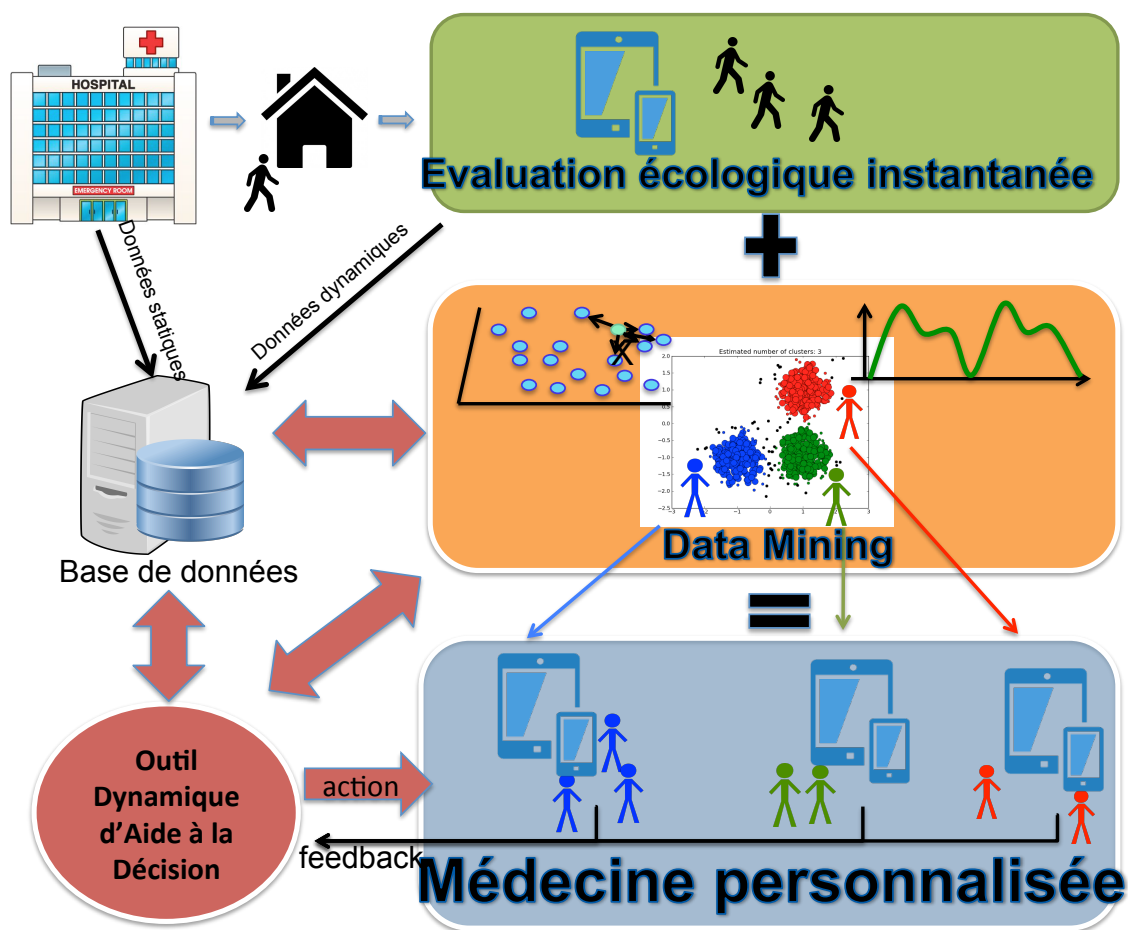


Fig. 5.2.: Système d'aide à la décision pour la prévention du risque suicidaire.

Dans le domaine de la santé mentale, et plus particulièrement de la gestion du risque suicidaire, ces techniques sont encore trop peu présentes [219] alors que de tels systèmes apporteraient un soutien ô combien précieux aux praticiens, parfois esseulés face à la complexité des situations : traitement le plus approprié, dose à prescrire, durée etc. La fonctionnalité première d'un CDSS serait d'identifier les risques encourus en terme de récurrence, voire de multi-récurrence, et ce dans une période de temps à venir. Le recueil de données, passives et actives, sur les patients, après leur prise en charge à l'hôpital, doit alimenter le CDSS et améliorer la connaissance des praticiens, et donc leur futures pratiques, en particulier sur des facteurs de risques et leur combinaison [175].

5.2 Fouille de données pour la prévention du risque suicidaire

Nous présentons ici une première contribution au domaine à travers l'adaptation d'un article en cours de soumission (Berrouiguet, Billot *et al.*) à une revue de psychiatrie. L'article regroupe les

principaux résultats d'un processus complet de fouille de données sur un échantillon de suicidants reçus dans des hôpitaux européens. L'objectif est d'identifier des groupes de patients similaires puis d'identifier un certain nombre de facteurs de risque associés au nombre de tentatives de suicide par patient. L'hypothèse est de considérer la fouille de données comme un outil apte à mettre en valeur des groupes pertinents de patients et leurs facteurs de risque associés pour certaines variables d'intérêt. Par rapport à la figure 5.2, le travail est relatif à la brique *data mining* (couleur orange), il n'intègre pas encore les données temps réel provenant d'une évaluation écologique instantanée (EMA).

5.2.1 Recrutement des patients et évaluation clinique

5.2.1.1 Recrutement des patients

La population est composée de patients âgés de 18 ans ou plus, reçus pour des tentatives de suicide dans deux hôpitaux universitaires à Madrid (hôpital Ramon y Cajal), Espagne, et Montpellier (hôpital Lapeyronie), France, entre 1994 et 2006. Les deux hôpitaux couvrent une population de 500 000 habitants à Madrid, et 400 000 à Montpellier. Ils participent, sous la houlette des professeurs Baca-Garcia et Courtet, au consortium européen EURECA, qui a recueilli des données sur plus de 3000 suicidants au cours des dernières années [276]. Les équipes médicales du consortium se sont accordées pour employer les mêmes méthodes cliniques et des procédures d'évaluation comparables. Les données sociodémographiques et cliniques des patients ont été fusionnées dans une base de données commune qui intègre également les résultats des questionnaires d'évaluation. La définition d'une tentative de suicide retenue est la suivante : *"a potentially self-injurious behavior with a nonfatal outcome, for which there is evidence (either explicit or implicit) that the person intended at some (nonzero) level to kill himself/herself."* Les études ont été approuvées par les différents comités d'éthique et conduites selon les principes de la déclaration d'Helsinki, à propos des principes éthiques de la recherche médicale. Tous les participants ont signé un document de consentement après une explication précise des objectifs de l'étude et des procédures. Le jeu de données complet inclut des patients ne présentant pas d'historiques de tentatives de suicide, des donneurs de sang, et autres patients de "contrôle". Dans notre étude, nous nous sommes concentrés sur la seule population des suicidants.

5.2.1.2 Procédure clinique

Des psychiatres ou psychologues formés ont questionné les patients lors de leur prise en charge. Les procédures se fondent sur l'échelle Columbia (*Columbia Suicide History Form*), qui consiste en une série de questions permettant d'évaluer l'idéation suicidaire chez les patients à risque. Les données ont été collectées grâce à l'application MEmind (figure 5.1), qui a été développée pour recueillir des données électroniques sur les patients mais également des données écologiques (EMA). L'interface patients (EMA), n'a pas été utilisée dans cette étude, les cliniciens se contentant de compléter le dossier électronique des patients à l'aide des formulaires qui leur sont dédiés. L'application collecte des données sociodémographiques, de diagnostic, et pharmacologiques au sein du protocole d'évaluation du suicide. Les variables sociodémographiques incluent notamment l'âge au moment de l'épisode suicidaire, le genre, la profession, le statut marital, la situation professionnelle, le nombre d'enfants et le niveau d'éducation. Des informations sur les historiques familiaux quant à la question du suicide, l'âge de la première tentative, le caractère violent de l'acte, sont également récoltées. Les

versions française et espagnole du questionnaire neuropsychiatrique international MINI ([253]) ont été passées afin d'obtenir des diagnostics psychiatriques dans les catégories suivantes : troubles de l'humeur (ex : troubles bipolaires, dépression), anxiété, troubles obsessionnels compulsifs, drogues et alcool, troubles psychotiques, alimentaires, somatiques, etc. Le psychiatre en charge du diagnostic complète à chaque fois les informations à l'aide du dossier médical et potentiellement d'informations venant des proches. Le risque suicidaire a été évalué à l'aide de l'échelle SIS (*Suicide Intent Scale*, [43]), une échelle de risque semi structurée à 15 items qui produit un score globale de sévérité quant à l'intention suicidaire. La tentative en tant que telle est évaluée à l'aide de l'échelle RRRS, ou *Risk-Rescue Rating Scale*, qui est un questionnaire de 10 items qui mesure la sévérité de l'intention et du geste suicidaire au regard de la létalité et de la vraisemblance d'une intervention de sauvetage au moment du geste [459]. L'impulsivité du patient a été mesurée à l'aide l'échelle BIS10, pour *Barratt Impulsiveness Scale*, une série de 34 questions qui mesurent la propension du sujet à prévoir ses gestes, son comportement, dans diverses situations [346].

5.2.2 Méthodologie statistique

5.2.3 Qualification des données brutes et description du jeu de données final

Un processus robuste de qualification des données a été mené afin d'assurer la consistance et la complétude de la base avant la conduite de toute analyse statistique. Bien que les données sont censées être recueillies selon des processus identiques, une variabilité de la qualité est inévitable de par la diversité des équipes de recueil. De plus, le contexte particulier de l'accueil en urgence de personnes très fragiles ne facilite pas un recueil exhaustif. La base de données originale contenait 2802 patients et 263 variables. Dans un premier temps, la question des données manquantes a mobilisé notre attention. En accord avec les spécialistes, il a été décidé de ne garder que les variables qui satisfaisaient un taux de complétion minimum de 70 %. Dans cette base de données, il a été très difficile de saisir et modéliser la structure et la typologie des données manquantes. Pour cette raison, aucune méthode d'imputation n'a été envisagée et un simple seuil strict a été respecté. En revanche, de nouvelles variables ont pu être créées. Par exemple, les 34 questions du questionnaire d'impulsivité de Barratt (BIS10) ont été traitées afin de créer trois scores d'impulsivité : impulsivité motrice, attentionnelle, ou de non planification [346]. Pour chaque question du BIS10, le score prend une valeur entre 1 (faible impulsivité) et 4 (forte impulsivité). Le score total s'étend de 34 à 136 et les sous-scores s'inscrivent dans l'intervalle [11 – 44] pour les impulsivités motrice et attentionnelle, et [12 – 48] pour la dernière catégorie. Ainsi, trois nouvelles variables sont ajoutées au tableau de départ. En appliquant le filtre strict de 70% de complétion, 26 variables sont conservées. En ce qui concerne les patients, il a été décidé de ne conserver que les suicidants présentant un taux de complétion de 100% pour ces 26 variables. Dans une dernière étape, six variables ont été supprimées pour des questions de redondance ou de pertinence. Finalement, le processus de qualification conduit à un jeu de données final de 681 patients et 20 variables d'intérêt. Le tableau 5.1 rassemble les principales caractéristiques sociologiques du jeu de données final.

De manière générale, les patients sont plutôt jeunes (âge médian : 40 ans), plutôt des femmes, avec une situation professionnelle de salarié(e) et un statut marital varié. Le tableau 5.2 résume les caractéristiques cliniques des patients.

| Caractéristiques générales | | | |
|-----------------------------------|----------------------|-----------------------|-----------------------|
| Variables qualitatives | | | |
| Variable | Modalité | <i>n</i> | % |
| Genre | Féminin | 493 | 72.4 |
| | Masculin | 188 | 27.6 |
| Statut marital | Célibataire | 239 | 35.1 |
| | Marié(e) | 240 | 35.2 |
| | Séparé(e)/Divorcé(e) | 181 | 26.6 |
| | Veuf/Veuve | 21 | 3.1 |
| Enfants | Non | 272 | 39.9 |
| | Oui | 409 | 60.1 |
| Niveau d'éducation | Faible | 31 | 4.6 |
| | Intermédiaire | 368 | 54.0 |
| | Elevé | 282 | 41.4 |
| Situation professionnelle | Salarié(e) | 451 | 66.2 |
| | Sans emploi | 110 | 16.2 |
| | En incapacité | 41 | 6.0 |
| | Retraité(e) | 79 | 11.6 |
| Variables quantitatives | | | |
| Variable | Médiane | <i>Q</i> ₁ | <i>Q</i> ₃ |
| Age | 40.6 | 28 | 49.6 |

Tab. 5.1.: Caractéristiques sociologiques du jeu de données post-filtrage de 681 patients reçus après une tentative de suicide. *Q*₁ et *Q*₃ : 1er et 3ème quartiles, respectivement. BIS10 : *Barrat Impulsiveness Scale*. Echelle d'impulsivité de Barrat.

Comme le montre le tableau 5.2, la plupart des patients inclus dans l'analyse finale ont un historique de trouble mental, en particulier des épisodes sérieux de dépression (70.3 %) ou des troubles bipolaires (23 %).

5.2.4 Méthodes de fouille de données

Tout d'abord, des statistiques descriptives, comprenant des analyses à une ou deux dimensions des différentes variables qualitatives ou quantitatives, ont été effectuées. Des tests d'hypothèse de type Fisher-Snedecor sont appliqués si nécessaire pour décider d'une différence entre deux sous-groupes (ex : homme/femme) [381]. Ensuite, une approche non supervisée a été conduite afin d'extraire des structures homogènes du jeu de données, sans hypothèse *a priori*. L'objectif est de mettre en lumière, potentiellement, des groupes de patients partageant des caractéristiques similaires, puis d'interpréter ces groupes avec les experts en psychiatrie. La méthode s'appuie sur deux étapes. En premier lieu, une analyse des correspondances multiples (ACM) a été effectuée sur les seules variables qualitatives. Cette méthode factorielle, bien adaptée à l'analyse de questionnaires, représente les individus dans un nouvel espace où chaque dimension est une combinaison des variables de départ. Les variables quantitatives comme l'âge ne sont pas utilisées pour le calcul des composantes principales, mais projetées ensuite sur le plan factoriel. Cette méthode est très utile en tant que technique de réduction de dimension. Elle permet également de représenter les individus dans un nouvel espace auquel on peut facilement associer une métrique, distance entre individus. C'est précisément ce que fait la méthode HCPC (*Hierarchical Clustering on Principal Components*), en partant d'une représentation des individus dans le nouvel espace factoriel. Une classification ascendante hiérarchique permet de regrouper itérativement les paires de clusters les plus proches, en partant de singletons (chaque individu est seul dans son groupe). La méthode construit un arbre

| Caractéristiques cliniques | | | |
|--|----------|-----------------------|-----------------------|
| Variables qualitatives | | | |
| Variable | Modalité | <i>n</i> | % |
| Trouble mental connu | Non | 6 | 0.9 |
| | Oui | 675 | 99.1 |
| Comportement suicidaire familial connu | Non | 424 | 62.3 |
| | Oui | 257 | 37.7 |
| Episode dépressif connu | Non | 199 | 29.2 |
| | Oui | 482 | 70.8 |
| Trouble bipolaire connu | Non | 521 | 76.5 |
| | Oui | 160 | 23.5 |
| Trouble dysthymique connu | Non | 651 | 95.6 |
| | Oui | 30 | 4.4 |
| Trouble Compulsif Obsessionnel connu | Non | 623 | 91.5 |
| | Oui | 58 | 8.5 |
| Trouble alimentaire connu | Non | 571 | 83.8 |
| | Oui | 110 | 16.2 |
| Prise d'alcool/drogue par le passé | Non | 465 | 68.3 |
| | Oui | 216 | 31.7 |
| Prise de substance | Non | 586 | 86.0 |
| | Oui | 95 | 14.0 |
| Alcoolisme | Non | 503 | 73.9 |
| | Oui | 178 | 26.1 |
| Variables quantitatives | | | |
| Variable | Médiane | <i>Q</i> ₁ | <i>Q</i> ₃ |
| Nombre de TS | 2 | 1 | 3 |
| BIS10 impulsivité motrice | 26 | 22 | 30 |
| BIS10 attention | 27 | 23 | 30 |
| BIS10 non-planification | 28 | 24 | 31 |

Tab. 5.2.: Caractéristiques cliniques du jeu de données final.

hiérarchique nommé dendrogramme, qui permet une interprétation visuelle du jeu de données et de la proximité entre individus. La visualisation du dendrogramme montre l'émergence de groupes de patients partageant des facteurs de risque similaires. Cet outil, plus que toute autre méthode sophistiquée, s'impose naturellement lors de l'interaction avec des non statisticiens, experts du domaine d'application qui peuvent notamment orienter le choix du nombre final de groupes à retenir. Chaque cluster peut ensuite être interprété à travers la significativité de son association avec les modalités des différentes variables qualitatives et quantitatives de départ (*v-test*). Un deuxième objectif a consisté à s'intéresser à une variable d'intérêt à expliquer, en l'occurrence le nombre de tentatives de suicide. Un partitionnement récursif du jeu de données de départ est effectué en séparant itérativement les patients en sous-populations, en fonction de seuils sur les variables les plus discriminantes. Un arbre de régression est obtenu et le nombre moyen de tentatives de suicide par sous-groupes de patients est obtenu par une série de tests binaires sur les variables prédictives de départ.

5.2.5 Résultats

5.2.5.1 Clustering des patients

Le premier type de résultats concerne une approche de réduction de dimension, non supervisée, pour faire apparaître des groupes de patients homogènes. L'ACM mène à une représentation des individus en fonction de nouvelles dimensions combinant les variables de départ. Les variables qualitatives sont donc transformées en variables continues et la méthode HCPC opère un clustering hiérarchique sur les coordonnées factorielles. La figure 5.3 montre le dendrogramme résultant de l'approche hiérarchique ainsi qu'une projection en deux dimensions des individus sur le premier plan factoriel, incluant une visualisation des clusters. La structure de l'arbre (en terme de gain d'inertie dans la hiérarchie) et une discussion avec les experts en santé mentale a mené à la sélection de trois clusters.

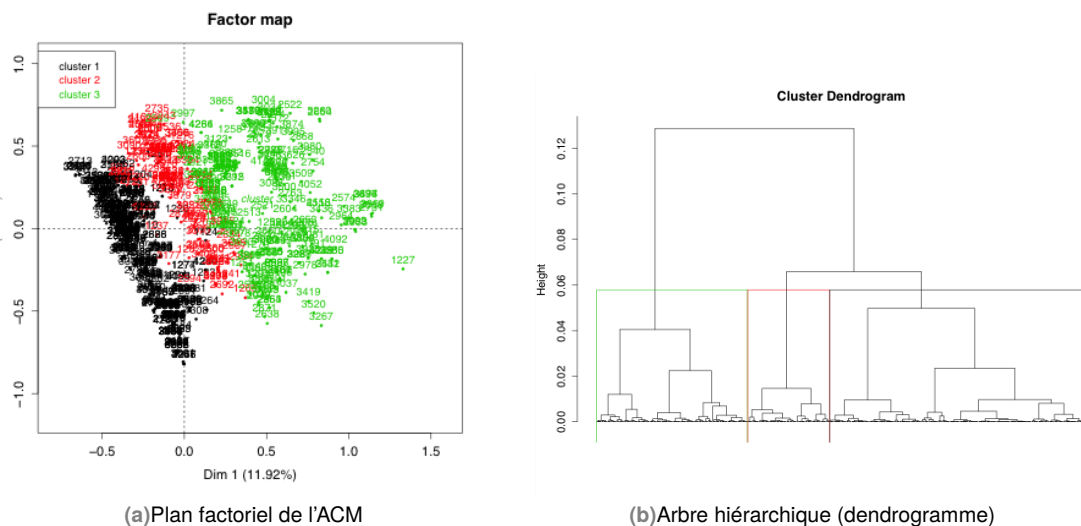


Fig. 5.3.: Approche de réduction de dimension : projection des individus statistiques sur les deux premières dimensions et dendrogramme menant au choix de trois groupes.

Une analyse approfondie des trois groupes a été menée pour la phase d'interprétation. Des tests d'association mettent en exergue les modalités sur- ou sous- représentées au sein des trois groupes. Les principales conclusions sont les suivantes : le groupe 1 correspond à un profil moyen plutôt féminin ($p < 0.001$), sans troubles bipolaires ni prises de drogues, substance, ou consommation d'alcool ($p < 0.001$). En revanche, cette population a en moyenne déjà subi des épisodes de dépression ($p < 0.001$) ou des troubles mentaux divers ($p < 0.05$). Le troisième groupe s'oppose au premier en exhibant un profil de patients plutôt masculin ($p < 0.001$), consommateur d'alcool, substances, ou drogues ($p < 0.001$). Les patients de ce groupe sont souvent célibataires et sans enfants ($p < 0.05$), et ne sont pas associés à un historique d'épisode dépressif connu ($p < 0.05$). Entre ces deux groupes, on trouve le groupe 2, neutre en termes de genre, mais marqué par des personnes présentant une incapacité au travail ($p < 0.05$), un faible niveau d'éducation ($p < 0.05$) avec de possibles troubles bipolaires ($p < 0.001$) mais aucune consommation des produits évoqués précédemment ($p < 0.001$). D'autre part, ces personnes ne présentent pas d'autres troubles de la santé mentale ou d'épisodes dépressifs connus ($p < 0.001$).

Ces premiers résultats sont d'intérêt pour les spécialistes car ils imposent le genre comme un marqueur significatif des populations de suicidants. D'autre part, des aspects connus empiriquement se retrouvent dans les modalités significatives, en particulier concernant les personnes en incapacité. Les conclusions de l'approche non supervisée valident d'ores et déjà la pertinence du *data mining* pour l'identification des facteurs de risque, ou de sous populations de patients qui pourraient faire l'objet de stratégies d'intervention et de prévention ciblées.

5.2.5.2 Identification des facteurs de risques pour le nombre de TS par arbres de régression

La deuxième phase de l'analyse a résidé dans l'identification de facteurs associés à un risque élevé de multi-récidives. La variable à expliquer est le nombre de tentatives de suicides déjà effectuées, en comptant celle pour laquelle le patient est reçu dans les services hospitaliers. Au vu des résultats du clustering précédent, nous avons choisi de dupliquer l'analyse sur deux sous-populations constituées des hommes et des femmes. La figure 5.4 montre l'arbre de régression résultant pour la population masculine tandis que la figure 5.5 dépeint un arbre sensiblement différent pour la population féminine.

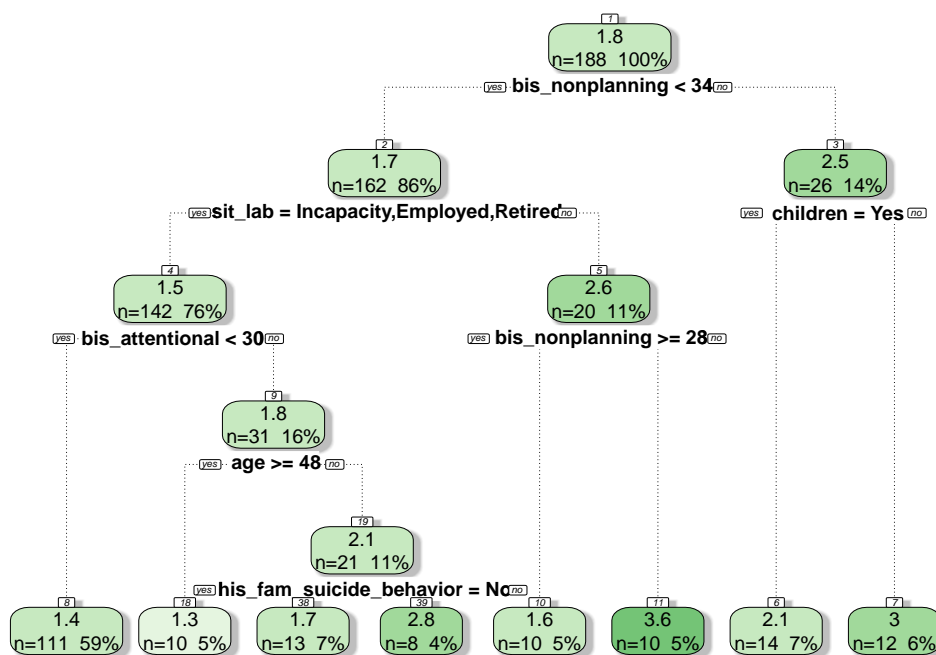


Fig. 5.4.: Arbre de régression de la variable "nombre de TS" pour la population masculine.

L'arbre de régression des patients masculins fait apparaître des facteurs de risque propres aux hommes comme le statut professionnel ou un historique de comportement suicidaire dans la famille (nombre de TS moyen 2.8 vs. 1.7 sinon, $p < 0.05$). D'autre part, le fait de ne pas avoir d'enfants entraîne un sur-risque, qui n'apparaît pas chez les femmes.

Pour les deux sous-populations, il est notable de voir que les scores d'impulsivité sont des facteurs explicatifs du nombre de TS par patients. Pour les femmes, la notion de troubles alimentaires

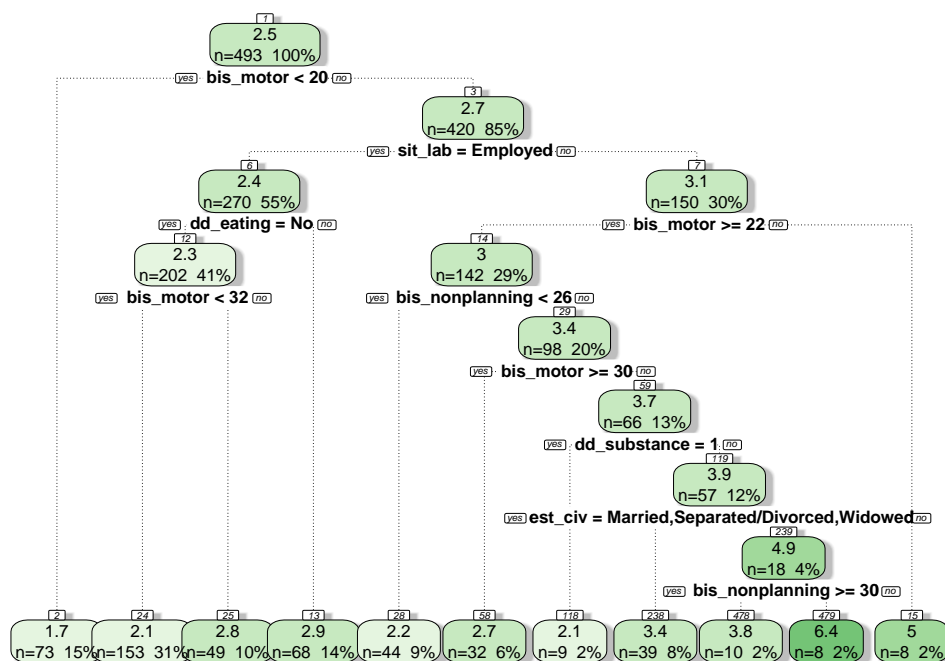


Fig. 5.5.: Arbre de régression de la variable "nombre de TS" pour la population féminine.

émerge, ce qui constitue un résultat important. En effet, pour les femmes salariées présentant une impulsivité motrice supérieure au score moyen, le nombre moyen de TS est de 2.9 pour les 68 femmes présentant des troubles alimentaires contre 2.3 pour les 202 autres ($p < 0.05$).

5.3 Discussion

5.3.1 Limites et perspectives du travail

Une évaluation systématique des patients pendant leur prise en charge a permis de constituer, au fil des années, une base de données qui a atteint une taille critique pour rendre significatif l'emploi de techniques de fouille de données. Dans la contribution présentée dans le manuscrit d'HDR, nous avons été capables d'identifier des clusters de patients suicidants puis d'expliquer le nombre de TS effectuées au cours d'une vie en fonction d'un certain nombre de facteurs de risque. Ces analyses ont déjà permis d'apporter un écot à l'état de l'art très riche sur la prévention du risque suicidaire. Certains facteurs de risque identifiés, comme les troubles alimentaires pour les femmes, ne sont pas aujourd'hui présents dans tous les interrogatoires cliniques.

Les limites de cette première étude sont de plusieurs ordres. Tout d'abord, on pourrait critiquer le fait que cette dernière ne se limite qu'à une population de suicidants : en effet, la base de données n'inclut que des patients ayant véritablement tenté de mettre fin à leurs jours. Nous pensons en effet, à l'instar de Zalsman *et al.* que les suicidants constituent une population toute indiquée pour ce genre d'études en ce sens qu'elle requiert l'attention maximale du personnel hospitalier à cause du risque de récurrence [475]. D'autre part, les procédures de traitement et de suivi sont très bien définies

pour cette population et un traitement approprié a pu être trouvé pour chacun des patients. Appliquer les mêmes analyses à une population de patients à risque (niveau élevé d'idéation suicidaire) mais vierge de toute TS aurait pu être possible, mais en la considérant comme une population à part, qui nécessite ses propres modèles. Dans les perspectives actuelles du travail, une nouvelle base de données de suicidants devrait nous permettre d'analyser la temporalité des TS plutôt que leurs valeurs absolues. Cette nouvelle finesse dans les données devrait ouvrir de nouveaux champs d'investigations.

