

N° d'ordre NNT : 2016LYSET013

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

l'Ecole Nationale des Travaux Publics de l'Etat

École Doctorale N° 162

MEGA (Mécanique, Energétique, Génie Civil et Acoustique)

Spécialité / discipline de doctorat : Génie Civil

proposée pour soutenance publique le 16 Décembre 2016, par :

Pierre-Antoine LAHAROTTE

Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier

---

Devant le jury composé de :

VINCENT AGUILERA,	Directeur de Recherches (École des Ponts ParisTech)	*Rapporteur*
LIMING CHEN,	Professeur (École Centrale Lyon)	*Rapporteur*
LATIFA OUKHELLOU,	Directrice de Recherches (Université de Marne-la-Vallée)	*Examinatrice*
PHILIPPE GOUDAL,	Responsable Innovations (MEDIAMOBILE)	*Examinateur*
ROMAIN BILLOT,	Maitre de Conférences (Telecom Bretagne)	*Co-Directeur de Thèse*
NOUR-EDDIN EL FAOUZI,	Directeur de Recherches (Université de Lyon)	*Directeur de Thèse*

Thèse préparée au LICIT (Laboratoire d'Ingéniérie Circulation Transport)

Pierre-Antoine LAHAROTTE

Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier

16 Décembre 2016

Rapporteurs : Vincent AGUILERA et Liming CHEN

Encadrants : Romain BILLOT et Nour-Eddin EL FAOUZI





Thèse préparée au laboratoire

LICIT

Laboratoire Ingéniérie CIrculation Transports

UMR T_E IFSTTAR-ENTPE

Université de Lyon

IFSTTAR

COSYS/LICIT

25 Avenue François Miterrand

Case 24

Cité des mobilités

69675 Bron Cedex, France

ENTPE

LICIT

Rue Maurice Audin

69518 Vaulx-en-Velin Cedex, France

À Mademoiselle Pasquotte,
Citadelle sans qui ce
Chemin eût été fort
Taciturne à suivre

Résumé

La maturité de la télématique et des technologies de l'information et la communication (TIC), ainsi que l'avènement du big data dans le transport ont conduit à des développements foisonnants dans le domaine des systèmes de transports intelligents (ITS), aussi bien sur le plan des technologies de recueil que du traitement innovant de l'information. Il est désormais possible de connaître les conditions de circulation et les états de trafic sur la plupart des sections d'un réseau routier sans avoir recours à des infrastructures intrusives de collecte de données, de transmettre l'information résultante via des réseaux sans fil et de traiter rapidement toutes ces données multi-sources disponibles.

La constitution de grandes bases de données a naturellement fait évoluer la pratique de gestion du trafic et plus particulièrement les méthodes de prévision. Ces méthodes ont connu un renouveau en s'inspirant des travaux produits en apprentissage statistique. Néanmoins, la façon d'appréhender le problème de la prévision est restée à une échelle locale. Pour chaque section de route, un modèle de prévision est adapté et optimisé. Notre travail de thèse présente un cadre de prévision du trafic routier qui aborde la question à l'échelle du réseau. L'étude menée au sein de ces travaux de thèse vise à exposer et évaluer cette nouvelle approche, dite globale, au regard d'approches usuelles, puis à analyser sa sensibilité vis-à-vis de divers facteurs.

Après un positionnement par rapport à l'état de l'art en théorie du trafic, le cadre prédictif fondé sur des méthodes de prévision multi-variées par apprentissage est détaillé. Une version multidimensionnelle des k plus proches voisins, modèle parcimonieux et simple, est évaluée sur divers cas d'études. L'originalité réside dans l'exploitation de données issues de méthodes innovantes de collecte (e.g. Bluetooth, véhicules traceurs, véhicules connectés).

Par la suite, les performances de l'approche initiale sont comparées à d'autres méthodes d'apprentissage. Un effort particulier est porté sur l'adaptation de méthodes à noyaux au cadre prédictif global. Les performances obtenues laissent entrevoir une typologie des méthodes en fonction des caractéristiques spatiotemporelles du réseau. Afin d'améliorer les performances en prévision et de réduire les temps de calcul, une méthode d'identification et de sélection des sections critiques du réseau est proposée. Les résultats prouvent qu'un sous-ensemble restreint de sections est en effet suffisant pour garantir des performances satisfaisantes en généralisation.

Enfin, la résilience du cadre prédictif est évaluée au regard des événements non récurrents affectant le fonctionnement nominal du réseau, comme des incidents ou des conditions météorologiques dégradées. Les résultats soulignent l'impact de ces conditions non récurrentes sur la prévision temps-réel de la dynamique court-terme d'un réseau et permettent de dresser une feuille de route pour l'élaboration d'un cadre prédictif résilient et opérationnel.

Cette nouvelle vision de la prévision s'inscrit dans les perspectives actuelles en termes d'applications sur les modules embarqués et les objectifs des gestionnaires d'infrastructures.

Mots-clés

Big Data, Bluetooth, *Floating Car Data* (FCD), prévision, apprentissage, réseau, résilience, événement non récurrent, sections critiques.

Contributions to the short-term, multi-variate and multi-scale prediction of traffic states based on machine learning methods

Abstract

The maturity of information and communication technologies and the advent of Big Data have led to substantial developments in intelligent transportation systems (ITS) : from data collection to innovative processing solutions. Knowledge of current traffic states is available over most of the network range without the use of intrusive infrastructure-side collection devices, instead relying on wireless transmission of multi-source data.

The increasing use of huge databases had a strong influence on traffic management, including forecasting methods. These approaches followed the recent trend towards innovative works on statistical learning. However, the prediction problem remains mainly focused on the local scale. The prediction for each road link relies on a dedicated, optimized and adapted prediction model. Our work introduces a traffic-forecasting framework able to tackle network scale problems. The study conducted in this thesis aims to present and evaluate this new "global" approach, in comparison to most-used existing works, and then to analyze its sensitivity to several factors.

The traffic-forecasting framework, based on multi-variate learning methods, is detailed after a review of the literature on traffic flow theory. A multi-dimensional version of the k nearest-neighbors, a simple and sparse model, is evaluated through several use cases. The originality of the work stands on the processing approach, applied to data collected through new measurement process (e.g. Bluetooth, floating car data, connected vehicles).

Then, the performance of our primary approach is compared to other learning-based methods. We propose an adaptation of kernel-based methods for the global prediction framework. The obtained results show that global approaches perform as well as usual approaches. The spatial and temporal specificities of the methods are highlighted according to the prediction accuracy. To improve the forecasting accuracy and reduce the computation time, we propose an identification and selection method targeting critical links. The results demonstrate that the use of a restricted subset of links is sufficient to ensure acceptable performances during validation tests.

Finally, the prediction framework resilience is evaluated with respect to non-recurrent events as incidents or adverse weather conditions affecting the nominal network operations. The results highlight the impact of these non-recurrent conditions on real-time forecasting of short-term

network dynamics. This enables the design of a further operational and resilient prediction framework.

This perspective of forecasting matches the current applications relying on embedded systems and addressing the traffic network supervisor's expectations.

Keywords

Big Data, Bluetooth, Floating Car Data, prediction, forecasting, learning, network, non-recurrent event, critical links.

Remerciements

Mes premiers remerciements s'adressent à vous, courageux lecteur. Je vous remercie de l'intérêt que vous portez à mes travaux de thèse et vous souhaite de pouvoir y trouver des informations utiles, voire d'y puiser un peu d'inspiration.

Dans cette même veine, je tiens particulièrement à remercier l'ensemble des membres de mon jury : mes deux rapporteurs, messieurs Vincent Aguilera et Liming Chen, pour leur lecture assidue et les questions pointues qu'ils ont soulevées ; mes deux examinateurs, madame Latifa Oukhellou et monsieur Philippe Goudal, pour la curiosité, dont ils ont fait preuve à l'égard de mes travaux et, enfin, mes encadrants et directeurs de thèse, messieurs Romain Billot et Nour-Eddin El-Faouzi, dont les relectures se sont toujours révélées pertinentes.

Dans la famille mentor, je voudrais remercier papa-mentor, alias Nour-Eddin El-Faouzi, difficile à attraper, souvent surchargé, mais contribuant à la réflexion toujours avec une joie de vivre communicative, et grand frère-mentor, alias Romain Billot, qui m'a véritablement encadré aussi bien sur le plan professionnel que personnel en m'aidant à me surpasser dans de nombreux domaines. Je voudrais également les remercier pour la confiance qu'ils ont eu à mon égard, même dans les moments où je doutais de moi-même, et leur soutien.

Je tiens maintenant à remercier mes compagnons de thèse, Raphaël Delhome et Maxime Guériau, pour tous les moments passés au boulot comme à l'extérieur. Contrairement à nos parties de jeux coopératifs, cette expérience de thèse est loin d'avoir été un "échec collectif". Je suis content d'avoir partagé ces trois intenses années avec vous, à se supporter mutuellement dans tous les sens du terme. Vous êtes mes deux râleurs préférés. Un remerciement particulier à Maxime, qui a eu le courage de partager son bureau avec moi, ainsi que notre passion commune pour flamber notre argent chez Decathlon.

Viens maintenant le tour de mes collègues du LICIT.
Je commencerais par ceux de l'équipe MOMI avec qui je partageais mon quotidien et de nombreuses pauses café. Je pense notamment à : Junior, alias Aurélien Clairais, qui a fait une tentative de putsch de bureau avortée juste à temps, à son encadrant, Aurélien Duret, qui ma fait déposer un poumon sur le bord du chemin lors de notre dernier entraînement de course, au couple d'inséparables : Fouad Baouche et Bernard Schnetzler, à Colette Minsta-Eya, avec qui j'ai partagé les cours de statistiques, à smart Jean-Luc Ygnace, qui est capable de faire ressortir tous les bugs sous Mac, rien qu'en en allumant un, à Anne-Christine Demanny, qui

a fait explosé notre glycémie chaque lundi et a eu le courage de relire nos thèses et articles, à Angelo Furno, qui va connaître les joies de l'exploitation de données Bluetooth, à Vincent Grébert, qui débute cette belle expérience de la thèse, à Julien Salotti, qui partage la même passion pour les méthodes de prévision, et à Anastasia, qui amène un peu de la Grèce dans ce monde de brutes.
Côté AMMET, sans partager votre quotidien, les événements sociaux ont contribué à faire plus ample connaissance. Je remercie donc Ludovic Leclercq, Christine Buisson, Nicolas Chiabaut, Cécile Becarie, Delphine Leijri, Jean Krug, Arthur Burianne et Sonia Cenille, qui m'a aidé à survivre dans le monde impitoyable des papiers administratifs, ainsi que l'ensemble des doctorants : Clélia Lopez, qui adore les tartes au citron et sur qui notre canapé à un effet soporifique immédiat, Nicole Schiper et sa voix, Guilhem Mariotte, qui a su charmer nos oreilles avec son ukulélé, Sergio Assuncao, portugais et fier de l'être, Johana Cattin, qui possède des bureaux transitoires dans tout Lyon, et les nouveaux : Chuanlin Zhao et Humberto Gonzalez.

Je remercie ceux qui ont été de passage au Licit et ont partagé pour un temps mon quotidien : Anahid, Sergey, Matthieu, Fathia, Fabien, Frédéric, Camille et Guillaume entres autres.

Je tiens à remercier quelques amis : Etienne, qui me précède d'un an sur les études comme la carrière professionnelle, Marine, ex-coloc qui m'a donné le secret des meilleures tartes au citron, Justine Dinh et tous ceux qui le sont devenus au cours de ces trois dernières années : Maxime, Raphaël, Junior, Amandine, Jules, Marie, Guillaume, Lucie, Guillaume, Vanina et tous ceux que j'oublie.

Le sport m'a permis de beaucoup me défouler et de tenir le rythme tout du long. Je tiens donc à remercier mon maître d'arme, Laurence, et tous les membres du club d'escrime, ainsi que Romain et Aurélien pour les courses partagées, sans oublier, Maxime pour les parties de tennis du début de thèse.

Enfin, je voudrais également remercier mes parents, Catherine et Jean-Marie, et grand-parents, Charles & Denise et Jacqueline & Claude, sans qui rien n'aurait été possible, ainsi que mes six frères et soeurs, Pierre-Marie, Claire-Océane, Corentin, Coraline, Tiphaine et Gwendoline, qui m'ont soutenu et distrait tout au long de ces trois années. Vous êtes géniaux.

Le plus tendre pour la fin. Mes ultimes remerciements vont à celle qui partage ma vie depuis maintenant deux années et qui m'a apporté un soutien sans faille à tout moment : madame Anaïs Pasquier. Merci d'avoir été là pour moi, de m'avoir aidé à me changer les idées et de m'avoir écouté dans les moments de doute. Je suis heureux de vivre avec toi et de partager tous ces moments, bons comme mauvais, pour profiter pleinement de l'intensité de la vie. Merci d'être toi. Je ne manquerais pas de te soutenir du mieux tout au long de la prochaine année pour mener à bien cette aventure de la thèse. Aventure qui s'achèvera pour laisser la place à de nouvelles, remplies de voyages, de balades, de plongées et de moments complices à partager au sein de cette famille que nous construisons petit à petit.

Table des matières

Table des figures **xvii**

Liste des tableaux **xxi**

1 Supervision et prévision du trafic sur les réseaux routiers **3**
1.1 Caractérisation du réseau routier 4
1.1.1 Définition de la notion de réseau 4
1.1.2 Définition opérationnelle 5
1.1.3 Caractéristiques de trafic 6
1.2 Supervision du réseau routier 7
1.2.1 Différentes méthodes de recueil 8
1.2.1.1 Méthodes de recueil liées à l'infrastructure 8
1.2.1.2 Méthodes de recueil liées à l'infrastructure et embarquées . . . 11
1.2.1.3 Méthodes de recueil embarquées 12
1.2.2 Enjeux de la supervision du trafic 13
1.2.3 Cas étudiés 14
1.2.3.1 Réseau de Brisbane (Australie) et technologie Bluetooth 15
1.2.3.2 Réseau nantais et technologie GPS 19
1.3 Prévision court-terme à l'échelle du réseau routier 20
1.3.1 Enjeux de la prévision du trafic 20
1.3.2 Prévision à partir de modèles (*model-driven*) 21
1.3.2.1 Approches macroscopiques 21
1.3.2.2 Les approches microscopiques 23
1.3.2.3 Les approches mésoscopiques 23
1.3.3 Prévision à partir des données (data-driven) 24
1.3.3.1 Évolutions du paramétrique à l'apprentissage 24
1.3.3.2 Approches spatio-temporelles 31
1.3.3.3 Unification des modèles à l'échelle d'un réseau : transition d'une approche de décomposition à une vision systémique 34
1.4 Verrous scientifiques et démarche proposée 36
1.4.1 Positionnement de l'étude 36
1.4.2 Démarche adoptée 38

2 Introduction d'un cadre prédictif à l'échelle du réseau **41**
2.1 Exigences associées à la démarche globale . . . 43
2.2 Prévision par apprentissage . . . 45
2.2.1 Cadre prédictif usuel . . . 45
2.2.2 Adaptation du cadre prédictif . . . 46
2.2.3 Composantes fonctionnelles du cadre prédictif à adapter à l'approche globale . . . 48
2.2.3.1 Définition des variables d'entrée et de sortie . . . 48
2.2.3.2 Méthode de prévision . . . 49
2.2.3.3 Erreur de prévision . . . 51
2.2.4 Composantes fonctionnelles classiques du cadre prédictif . . . 53
2.2.4.1 Gestion des données manquantes . . . 53
2.2.4.2 Gestion de la taille de l'historique . . . 56
2.3 Mise en oeuvre de la stratégie d'optimisation des hyperparamètres . . . 58
2.3.1 Recherche du compromis entre représentativité et généralisation . . . 58
2.3.2 Procédés d'optimisation des hyperparamètres . . . 59
2.3.2.1 Stratégies d'échantillonnage et validation croisée . . . 60
2.3.2.2 Choix d'une stratégie d'optimisation . . . 61
2.4 Tests sur deux cas d'étude . . . 64
2.4.1 Prévision sur le réseau de Brisbane à partir de données Bluetooth . . . 65
2.4.1.1 Optimisation des hyperparamètres . . . 65
2.4.1.2 Performances de prévision . . . 67
2.4.2 Prévision sur le réseau de Nantes à partir des données FCD . . . 70
2.5 Conclusions et perspectives . . . 73

3 Adaptation de méthodes à noyaux et analyse comparative **75**
3.1 Adaptation des méthodes à noyaux aux exigences de l'approche globale . . . 76
3.1.1 Régression non paramétrique : méthodes à noyaux . . . 76
3.1.1.1 Méthodes de régression à noyaux . . . 76
3.1.1.2 Formulation pour l'approche globale . . . 78
3.1.1.3 Optimisation des hyperparamètres . . . 78
3.1.2 Régression par machines à Vecteurs Supports (SVR) . . . 80
3.1.2.1 Méthode à Vecteurs Supports usuelle . . . 80
3.1.2.2 Adaptation de la Régression SVR à l'approche globale . . . 85
3.1.2.3 Optimisation des hyperparamètres du MSVR . . . 88
3.2 Analyse comparative des performances des méthodes à noyaux sur la zone de Cheviré . . . 90
3.2.1 Performances en termes de temps d'estimation en temps-réel . . . 90
3.2.2 Performances globales en fonction de l'horizon de prévision . . . 91
3.2.3 Analyse des performances suivant les périodes de la journée . . . 93
3.2.4 Analyse des performances suivant les sections considérées . . . 94
3.2.5 Typologies spatiales et temporelles des meilleures méthodes . . . 96

3.3 Analyse comparative pour le réseau très maillé d'une zone urbaine (centre-ville nantais) . . . 98
3.3.1 Évolution de la performance quant à la complexité temporelle . . . 99
3.3.2 Évaluation des performances vis-à-vis de l'erreur de prévision . . . 100
3.3.2.1 Analyse de l'erreur globale sur la période de validation . . . 100
3.3.2.2 Analyse spatiale et temporelle . . . 101
3.3.2.3 Identification des typologies spatio-temporelles des meilleures méthodes . . . 102
3.4 Confrontation à des méthodes alternatives de l'état de l'art . . . 104
3.4.1 Introduction succincte aux méthodes de l'état de l'art testées . . . 104
3.4.1.1 Méthode paramétrique simple . . . 105
3.4.1.2 Deux approches ensemblistes apprenant la structure de l'historique . . . 105
3.4.2 Analyse comparative des performances . . . 107
3.5 Conclusions et perspectives . . . 109

4 Sélection de sections critiques sur le réseau 111
4.1 Positionnement des démarches envisagées au sein de la littérature . . . 113
4.1.1 Traitement de la démarche d'identification des sections critiques dans la littérature . . . 113
4.1.2 Traitement de la démarche de réduction de la dimensionnalité dans la littérature . . . 114
4.2 Identification de sections critiques pour la sélection de variables . . . 116
4.2.1 Phase d'encodage par décomposition physique du système : les configurations locales . . . 118
4.2.1.1 Analyse du système local . . . 118
4.2.1.2 Encodage des configurations d'état . . . 119
4.2.2 Phase d'identification et de tri des motifs latents à partir de l'Allocation de Dirichlet Latente (LDA) . . . 122
4.2.2.1 Le LDA, un modèle génératif probabiliste . . . 123
4.2.2.2 Sélection des sections critiques à partir du LDA . . . 129
4.2.2.3 Illustration de la phase de détermination du critère de criticité par le LDA sur un cas simple sur le réseau du Pont de Cheviré . 134
4.3 Évaluation du procédé de sélection de variables sur les réseaux nantais . . . 139
4.3.1 Introduction de critères alternatifs de classement des sections . . . 139
4.3.1.1 Critère M1 . . . 139
4.3.1.2 Critère M1Bis . . . 139
4.3.1.3 Critère M2 . . . 140
4.3.1.4 Critère M3 . . . 140
4.3.1.5 Critère LDA64 . . . 141
4.3.1.6 Critère MLasso . . . 142
4.3.2 Comparaison des critères de criticité sur la zone du Pont de Cheviré . . . 142

4.3.3 Application à une zone urbaine étendue . . . 145
4.3.4 Vers une application en temps-réel du procédé de sélection de variables . 147
4.4 Conclusions et perspectives . . . 148

5 Identification et intégration de phénomènes non récurrents perturbant le comportement nominal du réseau 151
5.1 Les événements non récurrents dans la littérature du trafic . . . 152
5.1.1 Identification des événements non récurrents . . . 152
5.1.2 De la mesure à l'anticipation d'impact des incidents . . . 154
5.2 Objectifs méthodologiques : un procédé en deux phases . . . 155
5.3 Phase d'identification des incidents . . . 156
5.3.1 Une méthode originale combinant classification et seuils . . . 157
5.3.1.1 Hypothèse et méthodologie . . . 157
5.3.1.2 Usage de la perplexité . . . 157
5.3.1.3 Construction d'individus suivant les contraintes . . . 158
5.3.1.4 Application du LDA . . . 159
5.3.1.5 Vérification de la pertinence de l'hypothèse fondamentale . . . 160
5.3.1.6 Application du filtrage par seuil . . . 161
5.3.2 Introduction d'approches alternatives . . . 164
5.3.2.1 Méthode de référence pour les approches physiques par seuils. 164
5.3.2.2 Méthode de référence pour les approches par classification : une machine à vecteurs supports (SVM). . . . 165
5.3.2.3 Méthode alternative : le SVM novelty. . . . 165
5.3.3 Comparatif des méthodes de détection d'incidents . . . 166
5.3.3.1 Comparaison des performances via des outils usuels . . . 166
5.3.3.2 Comparatif sur l'axe principal de la zone du Pont de Cheviré . . . 167
5.4 Phase de validation et d'intégration temps-réel des incidents . . . 170
5.4.1 Procédé d'adaptation du cadre prédictif pour intégration des incidents . 171
5.4.2 Passage à l'échelle du réseau de la détection locale d'incidents . . . 173
5.4.3 Limitation des itérations d'informations au sein de l'historique d'incidents 173
5.4.3.1 L'erreur de prévision comme qualificatif d'impact . . . 174
5.4.3.2 Caractérisation formelle des incidents à fort impact prédictif . . . 174
5.4.4 Evaluation des performances liées à l'intégration d'un historique permanent d'incidents . . . 175
5.5 Perspectives et discussion . . . 177

6 Vers une intégration de l'effet des conditions météorologiques dans le cadre prédictif 179
6.1 La météorologie, un facteur exogène récurrent de la littérature du trafic . . . 181
6.1.1 Estimation des effets météorologiques sur le trafic . . . 181
6.1.2 Intégration des effets météorologiques aux méthodes de prévision . . . 183
6.2 Qualification des informations météorologiques . . . 185
6.2.1 Descriptif des données météorologiques et cas d'étude . . . 185

6.2.2 De l'information disponible à l'information exploitable 186
6.2.2.1 Analyse de la répartition des observations sur les catégories pour évaluation de la significativité 187
6.2.2.2 Mesure de significativité de l'impact par analyse des variances . 188
6.3 Quantification des effets météorologiques sur la vitesse 190
6.3.1 Analyse d'impact par période de la journée 190
6.3.2 Confrontation de journées similaires 193
6.4 Évaluation de l'impact du degré de précipitation sur l'erreur de prévision 195
6.4.1 Analyse statistique de l'erreur de prévision en fonction des conditions météorologiques 195
6.4.2 Vers une intégration des phénomènes météorologiques au sein du cadre prédictif 198
6.5 Discussion et perspectives 200

Conclusion générale et perspectives **203**

Articles de l'auteur et valorisation **209**

Bibliographie **211**

A Mise en oeuvre algorithmique du procédé d'optimisation des hyperparamètres **231**

B Implémentation sous Matlab du modèle MSVR **233**

Table des figures

1.1 Représentations classiques des variables macroscopiques du trafic 8
1.2 La supervision du réseau nécessite une couche de capteurs positionnés sur le réseau. 9
1.3 Illustration des typologies de recueil de données 11
1.4 Illustration des deux réseaux étudiés 15
1.5 Illustration des difficultés rencontrées à partir des données Bluetooth 16
1.6 Illustration du procédé de filtrage adopté sur le réseau de Brisbane (technologie Bluetooth) 17
1.7 Fiabilité des capteurs Bluetooth 18
1.8 D'une approche série temporelle à une méthode spatio-temporelle 32
1.9 D'une approche série spatio-temporelle (locale) à approche globale 38
1.10 Illustration de la démarche méthodologique adoptée 40

2.1 Procédé de prévision depus une méthode par apprentissage supervisé 42
2.2 Illustration des dimensions des variables d'entrée et de sortie pour l'approche globale. 44
2.3 Cadre prédictif usuel 45
2.4 Application de la règle de décision des k plus proches voisins à un cliché du réseau (μ=0) 50
2.5 Illustration du procédé de complétion des données manquantes à partir de profils nominaux 55
2.6 Illustration du lissage des valeurs se substituant aux données manquantes 56
2.7 Gestion usuelle de la base de données dans le cadre de l'apprentissage supervisé . 56
2.8 Evaluation de la taille d'historique glissant optimale 57
2.9 Illustration des phénomènes de sur - et sous-apprentissage 59
2.10 Recherche d'un compromis entre généralisation et représentativité 60
2.11 Confrontation du procédé de validation croisée à m-échantillons à celle spécialement adaptée aux séries temporelles. 61
2.12 Illustration des deux principales approches d'optimisation 63
2.13 Représentation de la cartographie de performance de MKNN depuis une analyse par grille sur le réseau de Brisbane 66
2.14 Confrontation des performances à l'échelle du réseau ($MAPE_{global}$) sur le mois de validation 68

2.15 Distribution de l'erreur de prévision ($RMSE_{TimeStep}$) en fonction des périodes de la journée . . . 69
2.16 Distribution de l'erreur de prévision en fonction des caractéristiques des sections . 70
2.17 Analyse focalisée sur la propagation d'une congestion non récurrente sur le réseau de Brisbane. . . . 71
2.18 Comparaison des performances des diverses méthodes et approches de prévision par KNN considérées . . . 72
2.19 Analyse des distibutions de l'erreur de prévision par les diverses méthodes KNN sur la zone de Cheviré . . . 73

3.1 Analogie de fonctionnement entre la régression non paramétrique des K plus proches voisins (KNN) et celle à fonctions noyaux (Kernel de type RBF). . . . 77
3.2 Illustration des différentes étapes du processus d'optimisation des hyperparamètres pour la méthode MKernel sur le Pont de Cheviré . . . 79
3.3 Illustration des deux idées directrices du processus de classification par la méthode SVM. . . . 81
3.4 Illustration du procédé de recherche de l'hyperplan optimal avec relâche des contraintes liées à la marge maximale dans l'espace à haute dimensions résultant de ϕ. . . . 82
3.5 Illustration de la transition d'un problème de classfication SVM à un procédé de régression SVR (Représentation dans l'espace à haute dimension). . . . 83
3.6 Illustration des résultats du procédé ACO d'optimisation des hyperparamètres pour la méthode MSVR sur le Pont de Cheviré . . . 89
3.7 Performances globales de prévision des 7 méthodes testées suivant l'horizon de prévision . . . 92
3.8 Analyse locale et par pas de temps ($MAPE_{TimeStep}$) des erreurs de prévision . . 94
3.9 Analyse locale et par section ($MAPE_{section}$) des erreurs de prévision . . . 95
3.10 Analyse de la répartition spatiale de l'erreur de prévision ($MAPE_{section}$) . . . 95
3.11 Mise en concurrence des sept méthodes de prévision . . . 97
3.12 Représentation spatiale des meilleures méthodes en fonction de la période de la journée . . . 98
3.13 Performances globales de prévision des 6 méthodes testées suivant l'horizon de prévision sur la zone urbaine . . . 100
3.14 Analyse locale ($MAPE_{TimeStep}$ et $MAPE_{section}$) des erreurs de prévision . . . 101
3.15 Représentation spatiale des meilleures méthodes en fonction de la période de la journée sur la zone urbaine de Nantes . . . 103
3.16 Performances globales de prévision des méthodes de l'état de l'art testées suivant l'horizon de prévision sur le Pont de Cheviré . . . 108

4.1 Représentation du fonctionnement local autour d'une section : identification des pères et fils. . . . 118
4.2 Illustration de la modélisation locale aboutissant au concept de configuration d'état. 120

4.3 Illustration du procédé d'encodage et de distribution de configurations. 121
4.4 Représentation d'un modèle génératif probabiliste à deux niveaux. 125
4.5 Représentation d'un modèle génératif probabiliste à trois niveaux : le LDA. 128
4.6 Application du LDA au réseau de Cheviré (8 configurations et 4 périodes temporelles) 134
4.7 Évolution des performances associées aux jeux de variables produits par les $K_{lda} = 4$ motifs en fonction du nombre n de sections sélectionnées 136
4.8 Identification de 44 sections critiques (cyan) suivant la période temporelle tp et le motif latent 2 (LDA à 8 configurations). 138
4.9 Évolution des performances associées aux divers critères testés en fonction du nombre n de sections sélectionnées 143
4.10 Évolution des performances associées aux divers critères testés en fonction du nombre n de sections sélectionnées sur la zone urbain de Nantes 146
4.11 Identification de 132 sections critiques (cyan) suivant la période temporelle tp et le motif latent 2 (LDA à 8 configurations) sur la zone urbaine de Nantes. 146