L'étude présentée illustre d'ailleurs le besoin de données de très haute qualité afin d'extraire des profils robustes de patients à risque. À partir d'un jeu de données initial de 2802 patients et 263 variables, le processus de qualification et filtrage n'a conservé que 681 patients et 20 variables, pour lesquels les informations étaient complètes. Bien que ce volume de données assure d'ores et déjà une représentativité statistique, la masse de données filtrées justifie une meilleure harmonisation entre les différents services hospitaliers participants ou bien une refonte du questionnaire avec la suppression de certaines zones mal remplies.

Tout au long de ce chapitre, nous nous sommes bien gardés d'employer le mot "prédiction", même si la nature des méthodes envisagées, en particulier pour la régression, nous l'autoriseraient d'un point de vue de la sémantique statistique. Comme le rappelle très justement Peter dans un article récent [352], il ne faut pas confondre évaluation du degré de risque et prédiction du passage à l'acte. C'est précisément sur le premier aspect que nos recherches tentent, modestement, de contribuer, la prédiction d'un acte aussi complexe étant illusoire. En identifiant des groupes de patients et des facteurs de risque associés, nos études pourraient permettre d'atteindre la dernière brique de notre système dynamique d'aide à la décision clinique (figure 5.2) et de mettre en place un vrai plan de suivi personnalisé pour chaque patient.

5.3.2 Vers une application mobile française pour la prévention du risque suicidaire

Nos travaux se sont inscrits en parallèle d'un certain nombre d'initiatives fortes au niveau national, en particulier le projet PRINTEMPS (Programme de Recherche INTerventionnelle et Evaluative Mené pour la Prévention du Suicide), porté par l'équipe ECEVE UMR 1123 de l'INSERM, et pour lequel nous avons obtenu un contrat de recherche pour la rédaction du cahier des charges, afin de préparer un appel d'offres sous la forme de marché public. L'objectif principal du projet est de démontrer l'efficacité d'une intervention internet mobile de prévention primaire du suicide incluant une application pour smartphone et un site internet associé, promue par les villes et impliquant potentiellement les médecins généralistes. Les objectifs secondaires sont de déterminer l'effet de cette intervention sur la reconnaissance par l'utilisateur de sa douleur psychologique et sur sa capacité à chercher de l'aide, d'évaluer son efficacité en prévention secondaire du suicide, de mesurer la satisfaction des utilisateurs vis-à-vis du programme mis en place, d'étudier des stratégies de promotion, de déterminer son effet sur le développement de nouvelles mesures de prévention par les villes, d'examiner son effet sur la pratique des médecins généralistes envers les patients souffrant de douleur psychologique et/ou à risque suicidaire élevé, de réaliser une évaluation économique de l'intervention et de relever les obstacles potentiels à sa mise en œuvre. Outre des objectifs clairs de prévention primaire, une application mobile peut servir à une évaluation

écologique instantanée en proposant au patient une partie sécurisée qui peut comprendre, comme dans le projet PRINTEMPS :

- Un ensemble de questionnaires d'auto-évaluation (état de mal-être, risque suicidaire, niveau de dépression, etc.) qui peuvent être éventuellement transmis aux spécialistes.
- Une géolocalisation : l'utilisateur saisit le code postal de son lieu de résidence.
- Un annuaire des structures susceptibles d'apporter de l'aide. Cet annuaire est adapté à la localisation de l'utilisateur : les ressources médicales proches mais aussi les organismes sociaux, les réseaux de prévention primaire (les proches) etc.
- Quelques détails personnels pouvant être remplis par l'utilisateur s'il le souhaite (profil de l'utilisateur) : hobbies, animal de compagnie etc.
- Le plan de soutien adapté au niveau de risque. Ce plan de soutien (*safety plan*) correspond à un ensemble de recommandations et d'actions générées sous une forme synthétique et didactique pour gérer un risque de passage à l'acte.

A la suite du projet PRINTEMPS, nous avons poursuivi notre objectif de développement d'un système dynamique d'aide à la décision clinique en nous concentrant sur la prévention secondaire et le *design* d'une application dédiée à l'évaluation écologique instantanée (EMA). A travers MEmind ou un autre type d'application, notre stratégie repose sur la fusion entre données statiques provenant d'entretiens en face-à-face avec un spécialiste, et données dynamiques délivrées en temps réel ou différé au praticien afin de nourrir un modèle d'aide à la décision spécifique au patient. Nous pensons que seule une application dédiée et précise emportera un fort taux d'adhésion parmi les patients. Des études récentes ont montré que la plupart des applications de e-santé, pourtant foisonnantes aujourd'hui, ne sont pas validées par des évaluations scientifiques solides. La même conclusion a été faite pour les applications mobiles relatives au suicide [249]. Larsen *et al.* ont ainsi identifié, parmi 123 applications relatives au thème du suicide, 32% d'entre elles contenant un contenu inapproprié, voire contre-productif [249]. Le choix des questions, l'ordre, les échelles, doivent être choisis par rapport à l'état de l'art le plus récent. Cette efficacité est aussi valable du point de vue du praticien qui souhaite avoir une vue synthétique de l'état physique et mental de son patient, afin de prendre les meilleures décisions quant à des prescriptions ou des interventions. Sur la figure 5.6, nous montrons des copies d'écran d'une application développée à l'IMT Atlantique, que nous avons créé comme prototype d'une possible application de prévention du risque suicidaire au niveau national. L'application test a été développée en langue anglaise et reste pour le moment un prototype interne de l'IMT Atlantique.

Du fait des contraintes liées aux données manipulées par ce type d'application mobile, souvent accompagnée d'un site internet, les données doivent être stockées par un hébergeur agréé par l'Agence des Systèmes d'Information Partagés de santé (ASIP Santé). L'agrément concerne une prestation d'hébergement d'applications fournies par les clients et gérant des données de santé à caractère personnel. En particulier, des critères de haut niveau de service doivent être respectés, avec une architecture fondée sur une réplique asynchrone entre deux sites distants, un chiffrement sophistiqué, une protection et traçabilité de haute qualité. La continuité du service doit être assurée en toutes circonstances et un plan de reprise d'activité en cas de panne majeure doit être proposé du fait de la fragilité des utilisateurs de l'application mobile et du site internet qui ne doivent pas se retrouver sans soutien en cas d'urgence.

Le canevas d'évaluation de la dimension ergonomique respectera en particulier les critères de Bastien & Scapin, adaptés dans les items suivants [385] :

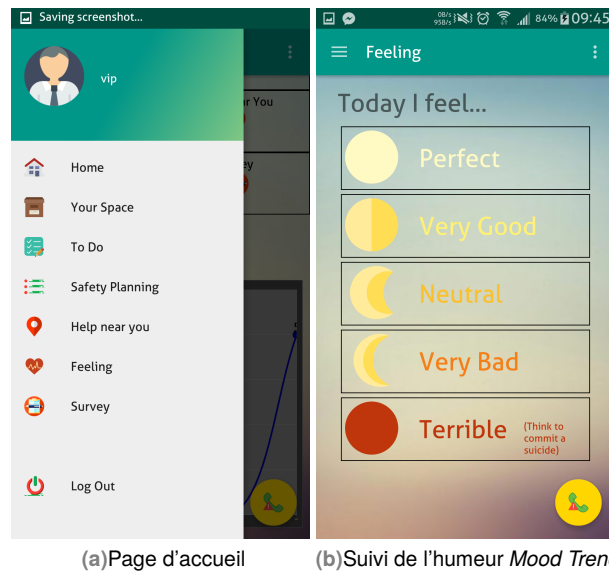


Fig. 5.6.: Application Beleaf de prévention du risque suicidaire. Projet d'étudiants à IMT Atlantique.

- **Guidage** : sur un sujet aussi sensible que la santé mentale, l'utilisateur doit se sentir pris en main et totalement informé des options de navigation possibles à chaque étape de la visite ou de l'utilisation de l'application mobile. Les différentes rubriques pourront être hiérarchisées, regroupées, distinguées par des formes graphiques ou codes couleurs discutés avec les experts. D'autre part, chaque action effectuée peut être associée à un feedback immédiat qui montre à l'utilisateur que l'action a été, ou n'a pas été possible (ex : effet visuel après création d'un compte utilisateur ou renseignement d'une information géographique, effet visuel et/ou sonore pour une action impossible). Un bouton d'urgence doit être accessible en un clic à n'importe quel moment de l'expérience de navigation ou d'utilisation de l'application. Il permet de mettre en contact l'utilisateur avec les services appropriés.
- **Charge cognitive** : l'utilisateur doit pouvoir accéder à n'importe quelle information du site ou de l'application en un nombre d'actions minimal. La densité des informations présentées sur une même page doit être raisonnable pour limiter la charge cognitive de l'utilisateur.
- **Contrôle explicite** : l'utilisateur doit conserver l'impression qu'il a un contrôle total sur le système. Ainsi, un soin particulier sera par exemple apporté aux opérations automatiques de mises à jour de contenu de l'application.
- **Clarté de l'information** : il est important que chaque information présentée soit comprise. L'utilisateur doit être en mesure de comprendre la dénomination d'un acronyme (ex : EMA ?), le contenu d'un lien, ou ce qui se cache derrière un bouton, notamment en déplaçant le curseur ou son doigt sur ces éléments.
- **Adaptabilité et compatibilité** : les sous-systèmes doivent s'adapter aux niveaux d'expérience des utilisateurs. Alors qu'un utilisateur débutant aura besoin d'être guidé pas à pas, par des bulles, des aides, un utilisateur confirmé voudra quant à lui personnaliser au maximum sa visite ou son utilisation de l'application mobile (ex : personnalisation des affichages du plan de soutien pour la partie privée). En termes de compatibilité, le bon fonctionnement du site avec les différentes technologies (iOS, Android) doit être garanti.
- **Gestion des erreurs** : des erreurs utilisateurs sont toujours possibles, à travers la création d'un compte, puis l'élaboration d'un profil ou encore la recherche d'information sur le site.

Ces erreurs doivent être prévenues, signalées, avec des aides à la saisie et des messages d'erreur de qualité. Le système doit être capable, dans la mesure du possible, d'indiquer à l'utilisateur les moyens de corriger ces erreurs (exemple : erreur lors de la saisie d'une adresse personnelle, saisie du profil non enregistrée, etc.).

Bilan et projet scientifique

” *Le futur, que nous réserve-t-il ?*

— **Assassin**
Groupe de rap.

Ce mémoire, présenté en vue de l'obtention de l'habilitation à diriger des recherches, a dressé un panorama de nos recherches effectuées depuis 6 ans à Brisbane, Lyon puis Brest. Le chapitre 1 peut être vu comme une synthèse auto-suffisante de nos recherches sur la mobilité intelligente. Ce chapitre a permis d'identifier trois grands défis de la recherche sur les systèmes de transports intelligents :

- Défi 1 : anticiper le déploiement de nouvelles technologies,
- Défi 2 : extraire des connaissances pour la caractérisation temps réel et la prévision court terme du trafic routier,
- Défi 3 : améliorer la planification des trajets pour les usagers des réseaux de transports.

À partir de ces défis, nous avons identifié trois grandes questions de recherche dans lesquelles nos contributions se sont naturellement inscrites :

- Q1 : modélisation et simulation des C-ITS,
- Q2 : approches statistiques pour la caractérisation et la prévision court terme du trafic routier,
- Q3 : optimisation des itinéraires sous contraintes complexes.

La première question de recherche se rapproche plus du domaine des mathématiques appliquées, avec de la modélisation du trafic inspirée de la physique cinétique ou de la mécanique des fluides, et de l'informatique, avec de la simulation et des modélisations à base d'agents. La deuxième question de recherche convoque le domaine de la fouille de données et des statistiques tandis que la dernière question de recherche a trait à des problèmes d'optimisation dans les graphes, donc à la recherche opérationnelle. Les trois chapitres centraux ont rassemblé nos principales contributions à ces trois questions de recherche. Le niveau de détails est volontairement faible et le lecteur pourra consulter les articles associés, et parfois les manuscrits de thèse des différents doctorants encadrés (notamment les thèses de Julien Monteil, Fouad Baouche, Maxime Guériau, Pierre-Antoine Laharotte, Raphaël Delhome) pour trouver des explications plus denses sur certaines notions. Le chapitre 5 constitue un chapitre d'ouverture à un nouveau domaine d'application, la santé connectée. Des premiers résultats sur la prévention du risque suicidaire y sont présentés. Les contributions manipulent clairement des outils statistiques similaires à ceux employés dans le chapitre 3 pour l'extraction de connaissances à partir de nouvelles sources de données. D'autre part, le développement d'un système d'aide à la décision dynamique pour une médecine personnalisée, impliquera forcément des approches issues de la recherche opérationnelle. On pense, de par la diversité des acteurs investis dans le suivi du patient, notamment à l'aide multi-critères à la décision. La suite et fin de thèse de Sofian Berrouiguet devrait permettre de consolider les premiers résultats et surtout nouer des collaborations scientifiques aux niveaux national et international, à travers de nouveaux projets.

6.1 Vision globale d'un projet de recherche à long terme

À l'aune de l'ensemble des travaux présentées, le chapitre 5 sur la santé connectée constitue une transition vers la présentation d'un projet de recherche à long terme qui n'est plus exclusivement centré sur la mobilité intelligente. En prenant du recul sur les deux domaines d'applications mis en avant dans ce manuscrit, force est de constater que nous les abordons de manière similaire, en terme de focale et d'enjeux. Le premier mot clé est celui de réseau, réseau de transports vs. réseau de soins, système complexe dans lequel l'objectif est d'améliorer la prise en charge d'un usager ou d'un patient. L'amélioration de cette prise en charge répond à des objectifs économiques, environnementaux, ou de santé publique. Dans le domaine de la mobilité, l'amélioration de la qualité de l'écoulement, grâce à des meilleurs temps de parcours individuels par exemple, apportent des bénéfices clairs à ces trois niveaux, y compris en santé publique grâce à une sécurité routière accrue. En santé, cet objectif fait écho à l'amélioration de la prise en charge du patient et finalement d'un temps de parcours de soin pour le patient, dans un contexte de fort encombrement des différents services. Le deuxième mot clé pourrait être celui de nouvelles technologies, que nous appréhendons par l'analyse des nouvelles masses de données générées : nouvelles données trafic issues de GPS ou véhicules connectés, données de santé connectée grâce à des capteurs ou des EMA. Le couplage de l'analyse de données à des méthodes d'optimisation doit permettre *in fine* de proposer des systèmes d'aide à la décision dynamiques dans ces réseaux intelligents. En rejoignant il y a un an l'équipe DECIDE du Lab-STICC, à l'IMT Atlantique (ex Télécom Bretagne), j'ai choisi de concrétiser cette évolution naturelle de mon parcours de recherche vers une approche plus générique, systémique, en le recentrant sur la fouille de données et l'aide à la décision, et en l'ouvrant à d'autres domaines d'application. Mon expertise sur les réseaux de transports (*smart transports*), transposable à d'autres formes de réseaux de capteurs (ex : *smart grids* pour les réseaux électriques), fait écho à des thématiques actuelles de l'équipe DECIDE, en témoigne la thèse récente de Julie Soulas sur la supervision d'activités de personnes âgées dans des maisons équipées de capteurs (*smart homes*). Les mutations technologiques de ces dernières années ont amené, dans beaucoup de ces domaines, une hyper-connectivité, des données ouvertes et de masse, qui ouvrent de formidables perspectives mais aussi une exigence nouvelle de la part des usagers de ces réseaux qui souhaitent tout connaître, tout de suite : supervision en temps réel de leurs activités, analyse et prévision de leur consommation électrique, informations ciblées sur leur loisirs, pertinence de leurs réseaux sociaux, etc. Il s'ensuit un besoin sans précédent de recherche afin de transformer ces masses d'information en connaissance. Nous proposons sur la figure 6.1 , une vision synthétique du projet de recherche que nous détaillerons dans les paragraphes suivants, avec quelques exemples d'applications possibles.

Notre projet, tel qu'il est présenté par la figure 6.1, opère un recentrement sur les disciplines de la fouille de données et de l'aide à la décision. La thématique de la modélisation (appliquée aux C-ITS) n'est pas abandonnée mais elle sera abordée sous l'angle de l'extraction de connaissances. Il nous semble, en effet, que le domaine des systèmes de transports intelligents entre aujourd'hui véritablement dans l'ère du *big data* et que des données pourront, à très court terme, être disponibles, y compris pour les véhicules du futur. L'enjeu se situerait ainsi moins du point de vue d'une modélisation théorique et prospective que de l'extraction de connaissance ou l'intégration de la donnée dans les approches existantes, pour les enrichir ou les confronter à la vérité terrain. Une de nos principales activités de recherche portera ainsi sur l'extraction de connaissances à partir des flots de données hétérogènes. L'expertise dans le domaine de la mobilité nous a permis

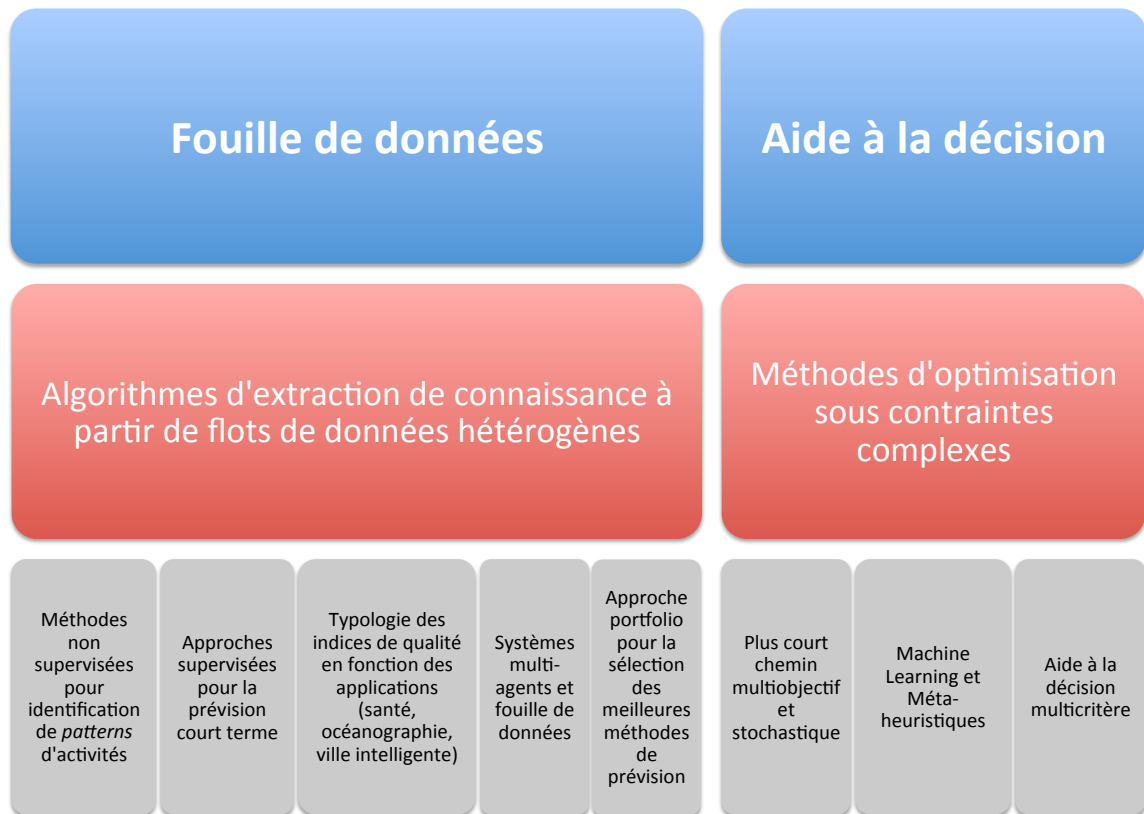


Fig. 6.1.: Structuration du projet de recherche.

d'identifier des verrous scientifiques que l'on retrouve dans pléthore de champs d'application, des besoins majeurs se situant notamment aux niveaux des aspects spatio-temporels, de la résilience des méthodes proposées aux événements non récurrents et de la validation des modèles. Nos contributions théoriques viseront en partie les domaines de la classification non supervisée et supervisée.

Si la coloration de nos activités de recherche penche indiscutablement vers la fouille de données, notre profil, nos travaux récents, et surtout les compétences fortes de l'équipe DECIDE sur ces thèmes, nous encouragent à conserver une composante en aide à la décision et optimisation. Finalement, comme nous allons le souligner, c'est en faisant dialoguer les différentes disciplines que devraient émerger les directions de recherche les plus originales.

6.2 Recherches futures

6.2.1 Faire l'interface entre la fouille de données, l'aide à la décision et la modélisation

Certains résultats présentés dans ce manuscrit combinent déjà plusieurs disciplines. Ainsi, nos travaux sur la modélisation des temps de parcours, très orientés statistiques et fouille de données, ont ensuite été utilisés pour évaluer des algorithmes de plus court chemin stochastique au regard de la distribution de probabilité utilisée. La contribution naît ainsi de l'interaction entre deux domaines,

à l'instar de nombreux projets de recherche actuels visant à faire tomber les cloisons entre les différentes chapelles. Surtout, il nous semble que le *machine learning*, poussé par le phénomène *big data*, se situe à une étape charnière d'ouverture et d'intégration aux autres disciplines. Nous souhaitons, à notre niveau, mettre en avant cette interfaçage multi-disciplinaire à travers trois pistes de recherche appliquées concrètes.

6.2.1.1 Approche portfolio pour la sélection des meilleures méthodes de prévision

Cette activité de recherche a été initiée dans le cadre de l'encadrement de la thèse de Julien Salotti (INSA Lyon), auquel nous contribuons, notamment à propos de l'utilisation de graphes causaux pour l'amélioration des méthodes de prévision de trafic. La motivation d'un tel sujet découle également des quelques résultats présentés au chapitre 3, où l'on voit que la performance des méthodes supervisées dépend du contexte, du type de réseau, de l'heure de la journée, etc. En effet, comme le rappellent Leyton *et al.*, différents algorithmes sont performants sur différents types d'instances [258]. L'approche portfolio peut se définir comme une méthodologie générique permettant de sélectionner le meilleur algorithme en fonction de la configuration du problème. On voit bien ici comment des méthodes issues de l'optimisation ou de l'aide à la décision multicritères pourraient venir enrichir le domaine du *machine learning*. En plus d'une méthodologie portfolio générique, Les applications, pour nos deux domaines de prédilection, seraient possiblement :

1. La sélection du meilleur algorithme de prévision du trafic routier en fonction de la taille du réseau, du volume de données manquantes, de l'heure de la journée, etc.
2. La sélection de la meilleure méthode de prévision d'une maladie (ex : cancer ou état de santé mentale) en fonction de la taille de l'historique d'apprentissage, de la rareté de la maladie à prédire, des descripteurs disponibles etc.

6.2.1.2 *Machine Learning* et métaheuristiques

Des échanges récents au sein de l'équipe DECIDE du lab-STICC ont contribué au montage d'une mini-équipe projet autour du thème *Machine Learning* et métaheuristiques, notamment en collaboration avec l'équipe lorientaise de Marc Sevaux. L'échange s'effectue ici dans l'autre direction puisqu'il s'agit d'évaluer dans quelle mesure des approches issues du *machine learning* pourraient améliorer, ou du moins accélérer, des métaheuristiques reconnues [55]. La méthode visée est la méthode GRASP *Greedy Randomized Adaptive Search Procedure*, qui réside en une technique itérative composée de deux phases [152]. Lors de la phase de construction, la solution est construite pas à pas, chaque nouvelle composante étant choisie dans une liste de candidats nommée RCL (*Restricted Candidate List*). Si l'approche est gloutonne, la sélection entre les candidats se fait de manière aléatoire. Une phase d'amélioration de la solution résultante est ensuite appliquée à travers une recherche locale. Nos recherches sur ce thème viseront à accélérer la résolution d'un problème en analysant "hors ligne" un ensemble d'apprentissage issu des premières itérations de l'algorithme. Des méthodes régressives sur le coût permettraient ainsi de déterminer les configurations favorables et d'orienter la constitution de la RCL et/ou le choix des composantes candidates parmi la RCL. Des premières analyses sont en cours sur un problème d'ordonnancement consistant à minimiser le retard total pondéré sur une seule machine, avec contraintes de dates de disponibilité et dates échues (problème SMTWTP pour *Single machine Total Weighted Tardiness Problem* [103]).

6.2.1.3 Fouille de données et systèmes multi-agents

Dans le chapitre 2, nous avons proposé une approche constructiviste visant à développer un module de prise de décision capable de s'adapter dynamiquement à un problème. Fondé sur le paradigme multi-agents, il s'agit de faire apprendre à un système autonome les meilleures actions à entreprendre pour un contrôle efficace de son environnement. Le système construit une abstraction des informations reçues par l'intermédiaire de ses capteurs, passant d'une perception bas niveau à haut niveau, à l'aide des représentations concurrentes d'une population d'agents. Parmi les perspectives immédiates du travail, l'apport de la fouille de données à la méthode initiale pourrait être évalué selon différents points de vue. Tout d'abord, au sein même de l'approche constructiviste, des algorithmes de clustering sont utilisés pour délimiter les espaces d'états potentiels (hypothèses) des différents agents discrétiseurs. Une analyse de sensibilité du choix des méthodes non supervisées puis un développement de méthodes ensemblistes pour combiner les agents constitueraient des évolutions naturelles de la méthode. D'autre part, une piste de recherche prometteuse consisterait à *booster* le système en le faisant apprendre, là encore, hors ligne, sur un jeu de données de simulation ou du monde réel. Cet apprentissage hors ligne permettrait de pondérer les différentes actions et guider la sélection lors du processus exploration-exploitation.

6.2.2 Méthodes non supervisées pour l'identification de pattern d'activités

Dans le domaine non supervisé, il s'agit de faire apparaître des structures naturelles, groupes homogènes, dans des grands ensembles de données, afin de mieux en comprendre les mécanismes ou dynamiques internes. Les potentiels applicatifs de ces méthodes sont variés, citons par exemple la mise en évidence de groupes de personnes proches par rapport à leurs patterns d'activités. Les flots de données hétérogènes impliquent le développement de processus algorithmiques complexes de qualification, fusion et traitement de la donnée brute, prenant en compte la nature spatio-temporelle et massive intrinsèque aux masses d'information. Dans la continuité des travaux entrepris en mobilité intelligente, le clustering spatio-temporel d'un réseau routier reste un défi à relever, notamment dans la perspective d'un contrôle distribué de zones homogènes résultantes de telles méthodes non supervisées. Un montage de projet en cours, impliquant Rennes Métropole et des partenaires privés, prévoit de remobiliser nos contributions sur les capteurs Bluetooth pour une approche plus ambitieuse à l'échelle de la métropole rennaise, qui couple données trafic et données environnementales. Le clustering de la ville fusionnerait ainsi des exigences de mobilité avec des futures contraintes environnementales liées à des zones de circulation restreintes. En santé connectée, la dimension temporelle soulève des difficultés méthodologiques, des collaborations avec des services de psychiatrie et de médecine générale nous ayant permis de mettre en place des projets concrets à court et moyen terme :

- Analyse longitudinale de patients présentant un risque suicidaire. Données *follow-up* de l'hôpital de Madrid (2018-2019). L'objectif est de mettre en valeur des groupes de patients à risque présentant des caractéristiques communes, en intégrant la dimension temporelle.
- Une approche *Big Data* pour la prévention de la décompensation chez les patients multimorbides suivis en soin primaires (2018). Ce projet s'effectuera en collaboration avec le CHU de Brest. Un aspect innovant du projet concerne la collecte de données de santé connectée auprès du patient, notamment à partir d'applications mobiles ou capteurs (ex : podomètre), après consentement. Des méthodes non supervisées seront utilisées pour faire émerger

des typologies de patients et identifier des liens significatifs entre les différents thèmes de la multimorbidité et la survenue d'une décompensation.

6.2.3 Indices de qualité et leur typologie

Une question de recherche prédominante s'intéresse à l'évaluation des modèles et la validation des groupes proposés. Dans le domaine non supervisé, il n'y a pas de juge de paix et l'évaluation se ramène à l'utilisation d'indices de qualité, par exemple fondés sur des critères géométriques. Sur ce thème, nous souhaitons poursuivre les travaux de l'équipe DECIDE sur les règles d'association [254] en étendant les recherches à la question de l'évaluation d'un clustering, question que nous avons déjà abordée dans [54]. Au niveau théorique, le sujet est toujours très actuel mais la plupart des contributions se cantonnent à l'utilisation de jeux de données synthétiques ou test comme ceux de l'UCI [14]. En développant une approche axiomatique des indices de qualité [2], nos contributions présenteraient l'intérêt de combler des lacunes sur l'applicabilité des différents indices de qualité, par le prisme des projets ciblés par l'équipe DECIDE (ex : santé, océanographie, industrie, web, ville numérique), en montrant que la pertinence d'un indice peut s'appréhender à l'aune de l'application visée. D'ailleurs, un tel travail soulèverait la question d'une analyse multicritère des indices de validation. L'on voit bien ici les passerelles entre les compétences fouille de données de l'équipe DECIDE et celles plus tournées vers l'aide à la décision multi-objectif.