5.1 Illustration de la démarche de détection et intégration des événements non récurrents au sein d'une approche globale de prévision. 156
5.2 Les $K_{lda} = 6$ thèmes latents produits lors de la phase d'apprentissage du LDA. . . 159
5.3 Représentation des distributions de perplexité en fonction du pas de temps de la journée et de la présence ou non d'incidents 160
5.4 Courbe de tendances linéaires liant la perplexité à l'erreur de prévision par MKNN 161
5.5 Illustration du procédé de sélection du paramètre a^{lda} optimal. Le nombre d'incidents détectés en fonction du a^{lda} et le nombre d'incidents réels et détectés sont conjointement représentés. 162
5.6 Représentation des distributions de perplexités en fonction du pas de temps de la journée 163
5.7 Représentation des distributions de vitesses en fonction du pas de temps de la journée 164
5.8 Comparaison des performances d'identification d'incidents distincts 168
5.9 Comparaison des méthodes à la réalité terrain à partir des diagrammes espace-temps sur la zone réduite de Cheviré 169
5.10 Représentations en fonction du pas de temps de la journée des distributions d'erreur ou de vitesse sur la section 105 170
5.11 Représentations des performances globales de prévision du MKNN suivant l'horizon de prévision et l'intégration ($MKNN_{Incident}$) ou non ($MKNN_{Usual}$) de l'historique d'incidents 175
5.12 Ciblage de l'analyse sur les performances de prévision durant les périodes recueillant des phénomènes non récurrents 176
5.13 Pourcentage de réduction du nombre d'incidents à intégrer à l'historique permanent d'incidents (à fort impact prédictif) entre prévision avec & sans historique d'incidents. 177

6.1 Transition de l'analyse d'impact à l'intégration de facteurs exogènes. 180
6.2 Camemberts illustrant l'interaction entre météorologie dégradée et phénomènes non récurrents sur le mois de septembre 2013. a) Occurence de pluies durant des événements non récurrents b) Occurence d'événements non récurrents durant les périodes pluvieuses. 180
6.3 Positionnement des 6 capteurs MétéoFrance sur le réseau du Pont de Cheviré à Nantes. 185
6.4 Camemberts illustrant la proportion de données recueillies au sein de chacune des catégories : a) pour l'intensité pluviométrique b) pour la température au sol c) pour la force des rafales de vent d) pour la visibilité croisée avec la pluie. (Période de septembre à novembre 2013) 188
6.5 Résultats d'analyse de variances (ANOVA) menée sur les 177 sections du Pont de Cheviré entre septembre et octobre 2013 189
6.6 Identification d'impacts du temps sec ou humide sur la vitesse du flux par période de la journée à partir de 55 sections du Pont de Cheviré entre septembre et octobre 2013 191
6.7 Représentation des distributions de vitesses en fonction des catégories associées aux variables météorologiques sur la section 176 (septembre-octobre 2013) . . . 192
6.8 Confrontation de journées recevant des demandes de trafic similaires, mais expérimentant des conditions météorologiques distinctes 194
6.9 Résultats d'analyses de variances menées sur l'erreur de prévision sur l'intégralité de la zone du Pont de Cheviré 196
6.10 Distributions des erreurs de prévision en fonction de l'intensité de précipitation et de l'heure de la journée 197
6.11 Illustration des performances des tentatives d'intégration de facteurs météorologiques au sein du cadre prédictif 199
6.12 Illustration des principales contributions et objets d'étude rencontrés au cours de cette thèse 205

Liste des tableaux

2.1 Valeurs optimales des hyperparamètres obtenues pour application du MKNN au réseau de Brisbane. 66
2.2 Comparaison des temps d'optimisation des hyperparamètres depuis la méthode KNN : confrontation approche locale et approche globale lors d'une analyse par grille. 67
2.3 Comparaison des temps de prévision par la méthode KNN pour 1 mois de données : confrontation de l'approche locale à l'approche globale. 67

3.1 Comparaison des temps d'apprentissage et de prévision en fonction de la méthode de prévision pour le mois de novembre 2013 (1440 évaluations) sur le Pont de Cheviré : confrontation de l'approche locale à l'approche globale. 91
3.2 Table comparative des ordres de grandeur des temps d'apprentissage et de prévision en fonction de la méthode de prévision pour le mois de novembre 2013 (1440 évaluations) sur la zone urbaine de Nantes : confrontation de l'approche locale à l'approche globale. 99

4.1 Identification des variables mises en jeu pour la construction d'un objet complexe de type texte. 124
4.2 Amélioration de la complexité temporelle liée à la sélection de variables. 137

5.1 Performance des 4 méthodes suivant les indicateurs usuels (en pourcentages). SVMn est noté pour SVM novelty. 167

6.1 Tableau croisé de proportions pour chacune des catégories météorologiques sur la période septembre - novembre 2013. 189

Table des notations

Les présentes notations sont introduites au cours de cette thèse.

Notions propres aux réseaux ou graphes :

$\mathcal{G}$	Graphe représentant un réseau
$\mathcal{V}$	Ensemble des nœuds du graphe $\mathcal{G}$
$\mathcal{E}$	Ensemble des arcs du graphe $\mathcal{G}$
N	Nombre d'arcs du graphe $\mathcal{G}$ ou nombre de sections du réseau
l	Section (ou lien) du réseau routier

Notions propres à la théorie du trafic :

Q_d	Débit
V	Vitesse
V_T	Vitesse moyenne temporelle
V_L	Vitesse moyenne spatiale ou vitesse harmonique
TT	Temps de parcours
K_d	Densité
L	Longueur d'une section

Notions générales de classification :

X	Individu à classer
Y	Label observé et associé à l'individu X
$\widehat{Y}$	Label prévu par le modèle et associé à l'individu X
f(.)	fonction de classification ou modèle de prévision : $\widehat{Y}$ = f(X)
$\mathcal{L}(.)$	Fonction de perte

Notions générales de prévision à partir de séries temporelles :

t ..... Instant ou pas de temps issu d'une série temporelle

T ..... Période temporelle indifférenciée ou nombre de pas de temps

T_{data} ..... Période temporelle ou nombre de pas de temps au sein des données

T_{hist} ..... Période temporelle ou nombre de pas de temps pour l'apprentissage

T_{test} ..... Période temporelle ou nombre de pas de temps pour le test

h ..... Horizon de prévision

N_h ..... Nombre d'horizons de prévision

f(.) ..... fonction de régression ou de prévision : $\widehat{Y_{t+h}} = f(X_t)$

$\mathcal{L}(.)$ ..... Fonction de perte

X_t ..... Individu d'entrée du modèle (contexte série temporelle)

Y_{t+h} ..... Label observé à l'horizon $t+h$ (contexte série temporelle)

$\widehat{Y_{t+h}}$ ..... Label de sortie prévu par le modèle (contexte série temporelle)

Z_t^l ..... Caractéristique de trafic mesurée à intervalle régulier sur la section l

w^X ..... Pondération des mesures Z_t^l pour construction de X_t

Ω ..... Voisinage spatial d'une section l

N_Ω ..... Nombre de voisins de la section l

μ ..... Longueur de mémoire temporelle

M ..... Période temporelle ou taille de l'historique glissant

Notions générales sur l'erreur de prévision :

$Error_t^{l,t+h}$ ..... Erreur de prévision à l'horizon h sur la section l

$AE_t^{l,t+h}$ ..... Erreur absolue de prévision à l'horizon h sur la section l

$RMSE_{section}$ ..... Erreur moyennée sur une période temporelle pour une section

$RMSE_{TimeStep}$ ..... Erreur moyennée sur les sections à un instant donné

$RMSE_{global}$ ..... Erreur moyennée sur les sections et les pas de temps

$APE_t^{l,t+h}$ ..... Pourcentage d'erreur absolue de prévision à l'horizon h sur la section l

$MAPE_{section}$ ..... Pourcentage d'erreur moyennée sur les pas de temps pour une section

$MAPE_{TimeStep}$ .... Pourcentage d'erreur moyennée sur les sections à un instant donné

$MAPE_{global}$ ..... Pourcentage d'erreur moyennée sur les sections et les pas de temps

Notions propres au procédé de lissage :

$\tilde{Z}_t^l$ ..... Valeur lissée se substituant à la valeur originale ou manquante

w^{smooth} ..... Pondération propre au lissage temporel

Notions propres aux méthodes de prévision :

► Méthode de régression KNN :

K ..... Nombre de voisins

w^{KNN} ..... Pondération propre à la méthode KNN

► Méthode de régression à noyaux (Kernel) :

$\mathcal{K}(.)$ ..... Fonction noyau ou *Kernel*

γ ..... Largeur de bande : paramètre de la fonction à noyaux gaussiens

w^{kernel} ..... Pondération résultant de la fonction à noyaux

► Méthode des Machines à Vecteurs Supports (SVM - SVR) :

ϕ**(.)** ..... Fonction de passage en haute dimension (version matricielle Φ)

W ..... Vecteur directeur de la régression (ou Matrice directrice)

b ..... Valeur ou vecteur affine

$\mathcal{K}(.)$ ..... Fonction noyau ou *Kernel*

γ ..... Largeur de bande : paramètre de la fonction à noyaux gaussiens

C ..... Coût ou pénalité d'écart à la marge (Hyperparamètre)

ϵ ..... Largeur de marge (Hyperparamètre)

ξ_k ..... Variable ressort ($k^{ième}$)

ζ_ϵ ..... ϵ-fonction de perte insensible

$\alpha_k^{\pm}$ ..... Multiplicateur lagrangien portant sur la marge (version matricielle D_α)

$\eta_k^{\pm}$ ..... Multiplicateur lagrangien portant sur la variable souple

D_{ual}**(.)** ..... Formulation duale

ρ ..... Coefficients multiplicatifs définissant W comme combinaison linéaire

► Méthode alternatives testées :

W ..... Vecteur directeur de la régression (ou Matrice directrice)

b ..... Valeur ou vecteur affine

J ..... Nombre d'arbres décisionnels construits

F_j**(.)** ..... Fonction cumulatrice de modèles simples

f_j**(.)** ..... Modèle associé à un modèle simple (arbre décisionnel ou *Gradient Tree Boosting*)

w^{RF} ..... Pondération propre aux forêts d'arbres décisionnels

w^{GTB} ..... Pondération propre au *Gradient Tree Boosting*

Notions propres à l'encodage en configurations spatiales :

F ..... Nombre de pères pour une section l

f ..... Indice associé à un des F pères potentiels

S ..... Nombre de sections Fils pour une section l

s ..... Indice associé à un des S fils potentiels

τ ..... Décalage temporel décliné pour chaque père τ_f ou fils τ_s

w^{config} ..... Pondération ou corrélation temporelle d'un père ou fils

δ_l ..... Triplet $(\delta_{Fathers}, \delta_{Link}, \delta_{Sons})$

$\delta_{Fathers}$ ..... Signal représentant la variation de vitesse sur les pères

δ_{Link} ..... Signal représentant la variation de vitesse sur la section étudiée

δ_{Sons} ..... Signal représentant la variation de vitesse sur les fils

$c_{t,l}$ ..... Configuration sur la section l à l'instant t

c ..... Configuration ou indice de configuration parmi les N_C possibles

N_C ..... Nombre de configurations

r ..... Nombre de catégories discrètes partageant la variable δ_l

Δ_l ..... Triplet discret associé à δ_l

$\Delta_{Fathers}$ ..... Catégorie discrète associée au signal $\delta_{Fathers}$

Δ_{Link} ..... Catégorie discrète associée au signal δ_{Link}

Δ_{Sons} ..... Catégorie discrète associée au signal δ_{Sons}

TP ..... Nombre de périodes temporelles

tp ..... Indice associé à une des TP périodes temporelles

$\phi_{tp,l}$ ..... Distribution de probabilité sur les N_C configurations possibles

Notions propres à l'Allocation de Dirichlet Latente :

K_{lda} ..... Nombre de motifs latents (Paramètre du LDA)

m ..... Mot

Voc ..... Vocabulaire ou ensemble de mots

N_{Voc} ..... Nombre de mots dans le vocabulaire

Q ..... Ensemble des variables latentes

q_k ..... Variable associée au $k^{ième}$ motif latent

D ..... Document ou ensemble de mots

N_d ..... Nombre de mots composants le document D

E ..... Corpus de Documents D

N_E ..... Nombre de Documents au sein du corpus C

Λ ..... Matrice de distributions sur l'ensemble des mots du vocabulaire

θ ..... Vecteur caractérisant la proportion de chacun des motifs dans D

β^{lda} ..... Paramètres de la loi de Dirichlet régulant la distribution sur les mots

α^{lda} ..... Paramètres de la loi de Dirichlet régulant la distribution sur les thèmes

Notions propres à la sélection de variables :

n ..... Nombre de sections considérées pour construire les variables d'entrée

$\hat{n}$ ..... Nombre optimal de sections n en entrée

$\hat{k}$ ..... Motif optimal

Notions propres à l'identification d'incidents :

$\iota_{l,t}$ ..... Individu de trafic à l'instant t sur la section l

$p_{l,t}$ ..... Indicateur de difficulté de classement associée aux individus de trafic

$g_{l,t}$ ..... Seuil de qualification des événements : usuel/ incidents

sw ..... Fenêtre glissante pour construction des individus de trafic

$\mathbf{T}_{typical}$ ..... Période temporelle ne recueillant aucun événement non récurrent

u ..... Classe de vitesses centrées-réduites

U_{Inf} ..... Indicateur de la classe inféreure de vitesses centrées-réduites

U_{Sup} ..... Indicateur de la classe supérieure de vitesses centrées-réduites

$\Upsilon_{l,t}$ ..... Jeu de classes de vitesses centrées-réduites pour construire $\iota_{l,t}$

T_{day} ..... Période temporelle ou nombre de pas de temps dans une journée (48)

a^{lda} ..... Paramètre fixant la borne supérieure du seuil $g_{l,t}^{lda}$

a^{mad} ..... Paramètre fixant la borne inférieure du seuil $g_{l,t}^{mad}$

ψ ..... Paramètre du $SVM^{novelty}$ pour le taux maximum de valeurs aberrantes

$\alpha^{incident}$ ..... Pourcentage de sections identifiées comme victimes d'incidents

$Bound_t$ ..... Seuil d'erreur pour identification des incidents à fort impact prédictif

$\beta^{incident}$ ..... Paramètre fixant la borne supérieure du seuil

Notions propres à la météorologie :

$\mathbf{I}$ ..... Intensité de pluie

T_{sol} ..... Température au sol

B_R ..... Force des rafales de vent

Vis ..... Visibilité

$p-value$ ..... Probabilité de trouver une valeur plus extrême que celle observée

Table des abbréviations

Les présentes abbréviations sont introduites, par ordre d'apparition, au cours de cette thèse :

GNSS	*Global Navigation Satellite System*
GPS	*Global Positioning System*
FCD	*Floating Car Data*
ITS	Systèmes de Transports Intelligents
BT	Bluetooth
LASER	*Light Amplification By Stimulated Emission of Radiation*
LAPI	Lecture Automatique de Plaques d'Immatriculation
MAC	Media Access Control
P-L	Poids-Lourds
V-L	Véhicules-Légers
D-R	Deux-Roues
CNIL	Commission Nationale de l'Informatique et des Libertés
CBD	*Central Business District*
FD	*Fundamental Diagram*
MFD	*Macroscopic Fundamental Diagram*
ARIMA	modèle prédictif : *AutoRegressive Integrated Moving Average*
ANN	*Artificial Neural Network*
KNN	*K Nearest Neighbors*
LASSO	*Least Absolute Shrinkage and Selection Operator*
MIMO	*MultiInput - MultiOutput*
ACP	*Analyse en Composantes Principales*
RMSE	*Root Mean Squared Error*
MAPE	*Mean Absolute Percentage of Error*
ACO	*Ant Colony Optimization*
E-M	*Expectation - Maximisation*
V-E-M	*Variational Expectation Maximisation*
DR	*Detection Rate*
FAR	*False Alarm Rate*
GCR	*Good Classification Rate*
ANOVA	*Analyse des Variances*

Introduction générale

Le passage à l'ère internet et le développement technologique de ces dernières décennies ont conduit à une modification profonde du comportement et des pratiques des citoyens. Ces changements ont marqué la façon d'appréhender la mobilité et les déplacements. La carte routière a laissé la place aux systèmes de navigation embarqués s'appuyant sur les GNSS (*Global Navigation Satellite System*) et, en particulier le GPS (*Global Positioning System*). De manière générale, les véhicules se distinguent aujourd'hui par la quantité de technologie embarquée (Bluetooth, Wifi).