6.2.4 Approches supervisées pour la prévision court terme

Dans le domaine de l'apprentissage supervisé, le verrou scientifique se situe au niveau de méthodes de prévision court terme de l'état d'un réseau, et de la résilience de ces méthodes à des phénomènes non récurrents : panne de capteurs, incident, coupure ou autres externalités (ex : événement météorologique). Alors que des méthodes populaires comme le *deep learning* pourront être adaptées à des problèmes de prévision en mobilité ou santé, il nous semble que le salut pourrait venir des méthodes ensemblistes comme le *Gradient Boosting Machine*, ce que nos recherches devront montrer sur des instances de grande taille. Nous avons également vu que l'erreur de prévision était dépendante de la taille, de la structure et de la découpe spatiale d'un réseau. La recherche du découpage optimale du réseau pour garantir des performances de prévision satisfaisantes est un axe de recherche pertinent qui trouverait des terrains d'application dans plusieurs domaines. D'autre part, dans le chapitre 3, une analogie entre analyse textuelle (*text mining*) et fréquentation d'un réseau routier nous a permis d'assimiler avec succès des thèmes d'un document à des motifs de mobilité et d'utilisation d'un réseau (pics de congestion, sorties de loisirs) [245]. Cet axe de recherche, encore à ses prémices, possède un potentiel indéniable que les compétences des membres de DECIDE aideront à faire fructifier [189]. Ainsi, l'analyse de la propagation d'un incident sur un réseau quelconque (ex : communication, électricité) pourrait être assimilée à la formalisation d'une syntaxe. La maîtrise de ces syntaxes permettrait de mieux saisir les situations accidentogènes et de les classifier en conséquence. L'hypothèse d'un essor constant du volume de données récoltées, si elle se vérifiait pour les terrains d'application envisagés, mènerait à une granularité tellement précise que la discrétisation spatio-temporelle du réseau garantirait une supervision et prévision très réaliste des phénomènes visés.

6.2.5 Plus court chemin multi-objectif et stochastique

Dans le domaine de la recherche opérationnelle, nous traiterons des problèmes d'optimisation sous contraintes complexes dans les réseaux. Nos efforts viseront à proposer des solutions optimales ou approchées prenant en compte les aspects dynamiques ou stochastiques, revêtant une importance cruciale dans beaucoup de domaines d'application où les décideurs réclament des outils temps réel capables de s'adapter à un contexte incertain lié à la variabilité des flots de données spatiotemporelles. Par ailleurs, mon expérience de post-doctorat m'a permis de dresser un panorama complet des besoins de recherche sur l'optimisation multi-objectif du plus court chemin. Si nous avons proposé une approche lexicographique pour les réseaux multimodaux [240], le travail manque encore de maturité pour un réel passage à l'échelle. Quelle que soit l'application visée, ce thème nécessite une approche pluridisciplinaire, dont la présence de sociologues au sein du département LUSI de l'IMT Atlantique faciliterait la mise en œuvre. L'on pense notamment à la définition des objectifs, en fonction des groupes de personnes visés, des différents acteurs et du contexte de l'application. Un dernier sujet de recherche, théorique mais très lié à des contraintes opérationnelles, concerne le problème des techniques d'accélération du calcul du plus court chemin. Des contributions dans le domaine des contractions hiérarchiques, notamment à la lumière des aspects stochastiques et dynamiques, sont attendues dans la suite des résultats présentés dans ce manuscrit. On pense au passage à l'échelle des méthodes sur des réseaux de grande taille. Une nouvelle fois, des passerelles entre fouille de données et recherche opérationnelle sont à envisager. Une piste prometteuse concerne la sélection du paramétrage optimal pour la méthode des contractions hiérarchiques, qui pourrait être effectué sur une base d'apprentissage de solutions générées en simulation. Dans ce cadre, il serait pertinent de formaliser comment le caractère stochastique d'un arc influence ou non sa contraction. Une autre activité, qui nous tient à cœur, en collaboration avec l'IFSTTAR, viserait à intégrer les algorithmes présentés dans le chapitre 4 dans la plateforme TEMPUS : PCC pour véhicules électriques, PCC stochastique, techniques d'accélération. À partir de la rentrée 2017-2018, je co-dirigerai la thèse de Flavien Lucas avec Marc Sevaux, de l'université Bretagne Sud à Lorient. Le sujet de thèse présente une double vision novatrice optimisation/fouille de données qui est très rare dans les travaux actuels dans le domaine du transport. Beaucoup d'algorithmes proposés dans la communauté "recherche opérationnelle" sont intéressants d'un point de vue théorique mais ne sont pas capables de passer à l'échelle. L'ambition de cette thèse est de faire le pont entre des problématiques terrain très concrètes (ex : quelle est la meilleure stratégie de tournées pour une flotte de véhicules électriques de livraison étant donné les conditions de trafic prévues dans 1 heure ?) et des approches plus formelles de la communauté RO.

6.2.6 Aide Multicritère à la Décision (AMCD)

L'AMCD est une branche de la recherche opérationnelle qui vise à améliorer la qualité du processus décisionnel par une meilleure prise en compte des points de vue des différents acteurs du système. Il nous semble qu'une collaboration avec les chercheurs de l'équipe DECIDE actifs sur ce thème permettrait d'apporter des contributions significatives à plusieurs activités de recherche décrites précédemment. Pour le problème du plus court chemin, il s'agirait par exemple de raffiner nos approches multi-objectifs par un véritable apprentissage des préférences à partir de la donnée. Par exemple, des expérimentations réelles sur des choix d'itinéraires par les usagers permettraient de classer de façon pertinente l'importance des objectifs dans une approche lexicographique tandis que l'AMCD pourrait être combinée avec des techniques ambitieuses comme les intégrales de Choquet

[296]. Dans le problème du PCC stochastique, l'AMCD faciliterait l'exploration de la frontière de Pareto et aurait le potentiel d'accélérer de manière intelligente les algorithmes. Le choix de la bonne méthode va dépendre du type de données, des dimensions en entrée, de la disponibilité de l'humain dans le processus etc. D'autre part, dans le domaine de la santé connectée, on voit bien ce que l'AMCD pourrait apporter au développement d'un système dynamique d'aide à la décision pour la prévention du risque suicidaire. La multiplicité des acteurs, les différences entre préférences individuelles des patients et des médecins, pourraient être traitées par des approches AMCD [295]. Plutôt qu'une fusion sophistiquée de modèles de préférences individuelles, un travail approfondi devrait mener à l'élaboration d'un modèle de préférences de groupes, à partir duquel émergerait un nouveau processus de décision pour un suivi personnalisé des patients.

6.3 Exemple d'un projet type ANR en santé connectée

Nous proposons dans cette ultime section le canevas d'un pré-projet de type ANR, en cours de montage, dans le domaine de la santé connectée. Le projet propose de consolider une collaboration entre notre équipe CNRS de chercheurs en extraction de connaissances, l'UMR 6285 Lab-STICC à IMT Atlantique, et des experts en santé mentale au sein des équipes du Pr. Vaiva (CHU Lille/UMR CNRS 9193 PSYCHIC-SCA) et du Pr. Walter (CHRU Brest). L'initiative, nommée BigVigilanS, a pour ambition de faire avancer la recherche en psychiatrie autour du thème de la prévention du suicide. Le programme de recherche s'appuie sur un dispositif existant de veille post-hospitalière des suicidants (VigilanS), proposant des médias variés de recontact ainsi que la future mise à disposition d'une web intervention (portail + appli mobile). Les objectifs sont de qualifier les données existantes, proposer une architecture *big data* flexible et efficace pour le futur recueil de données et d'extraire des nouvelles connaissances sur le suicide à partir des données déjà générées.

6.3.1 Contexte

Comme nous l'avons souligné dans le chapitre 5, le geste suicidaire est un phénomène trop complexe pour être gouverné par des stratégies d'intervention simples et efficaces. Les essais positifs peuvent être distingués en deux familles : les dispositifs d'intervention intensive et les dispositifs de veille (surveillance ou *case management* en anglais). Notre projet s'inscrit dans la deuxième famille des dispositifs de veille, qui n'envahissent pas le quotidien du suicidant, qui opèrent en parallèle d'une éventuelle prise en charge quand elle est indiquée, et qui proposent des outils ou des recours fiables et efficaces en cas de crise suicidaire. L'on peut citer des expériences prometteuses de recontact téléphonique des suicidants [432], d'envois de SMS [47], de délivrance d'une carte "ressource prévention" [145], d'envoi de cartes postales [70]. L'équipe du Pr. Vaiva a convaincu l'Agence Régionale de Santé du Nord-Pas de Calais de proposer une expérience pilote de dispositif de veille à l'échelle de tout un territoire (5,3 millions d'habitants, 12.000 TS et 800 suicides par an). L'expérience, baptisée Vigilans, menée depuis début 2015, concerne les 28 centres de la région et porte aujourd'hui sur plus de 4 500 sujets. Les principes du dispositif sont brièvement décrits ci-après :

- L'équipe vigilans est composée de médecins, infirmiers et psychologues spécialisés dans la prise en charge et le suivi des patients à risque suicidaire. Elle assure le déploiement du dispositif dans l'ensemble des centres accueillant des suicidants (urgences, services

d'hospitalisation) ainsi que le recontact et la réception des appels des patients inclus dans le dispositif.

- À la sortie de l'hôpital, l'équipe qui a assuré la prise en charge propose au patient une carte ressources (numéros à contacter). S'il accepte, il est alors inclus dans la veille. Vigilans informe le médecin traitant et/ou le psychiatre référent par courrier de la mise en place du dispositif.
- Une fois inclus dans la veille, le patient est informé qu'il sera recontacté par téléphone. Le nombre d'appels varie en fonction des antécédents de tentatives de suicide du patient. Suite aux appels, des actions (nouvel appel, consultation en urgence, etc.) sont programmées selon le niveau de risque constaté. Les patients injoignables sont contactés par voie postale, au moyen de cartes personnalisées sous enveloppe.
- Au bout de six mois, pour tous les patients, un entretien téléphonique est effectué. Il comprend une évaluation psychopathologique, un relevé des conduites suicidaires ainsi qu'un questionnaire de satisfaction sur le dispositif. Il est proposé au patient de reconduire la veille s'il va mal. Si le patient va mieux, il sort de la veille. Il peut toutefois se remettre en lien avec Vigilans en cas de besoin.

Les premiers résultats concernant le passage aux urgences de conduites suicidaires sont prometteurs et montrent une diminution, passé un certain seuil, proportionnelle à la pénétrance du dispositif sur le site. Fort de ces premiers succès, le programme a été étendu dès 2016 à d'autres contextes régionaux. Ces extensions s'accompagneront du développement d'un portail WEB et d'une application mobile, VIGIKIT. L'application mobile permet une mise en réseau avec des professionnels ou des personnes ressources, elle donnera des informations générales sur le suicide, mais permettra également au patient de s'auto-évaluer avec des questionnaires. À ce jour, il n'existe aucune application mobile destinée à la prévention du suicide en France, ni en langue française, au contraire d'autres pays développés ou d'autres langues. Les meilleures applications de langue anglaise proposent soit des mini réseaux sociaux (My3), soit des ressources d'aide en cas de crise suicidaire, soit un lien privilégié avec une équipe de soin, mais rarement une intégration de ces dispositifs. VIGIKIT serait une application intégrative de ces dispositifs, tous individuellement plébiscités par leurs publics anglo-saxons. VIGIKIT sera proposée à tous suicidant hospitalisé au moins une nuit après l'accueil aux urgences. Un temps de présentation et de personnalisation de l'application est en effet souhaitable, et une équipe soignante aux urgences n'est pas en possibilité de mener cette action. Cette application mobile permettra d'offrir des outils pragmatiques aux sujets en crise (*safety plan* personnalisable, micro-réseau social paramétrable, etc. . .) mais aussi d'alimenter des bases de données destinées à la recherche en suicidologie :

- *Ecological Momentary Assessment* (EMA),
- Analyse comportementale,
- Analyse de discours.

6.3.2 Verrous scientifiques

Le projet Vigilans soulève plusieurs verrous scientifiques importants, tant du point de vue de la recherche en prévention du suicide qu'en analyse des données. Tout d'abord, les données déjà récoltées en Nord- Pas de Calais n'ont pas encore été analysées à travers le prisme de méthodes récentes de fouille de données, peu employées en recherche sur la santé mentale. Des approches non supervisées (ex : clustering ou règles d'association) pourraient permettre de faire apparaître

des groupes naturels de patients tandis que des approches supervisées mettraient en lumière de nouveaux facteurs de risque pour la récurrence suicidaire. Mais c'est surtout par la notion de temporalité de la récurrence suicidaire que des contributions innovantes pourraient naître de l'analyse des données Vigilans. En effet, le dispositif permet d'avoir des nouvelles de l'état de santé mentale du patient à J+1, J+10, J+20, J+180 après la tentative de suicide. Les méthodes de fouille de données temporelles peuvent aider à mieux comprendre le moment de la récurrence, mais également à améliorer le modèle de prise de décision de Vigilans et les orientations des patients vers des consultations, ou arrêts du suivi. De plus, des problématiques *big data* émergent de l'incorporation d'une application mobile dans le dispositif. Une telle application peut permettre de récolter, quasi en temps réel, des données complexes, volumineuses, et hétérogènes sur les patients. Ces nouveaux flux de données se prêtent particulièrement bien à un traitement *big data* et les résultats devraient faire avancer la recherche en prévention du suicide et en traitement de flux de données hétérogènes [243]. La proposition d'une architecture appropriée doit s'accompagner d'un véritable audit sur la qualité des données du dispositif Vigilans Nord- Pas de Calais, avant le déploiement sur le reste du territoire, les erreurs sur les données pouvant être d'origine humaine (saisie par les infirmiers, mauvaises interprétations) ou logicielle (mauvais formats, données tronquées).

6.3.3 Axes de recherches

Les axes de recherches développés dans cette proposition sont originaux et viennent s'ajouter au plan d'action Vigilans. Les verrous scientifiques découlant des masses de données de patients inclus dans le dispositif Vigilans, déjà disponibles et à venir, soulèvent des perspectives de recherche sur les trois à cinq prochaines années, que nous structurons en 3 axes :

6.3.3.1 Axe 1 : audit sur la qualité des données Vigilans

Au sein du dispositif Vigilans, les données sont très hétérogènes : données patients classiques suite à une hospitalisation, entretiens téléphoniques donnant lieu à des évaluations psychopathologiques et un relevé des conduites suicidaires, puis données liées à l'utilisation d'une application mobile. Le premier axe de recherche consiste à réaliser un diagnostic de la qualité de ces données hétérogènes, lister les sources d'erreurs (humaines, logicielles, réseau ?), mettre en place des méthodes automatiques de qualification de la donnée ainsi qu'émettre des recommandations avant le déploiement du dispositif au niveau national.

6.3.3.2 Axe 2 : une architecture *big data* pour un plan Vigilans national

L'extension du dispositif Vigilans au niveau national, couplée à l'utilisation d'une application mobile, va poser des problèmes de collaboration entre des systèmes d'informations et entrepôts de données propres à chaque hôpital. D'autre part, la génération de nouveaux flux de données dynamiques et hétérogènes va remettre en cause bon nombre d'architectures existantes. Un axe de recherche majeur consiste à proposer une méthodologie de collecte et d'incrémentation dynamique de données de patients suicidants au sein d'une architecture flexible type *big data*. Plutôt que de créer un immense entrepôt de données qui nécessiterait un lourd travail de fusion des systèmes d'information existants, il est proposé la mise en place d'une interface (API) qui interroge à la demande les entrepôts disponibles, et agrège en temps réel les flux de données du suivi des patients. Ceci permettrait de constituer de manière dynamique une base de données des patients suicidants. L'architecture

se fonde sur un stockage distribué de type Hadoop/Spark, qui sépare intelligemment le stockage et le traitement des données collectées pour garantir un niveau de performance et souplesse adéquat, même en présence de grosses masses de données. D'autre part, les patients pris en charge utiliseront l'application mobile et partageront des informations à travers des applications de *quantified self* ou d'EMA. Cette collecte de données hétérogènes implique de dépasser une démarche classique d'entrepôts de données qui fusionnerait les différents systèmes en un format unique. L'architecture *big data* permet un couplage souple entre la gestion des données et leurs traitements.

6.3.3.3 Axe 3 : fouille de données pour une meilleure compréhension de la récurrence suicidaire.

Les données patients récoltées continuellement depuis 2015 constituent une mine d'information conséquente. Des milliers de variables sont disponibles pour des milliers de patients, le projet Vigilans ajoutant la dimension temporelle, l'analyse de ces données est un challenge notamment du point de vue de la dimensionalité. L'axe 3 vise d'abord à faire émerger une typologie de patients suicidaires grâce à des méthodes de clustering en grandes dimensions. Ensuite des méthodes supervisées, comme les arbres de décision ou les forêts aléatoires qui sont appropriées pour une discussion interdisciplinaire, seront employées pour détecter de nouveaux facteurs de risque de la récurrence suicidaire. Les marqueurs temporels sur les données pourront modéliser des facteurs de risques d'aggravation de l'état général du patient à J+1, J+10, J+20 ou J+180 jours après la tentative.

6.3.4 Caractère innovant du projet

Le projet répond tout d'abord à un objectif de pluri-disciplinarité. Une rencontre entre deux disciplines éloignées se cristallise autour de la gestion et l'analyse de nouvelles masses de données disponibles sur les patients suicidaires. Le projet proposé répond à un certain nombre d'enjeux stratégiques :

- Sur la **gestion des données** : il s'agit d'anticiper la mise en place d'une architecture *big data* pour le dispositif Vigilans avant que son déploiement sur toute la France ne mette en difficulté les implémentations actuelles.
- Sur l'**extraction de connaissances** : de nouvelles méthodes de fouille appliquées à la prévention du suicide peuvent faire émerger de nouveaux facteurs de risques, et donc modifier par exemple les entretiens aux urgences après une TS. Et des données temporelles complexes et hétérogènes vont nécessiter des nouveaux traitements.
- Sur la **qualité des données** : de véritables routines de qualification de la donnée générée doivent être mises en place à toutes les échelles du dispositif.
- Sur les **problématiques sociales** : la question traitée est majeure du point de vue de la santé publique et l'utilisation de dispositifs innovants, récoltant toujours plus de données sur les patients, pose la question de l'acceptabilité et de la sécurité des données.

Bibliographie

- [1] A., H., HAMED, M., SADABADI, K. F., YOUNG, S., AND TARNOFF, S. Data collection of freeway travel time ground truth with bluetooth sensors. *Transportation Research Record : Journal of the Transportation Research Board Transportation Research Board of the National Academies, Washington D.C.*, 2160 (2010), 60–68. [Page 72]
- [2] ACKERMAN, M., BEN-DAVID, S., LOKER, D., AND SABATO, S. Clustering oligarchies. In *Artificial Intelligence and Statistics* (2013), pp. 66–74. [Page 152]
- [3] ADLER, J. D., MIRCHANDANI, P. B., XUE, G., AND XIA, M. The electric vehicle shortest-walk problem with battery exchanges. *Networks and Spatial Economics* 16, 1 (2016), 155–173. [Page 31]
- [4] ADLER, J. L., AND BLUE, V. J. A cooperative multi-agent transportation management and route guidance system. *Transportation Research Part C : Emerging Technologies* 10 (2002), 433 – 454. [Page 23]
- [5] ADLER, J. L., SATAPATHY, G., MANIKONDA, V., BOWLES, B., AND BLUE, V. J. A multi-agent approach to cooperative traffic management and route guidance. *Transportation Research Part B : Methodological* 39, 4 (2005), 297 – 318. [Page 23]
- [6] AERDE, M. V., HELLINGA, B., YU, L., AND RHAKA, H. Vehicle probes as real-time atms sources of dynamic od and travel time data. *Preprint Queen's university Department of Civil Engineering* (1993). [Page 26]
- [7] AGNELLO, P., LEE, J., AND CHEN, J. A comparison of bluetooth vs. traditional origin–destination survey data collection methods. *13th TRB National Planning and Applications Conference, Reno* (November 2011). [Page 27]
- [8] AHMED, S. A., AND COOK, A. R. *Application of time-series analysis techniques to freeway incident detection*. No. 841. 1982. [Page 29]
- [9] AHN, S., AND CASSIDY, M. J. Freeway traffic oscillations and vehicle lane-change maneuvers. In *International Symposium on Transportation & Traffic Theory* (2007), M. B. R. Alsop and B. Heydecker, Eds., vol. 1, pp. 691–710. [Page 22]
- [10] AIRPARIF. Surveillance & information sur la qualité de l'air en île-de-france – bilan 2015. Tech. rep., 2016. [Page 13]
- [11] ANDREWS, M., DOGRU, M. K., HOBBY, J. D., JIN, Y., AND TUCCI, G. H. Modeling and optimization for electric vehicle charging infrastructure. In *IEEE Innovative Smart Grid Technologies Conference* (2013). [Page 31]
- [12] ARAGHI, B. N., CHRISTENSEN, L. T., KRISHNAN, R., AND LAHRMANN, H. Application of bluetooth technology for mode-specific travel time estimation on arterial roads : Potentials and challenges.

Proceedings from the Annual Transport Conference at Aalborg University, Denmark (2012).
[Pages 27, 72]

- [13] ARAGHI, B. N., PEDERSEN, K. S., CHRISTENSEN, L. T., KRISHNAN, R., AND LAHRMANN, H. Accuracy of travel time estimation using bluetooth technology : Case study limfjord tunnel aalborg. *Proceedings of the 19th ITS World Congress, Vienna (october 2012)*. [Page 27]
- [14] ARBELAIZ, O., GURRUTXAGA, I., MUGUERZA, J., PÉREZ, J. M., AND PERONA, I. An extensive comparative study of cluster validity indices. *Pattern Recognition* 46, 1 (2013), 243–256. [Page 152]
- [15] ARTIGUES, C., HUGUET, M.-J., GUEYE, F., SCHETTINI, F., AND DEZOU, L. State-based accelerations and bidirectional search for bi-objective multi-modal shortest paths. *Transportation Research Part C : Emerging Technologies* 27 (2013), 233–259. [Page 33]
- [16] ARTMEIER, A., HASELMAYR, J., LEUCKER, M., AND SACHENBACHER, M. The shortest path problem revisited : Optimal routing for electric vehicles. In *Annual Conference on Artificial Intelligence (2010)*, Springer, pp. 309–316. [Pages 115, 116]
- [17] ASADA, M., HOSODA, K., KUNIYOSHI, Y., ISHIGURO, H., INUI, T., YOSHIKAWA, Y., OGINO, M., AND YOSHIDA, C. Cognitive developmental robotics : a survey. *IEEE Transactions on Autonomous Mental Development* 1, 1 (2009), 12–34. [Page 23]
- [18] AUER, P., CESA-BIANCHI, N., AND FISCHER, P. Finite-time analysis of the multiarmed bandit problem. *Machine learning* 47, 2-3 (2002), 235–256. [Pages 61, 64]
- [19] AW, A., AND RASCLE, M. Resurrection of "second order" models of traffic flow ? *SIAM journal on applied mathematics* 60 (2000), 916–938. [Pages 20, 21]
- [20] AZIMI A., ZHANG, Y. Categorizing freeway flow conditions by using clustering methods. *Transportation Research Record : Journal of the Transportation Research Board* 2173 (2010), 105–114. [Page 93]
- [21] BAÑOS, R., ORTEGA, J., GIL, C., FERNÁNDEZ, A., AND DE TORO, F. A simulated annealing-based parallel multi-objective approach to vehicle routing problems with time windows. *Expert Systems with Applications* 40, 5 (2013), 1696–1707. [Page 33]
- [22] BACKFRIEDER, C., MECKLENBRAUKER, C., AND OSTERMAYER, G. Traffsim - a traffic simulator for investigating benefits ensuing from intelligent traffic management. In *Modelling Symposium (EMS), 2013 European (Nov 2013)*, pp. 451–456. [Page 23]
- [23] BAJWA, S. U. I., CHUNG, E., AND KUWAHARA, M. Performance evaluation of an adaptive travel time prediction model. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, IEEE, pp. 1000–1005. [Page 28]
- [24] BALBO, F., AND PINSON, S. Toward a multi-agent modelling approach for urban public transportation systems. In *International Workshop on Engineering Societies in the Agents World (2001)*, Springer, pp. 160–174. [Page 23]
- [25] BALBO, F., AND PINSON, S. Dynamic modeling of a disturbance in a multi-agent system for traffic regulation. *Decision Support Systems* 41, 1 (2005), 131–146. [Page 23]
- [26] BANDO, M., HASEBE, K., NAKAYAMA, A., SHIBATA, A., AND SUGIYAMA, Y. Dynamical model of traffic congestion and numerical simulation. *Phys. Rev. E* 51 (1995), 1035–1042. [Pages 21, 40]
- [27] BAOUCHE, F. *Outils pour l'optimisation de la consommation des vehicules électriques*. PhD thesis, Ecole Nationale des Travaux Publics de l'Etat, 2015. [Page 30]

- [28] BAOUCHE, F., BILLOT, R., TRIGUI, R., AND EL FAOUZI, N.-E. Efficient allocation of electric vehicles charging stations : Optimization model and application to a dense urban network. *IEEE Intelligent Transportation Systems Magazine* 6, 3 (2014), 33–43. [Page 115]
- [29] BAOUCHE, F., TRIGUI, R., EL FAOUZI, N. E., AND BILLOT, R. Energy consumption assessment for electric vehicles. In *International symposium on recent advances in transport modeling* (2013), pp. 5–p. [Page 113]
- [30] BARCELÓ, J., AND CASAS, J. Dynamic network simulation with aimsun. In *Simulation approaches in transportation analysis*. Springer, 2005, pp. 57–98. [Page 23]
- [31] BARCELÓ, J., MONTERO, L., BULLEJOS, M., SERCH, O., AND CARMONA, C. A kalman filter approach for the estimation of time dependent od matrices exploiting bluetooth traffic data collection. *Transportation Research Board 91st Annual Meeting*, 12-3843 (2012). [Pages 27, 72]
- [32] BARCELÓ, J., MONTERO, L., MARQUÈS, L., AND C., C. Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation Research Record : Journal of the Transportation Research Board*, 2175(1) (2010), 19–27. [Pages 27, 89]
- [33] BARRACHINA, J., GARRIDO, P., FOGUE, M., MARTINEZ, F. J., CANO, J.-C., CALAFATE, C. T., AND MANZONI, P. Road side unit deployment : A density-based approach. *IEEE Intelligent Transportation Systems Magazine* 5, 3 (2013), 30–39. [Page 11]
- [34] BARRETT, C., JACOB, R., AND MARATHE, M. Formal language constrained path problems. p. 234–245. [Page 33]
- [35] BARRIGÓN, M. L., BERROUIGUET, S., CARBALLO, J. J., BONAL-GIMÉNEZ, C., FERNÁNDEZ-NAVARRO, P., PFANG, B., DELGADO-GÓMEZ, D., COURTET, P., AROCA, F., LOPEZ-CASTROMAN, J., ET AL. User profiles of an electronic mental health tool for ecological momentary assessment : Memind. *International Journal of Methods in Psychiatric Research* 26, 1 (2017). [Page 132]
- [36] BAST, H., DELLING, D., GOLDBERG, A., MÜLLER-HANNEMANN, M., PAJOR, T., SANDERS, P., WAGNER, D., AND WERNECK, R. F. Route planning in transportation networks. *arXiv preprint arXiv :1504.05140* (2015). [Pages 35, 125]
- [37] BAST, H., FUNKE, S., AND MATIJEVIC, D. Ultrafast shortest-path queries via transit nodes. In *The Shortest Path Problem* (2006), pp. 175–192. [Page 34]
- [38] BATES, J., POLAK, J., JONES, P., AND COOK, A. The valuation of reliability for personal travel. *Transportation Research Part E : Logistics and Transportation Review* 37, 2-3 (2001), 191–229. [Page 25]
- [39] BATZ, G. V. *Time-Dependent Route Planning with Contraction Hierarchies*. PhD thesis, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2014, 2014. [Page 35]
- [40] BATZ, G. V., GEISBERGER, R., SANDERS, P., AND VETTER, C. Minimum time-dependent travel times with contraction hierarchies. *ACM Journal of Experimental Algorithmics* 18, 1.4 (2013), 1–43. [Pages 35, 127]
- [41] BAZZAN, A. L., AND KLÜGL, F. A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review* 29, 03 (2014), 375–403. [Page 23]
- [42] BAZZAN, A. L. C. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. *Autonomous Agents and Multi-Agent Systems* 18, 3 (2009), 342–375. [Page 23]
- [43] BECK, R. W., MORRIS, J. B., AND BECK, A. T. Cross-validation of the suicidal intent scale. *Psychological reports* 34, 2 (1974), 445–446. [Page 136]