Outre l'apport technologique, le comportement des individus s'est transformé. Depuis le début des années 2000, l'individu est pensé comme un être connecté et communiquant avec ses concitoyens (Facebook, Youtube en sont des exemples). Cette connexion intense, libérée par le développement technologique et l'engouement des usagers, a conduit à des flux conséquents d'échanges d'information. Par extension, ces technologies de communication ont progressivement pris place dans notre quotidien et émettent en continu. Chaque individu laisse alors, dans son sillage, une trace des émissions produites en se connectant.

Le recueil de telles traces, permis par ces mêmes technologies et les progrès en matière de traitement informatique, a conduit à la mise à disposition de grandes bases de données exploitables et riches en information du point de vue de la mobilité des usagers. Ces bases de données constituent aujourd'hui la clé de la bonne compréhension des comportements des utilisateurs et de leur caractérisation. Ces données sont très largement exploitées à des fins commerciales ou pour proposer de nouveaux services adaptés aux besoins de l'usager. L'évolution récente a fait rapidement basculer le recueil de données trafic dans le paradigme *Big Data*.

Les applications en matière de gestion du trafic et d'estimation des états de trafic se sont rapidement démocratisées. Google, via Waze, et d'autres bien avant lui (*e.g.* MEDIAMOBILE, TomTom) ont très bien saisi le double intérêt de ces technologies : recueillir des informations sur les individus connectés (via la 3G et les GPS par exemple) et communiquer avec les usagers en transmettant, en temps-réel, les états de trafic sur l'intégralité du réseau. Cette dualité entre recueil et diffusion d'information aboutit à un jeu gagnant - gagnant liant fournisseur et usager : plus les usagers utilisent la technologie et accèdent au service, plus la quantité

d'information disponible est conséquente et donc significative, plus l'information fournie est précise et satisfait l'usager. L'application trafic de GoogleMaps, basée sur les données de Waze, V-Traffic ou Plans d'Apple témoignent de ce type de fonctionnement régissant cette relation usager / gestionnaire.

Les objectifs actuels des gestionnaires d'infrastructures sont de parvenir à tirer le meilleur de ces technologies, des grandes bases de données disponibles et de profiter de ces nouvelles sources de données pour communiquer des informations riches. La plus-value apportée par les gestionnaires réside dans leur capacité, à partir de l'information centralisée et brute, à informer les usagers, prévenir ou anticiper les phénomènes de congestion et intervenir en conséquence. Leur but est d'atteindre une gestion suivant un optimum global du trafic et nécessite une surveillance pointue des réseaux routiers.

Profondément ancrés dans ce contexte, nos travaux de thèse résultent de trois années de recherches effectuées au LICIT sur les thématiques de caractérisation des états de trafic par fouille de données et de prévision court-terme à partir de méthodes issues de l'apprentissage statistique. La question de recherche a été d'analyser en profondeur les apports et évolutions dues à l'introduction de nouvelles technologies fondées sur les transports intelligents (ITS) pour la gestion en temps-réel du trafic. La capacité des données à restituer des composantes spatio-temporelles du trafic est évaluée pour certaines technologies, puis les nouvelles opportunités en matière de prévision sont explorées.
La présentation des travaux débute par un **chapitre introductif**, qui expose en détail le contexte, la problématique et la démarche méthodologique structurant la suite du manuscrit. Les enjeux identifiés nous conduisent alors, dans le **deuxième chapitre**, à introduire un cadre prédictif original tirant parti des nouvelles ressources pour parvenir à des modèles parcimonieux, performants et unifiés à l'échelle du réseau. Les adaptations requises des méthodes de prévision s'inspirent de procédés issus du domaine de la classification. Les modifications apportées sont alors détaillées au cours du **troisième chapitre** aboutissant à une première mise en œuvre pour des cas d'étude de trafic. Les contributions sont alors confrontées aux techniques de l'état de l'art du domaine, évaluées et caractérisées par domaine de prédilection. Un affinement du cadre prédictif est proposé dans le **chapitre quatre** pour une prise en considération des interactions entre les sections du réseau routier à partir de l'identification des sections critiques. Enfin, la présentation des travaux s'achève sur une étude de l'impact des phénomènes non récurrents du trafic sur la prévision, **chapitre cinq**, puis ouvre, dans le **chapitre six**, les perspectives de l'intégration de facteurs exogènes, à savoir les conditions météorologiques.

1

Supervision et prévision du trafic sur les réseaux routiers

> *S'il y a tant d'accidents sur les routes, c'est parce que nous avons des voitures de demain, conduites par des hommes d'aujourd'hui sur des routes d'hier.*
>
> — **Pierre-Jean Vaillard**
> Le hérisson vert

Abstract :
This chapter introduces a literature review about network monitoring from a traffic management standpoint. Contributions of ITS (Intelligent Transportation Systems) to network supervision are highlighted and the ability of recent technologies to display traffic features is illustrated from 2 case studies : Bluetooth and GPS-based FCD (Floating Car Data). Main trends in traffic forecasting process are depicted and lead to notice that the prediction problem remains usually tackled at a local scale : a link on the network. The main concern of this study lies in the development and assessment of spatially unified strategies enabling to consider network scale prediction problems with parsimony. Such strategies are called "global" approaches in opposition to usual ones, named "local" approaches. Eventually the main components structuring this PhD thesis are outlined.

Les gestionnaires d'infrastructures visent à optimiser le fonctionnement des réseaux routiers en vue de satisfaire la majorité des usagers empruntant les routes. Pour ce faire, ils doivent être en mesure d'identifier rapidement l'apparition de congestions ou d'événements non-récurrents liés à des phénomènes extérieurs, de mettre en place des stratégies de régulation ou de faire intervenir les services adéquats. En conséquence, leur tâche de surveillance des réseaux routiers est parfaitement assimilable à celle, issue du domaine informatique, dite de *network monitoring* visant à prémunir toute rupture ou attaque malveillante sur le réseau. L'étude des réseaux constitue un champ de recherche académique en soi (*network science*), trouvant ses origines avec le problème des 9 ponts de Königsberg (Euler, 1736). La mise en œuvre du maintien de qualité de service exercée par les gestionnaires se décompose suivant 3 phases :

1. la modélisation du réseau étudié,
2. la supervision des états de trafic expérimentés par le réseau,
3. l'anticipation des états de trafic.

Cette démarche de contrôle qualité du réseau conduit à une phase d'intervention, via la mise en œuvre de stratégies dites de régulation. Cette dernière étape n'est pas étudiée dans le contexte de cette thèse et constitue un domaine d'étude à part entière au sein de la littérature, notamment par le biais des modèles de simulation de trafic.

Ce premier chapitre introduit le contexte opérationnel et les enjeux scientifiques environnant la gestion des réseaux routiers. Après une présentation de la modélisation des réseaux routiers et de leur appréhension, les diverses approches de maintien de la qualité de service sur le réseau routier sont introduites. Se focalisant sur la prévision du trafic sur un réseau routier, une revue des méthodes prédictives centrées sur les données est produite pour aboutir à la problématique ceignant cette thèse. La démarche méthodologique est alors dépeinte.

1.1 Caractérisation du réseau routier

1.1.1 Définition de la notion de réseau

Un réseau désigne un ensemble d'éléments interconnectés échangeant de l'information ou de la matière. Face à son problème des 9 ponts de Königsberg, Euler est le premier à avoir formalisé un réseau par un ensemble d'éléments discrets appelés **nœuds**, connectés les uns aux autres par des liens nommés **arcs**. Aujourd'hui, tout réseau est représenté par un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E}) \equiv \{\mathcal{V}_i, \mathcal{E}_i\}_{i \in [1, Card(\mathcal{V})]}$ composé de $Card(\mathcal{V})$ nœuds reliés par $N = \sum_{i=1}^{Card(\mathcal{V})} Card(\mathcal{E}_i)$ arcs, où N est borné par $Card(\mathcal{V})^2$ et $\mathcal{E}_i$ l'ensemble des arcs pointés vers le nœud $\mathcal{V}_i$. On parle de nœuds **adjacents**, si deux nœuds sont reliés par un arc.

La grande majorité des recherches actuelles portent sur les réseaux qualifiés de **complexes**. Ces derniers vérifient des propriétés d'irrégularités, de fonctionnement complexe et d'évolution

temporelle dynamique marquant leurs principaux enjeux. Bien que ces réseaux possèdent des caractéristiques communes, les réseaux complexes étudiés et modélisés se regroupent, d'après Newman [1], en quatre grandes familles :

- **Les réseaux sociaux**, qui relient un ensemble de personnes ou groupe de personnes les uns aux autres. Facebook, LinkedIn, les relations entre entreprises en sont les principaux exemples.
- **Les réseaux d'information** visant à connecter des informations les unes aux autres. Internet en est un parfait exemple.
- **Les réseaux biologiques**, qui servent à décrire des connexions physiques entre des entités biologiques. Exemple : réseau de neurones.
- **Les réseaux technologiques**, bâtis par l'homme en vue de transporter des biens, des personnes ou des informations. Exemple : le réseau électrique, internet ou le réseau routier.

La grande majorité de ces réseaux vérifient également les propriétés d'**invariance à l'échelle** (*scale free*) et du **petit monde** (*small world*). La propriété d'indépendance à l'échelle signifie qu'une petite quantité de nœuds se concentrent sur la grande majorité des liens. La propriété du petit monde soutient l'hypothèse que la distance entre deux nœuds croît de façon logarithmique avec la taille du réseau. Tout couple de nœuds est donc relié par un chemin mettant en œuvre un faible nombre d'arcs. Il a été montré, par exemple, que tout individu inscrit sur Facebook était, en moyenne, relié à un autre par environ 4.5 - 6 relations [2]. Ces propriétés ont conduit à une meilleure compréhension des réseaux étudiés et à d'importantes recherches sur la résilience des réseaux [3] et leur croissance [4]. L'étude de tels réseaux se nomme la science des réseaux (*network science*) et compte parmi les thématiques les plus actives du domaine.

Vérifiant l'ensemble des propriétés précédemment décrites [5], le réseau routier prend pleinement part à l'ensemble des réseaux nous environnant et, par extension, à la science associée. De par son enchâssement dans un espace muni de distances, le réseau routier appartient à la sous-classe des réseaux spatiaux, dont la particularité est d'introduire des notions physiques de propagation de l'information. Un tel positionnement au sein de la science des réseaux est assimilé dans toute étude de trafic routier. Comme l'illustre le problème des Ponts d'Euler, l'étude des réseaux routiers est même antérieure à l'apparition d'une telle science et certaines notions sont clairement établies depuis les années 60 [6].

1.1.2 Définition opérationnelle

Du point de vue opérationnel, le réseau routier est l'ensemble des infrastructures routières permettant l'écoulement de flux de véhicules. Dès lors, il peut être modélisé par le contenant comme par le contenu :

- le contenant est la **structure** physique du réseau routier, *i.e.* les infrastructures routières elles-mêmes. Elles déterminent les contraintes physiques portant sur l'écoulement du

trafic et sont hiérarchisées à diverses échelles (autoroute, nationale, départementale par exemple).

– le contenu caractérise **la dynamique** de fonctionnement du réseau routier, *i.e.* le flux de véhicules sillonnant le réseau physique. Il est assimilable au **transfert d'une information** se propageant à travers le contenant. Il s'agit véritablement d'un témoin de fonctionnement du réseau et ses caractéristiques sont associées à celles du réseau routier. Sans en changer la structure, les contraintes du contenant peuvent évoluer, dans le temps, ce qui modifie inexorablement le contenu. La fermeture de voies ou l'apparition d'accidents en constituent des exemples parlants.

Cette définition duale **structure-caractéristique dynamique** est commune à la plupart des réseaux, bien que la structure ne soit pas systématiquement physique.
La représentation graphique d'un réseau routier est usuellement basée sur sa **structure**, tandis que les caractéristiques **dynamiques** sont des informations portées sur les arcs et les nœuds. Le graphe $(\mathcal{V}, \mathcal{E})$ du réseau routier associe à chaque section de route un arc et un nœud à chacune de leurs extrémités. Les nœuds sont généralement les intersections ou carrefours routiers. Le réseau étant complètement connecté, le graphe construit est connexe : il est possible d'atteindre n'importe quel nœud à partir d'un autre.

Dans le contexte opérationnel, le gestionnaire d'infrastructures travaille avec une structure donnée du réseau routier. Aucune nouvelle route ne peut être construite ou ajoutée au réseau dans les horizons temporels considérés. Les études prospectives à long terme (souvent 20 ans) visent à analyser le bien-fondé de la construction de nouvelles infrastructures ou l'extension de celles existantes en évaluant l'évolution des déplacements et besoins en mobilité des usagers. Cet aspect est hors champs de cette thèse et des intérêts à court-terme des gestionnaires routiers, dans la mesure où leurs objectifs portent sur le fonctionnement dynamique du réseau appréhendé via les caractéristiques du trafic.

1.1.3 Caractéristiques de trafic

La dynamique des flux de véhicules se propageant sur le réseau routier constitue le véritable objet d'étude des gestionnaires d'infrastructures. Cette dynamique est appréhendée en diverses localisations du réseau via des caractéristiques du trafic bien définies dans le domaine. S'intéressant au flux de trafic et non aux véhicules (particules microscopiques) le composant, ces caractéristiques du trafic sont qualifiées de macroscopiques. Les états de trafic sont usuellement déterminés par quatre caractéristiques macroscopiques :

– **Le débit (Q_d)** définit le nombre de véhicules franchissant un point du réseau durant une période spécifique T. Le débit est exprimé en véhicules par heure (veh/h). Il caractérise l'intensité du flux traversant une portion du réseau durant une période spécifique T.