- [44] BELOMESTNY, D., JENTSCH, V., AND SCHRECKENBERG, M. Completion and continuation of nonlinear traffic time series : a probabilistic approach. *Journal of Physics A : Mathematical and General* 36, 45 (2003), 11369. [Page 27]
- [45] BERGE, C. Graphes et hypergraphes. [Page 33]
- [46] BERGER, A., GRIMMER, M., AND MÜLLER-HANNEMANN, M. Fully dynamic speed-up techniques for multi-criteria shortest path searches in time-dependent networks. In *Experimental Algorithms*. Springer, 2010, pp. 35–46. [Page 33]
- [47] BERROUIGUET, S., ALAVI, Z., VAIVA, G., COURTET, P., BACA-GARCÍA, E., VIDAILHET, P., GRAVEY, M., GUILLODO, E., BRANDT, S., AND WALTER, M. Siam (suicide intervention assisted by messages) : the development of a post-acute crisis text messaging outreach for suicide prevention. *BMC psychiatry* 14, 1 (2014), 294. [Page 154]
- [48] BERROUIGUET, S., BARRIGÓN, M. L., BRANDT, S. A., OVEJERO-GARCÍA, S., ÁLVAREZ-GARCÍA, R., CARBALLO, J. J., LENCA, P., COURTET, P., BACA-GARCÍA, E., GROUP, M. S., ET AL. Development of a web-based clinical decision support system for drug prescription : Non-interventional naturalistic description of the antipsychotic prescription patterns in 4345 outpatients and future applications. *PloS one* 11, 10 (2016), e0163796. [Page 132]
- [49] BERTOLETE, J. M., AND DE LEO, D. Global suicide mortality rates-a light at the end of the tunnel ? *Crisis* 33, 5 (2012), 249–253. [Page 131]
- [50] BHASKAR, A., AND CHUNG, E. Fundamental understanding on the use of bluetooth scanner as a complementary transport data. *Transportation Research Part C : Emerging Technologies* 37, 0 (2013), 42 – 72. [Pages 26, 72]
- [51] BHASKAR, A., QU, M., NANTES, A., MISKA, M., CHUNG, E., ET AL. Is bus overrepresented in bluetooth mac scanner data ? is mac-id really unique ? *International Journal of Intelligent Transportation Systems Research* 13, 2 (2015), 119–130. [Page 27]
- [52] BILLOT, R. *Analyse et modélisation de l'impact de la météorologie sur le trafic routier*. PhD thesis, Ecole Centrale Paris, 2010. [Pages 1, 20, 29]
- [53] BILLOT, R., EL FAOUZI, N.-E., AND DE VUYST, F. Multilevel assessment of the impact of rain on drivers' behavior : standardized methodology and empirical analysis. *Transportation Research Record : Journal of the Transportation Research Board*, 2107 (2009), 134–142. [Pages 25, 29]
- [54] BILLOT, R., SUCHIER, H.-M., AND LALLICH, S. Une approche ensembliste inspirée du boosting en classification non supervisée. In *EGC (2008)*, pp. 361–372. [Page 152]
- [55] BIRATTARI, M., AND KACPRZYK, J. *Tuning metaheuristics : a machine learning perspective*, vol. 197. Springer, 2009. [Page 150]
- [56] BLANDIN, S., EL GHAOUI, L., AND BAYEN, A. Kernel regression for travel time estimation via convex optimization. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on (2009)*, IEEE, pp. 4360–4365. [Page 28]
- [57] BLEI, D. M., NG, A. Y., AND JORDAN, M. I. Latent dirichlet allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022. [Pages 90, 92]
- [58] BLOGG, M., SEMLER, C., HINGORANI, M., AND TROUTBECK, R. Travel time and origin–destination data collection using bluetooth mac address readers. *33rd Australasian Transport Research Forum, Canberra, Australia (2010)*. [Page 27]
- [59] BOBÉE, B., PERREAULT, L., AND ASHKAR, F. Two kinds of moment ratio diagrams and their application in hydrology. *Stochastic Hydrology Hydraulic* 7 (1993), 41–65. [Pages 78, 80]

- [60] BÖRJESSON, M., ELIASSON, J., AND FRANKLIN, J. P. Valuations of travel time variability in scheduling versus mean-variance models. *Transportation Research Part B : Methodological* 46, 7 (2012), 855 – 873. [Page 25]
- [61] BOUSQUET, A. *Optimisation d'itinéraires multimodaux fondée sur les temps de parcours à l'échelle d'une agglomération urbaine dense*. PhD thesis, École Nationale des Travaux Publics de l'État, 2010. [Pages 18, 33]
- [62] BOX, G. E., JENKINS, G. M., REINSEL, G. C., AND LJUNG, G. M. *Time series analysis : forecasting and control*. John Wiley & Sons, 2015. [Page 28]
- [63] BOYD, D., AND CRAWFORD, K. Critical questions for big data : Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679. [Page 8]
- [64] BREIMAN, L. Random forests. *Machine learning* 45, 1 (2001), 5–32. [Page 28]
- [65] BROOKS, R. A. Intelligence without representation. *Artificial intelligence* 47, 1 (1991), 139–159. [Page 57]
- [66] BURMEISTER, B., HADDADI, A., AND MATYLIS, G. Application of multi-agent systems in traffic and transportation. In *Software Engineering. IEE Proceedings* (1997), vol. 144, IET, pp. 51–60. [Page 22]
- [67] CARPENTER, C., FOWLER, M., AND ADLER, T. J. Generating route-specific origin-destination tables using bluetooth technology. *Transportation Research Record : Journal of the Transportation Research Board vol. 2308*, no 1 (2012), p. 96–102. [Pages 26, 27]
- [68] CARRION, C., AND LEVINSON, D. Value of travel time reliability : A review of current evidence. *Transportation Research Part A : Policy and Practice* 46, 4 (2012), 720 – 741. [Page 25]
- [69] CARRION, C., AND LEVINSON, D. Valuation of travel time reliability from a gps-based experimental design. *Transportation Research Part C : Emerging Technologies* 35, 0 (2013), 305 – 323. [Page 25]
- [70] CARTER, G. L., CLOVER, K., WHYTE, I. M., DAWSON, A. H., AND ESTE, C. D. Postcards from the edge project : randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. *Bmj* 331, 7520 (2005), 805. [Page 154]
- [71] CASEY, B., BHASKAR, A., AND CHUNG, E. Data requirements and graph data structures for a multi-modal, multiobjective trip planner. 25th ARRB Conference - Shaping the future : Linking policy, research and outcomes, Perth, Australia, ARRB Group Ltd and Authors. [Page 33]
- [72] CASTELFRANCHI, C. Modelling social action for ai agents. *Artificial Intelligence* 103, 1 (1998), 157–182. [Page 22]
- [73] CASTELFRANCHI, C., AND FALCONE, R. Social trust : Cognitive anatomy, social importance, quantification and dynamics. In *Proceedings of the First International Workshop on Trust* (1998), pp. 35–49. [Page 46]
- [74] CASTILLO, E., CALVIÑO, A., SÁNCHEZ-CAMBRONERO, S., NOGAL, M., AND RIVAS, A. A percentile system optimization approach with and without path enumeration. *Computers & Operations Research* 40, 11 (2013), 2711–2723. [Page 32]
- [75] CASTRO-NETO, M., JEONG, Y.-S., JEONG, M.-K., AND HAN, L. D. Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert systems with applications* 36, 3 (2009), 6164–6173. [Page 28]
- [76] CELI, L. A., ZIMOLZAK, A. J., AND STONE, D. J. Dynamic clinical data mining : search engine-based decision support. *JMIR medical informatics* 2, 1 (2014), e13. [Page 134]

- [77] CHAN, M. K., BHATTI, H., MEADER, N., STOCKTON, S., EVANS, J., O'CONNOR, R. C., KAPUR, N., AND KENDALL, T. Predicting suicide following self-harm : systematic review of risk factors and risk scales. *The British Journal of Psychiatry* (2016), bjp-bp. [Page 132]
- [78] CHASSIAKOS, A. P., AND STEPHANEDES, Y. J. Smoothing algorithms for incident detection. *Transportation Research Record 1394* (1993), 8–16. [Page 29]
- [79] CHEBANA, F., AND EL ADLOUNI, S.-E. Mixed estimation method for halphen distributions with applications in extrem hydrologic events. *Stochastic Environmental Research and Risk Assessment 24* (2010), 359–376. [Page 79]
- [80] CHEBANA, F., EL ADLOUNI, S.-E., AND BOBÉE, B. Method of moment of the halphen distribution parameters. *Stochastic Environmental Research and Risk Assessment 22* (2008), 749–757. [Page 79]
- [81] CHEN, B., AND CHENG, H. H. A review of the applications of agent technology in traffic and transportation systems. *IEEE Transactions on Intelligent Transportation Systems 11* (2010), 485–497. [Page 22]
- [82] CHEN, B., CHENG, H. H., AND PALEN, J. Integrating mobile agent technology with multi-agent systems for distributed traffic detection and management systems. *Transportation Research Part C : Emerging Technologies 17*, 1 (2009), 1 – 10. [Page 23]
- [83] CHEN, B. Y., LAM, W. H., SUMALEE, A., LI, Q., SHAO, H., AND FANG, Z. Finding reliable shortest paths in road networks under uncertainty. *Networks and spatial economics 13*, 2 (2013), 123–148. [Page 32]
- [84] CHEN, B. Y., LAM, W. H., SUMALEE, A., LI, Q., AND TAM, M. L. Reliable shortest path problems in stochastic time-dependent networks. *Journal of Intelligent Transportation Systems 18*, 2 (2014), 177–189. [Page 32]
- [85] CHEN, B. Y., LAM, W. H., SUMALEE, A., AND LI, Z.-L. Reliable shortest path finding in stochastic networks with spatial correlated link travel times. *International Journal of Geographical Information Science 26*, 2 (2012), 365–386. [Page 32]
- [86] CHEN, C. Tests of fit for the three-parameter lognormal distribution. *Computational Statistics & Data Analysis 50*, 6 (2006), 1418 – 1440. [Page 80]
- [87] CHEN, H., CHIANG, R. H., AND STOREY, V. C. Business intelligence and analytics : From big data to big impact. *MIS quarterly 36*, 4 (2012), 1165–1188. [Page 8]
- [88] CHEN, R.-S., CHEN, D.-K., AND LIN, S.-Y. Actam : Cooperative multi-agent system architecture for urban traffic signal control. *IEICE Transactions 88-D*, 1 (2005), 119–126. [Page 23]
- [89] CHEN, S., AND WANG, W. Decision tree learning for freeway automatic incident detection. *Expert systems with applications 36*, 2 (2009), 4101–4105. [Page 29]
- [90] CHEN, S., WANG, W., AND VAN ZUYLEN, H. Construct support vector machine ensemble to detect traffic incident. *Expert systems with applications 36*, 8 (2009), 10976–10986. [Page 29]
- [91] CHEN, T. D., KOCKELMAN, K. M., KHAN, M., ET AL. The electric vehicle charging station location problem : a parking-based assignment method for seattle. In *Transportation Research Board 92nd Annual Meeting* (2013), vol. 340, pp. 13–1254. [Page 31]
- [92] CHEN, X.-Y., PAO, H.-K., AND LEE, Y.-J. Efficient traffic speed forecasting based on massive heterogenous historical data. In *Big Data (Big Data), 2014 IEEE International Conference on* (2014), IEEE, pp. 10–17. [Page 28]
- [93] CHOI, J., OH, S., AND HOROWITZ, R. Distributed learning and cooperative control for multi-agent systems. *Automatica 45*, 12 (2009), 2802–2814. [Page 52]

- [94] CHUNG, S. H., AND KWON, C. Multi-period planning for electric car charging station locations : A case of korean expressways. *European Journal of Operational Research* 242, 2 (2015), 677–687. [Page 31]
- [95] CHURCH, R., AND REVELLE, C. The maximal covering location problem. In *Papers of the Regional Science Association* (1974), vol. 32, Springer, pp. 101–118. [Page 31]
- [96] CLARK, S., AND WATLING, D. Modelling network travel time reliability under stochastic demand. *Transportation Research part B* 39 (feb 2005), 119–140. [Page 26]
- [97] COHEN, H., AND SOUTHWICK, F. On the measurement and valuation of travel time variability due to incidents on freeways. *Journal of Transportation and Statistics* 2 (1999), 123–132. [Page 25]
- [98] COME, E., RANDRIAMANAMIHAGA, N. A., OUKHELLOU, L., AND AKNIN, P. Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation : Application to vélib'bike sharing system of paris. In *TRB 93rd Annual meeting* (2014), TRANSPORTATION RESEARCH BOARD, p. 19p. [Page 91]
- [99] CORMEN, T. H. *Introduction to algorithms*. MIT press, 2009. [Page 33]
- [100] CORTES, C., AND VAPNIK, V. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297. [Page 28]
- [101] COSTELLOE, D., MOONEY, P., AND WINSTANLEY, A. Multi-objective optimisation on transportation networks. In *Proceedings of the 4th AGILE Conference on GIScience* (2001), pp. 523–560. [Page 34]
- [102] COUTINHO-RODRIGUES, J. M., CLÍMACO, J. C. N., AND CURRENT, J. R. An interactive bi-objective shortest path approach : searching for unsupported nondominated solutions. *Computers & Operations Research* 26, 8 (1999), 789–798. [Page 33]
- [103] CRAUWELS, H., POTTS, C. N., AND VAN WASSENHOVE, L. N. Local search heuristics for the single machine total weighted tardiness scheduling problem. *INFORMS Journal on computing* 10, 3 (1998), 341–350. [Page 150]
- [104] CREMER, M., AND LUDWIG, J. A fast simulation model for traffic flow on the basis of boolean operations. *Mathematics and Computers in Simulation* 28, 4 (1986), 297–303. [Page 21]
- [105] CURRENT, J. R., REVELLE, C. S., AND COHON, J. L. An interactive approach to identify the best compromise solution for two objective shortest path problems. *Computers & Operations Research* 17, 2 (1990), 187–198. [Page 33]
- [106] DAGANZO, C. F., GAYAH, V. V., AND GONZALES, E. J. Macroscopic relations of urban traffic variables : Bifurcations, multivaluedness and instability. *Transportation Research Part B : Methodological* 45, 1 (2011), 278 – 288. [Page 20]
- [107] DAGANZO, C. F., AND GEROLIMINIS, N. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B : Methodological* 42, 9 (2008), 771 – 781. [Page 20]
- [108] DASKIN, M. S. A maximum expected covering location model : formulation, properties and heuristic solution. *Transportation science* 17, 1 (1983), 48–70. [Page 31]
- [109] DASKIN, M. S. *Network and discrete location : models, algorithms, and applications*. John Wiley & Sons, 2011. [Page 31]
- [110] DASKIN, M. S., AND OWEN, S. H. Two new location covering problems : The partial p-center problem and the partial set covering problem. *Geographical Analysis* 31, 3 (1999), 217–235. [Page 31]

- [111] DAVIDSON, C. L., ANESTIS, M. D., AND GUTIERREZ, P. M. Ecological momentary assessment is a neglected methodology in suicidology. *Archives of suicide research* 21, 1 (2017), 1–11. [Pages 132, 133]
- [112] DAVIDSSON, P., HENESEY, L., RAMSTEDT, L., TÖRNQUIST, J., AND WERNSTEDT, F. An analysis of agent-based approaches to transport logistics. *Transportation Research Part C : Emerging Technologies* 13, 4 (2005), 255 – 271. [Page 22]
- [113] DAVIS, G. A., AND NIHAN, N. L. Nonparametric regression and short-term freeway traffic forecasting. *Journal of Transportation Engineering* 117, 2 (1991), 178–188. [Page 28]
- [114] DE BRUIN, D., KROON, J., VAN KLAVEREN, R., AND NELISSE, M. Design and test of a cooperative adaptive cruise control system. In *Intelligent Vehicles Symposium, 2004 IEEE* (2004), IEEE, pp. 392–396. [Page 22]
- [115] DEL CASTILLO, J. M. Three new models for the flow–density relationship : derivation and testing for freeway and urban data. *Transportmetrica* 8, 6 (2012), 443–465. [Page 20]
- [116] DELAFONTAINE, M., VERSICHELE, M., NEUTENS, T., AND DE WEGHE N., V. Analysing spatio-temporal sequences in bluetooth tracking data. *Applied Geography*, 34 (2012), 659–668. [Page 90]
- [117] DELHOME, R., BILLOT, R., DURET, A., AND EL FAOUZI, N.-E. The halphen distribution system, a toolbox for modeling travel time variability : some insights based on mesoscopic simulation. *Transportation Research Record : Journal of the Transportation Research Board* 2490 (2015), 32–40. [Pages 76, 84]
- [118] DELHOME, R., BILLOT, R., AND EL-FAOUZI, N. E. Moment-ratio diagram for travel time reliability : empirical study on urban and periurban links. In *Proc 6th International Symposium on Transportation Network Reliability (INSTR2015), Nara, Japan* (2015), Department of Urban Management, Kyoto University Kyoto, Japan. [Pages 76, 78]
- [119] DELHOME, R., BILLOT, R., AND EL FAOUZI, N.-E. Travel time statistical modeling with the halphen distribution family. *Journal of Intelligent Transportation Systems*, just-accepted (2017). [Pages 76, 79]
- [120] DEMAIZEAU, Y. From interactions to collective behaviour in agent-based systems. In *In : Proceedings of the 1st. European Conference on Cognitive Science. Saint-Malo* (1995), Citeseer. [Page 22]
- [121] DEMIRAY, H. An application of modified reductive perturbation method to long water waves. *International Journal of Engineering Science* 49, 12 (2011), 1397 – 1403. [Page 41]
- [122] DENG, W.-T., AND ZHAO, H.-B. Algorithm research for automatic incident detection based on wavelet analysis and contour method. In *Electrical and Control Engineering (ICECE), 2011 International Conference on* (2011), IEEE, pp. 539–543. [Page 29]
- [123] DESSALLES, J. L., FERBER, J., AND PHAN, D. *Intelligent complex adaptive systems*. Y. Shyan and A. Yang, 2007, ch. Emergence in Agent based Computational Social Science : conceptual, formal and diagrammatic analysis., pp. 255–299. [Page 22]
- [124] DIJKSTRA, E. W. A note on two problems in connexion with graphs. *Numerische mathematik* 1, 1 (1959), 269–271. [Page 31]
- [125] DING, J., GAO, S., JENELIUS, E., RAHMANI, M., HUANG, H., MA, L., PEREIRA, F., AND BEN-AKIVA, M. Routing policy choice set generation in stochastic time-dependent networks : Case studies for stockholm, sweden, and singapore. *Transportation Research Record : Journal of the Transportation Research Board* 2466 (2014), 76–86. [Page 32]

- [126] DINH, T.-U., BILLOT, R., PILLET, E., AND EL FAOUZI, N.-E. Real-time queue-end detection on freeways with floating car data : Practice-ready algorithm. *Transportation Research Record : Journal of the Transportation Research Board*, 2470 (2014), 46–56. [Page 75]
- [127] DION, F., AND RAKHA, H. Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B : Methodological*, 40(9) (2006), 745–766. [Page 27]
- [128] DONG, W., VU, H. L., NAZARATHY, Y., VO, B. Q., LI, M., AND HOOGENDOORN, S. P. Shortest paths in stochastic time-dependent networks with link travel time correlation. *Transportation Research Record : Journal of the Transportation Research Board* 2338, 1 (2013), 58–66. [Page 32]
- [129] DONIEC, A., ESPIÉ, S., MANDIAU, R., AND PIECHOWIAK, S. Multi-agent coordination and anticipation model to design a road traffic simulation tool. In *EUMAS* (2006). [Page 22]
- [130] DONIEC, A., MANDIAU, R., ESPIÉ, S., AND PIECHOWIAK, S. Comportements anticipatifs dans les systèmes multi-agents : application à la simulation de trafic routier. *Revue d'Intelligence Artificielle (RIA)* 21, 7 (2007), 183–221. [Page 22]
- [131] DONIEC, A., MANDIAU, R., PIECHOWIAK, S., AND ESPIÉ, S. A behavioral multi-agent model for road traffic simulation. *Engineering Applications of Artificial Intelligence* 21, 8 (2008), 1443 – 1454. [Page 22]
- [132] DRAZIN, P. G., AND JOHNSON, R. S. *Solitons : an introduction / P.G. Drazin, R.S. Johnson*. Cambridge University Press, Cambridge [England]; New York, 1989. [Page 41]
- [133] DRESCHER, G. L. *Made-up minds : a constructivist approach to artificial intelligence*. MIT press, 1991. [Page 23]
- [134] DRESNER, K., AND STONE, P. Multiagent traffic management : A reservation-based intersection control mechanism. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2* (2004), IEEE Computer Society, pp. 530–537. [Page 23]
- [135] DUDA, R. O., HART, P. E., AND STORK, D. G. *Pattern classification*. John Wiley & Sons, 2012. [Pages 28, 93, 95, 97, 100, 106]
- [136] EDIE, L. C. Discussion on traffic stream measurements and definitions. In *The 2nd International Symposium on the Theory of Traffic Flow* (1963), pp. 139–154. [Page 20]
- [137] EHRGOTT, M., WANG, J. Y. T., RAIH, A., AND VAN HOUTTE, C. A bi-objective cyclist route choice model. *Transportation Research Part A : Policy and Practice* 46, 4 (2012), 652–663. [Page 33]
- [138] EL FAOUZI, N.-E. Nonparametric traffic flow prediction using kernel estimator. In *International symposium on transportation and traffic theory* (1996), pp. 41–54. [Page 28]
- [139] EL FAOUZI, N.-E., BILLOT, R., AND BOUZEEDA, S. Motorway travel time prediction based on toll data and weather effect integration. *IET intelligent transport systems* 4, 4 (2010), 338–345. [Page 29]
- [140] EL FAOUZI, N.-E., AND MAURIN, M. Reliability metrics for path travel time under log-normal distribution. In *Proceedings of the 3rd International Symposium on Transportation Network Reliability (INSTR)* (2007). [Pages 25, 26]
- [141] EMAM, E., AND AI-DEEK, H. Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability. *Transportation Research Record : Journal of the Transportation Research Board* 1959 (2006), 140–150. [Pages 25, 26, 80]

- [142] ENGELSON, L., AND FOSGERAU, M. Additive measures of travel time variability. *Transportation Research Part B : Methodological* 45, 10 (2011), 1560 – 1571. [Page 25]
- [143] EPPSTEIN, D. Finding the k shortest paths. *SIAM Journal on Computing* 28, 2 (1998), 652–673. [Page 34]
- [144] ESPIÉ, S., AUBERLET, J. M., ET AL. Archisim : A behavioral multi-actors traffic simulation model for the study of a traffic system including its aspects. *International Journal of ITS Research*, n1 (2007). [Page 23]
- [145] EVANS, J., EVANS, M., MORGAN, H. G., HAYWARD, A., AND GUNNELL, D. Crisis card following self-harm : 12-month follow-up of a randomised controlled trial. *The British Journal of Psychiatry* 187, 2 (2005), 186–187. [Page 154]
- [146] FAN, Y., KALABA, R., AND MOORE II, J. Arriving on time. *Journal of Optimization Theory and Applications* 127, 3 (2005), 497–513. [Page 32]
- [147] FAN, Y., AND NIE, Y. Optimal routing for maximizing the travel time reliability. *Networks and Spatial Economics* 6, 3-4 (2006), 333–344. [Page 32]
- [148] FARRAHI, K., AND GATICA-PEREZ, D. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 1 (2011), 3. [Page 90]
- [149] FEBBRARO, A. D., RECAGNO, V., AND SACONE, S. INTRANET : a new simulation tool for intermodal transportation systems. *Simul. Pr. Theory* 4, 1 (1996), 47–64. [Page 33]
- [150] FELLENDORF, M. Vissim : A microscopic simulation tool to evaluate actuated signal control including bus priority. In *64th Institute of Transportation Engineers Annual Meeting* (1994), pp. 1–9. [Page 23]
- [151] FENG, J., SHEN, L., ZHANG, L., AND LIU, W. Combined method for traffic incident detection based on wavelet analysis and catastrophic theory. In *Intelligent Control and Automation (WCICA), 2010 8th World Congress on* (2010), IEEE, pp. 5604–5609. [Page 29]
- [152] FEO, T. A., AND RESENDE, M. G. Greedy randomized adaptive search procedures. *Journal of global optimization* 6, 2 (1995), 109–133. [Page 150]
- [153] FERBER, J. *Les Systèmes Multi-Agents : Vers une intelligence collective*. InterEdition, 1995. [Page 22]
- [154] FERBER, J. *Multi-agent systems : an introduction to distributed artificial intelligence*, vol. 1. Addison-Wesley Reading, 1999. [Pages 22, 49]
- [155] FINKELSTEIN, Y., MACDONALD, E. M., HOLLANDS, S., SIVILOTTI, M. L., HUTSON, J. R., MAMDANI, M. M., KOREN, G., AND JUURLINK, D. N. Risk of suicide following deliberate self-poisoning. *JAMA psychiatry* 72, 6 (2015), 570–575. [Page 132]
- [156] FOGEDBY, H. C. Solitons and diffusive modes in the noiseless burgers equation : Stability analysis. *Phys. Rev. E* 57 (Feb 1998), 2331–2337. [Page 41]
- [157] FOGUE, M., GARRIDO, P., MARTINEZ, F. J., CANO, J.-C., CALAFATE, C. T., MANZONI, P., AND SANCHEZ, M. Prototyping an automatic notification scheme for traffic accidents in vehicular networks. In *IFIP Wireless Days (WD)*. 2011. [Page 11]
- [158] FOSGERAU, M., AND KARLSTROM, A. The value of reliability. *Transportation Research Part-B : Methodological* 44, 1 (2010), 38–49. [Page 25]
- [159] FRADE, I., RIBEIRO, A., GONÇALVES, G., AND ANTUNES, A. Optimal location of charging stations for electric vehicles in a neighborhood in lisbon, portugal. *Transportation research record : journal of the transportation research board*, 2252 (2011), 91–98. [Page 31]

- [160] FRANK, H. Shortest paths in probabilistic graphs. *Operations Research* 17, 4 (1969), 583–599. [Pages 31, 32]
- [161] FREDERIX, R., VITI, F., AND TAMPÈRE, C. M. A hierarchical approach for dynamic origin-destination matrix estimation on large-scale congested networks. *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference* (October 2011), 1543–1548. [Page 27]
- [162] FU, L., AND RILETT, L. R. Expected shortest paths in dynamic and stochastic traffic networks. *Transportation Research Part B : Methodological* 32, 7 (1998), 499–516. [Page 32]
- [163] GABBAI, J., YIN, H., WRIGHT, W. A., AND ALLINSON, N. Self-organization, emergence and multi-agent systems. In *Neural Networks and Brain, 2005. ICNN B '05. International Conference on* (2005), vol. 3, pp. 1824–1863. [Page 22]
- [164] GAO, S., AND CHABINI, I. Optimal routing policy problems in stochastic time-dependent networks. *Transportation Research Part B : Methodological* 40, 2 (2006), 93–122. [Page 32]
- [165] GARROPO, R. G., GIORDANO, S., AND TAVANTI, L. A survey on multi-constrained optimal path computation : Exact and approximate algorithms. *Computer Networks* 54, 17 (2010), 3081–3107. [Page 33]
- [166] GE, H., CHENG, R., AND DAI, S. Kdv and kink and antikink solitons in car-following models. *Physica A : Statistical Mechanics and its Applications* 357, 30004 (2005), 466 – 476. [Page 41]
- [167] GE, H., DAI, S., AND DONG, L. An extended car-following model based on intelligent transportation system application. *Physica A : Statistical Mechanics and its Applications* 365, 2 (2006), 543 – 548. [Pages 22, 40, 41]
- [168] GE, H. X., DAI, S. Q., XUE, Y., AND DONG, L. Y. Stabilization analysis and modified korteweg–de vries equation in a cooperative driving system. *Phys. Rev. E* 71, 6 (Jun 2005), 066119. [Page 41]
- [169] GEISBERGER, R. Contraction hierarchies : Faster and simpler hierarchical routing in road networks. Master's thesis, Institut für Theoretische Informatik Universität Karlsruhe (TH), 2008. [Page 35]
- [170] GEISBERGER, R., SANDERS, P., SCHULTES, D., AND VETTER, C. Exact routing in large road networks using contraction hierarchies. *Transportation Science* 46, 3 (2012), 388–404. [Page 35]
- [171] GEROLIMINIS, N., AND DAGANZO, C. F. Existence of urban-scale macroscopic fundamental diagrams : Some experimental findings. *Transportation Research Part B : Methodological* 42, 9 (2008), 759–770. [Pages 85, 88]
- [172] GEROLIMINIS, N., AND SUN, J. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B : Methodological* 45, 3 (2011), 605 – 617. [Page 20]
- [173] GETTMAN, D., AND HEAD, L. Surrogate safety measures from traffic simulation models. *Transportation Research Record : Journal of the Transportation Research Board*, 1840 (2003), 104–115. [Page 53]
- [174] GIPPS, P. A behavioural car-following model for computer simulation. *Transportation Research Part B : Methodological* 15, 2 (1981), 105 – 111. [Page 21]
- [175] GLENN, C. R., AND NOCK, M. K. Improving the short-term prediction of suicidal behavior. *American Journal of Preventive Medicine* 47, 3 (2014), S176–S180. [Page 134]
- [176] GREENSHIELDS, B. D. The photographic method of studying traffic behavior. In *Highway Research Board Proceedings* (1934), vol. 13. [Page 20]

- [177] GUÉRIAU, M., ARMETTA, F., HASSAS, S., BILLOT, R., AND EL FAOUZI, N.-E. A constructivist approach for a self-adaptive decision-making system : application to road traffic control. In *Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on* (2016), IEEE, pp. 670–677. [Page 57]
- [178] GUÉRIAU, M., ARMETTA, F., HASSAS, S., BILLOT, R., AND EL FAOUZI, N.-E. Contrôle par apprentissage constructiviste et régulation du trafic coopératif. In *Journées Francophiles des Systèmes Multi-Agents 2017* (Caen, France, July 2017), G. B. Catherine Garbay, Ed., vol. 1602 of *Cohésion : fondement ou propriété émergente*, Cépaduès. [Page 57]
- [179] GUÉRIAU, M., BILLOT, R., EL FAOUZI, N.-E., HASSAS, S., AND ARMETTA, F. Multi-agent dynamic coupling for cooperative vehicles modeling. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (DEMO Track)* (2015), AAAI Press, pp. 4276–4277. [Page 49]
- [180] GUÉRIAU, M., BILLOT, R., EL FAOUZI, N.-E., MONTEIL, J., ARMETTA, F., AND HASSAS, S. How to assess the benefits of connected vehicles ? a simulation framework for the design of cooperative traffic management strategies. *Transportation Research Part C : Emerging Technologies* 67 (2016), 266 – 279. [Page 43]
- [181] GUERIN, F. Learning like a baby : a survey of artificial intelligence approaches. *The Knowledge Engineering Review* 26, 02 (2011), 209–236. [Page 23]
- [182] GUERRIERO, F., AND MUS, R. Label correcting methods to solve multicriteria shortest path problems. *Journal of Optimization Theory and Applications* 111, 3 (2001), 589–613. [Page 33]
- [183] GUESSOUS, Y., ARON, M., BHOURI, N., AND COHEN, S. Estimating travel time distribution under different traffic conditions. *Transportation Research Procedia* 3 (2014), 339–348. [Page 26]
- [184] HAKIMI, S. L. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations research* 12, 3 (1964), 450–459. [Page 30]
- [185] HALPHEN, E. Sur un nouveau type de courbe de fréquence. *Compte-Rendus de l'Académie des Sciences* 213 (1941), 633–635. [Page 76]
- [186] HALPHEN, É. *Les Fonctions factorielles*. Inst. de Statistique de l'Université de Paris, 1955. [Page 77]
- [187] HAMNER, B. Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. In *2010 IEEE International Conference on Data Mining Workshops* (2010), IEEE, pp. 1357–1359. [Page 28]
- [188] HANSEN, P. Bicriterion path problems. In *Multiple criteria decision making theory and application*, vol. 177. Springer, 1980, pp. 109–127. [Page 33]
- [189] HARALAMBOUS, Y., AND LENCA, P. Text classification using association rules, dependency pruning and hyperonymization. *arXiv preprint arXiv :1407.7357* (2014). [Page 152]
- [190] HART, P. E., NILSSON, N. J., AND RAPHAEL, B. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2 (1968), 100–107. [Page 34]
- [191] HAWORTH, J. *Spatio-temporal forecasting of network data*. PhD thesis, UCL (University College London), 2014. [Page 96]
- [192] HAWORTH, J., AND CHENG, T. Non-parametric regression for space-time forecasting under missing data. *Computers, Environment and Urban Systems* 36, 6 (2012), 538–550. [Pages 27, 28, 96]
- [193] HELBING, D. Traffic and related self-driven many-particle systems. *Reviews of modern physics* 73, 4 (2001), 1067. [Pages 20, 21]

- [194] HELBING, D. Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B* 70, 2 (2009), 229–241. [Page 20]
- [195] HELBING, D., AND BALIETTI, S. *Social Self-Organization*. Springer, 2012, ch. Agent-Based Modeling, pp. 25–70. [Page 43]
- [196] HELBING, D., AND TILCH, B. Generalized force model of traffic dynamics. *Phys. Rev. E* 58, 1 (Jul 1998), 133–138. [Page 40]
- [197] HENIG, M. I. Efficient interactive methods for a class of multiattribute shortest path problems. *Management Science* 40, 7 (1994), 891–897. [Page 33]
- [198] HERRING, R., HOFLEITNER, A., ABBEEL, P., AND BAYEN, A. Estimating arterial traffic conditions using sparse probe data. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on* (2010), IEEE, pp. 929–936. [Page 95]
- [199] HILGER, M., KÖHLER, E., MÖHRING, R. H., AND SCHILLING, H. Fast point-to-point shortest path computations with arc-flags. *The Shortest Path Problem : Ninth DIMACS Implementation Challenge 74* (2009), 41–72. [Page 34]
- [200] HOFLEITNER, A., HERRING, R., BAYEN, A., HAN, Y., MOUTARDE, F., AND DE LA FORTELLE, A. Large scale estimation of arterial traffic and structural analysis of traffic patterns using probe vehicles. In *Transportation Research Board 91st Annual Meeting (TRB'2012)* (2012). [Page 96]
- [201] HOLLANDER, Y. Travellers' attitudes to travel time variability : Intermodal and intra-modal analysis. In *Proceedings of the 3rd International SIIV Congress in Bari, Italy* (2005). [Page 25]
- [202] HOLLANDER, Y. Direct versus indirect models for the effects of unreliability. *Transportation Research Part A : Policy and Practice* 40, 9 (2006), 699 – 711. [Page 25]
- [203] HOOGENDOORN, S. P., AND BOVY, P. H. State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part I : Journal of Systems and Control Engineering* 215, 4 (2001), 283–303. [Page 21]
- [204] HUTSON, K. R., AND SHIER, D. R. Extended dominance and a stochastic shortest path problem. *Computers & Operations Research* 36, 2 (2009), 584–596. [Page 32]
- [205] HUYNH, T., FRITZ, M., AND SCHIELE, B. Discovery of activity patterns using topic models. *Proceedings of the 10th international conference on Ubiquitous computing, ACM New York, NY, USA* (2008), 10–19. [Page 90]
- [206] HUYNH, T. D., JENNINGS, N. R., AND SHADBOLT, N. R. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 13, 2 (2006), 119–154. [Page 53]
- [207] ILLNER, R., KIRCHNER, C., AND PINNEAU, R. A derivation of the aw-rasclé traffic models from fokker-planck type kinetic models. *Quant. Appl. Math* 27 (2009), 13–45. [Page 21]
- [208] ITALIANO, G. F., AND NANNI, U. Online maintenance of minimal directed hypergraphs. [Page 33]
- [209] JENELIUS, E. The value of travel time variability with trip chains, flexible scheduling and correlated travel times. *Transportation Research Part B : Methodological* 46, 6 (2012), 762 – 780. [Page 25]
- [210] JIANG, R., WU, Q., AND ZHU, Z. Full velocity difference model for a car-following theory. *Phys. Rev. E* 64 (Jun 2001), 017101. [Page 21]
- [211] JIN, J., AND RAN, B. Automatic freeway incident detection based on fundamental diagrams of traffic flow. *Transportation Research Record : Journal of the Transportation Research Board*, 2099 (2009), 65–75. [Page 29]

- [212] JINTANAKUL, K., CHU, L., AND JAYAKRISHNAN, R. Bayesian mixture model for estimating freeway travel time distributions from small probe samples from multiple days. No. 2136, Transportation Research Board of the National Academies, pp. 37–44. [Pages 25, 26]
- [213] JOERER, S., SOMMER, C., AND DRESSLER, F. Toward reproducibility and comparability of ivc simulation studies : a literature survey. *IEEE Communications Magazine* 50, 10 (2012), 82–88. [Page 23]
- [214] JOHNSON, D. B. Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM (JACM)* 24, 1 (1977), 1–13. [Page 117]
- [215] JOURQUIN, B., AND BEUTHE, M. Transportation policy analysis with a geographic information system : the virtual network of freight transportation in europe. *Transportation research part c : emerging technologies* 4, 6 (1996), 359–371. [Page 33]
- [216] JOZEFOWIEZ, N., SEMET, F., AND TALBI, E.-G. Multi-objective vehicle routing problems. *European Journal of Operational Research* 189, 2 (2008), 293–309. [Page 33]
- [217] KAMRAN, S., AND HAAS, O. A multilevel traffic incidents detection approach : Identifying traffic patterns and vehicle behaviours using real-time gps data. In *Intelligent Vehicles Symposium, 2007 IEEE* (2007), IEEE, pp. 912–917. [Page 29]
- [218] KAPARIAS, I., BELL, M. G. H., AND BELZNER, H. A new measure of travel time reliability for in-vehicle navigation systems. *Journal of International Transportation Systems* 12, 4 (2008), 202–211. [Page 26]
- [219] KARMAKAR, C., LUO, W., TRAN, T., BERK, M., AND VENKATESH, S. Predicting risk of suicide attempt using history of physical illnesses from electronic medical records. *JMIR mental health* 3, 3 (2016). [Page 134]
- [220] KAUFMAN, L., AND ROUSSEEUW, P. *Finding groups in data : an introduction to cluster analysis*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, 2005. [Page 94]
- [221] KERNER, B. S. Complexity of spatiotemporal traffic phenomena in flow of identical drivers : Explanation based on fundamental hypothesis of three-phase theory. *Phys. Rev. E* 85 (Mar 2012), 036110. [Page 41]
- [222] KERNER, B. S. Failure of generally accepted classical traffic flow theories. In *Breakdown in Traffic Networks*. Springer, 2017, pp. 123–186. [Page 21]
- [223] KESTING, A., AND TREIBER, M. Calibrating car-following models by using trajectory data : Methodological study. *Transportation Research Record : Journal of the Transportation Research Board* 2088 (2008), 148–156. [Page 21]
- [224] KESTING, A., TREIBER, M., AND HELBING, D. General lane-changing model mobil for car-following models. *Transportation Research Record : Journal of the Transportation Research Board* 1999 / 2007 *Traffic Flow Theory 2007* (2007), 86–94. [Pages 21, 45, 46, 50]
- [225] KESTING, A., TREIBER, M., AND HELBING, D. *Multi-Agent Systems : Simulation and Applications*. 2009, ch. Agents for Traffic Simulation, pp. 325–356. [Pages 23, 24]
- [226] KESTING, A., TREIBER, M., AND HELBING, D. Connectivity statistics of store-and-forward intervehicle communication. *Intelligent Transportation Systems, IEEE Transactions on* 11, 1 (2010), 172–181. [Page 23]
- [227] KESTING, A., TRIEBER, M., AND HELBING, D. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions of the Royal Society A : Mathematical Physical & Engineering Sciences* 368, 1928 (October 2010), 4585–4605. [Pages 21, 22, 45]

- [228] KHANI, A., AND BOYLES, S. D. An exact algorithm for the mean–standard deviation shortest path problem. *Transportation Research Part B : Methodological* 81 (2015), 252–266. [Page 32]
- [229] KHONDAKER, B., AND KATTAN, L. Variable speed limit : an overview. *Transportation Letters* 7, 5 (2015), 264–278. [Page 23]
- [230] KIM, J., AND MAHMASSANI, H. S. Compound gamma representation for modeling vehicle-to-vehicle and day-to-day travel time variability in a traffic network. In *Transportation Research Board 93rd Annual Meeting* (2014), no. 14-1616. [Page 25]
- [231] KIM, J., AND MAHMASSANI, H. S. Compound gamma representation for modeling travel time variability in a traffic network. *Transportation Research Part B : Methodological* 80 (2015), 40–63. [Pages 26, 83]
- [232] KIM, T., KIM, H., AND LOVELL, D. J. Traffic flow forecasting : overcoming memoryless property in nearest neighbor non-parametric regression. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.* (2005), IEEE, pp. 965–969. [Page 28]
- [233] KIRBY, H. R., WATSON, S. M., AND DOUGHERTY, M. S. Should we use neural networks or statistical models for short-term motorway traffic forecasting ? *International Journal of Forecasting* 13, 1 (1997), 43–50. [Page 28]
- [234] KIRCHLER, D., LIBERTI, L., AND WOLFLER CALVO, R. Efficient computation of shortest paths in time-dependent multi-modal networks. *Journal of Experimental Algorithmics (JEA)* 19, 2.5 (2014). [Page 33]
- [235] KIRSCHFINK, H. Knowledge-based system for the completion of traffic data. *European journal of operational research* 71, 2 (1993), 247–256. [Page 27]
- [236] KNOPP, S., SANDERS, P., SCHULTES, D., SCHULZ, F., AND WAGNER, D. Computing many-to-many shortest paths using highway hierarchies. In *2007 Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX)* (2007), SIAM, pp. 36–45. [Page 34]
- [237] KNOSPE, W., SANTEN, L., SCHADSCHNEIDER, A., AND SCHRECKENBERG, M. Empirical test for cellular automaton models of traffic flow. *Physical Review E* 70, 1 (2004), 016115. [Page 21]
- [238] KOLAR, D. R., BÜRGER, A., HAMMERLE, F., AND JENETZKY, E. Aversive tension of adolescents with anorexia nervosa in daily course : a case-controlled and smartphone-based ambulatory monitoring trial. *BMJ open* 4, 4 (2014), e004703. [Page 132]
- [239] KOSONEN, I. Multi-agent fuzzy signal control based on real-time simulation. *Transportation Research Part C : Emerging Technologies* 11, 5 (2003), 389 – 403. World Congress on Intelligent Transport Systems. [Page 23]
- [240] KOVALEV, S., AND BILLOT, R. An exact method for multi-objective multi-modal trip planning problem. In *28th European Conference on Operations Research (EURO)* (2016). [Pages 129, 153]
- [241] KRAENKEL, R. A., PEREIRA, J. G., AND MANNA, M. A. The reductive perturbation method and the korteweg-de vries hierarchy. *Acta Applicandae Mathematicae* 39 (1995), 389–403. 10.1007/BF00994645. [Page 41]
- [242] KRAJZEWICZ, D., ERDMANN, J., BEHRISCH, M., AND BIEKER, L. Recent development and applications of SUMO - Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements* 5, 3&4 (December 2012), 128–138. [Page 23]
- [243] KRAWCZYK, B., MINKU, L. L., GAMA, J., STEFANOWSKI, J., AND WOŹNIAK, M. Ensemble learning for data stream analysis : a survey. *Information Fusion* 37 (2017), 132–156. [Page 156]
- [244] LACOMME, P., PRINS, C., AND SEVAUX, M. *Algorithmes de graphes*, vol. 28. Eyrolles Paris, 2003. [Page 33]

- [245] LAHAROTTE, P.-A., BILLOT, R., COME, E., OUKHELLOU, L., NANTES, A., AND EL FAOUZI, N.-E. Spatiotemporal analysis of bluetooth data : Application to a large urban network. *IEEE transactions on intelligent transport systems* (2014). [Pages 26, 152]
- [246] LAHAROTTE, P.-A., BILLOT, R., AND EL FAOUZI, N.-E. Detection of non-recurrent road traffic events based on clustering indicators. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)* (2017). [Pages 95, 105]
- [247] LAHAROTTE, P.-A., BILLOT, R., FAOUZI, E., RAKHA, H. A., ET AL. Network-wide traffic state prediction using bluetooth data. In *Transportation Research Board 94th Annual Meeting* (2015), no. 15-3022. [Page 26]
- [248] LAM, W. H., SHAO, H., AND SUMALEE, A. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation research part B : methodological* 42, 10 (2008), 890–910. [Page 29]
- [249] LARSEN, M. E., NICHOLAS, J., AND CHRISTENSEN, H. A systematic assessment of smartphone tools for suicide prevention. *PLoS one* 11, 4 (2016), e0152285. [Page 143]
- [250] LAVAL, J. A., AND DAGANZO, C. F. Lane-changing in traffic streams. *Transportation Research Part B : Methodological* 40, 3 (2006), 251 – 264. [Page 21]
- [251] LEBLOND, H. The reductive perturbation method and some of its applications. *Journal of Physics B : Atomic, Molecular and Optical Physics* 41, 4 (2008), 043001. [Page 41]
- [252] LECLUYSE, C., VAN WOENSEL, T., AND PEREMANS, H. Vehicle routing with stochastic time-dependent travel times. *4OR* 7, 4 (2009), 363–377. [Page 32]
- [253] LECRUBIER, Y., SHEEHAN, D. V., WEILLER, E., AMORIM, P., BONORA, I., SHEEHAN, K. H., JANAVS, J., AND DUNBAR, G. C. The mini international neuropsychiatric interview (mini). a short diagnostic structured interview : reliability and validity according to the cidi. *European psychiatry* 12, 5 (1997), 224–231. [Page 136]
- [254] LENCA, P., MEYER, P., VAILLANT, B., AND LALLICH, S. On selecting interestingness measures for association rules : User oriented description and multiple criteria decision aid. *European journal of operational research* 184, 2 (2008), 610–626. [Page 152]
- [255] LENZ, H., WAGNER, C., AND SOLLACHER, R. Multi-anticipative car-following model. *The European Physical Journal B - Condensed Matter and Complex Systems* 7, 2 (1999), 331–335. [Pages 21, 22]
- [256] LESHEM, G., AND RITOV, Y. Traffic flow prediction using adaboost algorithm with random forests as a weak learner. In *Proceedings of World Academy of Science, Engineering and Technology* (2007), vol. 19, Citeseer, pp. 193–198. [Page 28]
- [257] LEVIN, M., AND KRAUSE, G. M. Incident detection : A bayesian approach. *Transportation Research Record* 682 (1978), 52–58. [Page 29]
- [258] LEYTON-BROWN, K., NUDELMAN, E., ANDREW, G., MCFADDEN, J., AND SHOHAM, Y. A portfolio approach to algorithm selection. In *IJCAI* (2003), vol. 3, pp. 1542–1543. [Page 150]
- [259] LI, K., AND IOANNOU, P. Modeling of traffic flow of automated vehicles. *Intelligent Transportation Systems, IEEE Transactions on* 5, 2 (2004), 99–113. [Page 22]
- [260] LI, Y., LI, Z., WANG, H., WANG, W., AND XING, L. Evaluating the safety impact of adaptive cruise control in traffic oscillations on freeways. *Accident Analysis & Prevention* 104 (2017), 137–145. [Page 23]
- [261] LI, Z., AND ZHANG, R. An extended non-lane-based optimal velocity model with dynamic collaboration. *Mathematical Problems in Engineering* 2013 (2013), 8. [Page 21]

- [262] LIGHTHILL, M. J., AND WHITHAM, G. B. On kinematic waves. I. flood movement in long rivers. In *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* (1955), vol. 229, The Royal Society, pp. 281–316. [Page 20]
- [263] LIGHTHILL, M. J., AND WHITHAM, G. B. On kinematic waves. II. a theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* (1955), vol. 229, The Royal Society, pp. 317–345. [Page 20]
- [264] LIN, S.-H. Finding optimal refueling policies : a dynamic programming approach. *Journal of Computing Sciences in Colleges* 23, 6 (2008), 272–279. [Page 31]
- [265] LIN, S.-H. Finding optimal refueling policies in transportation networks. In *International Conference on Algorithmic Applications in Management* (2008), Springer, pp. 280–291. [Page 31]
- [266] LIN, S. H., GERTSCH, N., AND RUSSELL, J. R. A linear-time algorithm for finding optimal vehicle refueling policies. *Operations Research Letters* 35, 3 (2007), 290–296. [Page 31]
- [267] LINAS, B. S., GENZ, A., WESTERGAARD, R. P., CHANG, L. W., BOLLINGER, R. C., LATKIN, C., AND KIRK, G. D. Ecological momentary assessment of illicit drug use compared to biological and self-reported methods. *JMIR mHealth and uHealth* 4, 1 (2016). [Page 132]
- [268] LIU, H. X., RECKER, W., AND CHEN, A. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A : Policy and Practice* 38, 6 (2004), 435 – 453. [Page 25]
- [269] LIU, J., AND GUAN, W. A summary of traffic flow forecasting methods [j]. *Journal of Highway and Transportation Research and Development* 3 (2004), 82–85. [Page 28]
- [270] LIU, Q., LU, J., AND CHEN, S. Traffic incident detection using random forest. In *Transportation Research Board 92nd Annual Meeting* (2013), no. 13-1610. [Page 29]
- [271] LIU, Y., LEI, Y., YI, Q., JIANQUAN, W., AND HUIMIN, W. Traffic incident detection algorithm for urban expressways based on probe vehicle data. *Journal of Transportation systems engineering and information technology* 8, 4 (2008), 36–41. [Page 29]
- [272] LOGI, F., AND RITCHIE, S. G. A multi-agent architecture for cooperative inter-jurisdictional traffic congestion management. *Transportation Research Part C : Emerging Technologies* 10 (2002), 507 – 527. [Page 23]
- [273] LOMAX, T., SCHRANK, D., TURNER, S., AND MARGIOTTA, R. Selecting travel reliability measures. Tech. rep., Texas Transportation Institute, 2003. [Pages 25, 82]
- [274] LOMBARD, A., PERRONNET, F., ABBAS-TURKI, A., AND EL MOUDNI, A. On the cooperative automatic lane change : Speed synchronization and automatic “courtesy”. In *2017 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (2017), IEEE, pp. 1655–1658. [Page 22]
- [275] LONG CHEU, R., SRINIVASAN, D., AND HOON LOO, W. Training neural networks to detect freeway incidents by using particle swarm optimization. *Transportation Research Record : Journal of the Transportation Research Board*, 1867 (2004), 11–18. [Page 29]
- [276] LOPEZ-CASTROMAN, J., DE LAS MERCEDES PEREZ-RODRIGUEZ, M., JAUSSENT, I., ALEGRIA, A. A., ARTES-RODRIGUEZ, A., FREED, P., GUILLAUME, S., JOLLANT, F., LEIVA-MURILLO, J. M., MALAFOSSE, A., ET AL. Distinguishing the relevant features of frequent suicide attempters. *Journal of psychiatric research* 45, 5 (2011), 619–625. [Page 135]
- [277] LOSILLA, F., GARCIA-SANCHEZ, A.-J., GARCIA-SANCHEZ, F., GARCIA-HARO, J., AND HAAS, Z. J. A comprehensive approach to wsn-based its applications : A survey. *Sensors* 11, 11 (2011), 10220–10265. [Page 40]

- [278] LOUI, R. P. Optimal paths in graphs with stochastic or multidimensional weights. *Communications of the ACM* 26, 9 (1983), 670–676. [Page 32]
- [279] LOZANO, A., AND STORCHI, G. Shortest viable path algorithm in multimodal networks. *Transportation Research Part A : Policy and Practice* 35, 3 (2001), 225–241. [Page 34]
- [280] LOZANO, A., AND STORCHI, G. Shortest viable hyperpath in multimodal networks. *Transportation Research Part B : Methodological* 36, 10 (2002), 853–874. [Pages 33, 34]
- [281] LU, H. Short-term traffic prediction using rainfall. *Int. J. Signal Process. Syst* 2, 1 (2014), 70–73. [Page 29]
- [282] LU, J., BAN, X., QIU, Z., YANG, F., AND RAN, B. Robust route guidance model based an advanced traveler information systems. *Transportation Research Record : Journal of the Transportation Research Board* 1935, 1 (2005), 1–7. [Page 32]
- [283] LUNGARELLA, M., METTA, G., PFEIFER, R., AND SANDINI, G. Developmental robotics : a survey. *Connection Science* 15, 4 (2003), 151–190. [Page 23]
- [284] MAHMASSANI, H., HOU, T., AND DONG, J. Characterizing travel time variability in vehicular traffic networks : deriving a robust relation for reliability analysis. *Transportation Research Record : Journal of the Transportation Research Board*, 2315 (2012), 141–152. [Page 25]
- [285] MALINOVSKIY, Y., LEE, U.-K., WU, Y.-J., AND WANG, Y. Investigation of bluetooth-based travel time estimation error on a short corridor. *90th Annual Transportation Research Board Meeting, Washington, DC.* (January 2011). [Page 72]
- [286] MALINOVSKIY, Y., AND WANG, Y. Pedestrian travel pattern discovery using mobile bluetooth sensors. *Proceedings of the Transportation Research Board 91st Annual Meeting* (2012), 1–16. [Page 72]
- [287] MANDOW, L., AND PÉREZ DE LA CRUZ, J. L. A new approach to multiobjective A* search. In *IJCAI'05 Proceedings of the 19th international joint conference on Artificial intelligence Pages* (2005), pp. 218–223. [Page 33]
- [288] MARTINEZ, F. J., TOH, C. K., CANO, J.-C., CALAFATE, C. T., AND MANZONI, P. A survey and comparative study of simulators for vehicular ad hoc networks (vanets). *Wireless Communications and Mobile Computing* 11, 7 (2011), 813–828. [Page 23]
- [289] MARTINS, E. Q. V. On a multicriteria shortest path problem. *European Journal of Operational Research* 16, 2 (1984), 236–245. [Page 33]
- [290] MAZAC, S., ARMETTA, F., AND HASSAS, S. On bootstrapping sensori-motor patterns for a constructivist learning system in continuous environments. In *Alife 14 : Fourteenth International Conference on the Synthesis and Simulation of Living Systems* (2014). [Page 23]
- [291] MCHUGH, D. Traffic prediction and analysis using a big data and visualisation approach, 2015. [Page 29]
- [292] MEEDEN, L. A., AND BLANK, D. S. Introduction to developmental robotics. *Connection Science* 18, 2 (2006), 93–96. [Page 23]
- [293] MEGIDDO, N., ZEMEL, E., AND HAKIMI, S. L. The maximum coverage location problem. *SIAM Journal on Algebraic Discrete Methods* 4, 2 (1983), 253–261. [Page 31]
- [294] MENG, F. H., YIZHI, L., WAI, L. H., AND CHUIN, L. H. A multi-criteria, multi-modal passenger route advisory system. In *Proceedings of the IES-CTR International Symposium, Singapore* (1999). [Page 34]

- [295] MEYER, P., AND OLTEANU, A.-L. Integrating large positive and negative performance differences into multicriteria majority-rule sorting models. *Computers & Operations Research* 81 (2017), 216–230. [Page 154]
- [296] MEYER, P., AND ROUBENS, M. On the use of the choquet integral with fuzzy numbers in multiple criteria decision support. *Fuzzy Sets and Systems* 157, 7 (2006), 927–938. [Page 154]
- [297] MICHAU, G., BORGNAT, P., PUSTELNIK, N., ABRY, P., NANTES, A., AND CHUNG, E. Estimating link-dependent origin-destination matrices from sample trajectories and traffic counts. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (2015), IEEE, pp. 5480–5484. [Page 27]
- [298] MICHAU, G., NANTES, A., BHASKAR, A., CHUNG, E., ABRY, P., AND BORGNAT, P. Bluetooth data in an urban context : Retrieving vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems* (2017). [Page 27]
- [299] MILANÉS, V., SHLADOVER, S. E., SPRING, J., NOWAKOWSKI, C., KAWAZOE, H., AND NAKAMURA, M. Cooperative adaptive cruise control in real traffic situations. *IEEE Transactions on Intelligent Transportation Systems* 15, 1 (2014), 296–305. [Page 22]
- [300] MILLER-HOOKS, E. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. *Networks* 37, 1 (2001), 35–52. [Page 32]
- [301] MILLER-HOOKS, E. D., AND MAHMASSANI, H. S. Least possible time paths in stochastic, time-varying networks. *Computers & operations research* 25, 12 (1998), 1107–1125. [Pages 32, 121]
- [302] MILLER-HOOKS, E. D., AND MAHMASSANI, H. S. Least expected time paths in stochastic, time-varying transportation networks. *Transportation Science* 34, 2 (2000), 198–215. [Page 32]
- [303] MILLER-HOOKS, E. D., AND MAHMASSANI, H. S. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. *European Journal of Operational Research* 146, 1 (2003), 67–82. [Pages 32, 121]
- [304] MINDERHOUD, M. M., AND BOVY, P. H. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention* 33, 1 (2001), 89–97. [Page 53]
- [305] MIRCHANDANI, P. B. Shortest distance and reliability of probabilistic networks. *Computers & Operations Research* 3, 4 (1976), 347–355. [Page 31]
- [306] MIRZAEI, S. Node formulation and a heuristic algorithm for location routing problem of electric-powered vehicles. [Page 31]
- [307] MONTEIL, J., BILLOT, R., REY, D., AND EL FAOUZI, N.-E. Distributed and centralized approaches for cooperative road traffic. *Proceedings of TRA 2012, Athens, to be published Procedia - Social and Behavioral Sciences* (2012). [Page 40]
- [308] MONTEIL, J., BILLOT, R., SAU, J., BUISSON, C., AND EL FAOUZI, N.-E. Calibration, estimation, and sampling issues of car-following parameters. *Transportation Research Record : Journal of the Transportation Research Board*, 2422 (2014), 131–140. [Pages 21, 51]
- [309] MONTEIL, J., BILLOT, R., SAU, J., AND EL FAOUZI, N.-E. Linear and weakly nonlinear stability analyses of cooperative car-following models. *IEEE Transactions on Intelligent Transportation Systems* 15, 5 (2014), 2001–2013. [Pages 41, 42, 45]
- [310] MONTEIL, J., NANTES, A., BILLOT, R., AND EL FAOUZI, N.-E. Microscopic cooperative traffic based on ngsim data. *ITS World Congress, Vienna* (2012). [Page 40]
- [311] MONTOLIU, R. Discovering mobility patterns on bicycle-based public transportation system by using probabilistic topic models. *ISAmI* (2012), 145–153. [Page 91]