– **La vitesse (V)** caractérise la vélocité du flux en mouvement sur le réseau en kilomètres par heure (km/h) et distingue 2 définitions :

- la vitesse moyenne des véhicules franchissant une position fixe du réseau. On parle alors de **vitesse moyenne temporelle**. Elle est définie par $V_T(x) = \frac{1}{N_T} \sum_{i=1}^{N_T} v_i(x)$, où N_T est le nombre de véhicules ayant franchi la position durant une période spécifique T.
- la vitesse moyenne des véhicules à un instant t sur une portion de route de longueur L spécifiée. Il s'agit de la **vitesse moyenne spatiale**. Elle est définie par $V_L(t) = \frac{1}{N_L} \sum_{i=1}^{N_L} v_i(t)$, où N_L est le nombre de véhicules au sein de la zone de longueur L.
- la **vitesse du flot** de véhicules, définie de façon équivalente par la vitesse moyenne spatiale des véhicules ou la vitesse moyenne temporelle harmonique, exprimée par $\frac{1}{V_T^{harmonique}} = \frac{1}{N_T} \sum_{i=1}^{N_T} \frac{1}{v_i}$.

- **Le temps de parcours (TT)** définit le temps nécessaire par un véhicule du flux pour rallier deux points distincts du réseau. Le temps de parcours individuel est défini pour tout véhicule par la différence des temps de passage entre les points A et B du réseau : $TT = t_B - t_A$. Pour une section l de longueur L du réseau, le temps de parcours moyen à un instant donné est fourni par : $TT_l(t) = \frac{1}{N_L} \sum_{i=1}^{N_L} TT_i(t)$, où N_L est le nombre de véhicules sur la section de longueur L. La division de cette expression par L aboutit directement à l'expression de la vitesse harmonique.
- **La densité (K_d)** ou concentration définit le nombre de véhicules sur une longueur spécifique de route. Cette densité est caractérisée par la relation : $K_d = \frac{Q_d}{V_L}$ le liant au débit et à la vitesse moyenne spatiale. En pratique, la densité K_d n'est pas directement mesurable. Seule la mesure du taux d'occupation TO en un point du réseau permet de l'estimer.

Ces caractéristiques du flux de trafic sont liées par une relation phénoménologique, identifiée par Greenshields dès 1934 [7] et appelée **diagramme fondamental**. Adoptant diverses formes (Figure 1.1 b)), ces modèles traduisent certains liens logiques afférents au trafic. La relation linéaire liant, en régime fluide, la densité au débit, due à la faible variabilité de la vitesse moyenne spatiale de ce régime, en est un exemple.
Un outil graphique permet de reproduire la dynamique d'une portion de route (longueur L) sur une période donnée T à partir des trajectoires des véhicules. Il s'agit du **diagramme espace-temps** illustré en Figure 1.1 a).

1.2 Supervision du réseau routier

La supervision du réseau routier consiste en l'observation des caractéristiques définies en section 1.1.3 sur chacun des arcs du réseau routier considéré. L'objectif du gestionnaire est de disposer du maximum d'informations concernant l'état de fonctionnement du réseau et de les centraliser pour interprétation. Cependant, ces variables de trafic ne peuvent être connues en tout point et nécessitent des mesures par capteurs. Dans cette perspective, diverses technologies

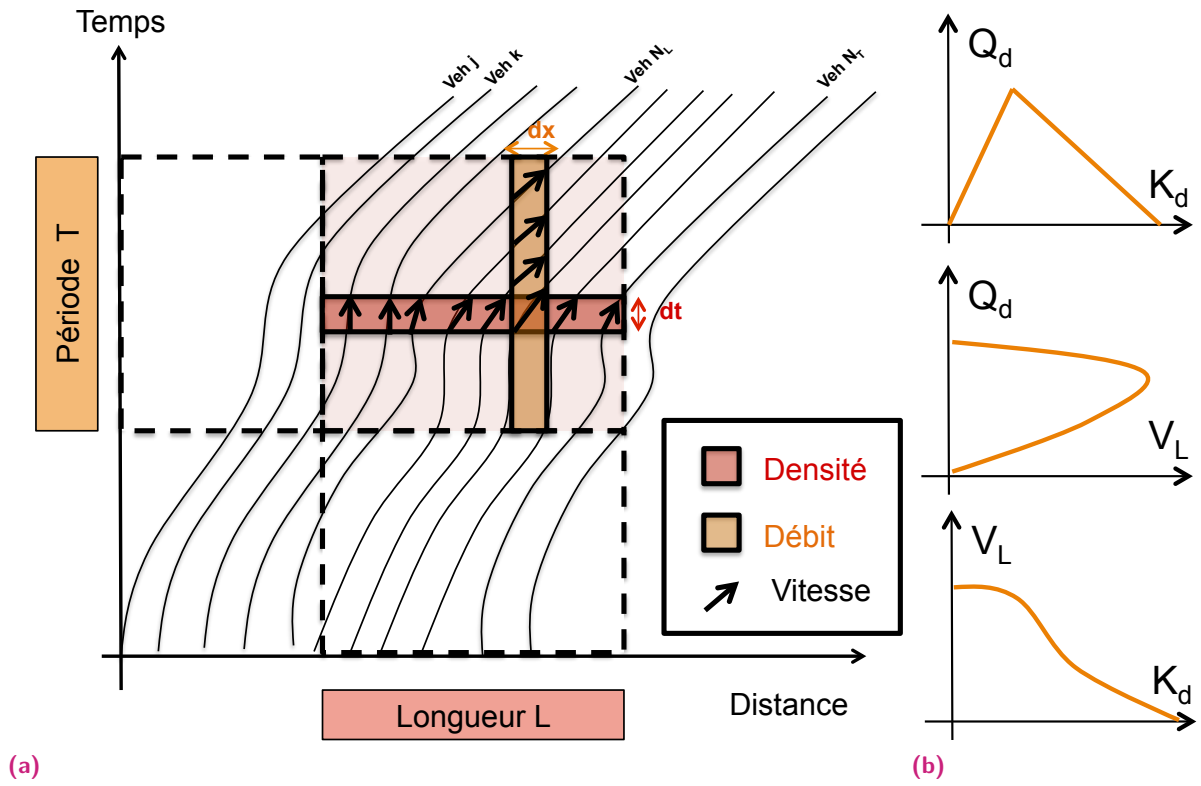


Fig. 1.1.: Représentations classiques des variables macroscopiques du trafic : a) Représentation spatio-temporelle du flux de véhicules (Diagramme Espace-Temps) b) Diagrammes Fondamentaux liant les variables macroscopiques du trafic, d'après le modèle de Greenshield.

ont vu le jour. A l'initiative des gestionnaires, elles peuplent, aujourd'hui, les infrastructures routières et spécifiquement les voies rapides et autoroutes [1].

1.2.1 Différentes méthodes de recueil

Trois grandes typologies de recueil de données se distinguent de par leur procédé d'installation et de fonctionnement : les capteurs liés à la seule infrastructure, les capteurs liés à l'infrastructure et au véhicule et les capteurs liés au véhicule.

1.2.1.1 Méthodes de recueil liées à l'infrastructure

Cette première typologie de recueil est également la plus ancienne, puisque ces procédés trouvent leur origine dans les années 60-70. Elle ne nécessite aucune technologie embarquée dans le véhicule. Toute l'information est tirée de la seule présence de véhicules (tous types : véhicule léger, poids-lourds ou deux-roues *etc*) sur l'infrastructure routière. Ces technologies

1. Pour de plus amples informations concernant les typologies de capteurs, le lecteur pourra consulter la source internet : `http://www.transport-intelligent.net/technologies/capteurs-77/#article419`

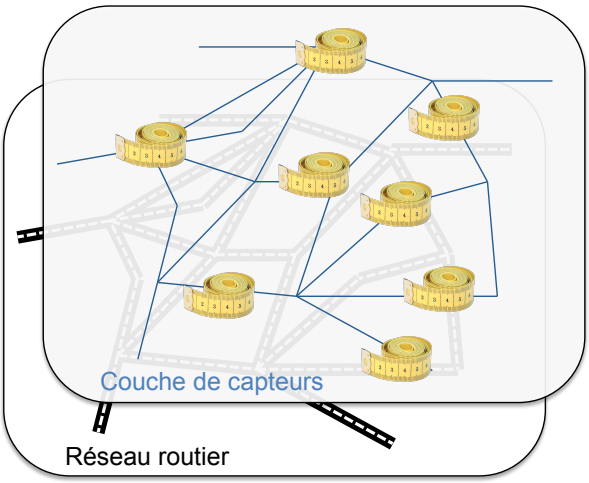


Fig. 1.2.: La supervision du réseau nécessite une couche de capteurs positionnés sur le réseau.

de recueil se partagent en deux sous-catégories : les capteurs intrusifs et les capteurs non-intrusifs.

1.2.1.1.1 Capteurs intrusifs

Les capteurs intrusifs regroupent l'ensemble des capteurs opérant des mesures très localisées et nécessitant une insertion au sein de la chaussée.

Concept de détection

Le capteur relève en permanence le profil d'une onde qu'il émet ou propage en continu au sein de la chaussée. Le signal prend diverses formes : électrique pour les boucles à induction, magnétique pour les magnétomètres (Sensys), mécano-électrique pour les capteurs à effet piézo-électrique et lumineux pour ceux à fibre optique.
Le principe de détection repose sur la modification du signal, voire sa transformation (de mécanique à électrique par exemple), lors du franchissement du capteur par un véhicule empruntant l'infrastructure. L'origine d'une telle altération du signal repose sur des fondements physiques tels que l'induction d'un champ électromagnétique lors du survol d'une boucle alimentée d'un faible courant ou la rupture d'un champ lumineux. À tout véhicule, est alors associée une perturbation du signal ou silhouette, qui est analysée et interprétée par une fonction de seuillage ou un petit ordinateur. Le décompte des véhicules est opéré en continu, tandis que leur silhouette fournit des informations sur leurs typologies (deux-roues, poids-lourds), la durée de passage *etc*.
Le plus connu et le plus répandu des capteurs vérifiant ces propriétés est la boucle inductive ou électromagnétique. Cette dernière consiste en une boucle cuivrée insérée à 7 cm sous la dalle de chaussée et dont les bornes sont alimentées par un signal à faible ampérage. Cette tension induit un champ magnétique rayonnant autour de la boucle et au-dessus de la chaussée. Lors de tout passage d'un véhicule, l'essieu métallique en mouvement induit un nouveau champ magnétique s'ajoutant au précédent et altérant la fréquence du signal électrique de la boucle.

Ce décalage de fréquence est traduit en tension et seuillé par une boucle de verrouillage de phase (filtre).

Mesures disponibles

Les capteurs opèrent des mesures spatialement ponctuelles, mais permanentes des états de trafic. Étant intrusives, les mesures manquantes sont rares, voire négligeables. Les caractéristiques macroscopiques du trafic (décrites en section 1.1.3) sont toutes mesurées, éventuellement en itérant deux capteurs à courte distance pour estimer la vitesse individuelle (boucles doubles). Outre les caractéristiques macrosocopiques usuelles du trafic, ces capteurs sont aptes à mesurer les comportements des individus, tels que le temps et la distance inter-véhiculaire moyens, ou les propriétés du véhicule (deux-roues, véhicule particulier *etc*).
Suivant les phénomènes physiques mis en jeu, des éléments extérieurs à la circulation routière sont susceptibles de perturber la sensibilité des capteurs ou d'induire de fausses détections. Les véhicules à faibles vitesses s'avèrent difficiles à détecter, puisque les altérations du signal sont également faibles. Leur principal défaut réside dans le coût d'installation et d'entretien. L'emprise au sol s'avère parfois problématique, d'autant que les vols de matériels sont à déplorer régulièrement (cuivre ou autres matériaux précieux).

1.2.1.1.2 Capteurs non intrusifs

Les capteurs non intrusifs regroupent l'ensemble des capteurs installés sur l'infrastructure, mais ne nécessitant pas de travaux de génie civil conséquents. Il s'agit une nouvelle fois de capteurs ponctuels.

Concept de détection

Le capteur est installé en accotement, sur des infrastructures extérieures à la chaussée telles que tout ouvrage d'art bordant la route ou antenne spécifiquement bâtie pour l'occasion. Ces procédés de recueil se partagent en deux sous-catégories :

- **les capteurs à réflexion d'ondes.** Que ce soient les capteurs à hyperfréquences utilisant l'effet Doppler ou à deux antennes, les capteurs LASER ou infrarouges, le procédé d'identification reste identique. Le capteur émet à fréquence régulière une onde électromagnétique. Les véhicules franchissant le champ d'émission réfléchissent l'onde qui est captée par un récepteur pouvant être confondu avec le capteur émetteur. Par effet Doppler, la vitesse des véhicules est connue et les véhicules sont décomptés.
- **les capteurs vidéo.** Le procédé d'identification se base sur des images du spectre visible ou infrarouge. Le principe repose sur des algorithmes de reconnaissance d'images : les véhicules se déplaçant sur l'infrastructure sont perçus par des groupes de pixels mobiles sur le fond fixe composé de l'infrastructure et du paysage environnant. Chaque groupe de pixels franchissant une zone spécifique de l'image de la caméra est comptabilisé. Les meilleurs capteurs sont en mesure de reconnaître les plaques d'immatriculation des usagers.

L'exemple le plus répandu et connu reste celui des capteurs à Lecture Automatique de Plaques d'Immatriculation (LAPI). En reconnaissant les plaques personnelles d'immatriculation, ils

offrent la possibilité de suivre les véhicules traversant un ensemble d'infrastructures et de déduire certains itinéraires.

Mesures disponibles et limites

Ces capteurs recueillent tous au moins les débits et vitesses des flux de véhicules. Certains sont en mesure de distinguer les véhicules légers des poids-lourds en caractérisant leur silhouette. Les véhicules proches les uns des autres, les faibles vitesses ou les véhicules à forte accélération sont autant de paramètres influençant la qualité des recueils de données. Pour les capteurs vidéo, la sensibilité aux conditions météorologiques dégradées ou aux luminosités intenses est également à regretter. Les limites de ces technologies conduisent à un taux de détection (ou taux de pénétration du flux de trafic) plus restreint que les procédés intrusifs, mais ils restent légèrement moins coûteux. On estime que 70-90% du trafic est recueilli par les LAPIs. Le taux d'appariement (*i.e.* de détections successives) est en revanche nettement plus faible.