- [312] MORLAT, G. Les lois de probabilités de halphen. *Revue de Statistique Appliquée* 3 (1956), 21–46. [Pages 76, 77]
- [313] MUGAN, J., AND KUIPERS, B. Autonomous representation learning in a developing agent. In *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Springer, 2013, pp. 63–80. [Page 58]
- [314] MULDER, R., NEWTON-HOWES, G., AND COID, J. W. The futility of risk prediction in psychiatry, 2016. [Page 132]
- [315] MÜLLER-HANNEMANN, M., AND SCHNEE, M. Finding all attractive train connections by multi-criteria pareto search. In *Algorithmic Methods for Railway Optimization*, F. Geraets, L. Kroon, A. Schoebel, D. Wagner, and C. D. Zaroliagis, Eds., no. 4359 in Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg, Jan. 2007, pp. 246–263. [Page 33]
- [316] MURAMATSU, M., AND NAGATANI, T. Soliton and kink jams in traffic flow with open boundaries. *Phys. Rev. E* 60 (Jul 1999), 180–187. [Page 41]
- [317] MURTHY, I., AND OLSON, D. L. An interactive procedure using domination cones for bicriterion shortest path problems. *European Journal of Operational Research* 72, 2 (1994), 417–431. [Page 33]
- [318] MYIN-GERMEYS, I., PEETERS, F., HAVERMANS, R., NICOLSON, N., DEVRIES, M., DELESPAUL, P., AND VAN OS, J. Emotional reactivity to daily life stress in psychosis and affective disorder : an experience sampling study. *Acta Psychiatrica Scandinavica* 107, 2 (2003), 124–131. [Page 132]
- [319] NAGEL, K., AND SCHRECKENBERG, M. A cellular automaton model for freeway traffic. *Journal de physique I* 2, 12 (1992), 2221–2229. [Page 21]
- [320] NAHUM, M., VAN VLEET, T. M., SOHAL, V. S., MIRZABEKOV, J. J., RAO, V. R., WALLACE, D. L., LEE, M. B., DAWES, H., STARK-INBAR, A., JORDAN, J. T., ET AL. Immediate mood scaler : Tracking symptoms of depression and anxiety using a novel mobile mood scale. *JMIR mHealth and uHealth* 5, 4 (2017). [Page 132]
- [321] NAJJAR, A., AND REIGNIER, P. Constructivist ambient intelligent agent for smart environments. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on* (2013), IEEE, pp. 356–359. [Page 23]
- [322] NANTES, A., MISKA, M. P., BHASKAR, A., CHUNG, E., ET AL. Noisy bluetooth traffic data ? *Road & Transport Research : A Journal of Australian and New Zealand Research and Practice* 23, 1 (2014), 33–43. [Page 72]
- [323] NATARAJAN, A., MOTANI, M., AND SRINIVASAN, V. Understanding urban interactions from bluetooth phone contact traces. In *in Proceedings of PAM* (2007), pp. 115–124. [Page 89]
- [324] NEWELL, G. F. A simplified car-following theory : a lower order model. *Transportation Research Part B : Methodological* 36, 3 (2002), 195 – 205. [Page 21]
- [325] NGODUY, D., HOOGENDOORN, S., AND LIU, R. Continuum modeling of cooperative traffic flow dynamics. *Physica A : Statistical Mechanics and its Applications* 388, 13 (2009), 2705–2716. [Page 22]
- [326] NGODUY, D., AND WILSON, R. Multianticipative nonlocal macroscopic traffic model. *Computer-Aided Civil and Infrastructure Engineering* 29, 4 (2014), 248–263. [Page 22]
- [327] NGUYEN, S., AND PALLOTTINO, S. Equilibrium traffic assignment for large scale transit networks. *European journal of operational research* 37, 2 (1988), 176–186. [Page 33]

- [328] NGUYEN VU, Q.-A., CANAL, R., GAUDOU, B., HASSAS, S., AND ARMETTA, F. Trustsets : using trust to detect deceitful agents in a distributed information collecting system. *Journal of Ambient Intelligence and Humanized Computing* 3, 4 (2012), 251–263. [Page 43]
- [329] NICHOLAS, J., LARSEN, M. E., PROUDFOOT, J., AND CHRISTENSEN, H. Mobile apps for bipolar disorder : a systematic review of features and content quality. *Journal of medical Internet research* 17, 8 (2015). [Page 132]
- [330] NIE, Y. M., AND WU, X. Shortest path problem considering on-time arrival probability. *Transportation Research Part B : Methodological* 43, 6 (2009), 597–613. [Pages 32, 120, 121, 122, 123]
- [331] NIEBLES, J. C., WANG, H., AND FEI-FEI, L. Unsupervised learning of human action categories using spatial-temporal words. *Int. J. Comput. Vision* 79, 3 (2008), 299–318. [Page 90]
- [332] NIKOLOVA, E. High-performance heuristics for optimization in stochastic traffic engineering problems. In *Large-Scale Scientific Computing*. Springer, 2010, pp. 352–360. [Page 32]
- [333] NOLAND, R. B., AND POLAK, J. W. Travel time variability : A review of theoretical and empirical issues. *Transport Review* 22 (nov 2002), 39–54. [Page 25]
- [334] NOLAND, R. B., AND SMALL, K. A. *Travel-time uncertainty, departure time choice, and the cost of the morning commute*. Institute of Transportation Studies, University of California, Irvine, 1995. [Page 25]
- [335] ÖNCÜ, S., VAN DE WOUW, N., HEEMELS, W. M. H., AND NIJMEIJER, H. String stability of interconnected vehicles under communication constraints. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)* (2012), IEEE, pp. 2459–2464. [Page 22]
- [336] ONISHI, H., AND MLINARSKY, F. Wireless technology assessment for automotive applications. In *Proceedings of the ITS World Congress* (2012). [Page 12]
- [337] OSSEN, S., AND HOOGENDOO, S. Multi-anticipation and heterogeneity in car-following empirics and a first exploration of their implications. In *2006 IEEE Intelligent Transportation Systems Conference* (2006), IEEE, pp. 1615–1620. [Pages 22, 40]
- [338] OSWALD, R. K., SCHERER, W. T., AND SMITH, B. L. Traffic flow forecasting using approximate nearest neighbor nonparametric regression. *Final project of ITS Center project : Traffic forecasting : non-parametric regressions* (2000). [Page 28]
- [339] PALLOTTINO, S., AND SCUTELLA, M. G. Shortest path algorithms in transportation models : classical and innovative aspects. Tech. rep., University of Pisa, 1997. [Pages 33, 34]
- [340] PAN, Y., SUN, L., AND GE, M. Finding reliable shortest path in stochastic time-dependent network. *Procedia-Social and Behavioral Sciences* 96 (2013), 451–460. [Page 32]
- [341] PAPADIMITRATOS, P., DE LA FORTELLE, A., EVENSSEN, K., BRIGNOLO, R., AND COSENZA, S. Vehicular communication systems : Enabling technologies, applications, and future outlook on intelligent transportation. *IEEE Communications Magazine* 47, 11 (2009), 84–95. [Page 11]
- [342] PAPAGEORGIU, M., DIAKAKI, C., DINOPOULOU, V., KOTSIALOS, A., AND WANG, Y. Review of road traffic control strategies. *Proceedings of the IEEE* 91, 12 (2003), 2043–2067. [Page 23]
- [343] PAPAGEORGIU, M., HADJ-SALEM, H., AND BLOSSEVILLE, J.-M. Alinea : A local feedback control law for on-ramp metering. *Transportation Research Record*, 1320 (1991). [Page 10]
- [344] PASCALE, A., MAVROEIDIS, D., AND LAM, H. T. Spatiotemporal clustering of urban networks : Real case scenario in london. *Transportation Research Record : Journal of the Transportation Research Board*, 2491 (2015), 81–89. [Page 96]

- [345] PASOLINI, G., AND VERDONE, R. Bluetooth for its ? In *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on* (2002), vol. 1, pp. 315–319 vol.1. [Page 26]
- [346] PATTON, J. H., STANFORD, M. S., ET AL. Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology* 51, 6 (1995), 768–774. [Page 136]
- [347] PAVERI-FONTANA, S. On boltzmann-like treatments for traffic flow : A critical review of the basic model and an alternative proposal for dilute traffic analysis. *Transportation Research* 9, 4 (1975), 225 – 235. [Page 21]
- [348] PAYNE, H. J. Models of freeway traffic and control. *Mathematical models of public systems* (1971). [Page 20]
- [349] PAYNE, H. J., AND TIGNOR, S. C. Freeway incident-detection algorithms based on decision trees with states. *Transportation Research Record*, 682 (1978). [Page 28]
- [350] PERREAULT, L., BOBEE, B., AND RASMUSSEN, P. F. Halphen distribution system i : Mathematical and statistical properties. *Journal of Hydrologic Engineering* 4, 3 (1999), 189–199. [Pages 76, 78, 210]
- [351] PERREAULT, L., BOBEE, B., AND RASMUSSEN, P. F. Halphen distribution system ii : Parameter and quantile estimation. *Journal of Hydrologic Engineering* 4, 3 (1999), 200–208. [Pages 79, 80]
- [352] PETER, F. Le geste suicidaire. *Perspectives Psy* 55, 2 (2016), 88–96. [Page 142]
- [353] PHAM, Q. D., DEVILLE, Y., AND VAN HENTENRYCK, P. Ls (graph) : a constraint-based local search for constraint optimization on trees and paths. *Constraints* (2012), 1–52. [Page 33]
- [354] PIAGET, J. The construction of reality in the child. *Journal of Consulting Psychology* 19, 1 (1955), 77. [Pages 23, 57]
- [355] PINYOL, I., AND SABATER-MIR, J. Computational trust and reputation models for open multi-agent systems : a review. *Artificial Intelligence Review* 40, 1 (2013), 1–25. [Pages 23, 47]
- [356] PLOEG, J., SCHEEPERS, B. T., VAN NUNEN, E., VAN DE WOUW, N., AND NIJMEIJER, H. Design and experimental evaluation of cooperative adaptive cruise control. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (2011), IEEE, pp. 260–265. [Page 22]
- [357] PORTER, J. D., KIM, D. S., PARK, S., SAEEDI, A., AND MAGANA, M. E. Wireless data collection system for travel time estimation and traffic performance evaluation. [Pages 26, 27]
- [358] PRAKASH, A. A., AND SRINIVASAN, K. A sample-based algorithm to determine minimum robust cost path with correlated link travel times. In *Transportation Research Board 93rd Annual Meeting* (2014). [Page 32]
- [359] PRIGOGINE, I., AND ANDREWS, F. C. A boltzmann-like approach for traffic flow. *Operations Research* 8 (1960), 789–797. [Page 21]
- [360] PRIGOGINE, I., HERMAN, R., AND SCHECHTER, R. Kinetic theory of vehicular traffic. *Systems, Man and Cybernetics, IEEE Transactions on* 2, 2 (1972). [Page 21]
- [361] PROVOST, J., KUIPERS, B. J., AND MIKKULAINEN, R. Self-organizing distinctive state abstraction using options. In *Proc. of the 7th International Conference on Epigenetic Robotics* (2007). [Page 58]
- [362] PU, W. Analytic relationships between travel time reliability measures. *Transportation Research Record : Journal of the Transportation Research Board* 2254 (2011), 122–130. [Page 25]

- [363] PUCKETT, D. D., AND VICKICH, M. J. Bluetooth-based travel time/speed measuring systems development. *UTCM*, 09-00-17 (2010). [Pages 27, 72]
- [364] PUGLIESE, L. D. P., AND GUERRIERO, F. A survey of resource constrained shortest path problems : Exact solution approaches. *Networks* 62, 3 (2013), 183–200. [Page 33]
- [365] QIN, Y., ZHANG, S., ZHU, X., ZHANG, J., AND ZHANG, C. Semi-parametric optimization for missing data imputation. *Applied Intelligence* 27, 1 (2007), 79–88. [Page 27]
- [366] QU, F., WANG, F.-Y., AND YANG, L. Intelligent transportation spaces : vehicles, traffic, communications, and beyond. *Communications Magazine, IEEE* 48, 11 (2010), 136–142. [Page 26]
- [367] QU, L., LI, L., ZHANG, Y., AND HU, J. Ppca-based missing data imputation for traffic flow volume : a systematical approach. *IEEE Transactions on intelligent transportation systems* 10, 3 (2009), 512–522. [Page 27]
- [368] RAITH, A., AND EHRGOTT, M. A comparison of solution strategies for biobjective shortest path problems. *Computers & Operations Research* 36, 4 (2009), 1299–1331. [Page 33]
- [369] RAKHA, H., EL-SHAWARBY, I., AND ARAFEH, M. Trip travel-time reliability : Issues and proposed solutions. *Journal of International Transportation Systems* 14, 4 (2010), 232–250. [Page 25]
- [370] RAKHA, H., FARZANEH, M., ARAFEH, M., AND STERZIN, E. Inclement weather impacts on freeway traffic stream behavior. *Transportation Research Record : Journal of the Transportation Research Board*, 2071 (2008), 8–18. [Page 29]
- [371] RAMEZANI, M., AND GEROLIMINIS, N. On the estimation of arterial route travel time distribution with markov chains. *Transportation Research part B* 46 (2012), 1576–1590. [Page 26]
- [372] REPEDE, J. F., AND BERNARDO, J. J. Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky. *European journal of operational research* 75, 3 (1994), 567–581. [Page 31]
- [373] REPORT OF THE FP7 PROJECT ECOMPASSI. D3.1 - New prospects in multi-modal route planning problems. Tech. rep., 2012. [Page 32]
- [374] REVELLE, C. S., AND LAPORTE, G. The plant location problem : new models and research prospects. *Operations Research* 44, 6 (1996), 864–874. [Page 31]
- [375] RICHARDS, P. I. Shock waves on the highway. *Operations research* 4, 1 (1956), 42–51. [Page 20]
- [376] RIETVELD, P., BRUINSMA, F., AND VAN VUUREN, D. Coping with unreliability in public transport chains : A case study for netherlands. *Transportation Research Part A : Policy and Practice* 35, 6 (2001), 539 – 559. [Page 26]
- [377] ROUSSEUW, P. J. Silhouettes : A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 0 (1987), 53 – 65. [Page 94]
- [378] SABATER, J., AND SIERRA, C. Review on computational trust and reputation models. *Artificial intelligence review* 24, 1 (2005), 33–60. [Page 23]
- [379] SABRAN, G., SAMARANAYAKE, S., AND BAYEN, A. Precomputation techniques for the stochastic on-time arrival problem. In *Proceedings of the Meeting on Algorithm Engineering & Experiments* (2014), Society for Industrial and Applied Mathematics, pp. 138–146. [Page 35]
- [380] SALOTTI, J., BILLOT, R., EL FAOUZI, N.-E., FENET, S., SOLNON, C., DUPRÉ, D., GILQUIN, L., CAPELLE, T., ARNAUD, E., AND PRIEUR, C. Vers l'utilisation de graphes de liens causaux pour l'amélioration de la prévision court-terme du trafic routier. In *RFP-IA 2016 Journée Transports Intelligents* (2016). [Page 104]

- [381] SAPORTA, G. *Probabilités, analyse des données et statistique*. Editions Technip, 2006. [Page 137]
- [382] SAU, J., MONTEIL, J., BILLOT, R., AND EL FAOUZI, N.-E. The root locus method : application to linear stability analysis and design of cooperative car-following models. *Transportmetrica B : Transport Dynamics* 2, 1 (2014), 60–82. [Pages 41, 43]
- [383] SAUVANET, G., AND NERON, E. Search for the best compromise solution on multiobjective shortest path problem. *Electronic Notes in Discrete Mathematics* 36 (2010), 615–622. [Page 33]
- [384] SAWANT, H., TAN, J., YANG, Q., AND WANG, Q. Using bluetooth and sensor networks for intelligent transportation systems. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on* (2004), pp. 767–772. [Page 26]
- [385] SCAPIN, D. L., AND BASTIEN, J. C. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & information technology* 16, 4-5 (1997), 220–231. [Page 143]
- [386] SCHAKEL, W. J., VAN AREM, B., AND NETTEN, B. D. Effects of cooperative adaptive cruise control on traffic flow stability. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on* (2010), IEEE, pp. 759–764. [Page 22]
- [387] SCHILLING, D., ELZINGA, D. J., COHON, J., CHURCH, R., AND REVELLE, C. The team/fleet models for simultaneous facility and equipment siting. *Transportation Science* 13, 2 (1979), 163–175. [Page 31]
- [388] SCHILLO, M., FUNK, P., AND ROVATSOS, M. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence* 14, 8 (2000), 825–848. [Pages 43, 47]
- [389] SCHNEIDER, M., STENGER, A., AND GOEKE, D. The electric vehicle-routing problem with time windows and recharging stations. *Transportation Science* 48, 4 (2014), 500–520. [Page 31]
- [390] SCHÖLKOPF, B., AND SMOLA, A. J. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT press, 2002. [Page 28]
- [391] SCHÖNHOF, M., TREIBER, M., KESTING, A., AND HELBING, D. Autonomous detection and anticipation of jam fronts from messages propagated by intervehicle communication. *Transportation Research Record : Journal of the Transportation Research Board* 1999, 1 (2007), 3–12. [Page 23]
- [392] SEN, S., PILLAI, R., JOSHI, S., AND RATHI, A. K. A mean-variance model for route guidance in advanced traveler information systems. *Transportation Science* 35, 1 (2001), 37–49. [Page 32]
- [393] SERVAT, D., PERRIER, E., TREUIL, J.-P., AND DROGOUL, A. When agents emerge from agents : Introducing multi-scale viewpoints in multi-agent simulations. In *Multi-Agent Systems and Agent-Based Simulation*, J. Sichman, R. Conte, and N. Gilbert, Eds., vol. 1534 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 1998, pp. 183–198. [Page 22]
- [394] SESHADRI, R., AND SRINIVASAN, K. K. Algorithm for determining most reliable travel time path on network with normally distributed and correlated link travel times. *Transportation Research Record : Journal of the Transportation Research Board* 2196, 1 (2010), 83–92. [Page 32]
- [395] SETHI, V., BHANDARI, N., KOPPELMAN, F. S., AND SCHOFER, J. L. Arterial incident detection using fixed detector and probe vehicle data. *Transportation Research Part C : Emerging Technologies* 3, 2 (1995), 99–112. [Page 29]
- [396] SHAHABI, M., UNNIKISHNAN, A., AND BOYLES, S. D. An outer approximation algorithm for the robust shortest path problem. *Transportation Research Part E : Logistics and Transportation Review* 58 (2013), 52–66. [Page 32]

- [397] SHAO, Q. Notes on maximum likelihood estimation for the three-parameter burr XII distribution. *Computational Statistics & Data Analysis* 45, 3 (2004), 675 – 687. [Page 80]
- [398] SHERALI, H. D., HOBEIKA, A. G., AND KANGWALKLAI, S. Time-dependent, label-constrained shortest path problems with applications. *Transportation Science* 37, 3 (2003), 278–293. [Page 33]
- [399] SHIFFMAN, S., STONE, A. A., AND HUFFORD, M. R. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32. [Page 132]
- [400] SIERPIŃSKI, G., CELIŃSKI, I., AND STANIEK, M. Using trip planners in developing proper transportation behavior. *International Science Index* 8, 11 (2014), 482–490. [Page 33]
- [401] SKABARDONIS, A., VARAIYA, P., AND PETTY, K. Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record : Journal of the Transportation Research Board*, 1856 (2003), 118–124. [Page 17]
- [402] SKRIVER, A. J. V. A classification of bicriterion shortest path (bsp) algorithms. *Asia-Pacific Journal of Operational Research* 17, 2 (2000), 199–212. [Page 33]
- [403] SMALL, K. A. The scheduling of consumer activities : Work trips. *American Economic Review* 72, 3 (June 1982), 467–79. [Page 25]
- [404] SMITH, B. L., AND DEMETSKY, M. J. Traffic flow forecasting : comparison of modeling approaches. *Journal of transportation engineering* 123, 4 (1997), 261–266. [Page 28]
- [405] SMITH, B. L., WILLIAMS, B. M., AND OSWALD, R. K. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C : Emerging Technologies* 10, 4 (2002), 303–321. [Page 28]
- [406] SOHN, K., AND KIM, D. Statistical model for forecasting link travel time variability. *Journal of Transportation Engineering* 135, 7 (2009), 440–453. [Pages 25, 26]
- [407] SOMMER, C., GERMAN, R., AND DRESSLER, F. Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis. *IEEE Transactions on Mobile Computing* 10, 1 (January 2011), 3–15. [Page 23]
- [408] SRINIVASAN, D., JIN, X., AND CHEU, R. L. Adaptive neural network models for automatic incident detection on freeways. *Neurocomputing* 64 (2005), 473–496. [Page 29]
- [409] SRINIVASAN, K. K., PRAKASH, A., AND SESHADRI, R. Finding most reliable paths on networks with correlated and shifted log–normal travel times. *Transportation Research Part B : Methodological* 66 (2014), 110–128. [Page 32]
- [410] STEWART, B. S., AND WHITE III, C. C. Multiobjective A^* . *Journal of the ACM (JACM)* 38, 4 (1991), 775–814. [Page 33]
- [411] STORANDT, S., EISNER, J., AND FUNKE, S. Enabling e-mobility : One way, return, and with loading stations. In *27th AAAI Conference on Artificial Intelligence* (2013). [Page 31]
- [412] SUN, S., AND CHEN, Q. Kernel regression with a mahalanobis metric for short-term traffic flow forecasting. In *International Conference on Intelligent Data Engineering and Automated Learning* (2008), Springer, pp. 9–16. [Page 28]
- [413] SUSILAWATI, S., TAYLOR, M. A., AND SOMENAHALLI, S. V. Distributions of travel time variability on urban roads. *Journal of Advanced Transportation* 47, 8 (2013), 720–736. [Pages 25, 26, 83]
- [414] SUZUKI, Y. A generic model of motor-carrier fuel optimization. *Naval Research Logistics (NRL)* 55, 8 (2008), 737–746. [Page 31]

- [415] SUZUKI, Y. A decision support system of dynamic vehicle refueling. *Decision Support Systems* 46, 2 (2009), 522–531. [Page 31]
- [416] SWEDA, T. M., AND KLABJAN, D. Finding minimum-cost paths for electric vehicles. In *Electric vehicle conference (IEVC), 2012 IEEE international* (2012), IEEE, pp. 1–4. [Page 31]
- [417] TALEBPOUR, A., MAHMASSANI, H., AND HAMDAR, S. Speed harmonization : evaluation of effectiveness under congested conditions. *Transportation Research Record : Journal of the Transportation Research Board*, 2391 (2013), 69–79. [Page 23]
- [418] TALEBPOUR, A., MAHMASSANI, H. S., AND HAMDAR, S. H. Modeling lane-changing behavior in a connected environment : A game theory approach. *Transportation Research Part C : Emerging Technologies* 59 (2015), 216–232. [Page 21]
- [419] TAN, M.-C., WONG, S., XU, J.-M., GUAN, Z.-R., AND ZHANG, P. An aggregation approach to short-term traffic flow prediction. *Intelligent Transportation Systems, IEEE Transactions on* 10, 1 (2009), 60–69. [Page 28]
- [420] TARAPATA, Z. Selected multicriteria shortest path problems : an analysis of complexity, models and adaptation of standard algorithms. *International Journal of Applied Mathematics and Computer Science* 17, 2 (2007), 269–287. [Page 33]
- [421] TAYLOR, M. A., AND SUSILAWATI, S. Modelling travel time reliability with the burr distribution. *Procedia-Social and Behavioral Sciences* 54 (2012), 75–83. [Pages 26, 80]
- [422] THIEMANN, C., TREIBER, M., AND KESTING, A. Longitudinal hopping in intervehicle communication : Theory and simulations on modeled and empirical trajectory data. *Phys. Rev. E* 78 (Sep 2008), 036102. [Page 23]
- [423] TOLEDO, T., KOUTSOPOULOS, H. N., AND BEN-AKIVA, M. E. Modeling integrated lane-changing behavior. *Transportation Research Record : Journal of the Transportation Research Board* 1857 (2003), 30–38. [Page 21]
- [424] TREIBER, M., HENNECKE, A., AND HELBING, D. Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E* 62 (2000), 1805–1824. [Pages 21, 45]
- [425] TREIBER, M., AND KESTING, A. Evidence of convective instability in congested traffic flow : A systematic empirical and theoretical investigation. *Procedia - Social and Behavioral Sciences* 17, 0 (2011), 683 – 701. Papers selected for the 19th International Symposium on Transportation and Traffic Theory. [Page 41]
- [426] TREIBER, M., AND KESTING, A. Validation of traffic flow models with respect to the spatio-temporal evolution of congested traffic patterns. *Transportation Research Part C : Emerging Technologies* 21, 1 (2012), 31 – 41. [Page 41]
- [427] TREIBER, M., AND KESTING, A. *Traffic Flow Dynamic : Data, Models and Simulation*. 2013. [Pages 21, 23, 40, 45]
- [428] TREIBER, M., KESTING, A., AND HELBING, D. Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts. *Transportation Research Part B : Methodological* 44, 8 (2010), 983–1000. [Pages 23, 48]
- [429] TRIGUI, R., JEANNERET, B., AND BADIN, F. Systemic modelling of hybrid vehicles in order to predict dynamic performance and energy consumption building the vehlib library of models. *RTS-Recherche Transports Securite*, 83 (2004), 129–150. [Pages 112, 113]
- [430] TSIRIGOTIS, L., VLAHOGIANNI, E. I., AND KARLAFTIS, M. G. Does information on weather affect the performance of short-term traffic forecasting models? *International Journal of Intelligent Transportation Systems Research* 10, 1 (2012), 1–10. [Page 29]

- [431] TSUBOTA, T. BHASKAR, A., CHUNG, E., AND BILLOT, R. Arterial traffic congestion analysis using bluetooth duration data. *Australian Transport Research Forum 2011 Proceedings, Adelaide, Australia* (September 2011). [Pages 26, 27, 72]
- [432] VAIVA, G., DUCROCQ, F., MEYER, P., MATHIEU, D., PHILIPPE, A., LIBERSA, C., AND GOUEMAND, M. Effect of telephone contact on further suicide attempts in patients discharged from an emergency department : randomised controlled study. *Bmj* 332, 7552 (2006), 1241–1245. [Page 154]
- [433] VAN AREM, B., VAN DRIEL, C. J., AND VISSER, R. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *Intelligent Transportation Systems, IEEE Transactions on* 7, 4 (2006), 429–436. [Pages 12, 22, 52]
- [434] VAN DER VOORT, M., DOUGHERTY, M., AND WATSON, S. Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C : Emerging Technologies* 4, 5 (1996), 307–318. [Page 28]
- [435] VAN KATWIJK, R., AND VAN KONINGSBRUGGEN, P. Coordination of traffic management instruments using agent technology. *Transportation Research Part C : Emerging Technologies* 10 (2002), 455 – 471. [Page 23]
- [436] VAN LINT, J. W. C., VAN ZUYLEN, H. J., AND TU, H. Travel time unreliability on freeways : Why measures based on variance tell only half the story. *Transportation Research Part-A : Policy and Practice* 42, 1 (JAN 2008), 258–277. [Pages 25, 85]
- [437] VAN OS, J., DELESPAUL, P., BARGE, D., AND BAKKER, R. P. Testing an mhealth momentary assessment routine outcome monitoring application : a focus on restoration of daily life positive mood states. *PLoS One* 9, 12 (2014), e115254. [Page 132]
- [438] VAN WOENSEL, T., KERBACHE, L., PEREMANS, H., AND VANDAELE, N. A queueing framework for routing problems with time-dependent travel times. *Journal of Mathematical Modelling and Algorithms* 6, 1 (2007), 151–173. [Page 32]
- [439] VANAJAKSHI, L., AND RILETT, L. R. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In *Intelligent Vehicles Symposium, 2004 IEEE* (2004), IEEE, pp. 194–199. [Page 28]
- [440] VAPNIK, V. Pattern recognition using generalized portrait method. *Automation and remote control* 24 (1963), 774–780. [Page 28]
- [441] VAPNIK, V. *The nature of statistical learning theory*. Springer Science & Business Media, 2013. [Page 28]
- [442] VASIRANI, M., AND OSSOWSKI, S. A market-inspired approach to reservation-based urban road traffic management. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (2009), International Foundation for Autonomous Agents and Multiagent Systems, pp. 617–624. [Page 23]
- [443] VLAHOGIANNI, E. I., GOLIAS, J. C., AND KARLAFTIS, M. G. Short-term traffic forecasting : Overview of objectives and methods. *Transport reviews* 24, 5 (2004), 533–557. [Page 28]
- [444] VLAHOGIANNI, E. I., KARLAFTIS, M. G., AND GOLIAS, J. C. Optimized and meta-optimized neural networks for short-term traffic flow prediction : a genetic approach. *Transportation Research Part C : Emerging Technologies* 13, 3 (2005), 211–234. [Page 28]
- [445] VLAHOGIANNI, E. I., KARLAFTIS, M. G., AND GOLIAS, J. C. Short-term traffic forecasting : Where we are and where we're going. *Transportation Research Part C : Emerging Technologies* 43 (2014), 3–19. [Page 96]

- [446] VOIGT, A. Collecting external data using bluetooth technology. *Texas Transportation Institute, Houston, TRB 90 th Annual Meeting 175* (2011). [Page 27]
- [447] WAKABAYASHI, H., AND MATSUMOTO, Y. Comparative study on travel time reliability indexes for highway users and operators. *Journal of Advanced Transportation* 46, 4 (2012), 318–339. [Page 25]
- [448] WANG, F., KEATS, J., AND ZIMMER, W. J. Maximum likelihood estimation of the burr xii parameters with censored and uncensored data. *Microelectronics Reliability* 36, 3 (1996), 359 – 362. [Pages 80, 83]
- [449] WANG, J. Y., WONG, K. I., AND CHEN, Y. Y. Short-term travel time estimation and prediction for long freeway corridor using nn and regression. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, IEEE, pp. 582–587. [Page 28]
- [450] WANG, M., DAAMEN, W., HOOGENDOORN, S. P., AND VAN AREM, B. Rolling horizon control framework for driver assistance systems. part i : Mathematical formulation and non-cooperative systems. *Transportation research part C : emerging technologies* 40 (2014), 271–289. [Pages 22, 52]
- [451] WANG, M., DAAMEN, W., HOOGENDOORN, S. P., AND VAN AREM, B. Rolling horizon control framework for driver assistance systems. part ii : Cooperative sensing and cooperative control. *Transportation research part C : emerging technologies* 40 (2014), 290–311. [Page 22]
- [452] WANG, W., CHEN, S., AND QU, G. Comparison between partial least squares regression and support vector machine for freeway incident detection. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE (2007)*, IEEE, pp. 190–195. [Page 29]
- [453] WANG, Y.-W. An optimal location choice model for recreation-oriented scooter recharge stations. *Transportation Research Part D : Transport and Environment* 12, 3 (2007), 231–237. [Page 31]
- [454] WANG, Y.-W. Locating battery exchange stations to serve tourism transport : A note. *Transportation Research Part D : Transport and Environment* 13, 3 (2008), 193–197. [Page 31]
- [455] WARBURTON, A. Approximation of pareto optima in multiple-objective, shortest-path problems. *Operations Research* 35, 1 (1987), 70–79. [Page 33]
- [456] WARFFEMIUS, P. Using the standard deviation of the travel time distribution as an indicator for valuing the reliability of travel time. In *Young Researchers Seminar of the European Conference of Transport Research Institutes, The Hague, Netherlands (2005)*. [Page 25]
- [457] WASSON, J. S., STURDEVANT, J. R., AND BULLOCK, D. M. Real-time travel time estimates using media access control address matching. *ITE Journal*, 78(6) (2008). [Pages 27, 72]
- [458] WATKINS, A. J. An algorithm for maximum likelihood estimation in the three parameter burr XII distribution. *Computational Statistics & Data Analysis* 32, 1 (1999), 19 – 27. [Page 80]
- [459] WEISMAN, A. D., AND WORDEN, J. W. Risk-rescue rating in suicide assessment. *Archives of General Psychiatry* 26, 6 (1972), 553–560. [Page 136]
- [460] WHITLOCK, M. E., AND QUEEN, C. M. Modelling a traffic network with missing data. *Journal of Forecasting* 19, 7 (2000), 561–574. [Page 27]
- [461] WILMINK, I., KLUNDER, G., AND VAN AREM, B. Traffic flow effects of integrated full-range speed assistance (irsa). In *2007 IEEE Intelligent Vehicles Symposium (2007)*, IEEE, pp. 1204–1210. [Pages 22, 40]

- [462] WILSON, R. E., AND WARD, J. A. Car-following models : fifty years of linear stability analysis - a mathematical perspective. *Transportation Planning and Technology* 34, 1 (2011), 3–18. [Page 41]
- [463] WINNER, H., WITTE, S., UHLER, W., AND LICHTENBERG, B. Adaptive cruise control system aspects and development trends. Tech. rep., SAE Technical Paper, 1996. [Page 22]
- [464] WU, X. Finding reliable shortest paths in dynamic stochastic networks. *Transportation Research Record : Journal of the Transportation Research Board* 2333, 1 (2013), 80–90. [Page 32]
- [465] WU, X., ZHU, X., WU, G.-Q., AND DING, W. Data mining with big data. *IEEE transactions on knowledge and data engineering* 26, 1 (2014), 97–107. [Page 8]
- [466] XI, X., SIOSHANSI, R., AND MARANO, V. Simulation–optimization model for location of a public electric vehicle charging infrastructure. *Transportation Research Part D : Transport and Environment* 22 (2013), 60–69. [Page 31]
- [467] XIA, J., AND CHEN, M. Defining traffic flow phases using intelligent transportation systems-generated data. *Journal of Intelligent Transportation Systems* 11, 1 (2007), 15–24. [Pages 26, 93]
- [468] XING, T., AND ZHOU, X. Finding the most reliable path with and without link travel time correlation : A lagrangian substitution based approach. *Transportation Research Part B : Methodological* 45, 10 (2011), 1660 – 1679. [Page 32]
- [469] XU, Y., CAI, W., ECKHOFF, D., NAIR, S., AND KNOLL, A. A graph partitioning algorithm for parallel agent-based road traffic simulation. In *Proceedings of the 2017 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation* (2017), ACM, pp. 209–219. [Page 23]
- [470] XU, Z., JIN, P. J., JIANG, X., AND RAN, B. Macroscopic fundamental diagram (mfd) based on vehicle-to-infrastructure (v2i) connected vehicle data. In *Transportation Research Board 94th Annual Meeting* (2015), no. 15-0749. [Page 22]
- [471] YAN, X., GU, F., HU, X., AND ENGSTROM, C. Dynamic data driven event reconstruction for traffic simulation using sequential monte carlo methods. In *Simulation Conference (WSC), 2013 Winter* (Dec 2013), pp. 2042–2053. [Page 24]
- [472] YANG, L., AND ZHOU, X. Constraint reformulation and a lagrangian relaxation-based solution algorithm for a least expected time path problem. *Transportation Research Part B : Methodological* 59 (2014), 22–44. [Page 32]
- [473] YEN, J. Y. Finding the k shortest loopless paths in a network. *Management Science* 17, 11 (1971), 712–716. [Pages 34, 117]
- [474] YUAN, F., AND CHEU, R. L. Incident detection using support vector machines. *Transportation Research Part C : Emerging Technologies* 11, 3 (2003), 309–328. [Page 29]
- [475] ZALSMAN, G., HAWTON, K., WASSERMAN, D., VAN HEERINGEN, K., ARENSMAN, E., SARCHIAPONE, M., CARLI, V., HÖSCHL, C., BARZILAY, R., BALAZS, J., ET AL. Suicide prevention strategies revisited : 10-year systematic review. *The Lancet Psychiatry* 3, 7 (2016), 646–659. [Page 141]
- [476] ZHANG, C., YANG, X.-C., AND YAN, X. Automatic incident detection methodology for freeway using floating cars. *Journal of Wuhan University of Technology (Transportation Science and Engineering)* 30, 6 (2006), 973–975. [Page 29]
- [477] ZHANG, J., WANG, F.-Y., WANG, K., LIN, W.-H., XU, X., AND CHEN, C. Data-driven intelligent transportation systems : A survey. *Intelligent Transportation Systems, IEEE Transactions on* 12, 4 (2011), 1624–1639. [Page 26]

- [478] ZHANG, Y., AND HAGHANI, A. A gradient boosting method to improve travel time prediction. *Transportation Research Part C : Emerging Technologies* 58 (2015), 308–324. [Page 28]
- [479] ZHANG, Y., AND LIU, Y. Data imputation using least squares support vector machines in urban arterial streets. *IEEE Signal Processing Letters* 16, 5 (2009), 414–417. [Page 27]
- [480] ZHANG, Y., AND XIE, Y. Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record : Journal of the Transportation Research Board*, 2024 (2008), 92–99. [Page 28]
- [481] ZHAO, H., YU, L., GUO, J., ZHAO, N., WEN, H., AND ZHU, L. Estimation of time-varying od demands incorporating fcd and rtms data. *Journal of Transportation Systems Engineering and Information Technology*, 10(1) (2010), 72–80. [Page 27]
- [482] ZHENG, W., LEE, D.-H., AND SHI, Q. Short-term freeway traffic flow prediction : Bayesian combined neural network approach. *Journal of transportation engineering* 132, 2 (2006), 114–121. [Page 28]
- [483] ZHU, F., AND UKKUSURI, S. V. Accounting for dynamic speed limit control in a stochastic traffic environment : A reinforcement learning approach. *Transportation Research Part C : Emerging Technologies* 41 (2014), 30 – 47. [Page 23]
- [484] ZIKOPOULOS, P., EATON, C., ET AL. *Understanding big data : Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011. [Page 8]
- [485] ZILIASKOPOULOS, A., AND WARDELL, W. An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research* 125, 3 (2000). available at <http://www.sciencedirect.com/science/article/B6VCT-40NMSSJ-3/2/8b7c573f760b1a9cea3f3fe5558a4bfa>. [Page 33]
- [486] ZLATEV, J., AND BALKENIUS, C. Introduction : Why "epigenetic robotics" ? In *Proceedings of the First Conference on Epigenetic Robotics*. In : Balkenius, C., Zlatev, J., Kozima, H., Dautenhahn, K., Breazeal, C.(eds.) : *Lund University Cognitive Science Series* (2001), no. 85. [Page 57]
- [487] ZOCCAIE, A., NIE, Y. M., AND MAHMASSANI, H. S. Plan b : A simulation-based method for finding minimum travel time budget paths in stochastic networks with correlated link times. In *Transportation Research Board 93rd Annual Meeting* (2014), vol. 14-5615. [Page 32]
- [488] ZOHDY, I., AND RAKHA, H. Game theory algorithm for intersection-based cooperative adaptive cruise control (cacc) systems. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on* (Sept 2012), pp. 1097–1102. [Page 44]

Annexe : *Curriculum Vitae*



A.1 Données personnelles

Nom : Billot

Prénom : Romain

Date de naissance : 21 mai 1984 à Nantes (44)

Nationalité : Française

Adresse professionnelle :

Lab-STICC (UMR CNRS 6285)

Technopôle Brest-Iroise

CS 83818

29238 Brest Cedex 3

Téléphone : +33 2 29 00 15 60

Adresse électronique : romain.billot@imt-atlantique.fr

Adresse personnelle :

10 rue de Kerangoff

29217 Le Conquet

A.2 Expériences professionnelles

Depuis le 1er Mars 2016 : IMT Atlantique (ex Télécom Bretagne), Brest, France.

Maître de conférences en Data Science, département LUSSE, UMR CNRS 6285 Lab-STICC.

2011-2016 : Ecole Nationale des Travaux Publics de l'Etat (ENTPE), Lyon, France.

Chargé de recherche (CR1), Laboratoire d'Ingénierie Circulation Transports (LICIT), unité mixte de recherche avec l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR). Fonctionnaire titulaire dans le corps des chargés de recherche du ministère de l'écologie, du développement durable et de l'énergie.

2011 : Queensland University of Technology (QUT), Brisbane, Australie.

Chercheur post-doctoral, Smart Transport Research Centre. Recherche opérationnelle.

Depuis 2012, je suis chercheur associé à QUT.

2007-2010 : Ecole Centrale des Arts et Manufactures, Paris, France.

Doctorat en mathématiques appliquées, mention très honorable.

Sujet : analyse et modélisation de l'impact de la météorologie sur le trafic routier.

Financement : Institut National de Recherche sur les Transports et leur Sécurité (INRETS).

Soutenue le 8 décembre 2010 devant le jury composé de :

- Jacques Sau, professeur émérite, université Lyon 1 : président.
- Hani Mahmassani, professeur, université de Northwestern : rapporteur.
- Michel Rascle, professeur émérite, université de Nice : rapporteur.
- Gilles Fleury, professeur, SUPELEC : examinateur.
- Florian de Vuyst, professeur, Ecole Centrale : directeur de thèse.
- Nour-Eddin El Faouzi, directeur de recherche INRETS : directeur de thèse.

2006-2007 : Université de Technologie de Compiègne (UTC), Compiègne, France.

Master en Sciences et Technologie, spécialité Technologies de l'Information et des Systèmes, mention Excellent (major de promotion).

Sujet : une approche ensembliste inspirée du boosting en classification non supervisée.

Encadrement : Professeur Gérard Govaert, UTC et Professeur Stephane Lallich, Lyon 2.

2002-2007 : Université de Technologie de Compiègne (UTC), Compiègne, France.

Diplôme d'ingénieur en génie informatique, spécialité Fouille de Données.

A.3 Thèmes de recherche

Fouille de données et statistiques

- Extraction de connaissances à partir de flots de données hétérogènes.
- Apprentissage non supervisé : clustering, critères de qualité, approches ensemblistes.
- Apprentissage supervisé.

Recherche opérationnelle

- Algorithmes de plus court chemin : aspects stochastiques et techniques d'accélération.
- Optimisation multi-objectif.

Domaines d'application des recherches : mobilité intelligente et santé connectée.

A.4 Activités d'enseignement

Passionné par l'enseignement, je suis intervenu environ 100 heures par an à l'ENTPE, l'université Lyon 2 et maintenant IMT Atlantique. A l'ENTPE, j'ai été très impliqué dans l'équipe pédagogique en tant que responsable d'une voie d'approfondissement de 3ème année en transports. J'ai participé activement au montage d'un nouveau Master pour la rentrée 2016 avec le parcours "mobilité dans les méga-cités". A l'IMT Atlantique, j'ai pris en charge le niveau 1 de la filière Data Science en troisième année et je suis responsable des cours de statistiques, fouille de données, et clustering avancé. Je participe activement au montage des maquettes de la nouvelle école.

2016-2017 : enseignant chercheur, IMT Atlantique, Brest, France.

- Enseignant principal du cours statistiques en formation FIG, filière F3B (45h/an).
- Responsable de l'UV « approche intégrée de l'ingénierie des services » en filière F3B.
- Enseignements en cours de fouille de données, algorithmique, modèles graphiques probabilistes en FIG et FIP (environ 40h/an).

- Enseignant principal du cours classification non supervisée en Master CSDS. Intervention en cours *big data analytics*.
- Formation sur le logiciel de statistiques et d'analyse de données R à destination des enseignants chercheurs de TELECOM Bretagne (15h/an).
- Encadrement de projets de fin d'études, projets libres, inter-semester, et stages étudiants étrangers.

2012-2016 : enseignant, ENTPE, Vaulx-en-Velin, France.

- Enseignant principal du cours « méthodes statistiques pour l'ingénieur » en tronc commun. Gestion de 200 étudiants et d'une équipe pédagogique de 5 enseignants. Interventions en amphithéâtre (24h/an), en travaux dirigés (24h/an) et travaux pratiques sur le logiciel R (16h/an). Le cours porte sur la statistique inférentielle.
- Conférencier dans le cours de modélisation du trafic. Cours magistral sur la mobilité intelligente (4h/an).
- Enseignant principal du cours de 3ème année « traitements de données trafic ». Gestion d'une équipe de 8 enseignants et interventions sur 2 séances (8h/an).
- Interventions diverses dans les cours de 3ème année (4 à 8h par an).

2008-2016 : vacataire, université Lyon 2, Lyon, France.

- Enseignant principal du cours « recherche opérationnelle » (licence 3 Informatique Décisionnelle et Statistiques), Département Informatique. Le programme contient les bases de la théorie des graphes ainsi que des notions d'algorithmique et de théorie de la complexité (28h équivalent TD par an).
- Chargé de TD dans le cours « traitement informatisé des données » sous la responsabilité de Julien Jacques, interventions en TP, enseignement des bases de la statistique inférentielle et introduction au logiciel R (depuis 2014, 21h équivalent TD).
- Chargé de TD dans le cours « d'algèbre linéaire » (licence 3), Département Informatique, faculté d'économie.

Tableau récapitulatif des enseignements :

| Niveau | Type | 2008 | 2009 | 2010 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | TOTAL |
|--------|------|------|------|------|------|------|------|------|------|------|-------|
| L3 | CM | 15 | 15 | 15 | 15 | 19 | 19 | 19 | | | 117 |
| | TD | 10.5 | 14.5 | 10.5 | 7 | 7 | 28 | 28 | | | 117 |
| | TP | | | | | | | | 9 | 9 | 18 |
| M1 | CM | | | | 24 | 24 | 24 | 24 | 6 | 6 | 108 |
| | TD | | | | 24 | 24 | 24 | 24 | | | 96 |
| | TP | | | | 16 | 16 | 16 | 16 | 6 | 6 | 76 |
| M2 | CM | | | | 8 | 8 | 4 | 4 | 49 | 49 | 122 |
| | TD | | | | | | | | 6 | 6 | 12 |
| | TP | | | | | | | | 37 | 37 | 74 |

Tab. A.1.: Tableau de synthèse des enseignements jusqu'à l'année 2016-2017

A.5 Encadrement de la recherche

A.5.1 Etudiants en thèse

En cours :

Flavien Lucas, Université de Bretagne Sud (UBS) et IMT Atlantique, début en octobre 2017.

Sujet : Open Data, Optimisation and Simulation for smart mobility. co-encadrement à 50% avec Marc Sevaux (UBS).

Sofian Berrouiguet, IMT Atlantique (ex Télécom Bretagne), 2016-présent (Directeur de thèse officiel, taux d'encadrement 50%) Sujet : développement d'outils de bilan clinique dans la gestion du risque suicidaire pour des sujets potentiellement récidivistes. Co-direction avec Philippe Lenca.

Julien Salotti, INSA Lyon, 2014-présent (participation non officielle à l'équipe d'encadrement) Sujet : analyse spatiotemporelle de données trafic. Co-encadrement avec Nour-Eddin El Faouzi (IFSTTAR), Christine Solnon et Serge Fenet, LIRIS (INSA Lyon).

Thèses soutenues :

Maxime Guériaux, Université de Lyon I, 2013-2016 (Co-encadrant de thèse, taux d'encadrement 50%) Sujet : approches auto-organisationnelles pour la modélisation et la gestion d'un trafic coopératif. Co-direction avec Salima Hassas, LIRIS (Lyon 1).

Raphaël Delhome, ENTPE, 2013-2016 (Directeur de thèse officiel, taux d'encadrement 50%) Sujet : prise en compte de l'incertitude dans la modélisation des temps de parcours et les calculs de plus courts chemins. Co-direction avec Nour-Eddin El Faouzi.

Pierre-Antoine Laharotte, ENTPE, 2013-2016 (Directeur de thèse officiel, taux d'encadrement 50%) Sujet : caractérisation météo-sensible de l'état du trafic et méthodes de prévision. Co-direction avec Nour-Eddin El Faouzi. Prix ABERTIS 2017 de la meilleure thèse.

Fouad Baouche, ENTPE-Université de Lyon, 2012-2015 (Encadrant IFSTTAR, taux d'encadrement 30%) Sujet : optimisation des itinéraires de véhicules électriques. Thèse soutenue le 2 Juin 2015 (co-encadrement avec Nour-Eddin El Faouzi et Rochdi Trigui).

Julien Monteil, ENTPE-Université de Lyon, 2011-2014 (Participation non officielle mais significative à l'encadrement de la thèse avec plus de 12 publications communes). Sujet : investigating the effects of cooperative vehicles on highway traffic flow homogenization : analytical and simulation studies. Thèse soutenue le 29 Janvier 2014. Prix ABERTIS 2014 de la meilleure thèse.

A.5.2 Étudiants de Master (Stage de Fin d'Études, 6 mois)

Frédéric Aliaga, ENTPE, 2015.

Sujet : prévision de trafic par la méthode des k plus proches voisins en présence de données hétérogènes.

Pierre-Antoine Laharotte, ENTPE, 2013.

Sujet : utilisation des données Bluetooth pour la caractérisation du trafic. Le cas de Brisbane, Australie. Prix Abertis 2014 du meilleur travail de Master.

Justine Dinh, ENTPE, 2013.

Sujet : analyse de l'apport des données issues de véhicules traceurs pour la gestion du trafic.

Boris Ly, ENTPE, 2012.

Sujet : analyse des algorithmes multi-objectif pour le problème de plus court chemin.

Deborah Manenti, Université Lyon 2, 2009.

Sujet : analyse statistique des effets microscopiques de la pluie sur le trafic à partir de données suisses.

A.5.3 Étudiants de Master 1 et Licence (Stages de 4 mois)

Massinissa Méniche, Télécom Bretagne, 2016.

Sujet : data mining et twitter pour détecter les troubles de la santé mentale.

Nawat Vatanawood et Pitarn Harnpanichpun, Université de Bangkok, Thaïlande, 2016.

Sujet : développement d'une application mobile pour la prévention du risque suicidaire.

Fabien Baradel, ENSAI Rennes, 2015.

Sujet : méthode Support Vector Regression pour la prévision de trafic.

Thibaut Blanc, Université Lyon 2, 2013.

Sujet : estimation de temps de parcours à partir de données FCD.

Benjamin Corroy, Université Lyon 2, 2011.

Sujet : vers un calculateur d'itinéraire multimodal et multiobjectif pour les réseaux urbains.

Samuel Stocksieker, Université Lyon 2, 2010.

Sujet : algorithmes de plus courts chemins multimodaux pour les réseaux urbains.

Ye Xia, École Centrale Lyon, 2010.

Sujet : méthodes d'apprentissage supervisé pour la prédiction du trafic.

A.6 Activités de valorisation

A.6.1 Contrats de recherche

- 2016-2017 : projet PRINTEMPS. Programme de Recherche INTerventionnelle et Evaluative Mené pour la Prévention du Suicide.
- 2014-2016 : Approches biomimétiques pour le développement durable. Financement du ministère de l'écologie sur le thème « vers un trafic coopératif bio-inspiré : de la communication entre véhicules à des phénomènes d'auto-organisation » .

- 2014-2017 : FUI EDIT (leader : VOLVO). Rôle : modélisation de véhicules connectés.
- 2012-2017 : ORSI COMET et MESSIGEO. Opérations de Recherche Stratégique Incitatives liant l'IFSTTAR au CEREMA, je suis responsable de 3 axes dans ces deux ORSI.
- 2012-2015 : CROSSING BORDERS. Financement du ministère de la recherche en Autriche, collaboration avec Austrian Institute of Technology. Développement de méthodes d'optimisation pour véhicules électriques.
- 2010-2014 : projet FUI Co-Drive (leader VALEO). Thématique des véhicules connectés. Contributions sur la partie fouille de données et information trafic.
- 2011-2014 : OPTIMUM (Optimum ITS-based tools for Intelligent Urban Mobility). Projet Marie Curie FP7-PEOPLE-2010-IRSES. Développement de recherches sur les véhicules connectés en partenariat avec l'université de technologie du Queensland (Australie) et l'école polytechnique fédérale de Lausanne (Suisse).
- 2014 : VORAB. Contrat d'expertise pour ASFINAG, 1er opérateur autoroutier en Autriche. Développement de méthodes de prévision de temps de parcours.
- 2011 : Projet Real Time Traveller Information (RTPI), QUT, Australie. Création d'un calculateur d'itinéraires multimodal pour la ville de Brisbane. Partenaires industriels : Department of Transport and Main Roads, Queensland Motorways Limited, IBM, Queensland Cyber Infrastructure Foundation, Thales Australia and Brisbane Airport Corporation.
- 2008-2011 : Projet européen COST TU0702. Real-time Monitoring, Surveillance and Control of Road Networks under Adverse Weather Conditions. Membre du groupe de travail 1, Webmaster, organisation de meeting.
- 2007-2010 : Projet régional TPTEO. Temps de parcours sur autoroutes. INRETS-AREA. Financement : cluster Transports et Société de la région Rhône-Alpes.

A.6.2 Logiciels

Depuis 2013 : coordinateur scientifique du projet TEMPUS. Plateforme open-source de calcul d'itinéraires multimodaux (<http://tempus.ifsttar.fr/>). Collaboration avec l'entreprise OSLANDIA et le CEREMA. En 2015, TEMPUS a été sélectionné par l'entreprise MAPPY pour la refonte de leur calculateur d'itinéraires.

2015 : dans le cadre de la thèse de Maxime Guériau, développement d'une extension de MovSim, plateforme open-source de modélisation du trafic.

A.6.3 Organisation de conférence et comités de programme

- 2018 : membre du comité d'organisation de la ROADEF.
- 2016 : membre du comité de programme de la journée transports intelligents, PFIA.
- Membre du comité de programme de la conférence BICA, Annual International Conference on Biologically Inspired Cognitive Architectures.
- Novembre 2015 : organisation d'un workshop « approches biomimétiques pour le développement durable », en marge de la conférence BICA (avec Jean-Louis Deneubourg).
- Juin 2014 : membre du comité scientifique et d'organisation de la conférence ISTS (International Symposium on Transport Simulation).

- Septembre 2011 : organisation d'une conférence, 1st French-Australian Symposium on Transportation Research, Brisbane, Australie.

A.7 Rayonnement scientifique

A.7.1 Séjours académiques en tant que chercheur invité

- 2015 : Trinity College Dublin, Distributed Systems Group. Thème : véhicules connectés et apprentissage par renforcement (séjour d'une semaine).
- 2015 : Virginia Tech, Blacksburg, USA. Thème : méthodes d'apprentissage supervisé et non supervisé pour la prévision des conditions de circulation (séjour d'une semaine en collaboration avec Pr. Hesham Rakha).
- 2015 : Delft University of Technology, Delft, Pays bas. Thème : véhicules connectés (séjour d'une semaine).
- 2012 : Queensland University of Technology, Brisbane, Australie. Thème : méthodes bayésiennes pour l'estimation des conditions de circulation (1 mois).
- 2009 : Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse. Chercheur invité au laboratoire des voies et circulation (LAVOC). Mission Scientifique de Court Terme (STSM) financée par l'action COST TU0702 (projet européen). Thème : analyse statistique de données autoroutières suisses.

A.7.2 Séminaires

- Assemblée générale du Lab-STICC. Titre : fouille de données et optimisation pour la mobilité intelligente. 9 décembre 2016.
- Séminaire du groupe distributed systems du Trinity College Dublin. Titre : connected vehicles and reinforcement learning. Juillet 2015.
- Séminaire du département transports de l'université de technologie de Delft, Pays Bas. Titre : some contributions to smart mobility. Mai 2015.
- Séminaire mensuel du Virginia Transportation Institute (VTTI), Virginia Tech, USA. Titre : some contributions to smart mobility at LICIT. Mai 2015.
- Séminaire modélisation des réseaux de transports, Marne La Vallée, France. Titre : integration of connected vehicles into traffic modelling. Mai 2015.
- Journée de travail sur la formation et l'enseignement sur le thème des systèmes de transport intelligents. Ministère de l'Ecologie et du Développement Durable. Titre de la présentation : panorama des formation ITS en France. Mai 2015.
- École des Mines de Saint Etienne : présentation de nos travaux sur la mobilité intelligente. Décembre 2014.
- Séminaire de l'institut rhônalpin des systèmes complexes, Ecole Normale Supérieure de Lyon : approche décentralisée à base de systèmes multi-agents pour le trafic coopératif intelligent. Juin 2013.

- Journée du RNSGT “apprentissage, éaction, gouvernance des systèmes complexes”. FMSH, CNRS, Paris. Titre : la modélisation du trafic à l’aune des systèmes intelligents. Novembre 2012.
- Séminaire national IFSTTAR TISIC (Traitement de l’Information, Signal, Image, Connaissance), Marne La Vallée, France. Titre de la présentation : optimisation des plus courts chemins multimodaux. Décembre 2011.
- Séminaire national IFSTTAR TISIC (Traitement de l’Information, Signal, Image, Connaissance). Titre de la présentation : filtrage particulaire pour l’estimation en ligne des conditions de circulation. Juin 2010.

A.7.3 Conférencier invité

“Les Journées Nationales des Communications dans les Transports (JNCT)”, 29-30 Mai 2013, Nevers. Titre de la conférence plénière : intégrer les systèmes coopératifs dans la modélisation du trafic.

A.8 Activités éditoriales

Membre du comité éditorial de la revue open science ISTE Stochastic Modeling and Statistics.

Relecteur pour revues et conférences internationales :

- Transportation Research Records,
- Transportation Research Part C.,
- Transportation Research Part B,
- International Journal of Intelligent Transportation Systems Research,
- IEEE Transactions on ITS,
- Computational Intelligence,
- Computer-Aided Civil Engineering,
- Sensors,
- Annals of Mathematics and Artificial Intelligence,
- Transportation Research Board annual meeting,
- International IEEE Conference on Intelligent Transportation Systems,
- Annual International Conference on Biologically Inspired Cognitive Architectures.

A.9 Compétences linguistiques et informatique

Langues :

- Anglais : courant (enseigne régulièrement dans la langue).
- Allemand : courant (capable d’enseigner dans la langue).
- Espagnol : niveau scolaire

Informatique :

- Langages de programmation : Pascal, C, C++, bases en JAVA, Python.
- Base de données : PHP/MySQL, Oracle, Access, Méthodes UML et MERISE.
- Systèmes d'exploitation : Unix, Windows, Linux.
- Logiciels scientifiques : R, SAS, Matlab., Scilab.

Annexe : publications de l'auteur

B.1 Chapitres dans ouvrage collectif

Habart, M. et Janssen, J. (Ed.).
Billot, R., Bothorel C., Lenca, P.
Le big data pour les compagnies d'assurance.
ISTE Editions, 2017. À paraître.

J. Žak, Y. Hadas, R. Rossi (Ed.)
Aron, M. Billot, R., Bhourri, N. El Faouzi, N.-E., Seidowsky, R.
Traffic indicators, accidents and rain : some relationships calibrated on a French urban motorway network.
Advanced concepts, methodologies and technologies for transportation and logistics". Springer Series "Advance in Intelligent Systems and Computing, Vol 572. , pp309-333. 2018. À paraître.

El Faouzi, N.-E. (Ed.).
Real-time monitoring, surveillance and control of road networks under adverse weather conditions : State of the Art and best practices.
Collection Recherche, R.283, Co-édition INRETS-COST. Décembre 2010.

El Faouzi, N.-E. and Heilmann B. (Ed.)
Real-time monitoring, surveillance and control of road networks under adverse weather conditions : Advances in Modeling and weather-sensitive traffic management.
Collection Recherche, Co-édition INRETS-COST. Septembre 2012.

B.2 Articles dans des revues internationales à comité de lecture

Simonnet, M., Gourvennec, B., Billot, R. and Lenca, P.
Round-the-Clock activity and recovery monitoring. A Comparison between 3 heart rate sensors acceptabilities.
Soumis à JMIR mhealth and uhealth : mobile and ubiquitous health

Berrouiguet, Sofian, Billot Romain et al.
A data mining approach for suicide risk management.
Soumis à British Journal of Psychiatry, 2017.