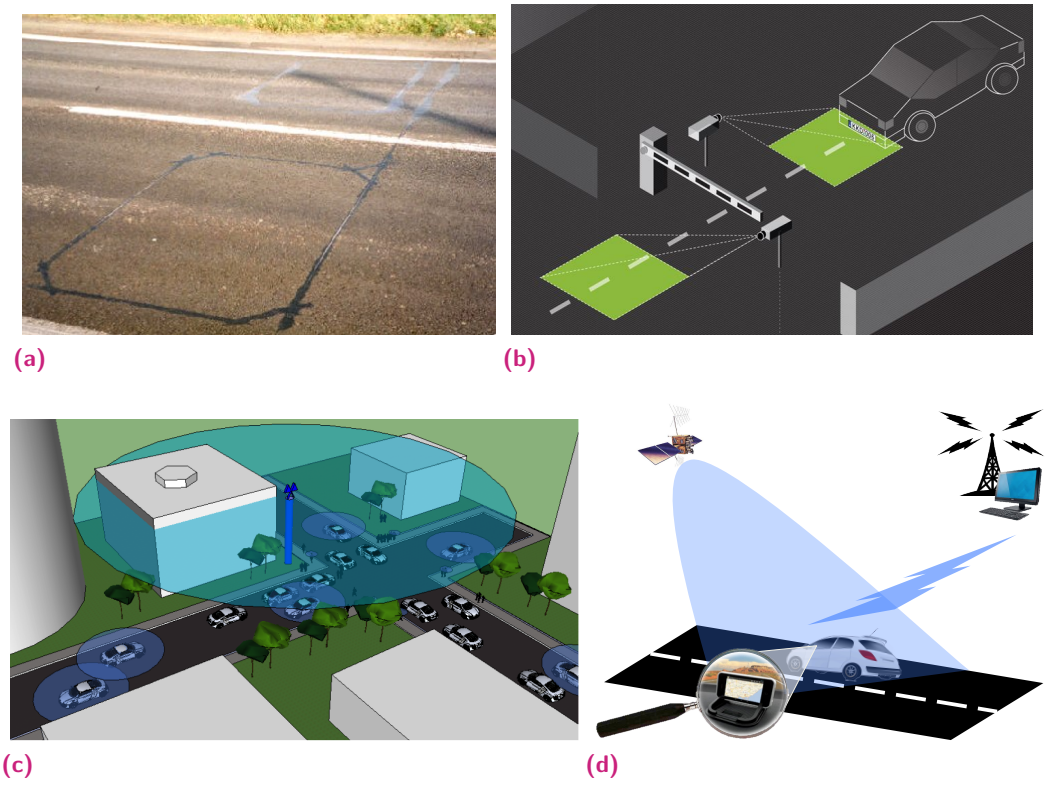


Fig. 1.3.: Illustration des typologies de recueil de données : a)Détection par capteur intrusif de type boucle électromagnétique b) Détection par capteur non intrusif : LAPI (source : CamShop.fr) c) Détection par capteur Bluetooth d) Suivi de traces de véhicules par GPS.

1.2.1.2 Méthodes de recueil liées à l'infrastructure et embarquées

Cette typologie de recueil nécessite l'emploi simultané de capteurs aussi bien sur l'infrastructure qu'embarqué dans les véhicules. Les technologies concernées sont essentiellement le Bluetooth et le Wifi.

1.2.1.2.1 Concept de détection

Le principe de détection [2] est illustré en Figure 1.3 c). Des récepteurs Bluetooth ou Wifi sont installés en bordure de route aux points clés du réseau et balaient à intervalle régulier une zone de détection. Le rayon de détection est propre à la catégorie de capteurs employés et atteint environ 100m pour la plupart des cas observés. Tout émetteur Bluetooth ou Wifi enclenché et embarqué à bord du véhicule est alors détecté et horodaté par le récepteur, puis stocké et centralisé dans une grande base de données. Les appareils émetteurs adoptent diverses formes, depuis l'autoradio jusqu'aux téléphones personnels. Chacun des appareils émetteurs possède une adresse MAC (Media Access Control) unique recueillie par le capteur et associée au temps de passage. Tout capteur en bord de route possède un unique identifiant, connu du gestionnaire. Ce dernier est en mesure, à partir du jeu de capteurs franchis par un émetteur sur le réseau, de reconstituer l'itinéraire réellement emprunté sur l'infrastructure par le véhicule.

1.2.1.2.2 Mesures disponibles et limites

Chacun des capteurs Bluetooth/Wifi est apte à décompter les véhicules équipés franchissant leur zone de détection. Le temps de parcours entre 2 capteurs est déduit de la différence des temps de passages à ces capteurs. La vitesse du flot et un débit sont ainsi mesurés. En outre, les itinéraires des usagers sont estimables à partir de leur suivi à travers le réseau de capteurs.
De même que pour les capteurs non intrusifs, le débit mesuré est partiel, puisqu'il dépend du taux d'équipement des véhicules circulant sur le réseau. Ce taux de pénétration est estimé à 15-20% pour la technologie Bluetooth sur des zones urbanisées, dont la flotte de véhicules est récente (Brisbane, Australie). Un tel échantillonnage de véhicules est suffisant pour aboutir à des vitesses du flot satisfaisantes. Comme nous l'avons montré dans un travail précédent [8], le taux de pénétration tend à varier dynamiquement, ce qui ne facilite pas l'estimation du débit réel. La thèse de Michau [9] développe un procédé de reconstruction des itinéraires aboutissant entre autre à l'estimation des débits et autres caractéristiques en tout point du réseau. Ces technologies présentent l'avantage indéniable d'être très bon marché avec une occupation de l'infrastructure routière extrêmement faible.

1.2.1.3 Méthodes de recueil embarquées

Cette typologie de recueil ne dépend que des équipements embarqués de l'usager. On parle de *crowd-sourcing* ou "approvisionnement par la foule".

1.2.1.3.1 Concept de détection

Les véhicules équipés d'instruments permettant leur géolocalisation à intervalle régulier sont suivis durant leur itinéraire et les informations relevées par le navigateur sont transmises en temps réel au gestionnaire d'infrastructure ou simplement centralisées. Le relevé d'informations, transmis en temps réel par une carte SIM (*Subscriber Identity Module*), consiste en l'interpré-

2. Pour de plus amples informations concernant le fonctionnement de la technologie Bluetooth et ses apports pour caractériser le trafic, consulter le rapport de Master de P.A. Laharotte

tation, par un navigateur, des signaux transmis par des objets de l'environnement (satellites *etc*). Les instruments embarqués se composent aussi bien de systèmes de navigation (GNSS abusivement appelés GPS) que de téléphones mobiles via le module de localisation interne et le service de données 3G. La Figure 1.3 d) illustre le procédé de recueil et de stockage des traces horodatées et géolocalisées associées à ces véhicules qualifiés de traceurs (*Probe Data*). Les données recueillies sont également nommées *Floating Car Data* (FCD), puisqu'une flotte de véhicules équipés est suivie.

1.2.1.3.2 Mesures disponibles et limites

À tout instant, la vitesse instantanée des véhicules est connue par mesure de l'effet Doppler lors de la géolocalisation. Contrairement aux méthodes introduites précédemment, il n'existe pas de capteur de mesure de l'état de trafic et donc pas de localisation de la mesure. Afin de donner du sens aux traces recueillies et aux mesures de vitesse effectuées, une phase dite de *map-matching* est nécessaire. Elle consiste à associer toute mesure à une section du réseau routier. Cette tâche incombe généralement aux opérateurs d'information routière et consiste en une projection des points de mesure sur les sections prédéfinies du réseau routier. Les caractéristiques macroscopiques du trafic sont alors déduites en procédant à des moyennes section par section sur de courtes périodes de temps. La vitesse du flot et le débit sont déterminés à pas de temps régulier. S'intéressant à un échantillon d'usagers équipés, le débit obtenu est partiel et sert d'indicateur de qualité de la mesure de vitesse. Cette ressource s'avère spécifiquement bien adaptée pour déterminer sur toute section du réseau, les vitesses du flot expérimentées. Se basant sur le matériel propre aux usagers, son coût d'installation et d'entretien des capteurs est nul.

1.2.2 Enjeux de la supervision du trafic

La superposition d'un réseau de capteurs au réseau routier permet d'approcher les états de trafic en tout point, mais n'est pas exempte d'imprécisions, voire d'erreurs. Trois enjeux bien identifiables sont sources de difficultés pour le gestionnaire routier lors de la phase de supervision :

– l'**adéquation entre les sections du réseau routier et la localisation des capteurs de trafic**. Une difficulté latente de la supervision réside dans la couverture du réseau par les capteurs. Une faible quantité de capteurs et/ou leur mauvais positionnement entraînent inexorablement des ruptures de connexité de l'information sur le graphe du réseau. Ces contraintes ont longtemps restreint l'étude du réseau routier à des approches très localisées sur des sections isolées ou des corridors bien définis. L'emploi des seuls procédés de relevés historiques (liés à l'infrastructure seule) favorise de telles difficultés. L'insertion des capteurs et/ou de la technologie employée est coûteux, ce qui restreint leur couverture du réseau. Ils requièrent souvent une paire de capteurs (mesure de la vitesse à partir de boucles électromagnétiques) et fournissent une information ponctuelle. De plus, ces capteurs sont caractérisés par le coût important des matériaux et les vols

fréquents. L'apparition des typologies de recueil alternatives et la dématérialisation des capteurs limitent ces désagréments. Une meilleure couverture du réseau est obtenue, certes, à l'heure actuelle, au détriment de la quantité de véhicules détectés.

– la **gestion des mesures aberrantes ou absentes** [3]. Des capteurs défectueux, déréglés ou même des phénomènes physiques parasites sont autant de causes probables de détections erronées ou absentes.
  Des procédés de filtrage et de complétion des données bruitées ont vu le jour au cours de ces dernières décennies. Le temps de parcours ou les mesures de vitesses individuelles sont les composantes du trafic propices à de tels filtrages. Les procédés reposent sur l'analyse des distributions de vitesses individuelles. Algorithmes à écart-type glissant [10], seuils de vitesses [11, 12], comparaison à des distributions usuelles lissées [13] et modélisation des distributions sont autant de procédés développés dans la littérature.
  L'estimation et assimilation de données est également un procédé observé et applicable à toute caractéristique du trafic. Le filtre de Kalman en est l'exemple le plus commun [14, 15, 16, 17]. Cet estimateur récursif repose sur un modèle prédictif s'appuyant sur des hypothèses de stationnarité du flux de trafic. Pratiques lors de données particulièrement bruitées ou absentes, les estimations seront erronées face à l'apparition d'incidents.

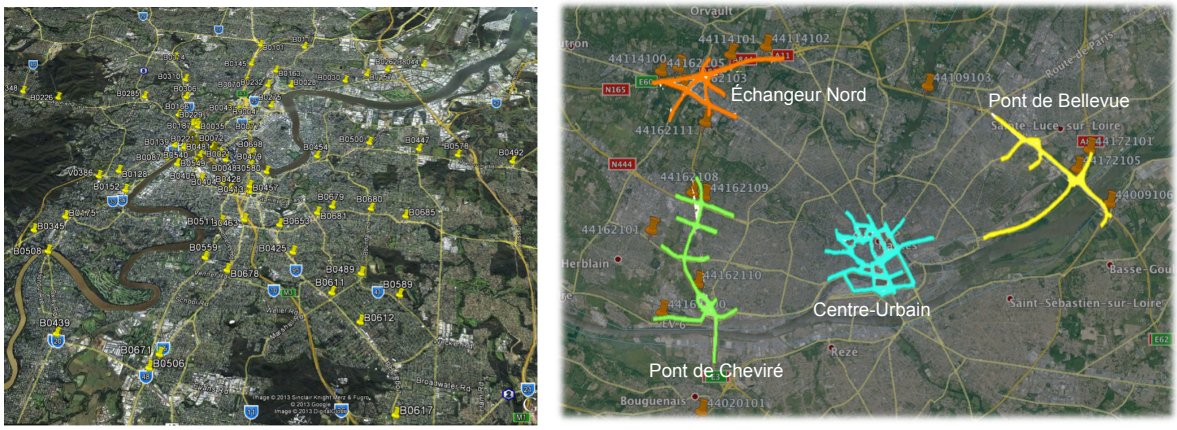
– le **temps d'estimation des états de trafic**. Suivant la technologie employée et des efforts requis pour éliminer le bruit des données, les caractéristiques de trafic peuvent être disponibles plus ou moins rapidement. Les interventions en temps réel des gestionnaires requièrent un temps d'estimation extrêmement faible pour l'ensemble du réseau traité.

1.2.3 Cas étudiés

Certaines technologies récentes apportent des solutions à la problématique de l'adéquation entre la structure du réseau routier et le positionnement des capteurs. C'est notamment le cas du Bluetooth et des données issues de GPS. Bien que l'information recueillie par ces technologies soit partielle (échantillonnage des véhicules détectables), l'information vitesse du flot est connue en tout point du réseau.
Les capteurs Bluetooth présentent peu d'emprise sur l'infrastructure routière et sont très peu coûteux, dans la mesure où ils peuvent être construits en interne (100$/capteur à Brisbane, Australie). Leur déploiement sur un réseau est envisageable à moindre coût et faible échelle temporelle. Le Brisbane City Council, en charge du réseau routier de cette ville australienne, a installé, en moins de quatre ans, près de 900 capteurs. Un tel réseau de recueil de données est unique au monde.
Les données FCD recueillent les traces laissées par les usagers sur l'infrastructure de transport. Les caractéristiques du réseau peuvent dès lors être approximées en tout point du réseau et à tout moment.
Compte tenu de leurs avantages indéniables en termes de couverture d'un réseau et de leurs

3. Cette problématique ayant d'ores et déjà fait l'objet de multiples études, aucune contribution ne sera présentée au cours de cette thèse.

potentialités pour le *monitoring* temps-réel du réseau, ces deux technologies ont été retenues pour servir de cas d'études par la suite. Deux réseaux aux typologies bien distinctes ont été traités : le réseau semi-urbain de Brisbane (Australie) et différents sous-réseaux de Nantes (France).



(a) (b)

Fig. 1.4.: Illustration des deux réseaux étudiés : a) Positionnement des capteurs Bluetooth sur le réseau de Brisbane b) Modélisation graphique des zones d'étude nantaises (données FCD de MEDIAMOBILE).

1.2.3.1 Réseau de Brisbane (Australie) et technologie Bluetooth

Depuis 2011, le *Brisbane City Council* de la ville de Brisbane s'est engagé dans le déploiement de capteurs Bluetooth sur les principales intersections de la ville australienne. Les informations recueillies par un jeu de ces capteurs sont exploités pour caractériser le trafic à Brisbane.

1.2.3.1.1 Descriptif du cadre d'étude

La zone d'étude considérée se compose du centre urbain et d'affaires de la ville de Brisbane et des principaux axes d'accès. La couverture du réseau par les capteurs Bluetooth est illustrée en Figure 1.4 a). Bien qu'aujourd'hui, près de 900 capteurs soient disposés à travers la ville de Brisbane, seuls 77 étaient disponibles en 2012. Ces derniers sont situés sur les principaux axes d'accès au centre-urbain et sur les artères urbaines les plus fréquentées. La période de recueil de données est de trois mois : janvier et février 2012, puis octobre 2012.