Delhome, R., Billot, R., & El Faouzi, N. E.
Travel time statistical modeling with the Halphen distribution family.

Journal of Intelligent Transportation Systems, (just-accepted). Doi 10.1080/15472450.2017.1326115. 2017.

Berrouiguet, Sofian, et al.

Ecological Assessment of Clinicians' Antipsychotic Prescription Habits in Psychiatric Inpatients : A Novel Web-and Mobile Phone–Based Prototype for a Dynamic Clinical Decision Support System. *Journal of medical Internet research* 19.1 (2017).

Guériau, M., Billot, R., El Faouzi, N.-E., Monteil, J.

How to assess the benefits of connected vehicles? A simulation framework for the design of cooperative traffic management strategies.

Transportation Research Part C : emerging technologies. 67, 266 – 279, 2016.

Delhome, R. Billot, R., Duret, A. & El-Faouzi, N.-E.

The Halphen Distribution System, a Toolbox for Modeling Travel Time Variability : some insights based on mesoscopic simulation.

In Transportation Research Record : Journal of the Transportation Research Board, 2490, 32-40, 2015.

Baouche, F., Billot, R., Trigui, R., & El Faouzi, N.-E.

Efficient Allocation of Electric Vehicles Charging Stations : Optimization Model and Application to a Dense Urban Network.

Intelligent Transportation Systems Magazine, IEEE, 6(3), 33-43. 2014.

Laharotte, P.A., Billot, R., Côme, E. , Oukhellou, L, Nantes, A., El Faouzi, N.-E. Spatiotemporal Analysis of Bluetooth Data : Application to a Large Urban Network.

IEEE Transactions on Intelligent Transportation Systems (99) : 1-10, 2014.

Dinh, T. U. J., Billot, R., Pillet, E., & El Faouzi, N.-E.

Real-Time Queue-End Detection on Freeways with Floating Car Data.

Transportation Research Record : Journal of the Transportation Research Board, 2470(1), 46-56. 2014.

Monteil, J., Billot, R., Sau, J., & El Faouzi, N.-E.

Linear and weakly nonlinear stability analyses of cooperative car-following models.

Intelligent Transportation Systems, IEEE Transactions on, 15(5), 2001-2013. 2014.

Sau, J., Monteil, J., Billot, R., & El Faouzi, N.-E.

The root locus method : application to linear stability analysis and design of cooperative car-following models.

Transportmetrica B : Transport Dynamics, 2(1), 60-82. 2014.

Monteil, J., Billot, R., Sau, J., Buisson, C., & El Faouzi, N.-E.

Calibration, Estimation, and Sampling Issues of Car-Following Parameters.

Transportation Research Record : Journal of the Transportation Research Board, (2422), 131-140. 2014.

Monteil, J., Nantes, A., Billot, R., & Sau, J.

Microscopic cooperative traffic flow : calibration and simulation based on a next generation simulation

dataset.

IET Intelligent Transport Systems, 8(6), 519-525. 2014.

Monteil, J., Billot, R., Sau, J., Armetta, F., Hassas, S., & El Faouzi, N.-E.

Cooperative Highway Traffic : Multiagent Modeling and Robustness Assessment of Local Perturbations.

Transportation Research Record : Journal of the Transportation Research Board, (2391), 1-10. 2013.

Billot, R., Chalons, C., De Vuyst, F., El Faouzi, N.-E., Sau, J.

A conditionally linearly stable second order traffic model derived from a Vlasov kinetic description.

Comptes Rendus de l'Académie des Sciences, CR. Mécanique. Doi :10.1016/j.crme.2010.07.018. 2010.

Billot, R., El Faouzi, N.-E., Sau, J., De Vuyst, F.

Integrating the Impact of Rain into Traffic Management : Online Traffic State Estimation Using Sequential Monte Carlo Techniques.

Transportation Research Records : Journal of the Transportation Research Board, vol 2169, ISSN 0361-1981, pp. 141-149, 2010.

El Faouzi, N.-E., De Mouzon, O., Billot, R., & Sau, J.

Assessing the changes in operating traffic stream conditions due to weather conditions.

Advances in Transp. Studies J. A, 21, pp. 33-46. 2010.

El Faouzi, N.-E., Billot, R., Bouzebda, S.

Motorway travel time prediction based on toll data and weather effect integration.

IET Intelligent Transport Systems, vol 4, issue 4, pp. 338-345, 2010.

Billot, R., El Faouzi, N.-E., De Vuyst, F.

A Multilevel Assessment of Rain Impact on Drivers' Behaviors : Standardized Methodology and Empirical Analysis.

Transportation Research Record : Journal of the Transportation Research Board, vol 2107., ISSN 0361-1981, pp. 134-142, 2009.

B.3 Autres revues à comité de lecture

El Faouzi, N.-E., Buton, J., & Billot, R.

Impacts de la météorologie sur la demande de trafic.

RTS : Recherche transports sécurité, (29). 2013.

Monteil, J., Billot, R., & El Faouzi, N.-E.

Véhicules Coopératifs Pour Une Gestion Dynamique Du Trafic : Approche Théorique Et Simulation.

Recherche Transports Sécurité, (29), p.47. 2013

Laharotte, P. A., Billot, R., El Faouzi, N.-E.

Flux de circulation : Bluetooth, une technologie aux dents longues !

Transports : revue bimestrielle, No. 486 (2014), pp. 5-13

B.4 Articles dans des conférences internationales avec comité de lecture

Salotti, J., Fenet, S., Billot R., Solnon, C., El Faouzi N.-E.

How does Causal Information Improve Short-Term Road Traffic Forecasting ? A Portfolio Approach.
Intelligent Data Analysis, IDA, 2017, London.

Laharotte, P.-A., Billot, R., El Faouzi, N.-E.

Detection of non-recurrent road traffic events based on clustering indicators.
Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning Bruges (Belgium), ESANN, 26 - 28 April 2017

Guériaux, M., Hassas, S., & Armetta, Billot, R., El Faouzi, N.-E.

A constructivist approach for a self-adaptive decision-making system : application to road traffic control.

28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Nov 6-8 2016, San Jose, California.

Guillodo, E., Berrouiguet, S., Simonnet, M., Conejero, I., Courtet, P., Baca-Garcia, E., Billot, R. & Walter, M.

Suicide Sleep Monitoring (SSleeM) : a feasibility and acceptability study of a wearable sleep tracking monitoring device in suicide attempters.

In ESSSB 2016 : 16th European Symposium on Suicide and Suicidal behaviour.

Simonnet, M., Gourvennec, B. & Billot, R.

Heart rate sensors acceptability : data reliability vs. ease of use.

In BSN : Body Sensor Network Conference. 14-17 june 2016 San Francisco, California, United States, pp. 94-98.

Simonnet, M., Gourvennec, B. & Billot, R.

Connected heart rate sensors to monitor sleep quality.

In CHASE : IEEE First Conference on Connected Health : Applications, Systems and Engineering Technologies, IEEE Computer Society, 27-29 june 2016, Washington, United States, pp. 344-345.

Berrouiguet, S., Billot, R., Lenca, P., Tanguy, P. et Baca-garcia, E., Simonnet, M. & Gourvennec, B
Toward E-Health Applications for Suicide Prevention.

In CHASE : IEEE First Conference on Connected Health : Applications, Systems and Engineering Technologies, IEEE Computer Society, 27-29 june 2016, Washington, United States, pp. 346-347.

Delhome R., Billot R., and El Faouzi N.-E.

Calibration of Time-Dependent Contraction Hierarchies based on Sensitivity Analysis.

In 28th European Conference on Operations Research (EURO), Poznan, Poland (2016).

Kovalev, S., Billot, R.

An exact method for multi-objective multi-modal trip planning problem,

In 28th European Conference on Operations Research (EURO), Poznan, Poland (2016).

- Billot, R.
TEMPUS : an open-source multi-modal trip planner.
European Conference of Operations Research (EURO), Glasgow, 12-16 July 2015.
- Guériau, M., Billot, R., El Faouzi, N.-E., Hassas, S., & Armetta
X2V-based information dissemination for highway congestion reduction.
Proceedings of the 2015 EWGT conference, Transportation Research Procedia, 2015.
- Laharotte, P. A., Billot, R., El Faouzi, N.-E., & Rakha, H. A.
Network-Wide Traffic State Prediction Using Bluetooth Data.
In Proceedings of the Transportation Research Board 94th Annual Meeting. 2015.
- Delhome, R., Billot, R. & El Faouzi, N.-E.
Moment-Ratio Diagram and Travel Time Reliability : Empirical Study on Urban and Periurban Links.
In Proceedings of the 6th International Symposium on Transportation Network Reliability (INSTR) in Nara, Japan. 2015.
- Delhome, R., Billot, R., Duret, A., & El Faouzi, N.-E.
The Halphen Distribution System, a Toolbox for Modeling Travel Time Variability : Some Insights Based on Mesoscopic Simulation.
In Proceedings of the Transportation Research Board 94th Annual Meeting. 2015.
- Guériau, M., Billot, R., El Faouzi, N.-E., Hassas, S., & Armetta, F.
Multi-Agent Dynamic Coupling for Cooperative Vehicles Modeling.
In Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.
- Gueriau, M., Billot, R., Monteil J., El Faouzi, N.-E, Hassas, S. & Armetta, F.
Agent-Based Cooperative Traffic Modeling and Simulation.
Proceedings of the Transportation Research Board 94th Annual Meeting. 2015.
- Billot, R., Faouzi, E., Guériau, M., & Monteil, J.
Can C-ITS lead to the Emergence of Traffic Management 2.0 ?
In Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on (pp. 483-488). IEEE. 2014.
- Baouche, F., Billot, R., Trigui, R., Faouzi, E., & El Faouzi, N.-E
Electric Vehicle Shortest Path Problem with Replenishment Constraint.
Proceedings of the IEEE ICCVE 2014 conference, Vienna, Austria.
- Guériau, M., Billot, R., Hassas, S., Armetta, F., & El Faouzi, N.-E.
An extension of MovSim for Multi-Agent Cooperative Vehicles Modeling.
2014 International Conference on Connected Vehicles & Expo. Vienne, Austria.
- Baouche, F., Billot, R., Trigui, R., El Faouzi, N.-E.
Electric Vehicle green routing with possible en-route recharging.
In Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on (pp. 2787-2792). IEEE. 2014.

Gueriau, M., Billot, R., & El Faouzi, N.-E.

Flocking behavior based microscopic cooperative traffic modeling : Simulation studies.

in Proceedings of the 4th International Symposium of Transport Simulation (ISTS), Ajaccio, France. 2014.

Laharotte, P. A., Billot, R., El Faouzi, N.-E.

How to filter transportation modes with Bluetooth data ? A case study on the Brisbane network.

In Proceedings of the 4th International Symposium of Transport Simulation (ISTS), Ajaccio, France. 2014.

Delhome, R., Billot, R. & El Faouzi, N.-E.

Investigating travel time reliability : empirical study based on Portland Bluetooth data.

In Proceedings of the 4th International Symposium on Transport Simulation (ISTS), Ajaccio, France. 2014

Monteil, J., Billot, R., Sau, J., El Faouzi, N.-E.

Calibration, Estimation, and Sampling Issues of Car-Following Parameters,

Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, USA, 2014.

Dinh, J., Billot, R., Pillet, E., El Faouzi, N.-E.

Real-Time Queue-End Detection on Freeways with FCD : Practice : A Practice-Ready Algorithm,

Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, USA, 2014.

Baouche, F., Billot, R., Trigui, R., El Faouzi, N.-E.

Optimization Model for Efficient Allocation of Electric Vehicle Charging Stations : Application to a Dense Urban Network.

Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, USA, 2014.

Nantes, A., Billot, R., Miska, M., Chung, E.

Bayesian Inference on Traffic Volume Based on Bluetooth data.

Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, USA, 2013.

Monteil, J., Billot, R., Sau, J., and El Faouzi, N.-E.

Cooperative highway traffic : multi-agent modelling and robustness assessment to local perturbations.

Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, USA, 2013.

Alfelor, R., Billot, R., El Faouzi, N.-E., & Pisano, P.

Approaches and Gaps in Weather Responsive Traffic Management – US and European Perspectives.

Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, USA, 2013.

Canaud, M., Mihaylova, L., El Faouzi, N.-E., Billot, R., & Sau, J.

A Bayesian Approach to Real-time Traffic State Estimation using Particle PHD with Appropriate Clutter Intensity.

Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, USA, 2013.

Sau, J., El Faouzi, N.-E., Billot, R., & Canaud, M.

Particle Filter Based Strategy For Online Calibration And Parameters' Estimation Of Motorway Traffic Model.

Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, USA, 2013.

Monteil, J., Nantes, A., Billot, R., Sau, J., & El Faouzi, N.-E.
Microscopic cooperative traffic based on NGSIM data.
In 19th ITS World Congress Proceedings. ITS International. 2012.

Monteil, J., Billot, R., Rey, D., & El Faouzi, N.-E.
Distributed and centralized approaches for cooperative road traffic dynamics.
Transportation Research Arena conférence, Procedia-Social and Behavioral Sciences, 48, 3198-3208. 2012

Billot, R., Rey, D. and Stocksieker, S.
Multiobjective shortest paths in multimodal networks.
Proceedings of the 19th Triennial Conference of the International Federation of Operational Research Societies (IFORS2011), 2011.

Monteil, J., Billot, R., El Faouzi, N.-E.
Towards cooperative traffic management : methodological issues and perspectives.
Australasian Transport Research Forum Proceedings 28 - 30 September 2011, Adelaide, Australia. 2011.

Tsubota, T., Chung, E., Bhaskar, A., Billot R.
Arterial traffic congestion analysis using Bluetooth Duration data.
Proceedings of the 34th Australasian Transport Research Forum. Adelaide, Australia. 2011

Billot, R., De Vuyst, F., El Faouzi, N.-E.
How does the fundamental diagram affect traffic characterization in a kinetic modeling ? Some evidence and weather effects integration.
Summer Meeting of the TRB Traffic Flow Theory Committee, Annecy, France, July 7-9, 2010.

Bousquet, A., Sau, J., El Faouzi ; N.-E., Billot, R.
Frequency-based characteristics of urban traffic and space-state modelling for travel time estimation.
5th IMA Conference on Mathematics in Transport , London (U.K.), 12-14 April, 2010.

El Faouzi, N.-E., Billot, R., Nurmi, P., Nowotny B.
Effects of adverse weather on traffic and safety : State of the art and the European initiative.
15th SIRWEC conference, Quebec City from February 5th to 7th 2010.

Billot, R., El Faouzi, N.-E., Sau, J., De Vuyst, F.
Integrating the Impact of Rain into Traffic Management : Online Traffic State Estimation Using Sequential Monte Carlo Techniques.
Proceedings of the 89th Transportation Research Board Annual Meeting (TRB), Washington (USA) , 10-16 January, 2010.

Billot R., N.-E. El Faouzi N.-E., De Vuyst F.
A Multilevel Assessment of Rain Impact on Drivers' Behaviors : Standardized Methodology and Empirical Analysis.
Proceedings of the 88th Transportation Research Board Annual Meeting (TRB), Washington (USA), 11-15 January, 2009.

El Faouzi, N.-E., Billot, R. , Bouzebda, S.
Spatio-temporal search algorithm and weather impact integration for travel time prediction based on motorway toll data.
Proceedings of the 16th ITS world congress, Stockholm, 21-25 September, 2009.

Billot, R.
Integrating the effects of adverse weather on traffic : methodology, empirical analyses and simulation.
ECTRI Young researcher seminar 2009, Torino, 5-9 June 2009.

El Faouzi N.-E., O. De Mouzon & R. Billot
Toward Weather-Responsive Traffic Management on French Motorways.
In proceedings of the fourth national conference on surface transportation weather, Transportation Research E-Circular Issue Number : E-C126, Indianapolis, 2008, pp 443-456.

Billot, R., El Faouzi, N.-E., Sau, J. & De Vuyst, F.
How does Rain affect traffic indicators ? Empirical study on a French interurban motorway.
In Proceedings of the Lakeside Conference. Klagenfurt, Austria, 2008.

Billot R., Suchier H.-M., Lallich S.
Toward an Unsupervised Boosting-Like Approach.
in Proceedings of the first joint meeting of the Société Francophone de Classification and the Classification and Data Analysis, Group of the Italian Society of Statistics, SFC CLADAG 2008, Caserta.

B.5 Articles dans des conférences nationales avec comité de lecture

Guériaux, M., Hassas, S., Armetta, F., Billot, R., El Faouzi, N.-E.
Contrôle par apprentissage constructiviste pour le trafic coopératif.
Journées Francophones des Systèmes Multi-Agents, JFSMA, 3 au 7 juillet 2017, Caen.

Salotti, J., Billot, R., El Faouzi, N. E., Fenet, S., Solnon, C.
Vers l'utilisation de graphes de liens causaux pour l'amélioration de la prévision court-terme du trafic routier.
In RFP-IA 2016 Journée Transports Intelligents.

Guériaux, M., Billot, R., Armetta, F., Hassas, S., & El Faouzi, N. E.
Un simulateur multiagent de trafic coopératif.
In 23es Journées Francophones sur les Systèmes Multi-Agents (JFSMA'15), (pp. 165-174). Cépaduès, 2015.

Delhome, R., Billot, R. & El-Faouzi, N.-E.
Recherche du plus court chemin stochastique avec des distributions de probabilité connues.
In Actes de la 16ème conférence de la société française de recherche opérationnelle (ROADEF2015), Marseille, France. 2015

Baouche, F., Billot, R., Trigui, R., & El Faouzi, N. E.
Electric Vehicle Charging Stations Allocation Model.
In ROADEF-15ème congrès annuel de la Société française de recherche opérationnelle et d'aide à la décision. 2014.

Billot R., Suchier H.M., Lallich S.
Une approche ensembliste inspirée du boosting en classification non supervisée.
actes 8e Conférence Extraction et Gestion des Connaissances, EGC 2008, Nice, Revue des Nouvelles Technologies de l'Information, RNTI. pp. 361-372. 2008

Autres publications

Billot R. Analyse et modélisation de l'impact de la météorologie sur le trafic routier.
Thèse de doctorat de l'Ecole Centrale de Paris. 2010.

Billot R. Une approche ensembliste inspirée du boosting en classification non supervisée.
Mémoire de Master Recherche, Université de Technologie de Compiègne (UTC). 2007.

Billot R. Performance Tests einer Wissensmanagement Applikation.
Rapport de stage (en langue allemande), Stage ingénieur longue durée pour PUMACY technologies (Berlin, Germany). 2006.

Monteil, J., Billot R. Techniques de stabilité pour la modélisation des effets des systèmes coopératifs sur le trafic.
Rapport technique, LICIT, ENTPE-IFSTTAR, Avril 2012.

Table des figures

| | | |
|------|--|----|
| 1.1 | Echange d'information dans les C-ITS. | 11 |
| 1.2 | Exemple d'un calculateur d'itinéraires : Rome2Rio | 14 |
| 1.3 | Complémentarité entre les échelles de travail | 19 |
| 1.4 | La modélisation du trafic routier. Une autre question d'échelles. | 20 |
| 1.5 | Défi 1 : thématiques de recherche | 24 |
| 1.6 | Défi 2 : thématiques de recherche | 29 |
| 1.7 | Défi 3 : thématiques de recherche | 35 |
| 1.8 | Organisation du manuscrit d'habilitation | 36 |
| 2.1 | Interprétation graphique des paramètres utilisés dans les lois de poursuites (crédit image : Maxime Guériau) | 40 |
| 2.2 | Représentation de la perception d'un véhicule équipé d'un système de contrôle basé sur le principe de la multi-anticipation (crédit image : Maxime Guériau). | 41 |
| 2.3 | Domaines de stabilité (le système est stable dans la partie supérieure aux courbes). | 42 |
| 2.4 | Développement d'ondes de congestion. A gauche 0% de véhicules coopératifs. A droite, 50% de véhicules équipés (<i>i.e</i> modèle BMA) | 42 |
| 2.5 | Root Locus : trafic instable avec prise en compte de 3 véhicules leaders (a) ou 6 véhicules leaders (b) | 43 |
| 2.6 | Modélisation conceptuelle par couplage de dynamiques | 44 |
| 2.7 | (a) Schéma général des systèmes coopératifs sur une section de route (b) véhicule non coopératif vs. véhicule coopératif. | 45 |
| 2.8 | Illustration de l'application de la règle de proximité. | 47 |
| 2.9 | (a) Capture d'écran de l'interface de MovSim. (b) Diagramme de classes de l'extension multi-agent de MovSim. Les classes bleues sont les classes du simulateur d'origine | 49 |
| 2.10 | Schéma d'opérations de la simulation | 50 |
| 2.11 | Trajectoires de véhicules pour différents taux de pénétration. (a) 0% (b) 20% (c) 40% (d) 60% (e) 80% (f) 100%. | 52 |
| 2.12 | Valeurs de désaccord de groupe pour différents taux de véhicules équipés : modèle multi-anticipatif bilatéral (BMA) vs. ACC. | 53 |
| 2.13 | Distributions des <i>PET</i> pour trois variantes de modèles de confiance. | 54 |
| 2.14 | Scénario d'insertion (a) instabilité du trafic avec 0% de véhicules connectés et (b) 40% de véhicules connectés. | 54 |
| 2.15 | Scénario d'insertion. Diagrammes fondamentaux de données agrégées par pas de 1min. (a) Voie 1 (b) Voie 2 (c) Voie 3 (d) | 55 |
| 2.16 | Scénario de fermeture de voie (a) instabilité du trafic avec 0% de véhicules connectés et (b) 50% de véhicules connectés. | 56 |

| | | |
|------|--|-----|
| 2.17 | Scénario de fermeture de voie. Diagrammes fondamentaux de données agrégées par pas de 1min. (a) Voie 1 (b) Voie 2 (c) Voie 3 (d) moyenne des 3 voies | 56 |
| 2.18 | Aperçu du modèle. | 59 |
| 2.19 | Processus de discrétisation, de classification et de sélection d'action. | 60 |
| 2.20 | Capture d'écran du scénario dans le simulateur. | 62 |
| 2.21 | Discrétisations générées par (a) D_1 [8 dimensions] et (b) D_2 [2 dimensions] sur le jeu de données d'apprentissage. | 65 |
| 2.22 | Résultats des 3 simulations : (a) Lundi 18, (b) Jeudi 21 et (c) Mardi 28. | 67 |
| 2.23 | Répartition entre les agents des consignes choisies par le modèle dans la simulation (c). 68 | |
| | | |
| 3.1 | Principe de détection de la technologie Bluetooth. | 72 |
| 3.2 | Cas d'études australien. | 73 |
| 3.3 | Description du site test de Portland et localisation des capteurs Bluetooth. | 73 |
| 3.4 | Cas d'étude de la rocade Est de Lyon. | 74 |
| 3.5 | Suivi de véhicules traceurs. | 75 |
| 3.6 | δ -MRD théorique [350]. | 78 |
| 3.7 | Boîtes à moustache des distributions de mesures de performance sur des échantillons d'une heure de données à Portland. | 81 |
| 3.8 | Boîtes à moustache des distributions de mesures de performance sur des échantillons de 15 minutes de données à Lyon. | 82 |
| 3.9 | Similarité entre les indicateurs de fiabilité classiques et les coordonnées du δ -MRD. | 83 |
| 3.10 | Comparaison des distributions de temps de parcours mesurés entre 16h et 17h à Portland, sur différents jours. | 84 |
| 3.11 | Représentation des échantillons de Portland et Lyon sur le δ -MRD. | 85 |
| 3.12 | Illustration des difficultés rencontrées à partir des données Bluetooth. | 86 |
| 3.13 | Représentation spatiale de matrices B-OD de volume à différentes heures pour la journée du 7 novembre 2011. À gauche : entre 3h et 4h du matin (période creuse). À droite : entre 8h et 9h du matin (pic de trafic). | 88 |
| 3.14 | Illustration des conditions de trafic pour un couple de capteurs Bluetooth : (a) Série temporelle de la vitesse horaire sur un mois. (b) Série temporelle du débit moyen sur un mois. (c) Diagramme fondamental Bluetooth (B-FD) débit/vitesse colorié par périodes temporelles ([bleu : soirée et nuit]; [rouge : matin]; [vert : midi]; [jaune : après-midi]). | 89 |
| 3.15 | Observation du taux de pénétration sur un jour pour différents couples de B-OD. | 90 |
| 3.16 | (a) Deux semaines de données Bluetooth. Nombre de détections générées par heure par chaque <i>pattern</i> . (b) Représentation spatiale des couples B-OD en fonction de leurs <i>pattern</i> associés (Novembre 2011). | 92 |
| 3.17 | Méthodologie de clustering de B-FD. | 93 |
| 3.18 | Illustration des résultats d'un clustering des diagrammes fondamentaux Bluetooth B-FD (Novembre 2011) : (a) formes typiques des B-FD quelques clusters. (b) Représentation spatiale du réseau à partir des clusters de B-FD. | 95 |
| 3.19 | Différence entre une prévision fondée sur des séries spatio-temporelles (approche locale) et une approche globale. | 97 |
| 3.20 | La méthode des k plus proches voisins adaptée à l'approche globale. | 98 |
| 3.21 | Processus d'apprentissage. | 99 |
| 3.22 | Evaluation de l'approche kNN globale. | 101 |
| 3.23 | Approches locale vs. globale : comparaisons sur un segment-témoin. | 102 |

| | | |
|------|--|-----|
| 3.24 | Prévision des états de trafic à partir d'une heure choisie à l'avance, pour les horizons successifs fixés respectivement à $t + 1$, $t + 2$, $t + 3$. | 102 |
| 3.25 | Résilience de la méthode kNN globale aux incidents. | 103 |
| 3.26 | Performances des méthodes sur le cas de Cheviré, Nantes. | 104 |
| 3.27 | Identification des sections critiques. Les Δ représentent les variations du signal de vitesse sur les liens adjacents en tenant compte de la physique spatiotemporelle du trafic routier. | 105 |
| 4.1 | TEMPUS : modèle conceptuel | 108 |
| 4.2 | Interface de TEMPUS | 109 |
| 4.3 | TEMPUS : exemple de requête multimodale | 109 |
| 4.4 | TEMPUS : A^* vs. Dijkstra | 110 |
| 4.5 | Application TEMPUS proposée lors de la conférence <i>ITS World Congress</i> , à Bordeaux en 2015. | 110 |
| 4.6 | Cycles de vitesse. | 112 |
| 4.7 | Approche systémique du modèle de consommation VEHLIB. | 113 |
| 4.8 | Carte du Grand Lyon. Lyon et ses 9 arrondissements (en rouge) et la métropole de Lyon (en orange) | 113 |
| 4.9 | Démarche méthodologique pour l'affectation des stations de recharge. | 114 |
| 4.10 | Construction des sous-ensembles <i>RCS</i> , <i>RD</i> et <i>SFP</i> . | 116 |
| 4.11 | Graphe de connectivité | 117 |
| 4.12 | Exemple de solution du problème de PCC énergétique. | 118 |
| 4.13 | Exemple de deux solutions à une même requête. PCC énergétique (gauche) vs. <i>Google Maps</i> (droite). | 119 |
| 4.14 | Exemple de reconstitution d'un itinéraire optimal entre u_1 et u_4 | 121 |
| 4.15 | Illustration de la dominance stochastique de premier ordre | 122 |
| 4.16 | Proportions des 5 situations constituant le premier indicateur pour un sous-ensemble de configurations. H : nombre de points support (discrétisation) ; V : niveau de variance de 1 (faible) à 3 (élevé) ; G : type du graphe ; D : niveau de distance entre origine et destination (ex : D9 = 9ème décile) ; Q : probabilité d'être à l'heure (ex : C90 = 90ème centile). | 125 |
| 4.17 | Configurations de hiérarchisation possibles | 127 |
| 4.18 | Type de chemin mis en valeur par la procédure des contractions hiérarchiques | 128 |
| 4.19 | Baisse du temps de calcul grâce aux contractions hiérarchiques | 129 |
| 5.1 | Exemple d'application mobile développée à Madrid : MEmind. | 133 |
| 5.2 | Système d'aide à la décision pour la prévention du risque suicidaire. | 134 |
| 5.3 | Approche de réduction de dimension : projection des individus statistiques sur les deux premières dimensions et dendrogramme menant au choix de trois groupes. | 139 |
| 5.4 | Arbre de régression de la variable "nombre de TS" pour la population masculine. | 140 |
| 5.5 | Arbre de régression de la variable "nombre de TS" pour la population féminine. | 141 |
| 5.6 | Application Beleaf de prévention du risque suicidaire. Projet d'étudiants à IMT Atlantique. | 144 |
| 6.1 | Structuration du projet de recherche. | 149 |

Liste des tableaux

| | | |
|-----|--|-----|
| 4.1 | Répartition des temps de calcul lors de la résolution du PCC énergétique | 117 |
| 4.2 | Temps d'exécution par groupe de requêtes de l'algorithme <i>SPOTAR</i> | 124 |
| 5.1 | Caractéristiques sociologiques du jeu de données post-filtrage de 681 patients reçus après une tentative de suicide. Q1 et Q3 :1er et 3ème quartiles, respectivement. BIS10 : <i>Barrat Impulsiveness Scale</i> . Echelle d'impulsivité de Barrat. | 137 |
| 5.2 | Caractéristiques cliniques du jeu de données final. | 138 |
| A.1 | Tableau de synthèse des enseignements jusqu'à l'année 2016-2017 | 191 |