1.2.3.1.2 Traitements appliqués aux données

En accord avec le procédé de détection introduit en section 1.2.1, tout véhicule équipé voit potentiellement son identifiant MAC (Media Access Control) recueilli par le capteur traversé. L'information est stockée, horodatée et centralisée. A chaque identifiant MAC est alors associé un jeu de détections composé de l'identifiant unique du capteur (localisation) et de la date de

passage. La différence entre deux détections successives d'un usager doit alors être associée au parcours d'une section ou arc du réseau et la différence de temps de passage à son temps de parcours effectif. Le débit moyen d'une section est estimé en décomptant le nombre d'usagers détectés pour le couple de détecteurs associé.
Le principe de détection du Bluetooth, mais surtout sa mise en pratique révèlent diverses problématiques illustrées en Figure 1.5 :

- identification de valeurs aberrantes ou bruitées, propres à tout procédé de recueil. Cette phase est d'autant plus importante que trois sources de valeurs aberrantes sont identifiées :
  - Les jeux de détections sont recueillis en continu, sans distinction en itinéraires.
  - Aucune distinction n'est faite entre les modes de déplacement durant la phase de détection. Tout usager équipé de la technologie Bluetooth est recueilli : piétons, cyclistes et deux-roues ou automobiles sont recensés sans distinction ou indication particulière.
  - Les rayons de détection des scanners Bluetooth balaient à fréquence régulière, mais leur rayon effectif de détection varie en conséquence. Les temps de parcours sont donc légèrement bruités.
- identification et complétion des itinéraires empruntés. Le jeu de capteurs rencontrés peut s'avérer incomplet, du fait de détections manquées. Une incertitude sur les chemins empruntés est alors introduite, l'estimation du temps de parcours n'a pas de sens et aucun arc n'est associé au set de détecteurs successifs. Nous avons proposé dans [8] une stratégie de réallocation basée sur l'imbrication d'algorithmes du plus court chemin. De même, Michau [9] ont suggéré une méthodologie de reconstitution des jeux de détecteurs véritablement rencontrés.

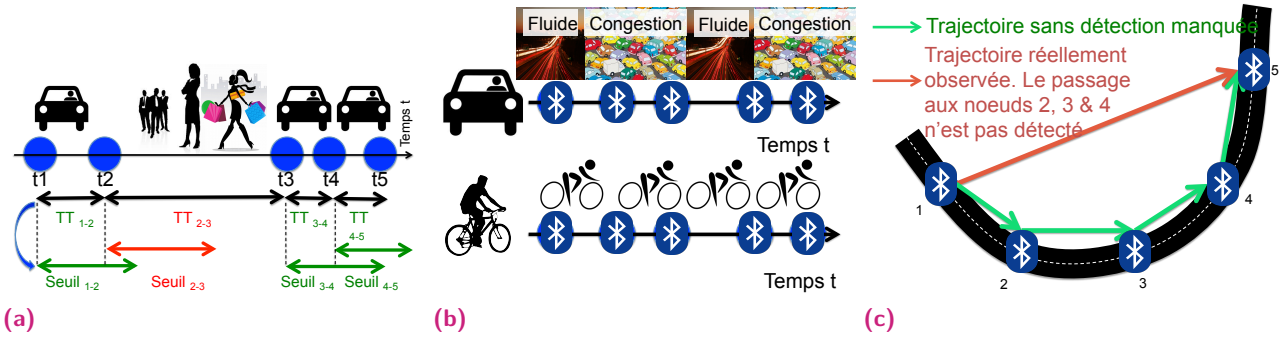


Fig. 1.5.: Illustration des difficultés rencontrées à partir des données Bluetooth : a) Scission en itinéraires du jeu de détecteurs franchi par un usager b) Distinction des modes de transports c) La problématique des détections manquées.

Filtrage des données

Plus qu'à l'exploitation des données Bluetooth, la plupart des études publiées s'intéressent à leur significativité et à la nécessité ou non de les filtrer. La polémique étant due à la partialité des données recueillies et à l'adoption de méthodes de filtrage usuelles et adaptées aux capteurs intrusifs [13, 18, 10] ou de méthodes plus douces [12, 11, 19]. L'enjeu réside dans la mise

en place d'un filtrage sélectif et dans la significativité de la quantité de valeurs restantes. Les méthodes basées sur des seuils de vitesses [12, 11] ou de classification non supervisée [19, 20] sont généralement adéquates, car elles limitent l'élimination de valeurs pertinentes.
Le procédé de filtrage appliqué repose lui aussi sur des seuils de vitesses et vise à pallier deux des trois sources de valeurs aberrantes : la scission des ensembles de détections en itinéraires et l'identification des typologies d'usagers détectés. La méthodologie est dépeinte dans la Figure 1.6 et procède en deux étapes :

- **Phase de partage en itinéraires** du jeu de détections expérimenté par un usager. Après analyse des distributions de vitesse section par section, une tendance est clairement identifiable : en deçà de 1m/s (3,6 km/h), la densité de vitesses tend anormalement à augmenter. Ces trajets reliant 2 détecteurs à très faible vitesse ne correspondent pas à des trajets véritablement expérimentés par l'usager, mais à une sortie et une entrée sur le réseau. Toute vitesse inférieure à 1 m/s est éliminée.
- **Phase d'identification des typologies d'usagers.** Il est supposé que les usagers motorisés d'une section de route sont soumis aux conditions de trafic. L'analyse des distributions de vitesses quotidiennes (voir Figure 1.6) montre que les véhicules motorisés observent globalement des chutes de vitesse lors des périodes d'affluence, ce qui se traduit par des embouteillages. Néanmoins, les véhicules non motorisés (cyclistes et piétons) ne sont pas soumis à de telles contraintes. L'identification d'une distribution particulière de vitesses indépendantes des congestions quotidiennes laisse supposer qu'il s'agit d'usagers empruntant de tels modes alternatifs. Un second seuil fixé à 4 m/s est appliqué sur les sections interurbaines uniquement.

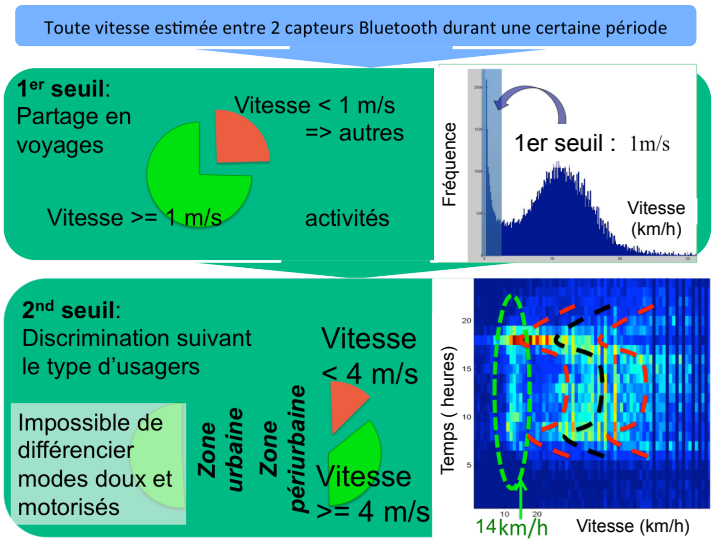


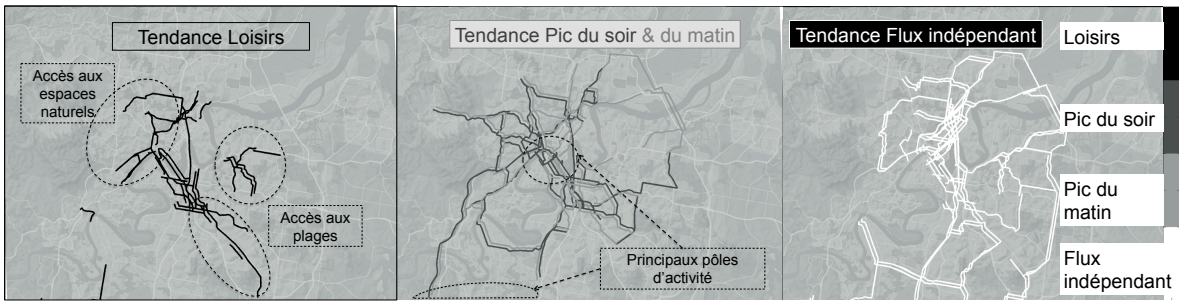
Fig. 1.6.: Illustration du procédé de filtrage adopté sur le réseau de Brisbane (technologie Bluetooth)

Une extension du procédé de distinction des modes a été proposée [21], mais non appliquée. Elle se base sur l'hypothèse qu'au cours de leur itinéraire, les usagers expérimentant de grandes variations de vitesses sont motorisés, tandis que ceux expérimentant des vitesses faibles et peu variables usent de modes alternatifs. Cette hypothèse est schématisée par la Figure 1.5 b).

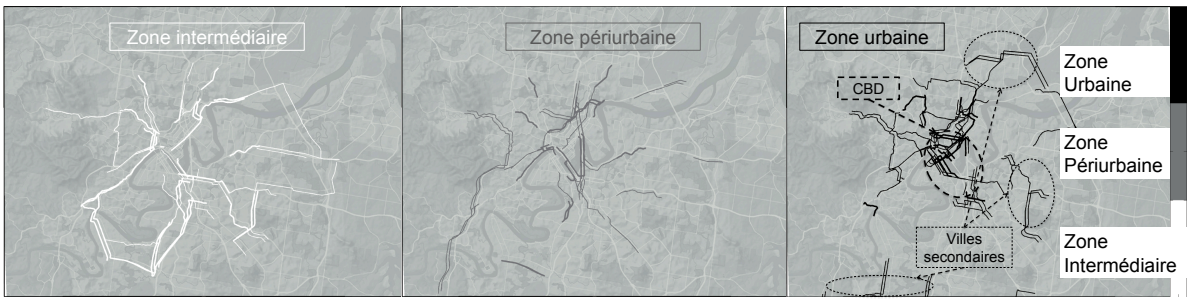
D'autres procédés s'appuyant sur des classifications non supervisées peuvent également être rencontrés [20].

Fiabilité des données

Un des grands questionnements concernant l'usage de la technologie Bluetooth (BT) porte sur sa fiabilité à représenter les états de trafic. La première partie de la thèse de Michau [9] et de notre travail de Master [8] fournissent un bon aperçu du potentiel de la technologie Bluetooth. Il est à retenir que le taux de pénétration sur le réseau de Brisbane varie d'une section à une autre entre 5% et 20%. Un tel pourcentage est suffisant pour produire des estimations de vitesses satisfaisantes, mais pose quelques difficultés quant à l'estimation d'autres variables macroscopiques du trafic comme le débit. L'étude approfondie de la capacité du Bluetooth à retranscrire les états de trafic nous a conduits à retrouver les principales caractéristiques structurelles et tendances journalières du réseau à partir des seules données recueillies [22]. Les résultats sont exposés en Figure 1.7. Les tendances lors des pics du matin et du soir sont bien inversées suivant le sens de circulation. De même, le CBD et les villes alentours sont identifiées en tant que zones urbaines. Diverses autres spécificités du réseau de Brisbane sont ainsi retrouvées. La qualité de retranscription des états de trafic par les données Bluetooth étant avérée, les caractéristiques macroscopiques (débit Bluetooth, vitesse) du trafic sont estimées avec un pas de 1 heure.



(a)



(b)

Fig. 1.7.: Illustration de la fiabilité des capteurs Bluetooth : a) Qualification des sections suivant leurs tendances temporelles identifiées par un procédé de classification génératif probabiliste b) Qualification des sections suivant les caractéristiques spatiales tirées de la forme des diagrammes fondamentaux.

1.2.3.2 Réseau nantais et technologie GPS

Ce cas d'étude a été permis grâce à la mise à notre disposition de données FCD par MEDIAMOBILE, opérateur de services d'information trafic routier en temps réel français, recueillant sur l'intégralité du réseau français des données trafic issues de FCD. L'obtention des données trafic sur la zone de Nantes résulte directement d'un partenariat avec l'opérateur.

1.2.3.2.1 Descriptif du cadre d'étude

Suite à une analyse succincte mais experte des conditions usuelles de trafic à partir du site internet V-Traffic (`http://www.v-traffic.com/`), 4 zones sensibles du réseau nantais ont été identifiées. Il s'agit du centre-ville urbain, de deux ponts, le Pont de Cheviré et de Bellevue, et d'une zone d'échangeurs autoroutiers reliant la Bretagne au reste de la France. Les réseaux tels qu'identifiés sont illustrés en Figure 1.4 b) et modélisés par MEDIAMOBILE sous forme de graphes orientés. Le centre-ville urbain décompte 796 arcs ou sections de routes, 177 pour la zone du Pont de Cheviré, 183 pour la zone de Bellevue et 127 pour l'échangeur Nord. Les traces issues de GPS sont recueillies en temps réel, puis projetées sur les graphes associés aux réseaux d'étude. Les traces sont reconstruites et l'information éventuellement propagée le long de ces traces. Par période de 3 minutes, les temps de parcours sont agrégés sur chacune des sections, puis la vitesse du flot est estimée. Cette phase a été accomplie par l'opérateur MEDIAMOBILE et aboutit à l'expression d'une vitesse du flot à tout pas de temps sur chacune des sections de la zone de Nantes. Trois mois de recueil sont disponibles pour ce cas d'étude : de septembre à novembre 2013.

1.2.3.2.2 Traitements appliqués aux données

A chaque section et pour tout pas de temps, est associé un ensemble de vitesses tirées des véhicules traceurs ayant traversé cette section durant la période de 3 minutes. Le procédé de mesure et la projection cartographique ne sont pas épargnés par le bruitage des données.

Filtrage des données

Le procédé de filtrage des vitesses recueillies a été opéré lors de la phase de *map-matching* et reste à la discrétion de l'opérateur MEDIAMOBILE.

Fiabilité des données

La fiabilité des données recueillies est estimable via le taux d'échantillonnage ou taux de pénétration. Sur la période de 3 mois considérée, le taux de pénétration de la technologie FCD n'est pas connu avec précision, mais supposé *a priori* faible. En effet, avec une fréquence de 3 minutes, environ 35% des valeurs sont inconnues. Le phénomène se traduit sous diverses formes : longues plages de données manquantes ou données ponctuellement inconnues. Les longues périodes sans information représentent 18% des valeurs manquantes. Afin de limiter la quantité de données absentes ponctuellement et d'augmenter la taille des échantillons, la période d'échantillonnage a été modifiée. La période d'agrégation et donc le pas de temps a été fixé à 30 minutes. Seules 22% des valeurs restent encore indéterminées. La gestion de

ces valeurs manquantes entre dans un cadre plus général exposé en section 2.2.4. Le passage d'une période de 3 minutes à 30 minutes agrège les informations sur les phénomènes les plus dynamiques du trafic, tels que les ondes de congestion. Cependant, de tels phénomènes sont extrêmement locaux et ne décrivent pas le comportement macroscopique de la section. Nous nous intéressons dès lors à la dynamique court-terme du réseau s'étalant de quelques minutes à quelques heures, se distinguant ainsi de l'étude des évolutions journalières. Le choix d'un pas de temps de 30 minutes constitue une période intermédiaire, suffisante au regard de l'échelle spatiale considérée. Un pas de temps de 30 minutes assure donc une bonne adéquation entre précision de la donnée et dynamique du trafic recherchée.

A partir de technologies récemment introduites dans le domaine du recueil d'information, les cas d'étude présentés fournissent une information sur toute section du réseau. Une bonne supervision des réseaux étudiés est ainsi envisageable.

1.3 Prévision court-terme à l'échelle du réseau routier

Pour la gestion d'un réseau, le gestionnaire ne peut pas se contenter de la supervision des états de trafic sur le réseau. La mission du gestionnaire d'infrastructures consiste également à intervenir en amont pour endiguer les propagations de la congestion et limiter les impacts à moyen terme. Pour être efficace, la stratégie adoptée doit se baser sur des états prévisionnels de trafic, d'où la nécessité de prévision court-terme.

La prévision court-terme consiste à prévoir, pour toute section du réseau, la caractéristique du trafic à l'horizon h depuis l'instant t. D'un point de vue formel, la prévision vise à rechercher la fonction f assurant la transition de $\underline{\boldsymbol{X_t}}$, l'état de trafic à l'instant t et éventuellement ses états antérieurs, à $\boldsymbol{Y_{t+h}}$, l'état de trafic à l'horizon h. La caractéristique de trafic prédite $\widehat{\boldsymbol{Y_{t+h}}}$ résulte de $\widehat{\boldsymbol{Y_{t+h}}} = f(\underline{\boldsymbol{X_t}})$.

1.3.1 Enjeux de la prévision du trafic

Comme on peut s'en douter, anticiper les états de trafic sur le réseau n'est pas tâche aisée. Outre la mesure des caractéristiques de trafic introduisant des incertitudes, le flux de trafic est lui-même extrêmement sensible aux perturbations externes (météo-sensibilité, événement sportif etc), comme internes (propagation de congestion etc). Ces variations se traduisent sur les caractéristiques du trafic et contribuent pleinement à ses **propriétés** :

– **non-linéarité** : les motifs du flux de trafic ne sont pas modélisables par des relations linéaires ou pouvant s'y ramener (exponentielles, logarithmiques). La variance des vitesses est dépendante de l'heure de la journée (hétéroscédasticité), de la section considérée et des interconnexions avec son voisinage (hétérogénéité) [23].

– **non-stationnarité** : Le flux de trafic est susceptible de subir des ruptures avec l'apparition d'événements non-récurrents et possède des tendances, notamment à travers les variations saisonnières.

Ces propriétés du trafic révèlent les enjeux auxquels sont confrontés les gestionnaires d'infrastructure :

– Maintien d'une prévision de qualité à toute heure de la journée ou période de l'année ;
– **Anticipation** des phénomènes pouvant **perturber le réseau** (chutes de neige, pluie, événement sportif *etc*) ;
– Anticipation **multi-horizon** des états de trafic pour intervention plus ou moins en amont ;
– Usage d'outils prédictifs adaptés : maintien d'un **temps de prévision compatible** avec les objectifs **temps réel** d'intervention et facilité de **prévision à l'échelle du réseau**.

En vue de répondre à ces objectifs, deux écoles de pensée se distinguent quant à la façon de déterminer la fonction prédictive f de transition :

– la prévision fondée sur des modèles reproduisant la physique du trafic ou "model-driven" ;
– la prévision fondée sur des composantes statistiques du flux de trafic et s'appuyant sur des données empiriques ou "data-driven".

La première approche suppose que les lois régissant le flux du trafic sont connues et décrites par la théorie du trafic. À partir de conditions initiales connues, l'application de ces lois permet de propager les caractéristiques du trafic considérées et donc de percevoir l'avenir. La seconde approche ne fait aucune supposition quant au fonctionnement du réseau, mais s'appuie sur la théorie statistique pour supposer que des conditions initiales communes ou proches entrainent des évolutions semblables. Le trafic y est perçu comme un signal se propageant dans des tubes contraints et reliés en réseau.
Des approches hybrides [24, 25, 26] sont également apparues ces dernières années, mais restent de l'ordre de l'exception.

1.3.2 Prévision à partir de modèles (*model-driven*)

Les modèles de trafic sont particulièrement bien adaptés à des études à l'échelle du réseau. À partir de conditions en entrée du réseau parfaitement connues et de profils de demande, ils sont en mesure de simuler la propagation du trafic à travers ce dernier. Les simulations s'appuient sur la théorie de la physique du trafic et trouvent leur source dans le modèle des diagrammes fondamentaux introduit par Greenshields [7] en 1934. Trois grandes familles de modélisation de la physique du trafic se distinguent suivant le niveau de détail désiré : les modèles macroscopiques, microscopiques et mésoscopiques.

1.3.2.1 Approches macroscopiques

Les modèles macroscopiques simulent le comportement moyen du flux de trafic en un point donné. Ils s'intéressent donc aux caractéristiques agrégées telles que la vitesse moyenne, le

débit et la densité.
Ces modèles trouvent leur fondement dans la théorie macroscopique du trafic introduite par Lighthill et Whitham, puis Richards, plus connue sous le nom de modèle **LWR** [27]. Le modèle s'appuie sur la théorie des **ondes cinématiques** représentant la propagation du flux dans une rivière. Richards s'est inspiré de la propagation d'ondes à la surface de l'eau pour compléter le modèle par la notion d'**ondes de congestion**, modélisant les remontées de congestions. Ce modèle est qualifié de modèle du 1er ordre, en référence au degré de la dérivée liant densité et vitesse. Les changements d'états de trafic sont supposés instantanés et immédiats, si bien que les ondes de choc et instabilités ne sont pas décrites.

Depuis ce modèle originel, trois principales variantes se sont distinguées :

- *le modèle LWR multi-classe*, évolution directe du modèle LWR [28, 29]. Cette approche reste du premier ordre, mais assouplit l'hypothèse de non dépassement des véhicules en introduisant des distributions de vitesses.
- *le modèle du Higher-Order de Payne [30]*, modèle du second ordre considérant la vitesse moyenne comme une variable indépendante et décrivant plus finement les changements de régime. Helbing et al. ont fait évoluer ce modèle grâce à des équations de gaz cinétiques [31].
- *le modèle à Transmission Cellulaire de Daganzo [32]*, résolution numérique du modèle LWR divisant la route en cellules entre lesquelles le flux de trafic circule en suivant les règles dictées par LWR.

La prévision à court-terme est possible à partir de tels modèles macroscopiques. Les équations différentielles de conservation du flux se voient alors résolues moyennant quelques approximations spatio-temporelles. En pratique, la prévision est obtenue par discrétisation de l'espace et du temps, puis par approximation de l'état à venir de chaque cellule spatio-temporelle grâce aux informations contenues dans ses voisines [33, 34, 14, 35].
Afin d'être efficaces, une grande quantité et qualité d'informations sont requises pour caler les conditions initiales des modèles. Les profils de demandes précis sur toutes les sections du réseau sont souvent nécessaires. Dans cette perspective, une assimilation des états courants depuis des mesures précises est entreprise (Filtre de Kalman [14, 17, 36]), ce qui requiert un temps d'exploitation non négligeable. Seules les infrastructures autoroutières, richement équipées en boucles électromagnétiques, sont usuellement modélisées. Les zones urbaines s'avèrent difficiles à prédire du fait de couvertures du réseau insuffisantes et de dynamiques locales dues aux feux tricolores [37, 38]. L'usage de ces modèles à titre prédictif n'est pas aisé et adaptable à toute typologie de réseau.
Une extension récente de ces modèles macroscopiques a été introduite par Daganzo et Geroliminis à travers le Diagramme Fondamental Macroscopique (MFD) [39] à l'échelle du réseau. Cette approche modélise non plus le fonctionnement et les interactions des sections d'un réseau, mais celui du réseau et ses divers degrés de saturation. Ils ont donc fait évoluer l'objectif de l'étude à l'échelle du réseau. Malheureusement, cette modélisation ne se prête pas à de la prévision court-terme des états de trafic sur des sections.

1.3.2.2 Les approches microscopiques

Ces modèles se placent à l'échelle du véhicule et simulent leurs trajectoires et vitesses. Ils supposent que les états de trafic agrégés résultent directement des hétérogénéités de comportement des conducteurs. Ils sont traditionnellement partagés en trois catégories [40] :

- les modèles de types lois de poursuite (*car-following*) [41, 42, 43], qui modélisent, en continu, la façon dont se suivent les véhicules en tenant compte du temps de réaction et du comportement des conducteurs.
- les modèles de types automates cellulaires (*cellular automaton*) [44, 45, 46], qui constituent la version discrète des lois de poursuite. Ils s'appuient sur un partage des voies en cases occupées ou non par un véhicule.
- les modèles à changement de voies [47], qui introduisent les dépassements et affectations de voies.

Tous ces modèles ont conduit à divers simulateurs, dont les plus connus sont AIMSUN [47], SYMUVIA [48]. Une revue détaillée des simulateurs est fournie par Jones *et al.* [49].
Ces modèles basés sur la simulation microscopique du trafic sont très largement utilisés par les gestionnaires de transport pour mener des études hors-ligne et mettre au point des stratégies de régulation. Simulant l'écoulement du flux de trafic, ils se prêtent particulièrement à la comparaison de scenarii de régulation en conditions parfaitement similaires. L'aspect prédictif de ces modèles en est une sortie naturelle, mais n'est que rarement exploitable en temps réel. Les temps de simulation s'avèrent élevés et se prêtent donc peu à une mise en œuvre pour une prévision dynamique du trafic.

1.3.2.3 Les approches mésoscopiques

Ces modèles se positionnent à une échelle intermédiaire entre la description complète du comportement des usagers, les uns après les autres, et la description du trafic en tant que flux. Ces approches modélisent le temps de parcours des véhicules sans spécification de leur comportement spatio-temporel. Elles présentent l'avantage de réduire le temps d'estimation par rapport aux modèles microscopiques tout en conservant l'essentiel de l'information. L'approche par cinétique des gaz développée par Prigogine [50] constitue la première tentative d'obtention d'un tel modèle approchant ces caractéristiques. Le modèle est apte à capturer les effets de dispersion des ondes de choc, ainsi que les régimes de transitions en incorporant des distibutions de vitesses et des dérivées secondes. Une formulation variationelle du LWR développée récemment par Duret *et al.* [17] aboutit à une approche mésoscopique alternative. Modélisant une portion du réseau réel de la DIRIF (Direction Interdépartementale des Routes d'Ile de France), le modèle LWR est résolu à l'échelle mésoscopique en s'appuyant sur un procédé de Monte-Carlo séquentiel.
Bien que prometteurs, les temps de calculs restent conséquents et peu enclins à une utilisation en temps réel. De plus, des hypothèses fortes quant au comportement des usagers sont intro-

duites. Leur usage en tant que méthode prédictive en ligne est donc envisageable à terme, mais reste à l'heure actuelle restreinte.

Les modèles issus de ces trois niveaux d'analyse souffrent des hypothèses fortes régulant le comportement des usagers. Mise à part des modèles spécifiques [47], l'hypothèse FIFO (*First In First Out*) est la plus commune et suppose aucun dépassement.
De plus, ces modèles n'aboutissent qu'à une reproduction partielle et limitée du comportement réel des usagers ou du flux de trafic. Seules les parts d'hétérogénéité et de variabilité du trafic connues sont modélisables, abandonnant une grande part des comportements non expliqués.
L'étude proposée se centre sur les méthodes alternatives basées sur les statistiques et la gestion de bases de données conséquentes. Faisant abstraction de la physique du système, ces modèles alternatifs n'exigent pas de description exhaustive du comportement des usagers. Ils s'appuient sur les phénomènes expérimentés par le réseau.

1.3.3 Prévision à partir des données (data-driven)

La prévision à partir d'événements expérimentés par le réseau (*data-driven*) a pris de l'ampleur au cours des 20 dernières années au point de venir concurrencer les méthodes mathématiquement fondées sur des modèles. Une telle évolution trouve son origine dans la combinaison d'une mise à disposition sans précédent de quantité de données traitées dans un temps record et du développement de l'intelligence artificielle conduisant à l'émergence de méthodes novatrices.
Dans ce domaine, le flux de trafic traversant une section y est perçu comme une série temporelle. Les relations statistiques sont extraites de la base de données et employées pour déduire, à partir des états antérieurs, ceux à venir.
Classiquement, est associé, à toute section, un modèle de prévision de la forme $\widehat{Y_{t+h}} = f(\boldsymbol{X_t})$, où $\widehat{Y_{t+h}}$ est la variable expliquée (ou de sortie) et $\boldsymbol{X_t}$ représente l'ensemble des variables explicatives (ou d'entrée).

Deux importantes évolutions ont particulièrement marqué le développement d'une telle approche au cours de ces dernières années :

- le passage de méthodes statistiques classiques à des méthodes issues de l'intelligence artificielle, grâce à la mise à disposition de grandes bases de données et des capacités de calcul adaptées à leur traitement ;
- le passage d'une analyse par séries temporelles à des séries spatio-temporelles, grâce à l'apparition de nouveaux Systèmes de Transports Intelligents (ITS).

1.3.3.1 Évolutions du paramétrique à l'apprentissage

La mise à disposition de grandes bases de données, conjuguée à l'émergence de nouvelles techniques de prévision (apprentissage statistique et apprentissage profond), a profondément marqué le domaine ces dernières années.

1.3.3.1.1 Méthodes statistiques, approches paramétriques originelles

Tandis que les modèles de trafic visent à décrire les phénomènes physiques du trafic, les approches paramétriques statistiques envisagent de décrire de façon exhaustive les interactions statistiques liant l'état prédit $\widehat{Y_{t+h}}$ aux variables d'entrée. Ces dernières se composent des états antérieurs sur la section $\boldsymbol{X_t}$ ou de variables externes (météorologiques par exemple). Pour un modèle statistique donné, la structure de la relation liant la prévision à ses variables explicatives est entièrement déterminée par ses paramètres préalablement calés à partir d'états expérimentés. Le terme **paramétrique** fait référence aux hypothèses quant à la forme des règles régissant le système d'étude. Ces méthodes sont largement employées dans le domaine du trafic, car simples et explicites de formulation. L'interprétation des relations entre les variables étudiées est aisée.

La famille de méthodes **ARIMA** est la plus répandue dans la littérature du trafic et plus largement dans le domaine des séries temporelles [51]. Toutes les méthodes de la famille intègrent deux composantes : un modèle Auto-Régressif (AR) et un modèle à moyenne glissante (MA) liés au sein d'une même équation. La composante AR prévoit l'état à venir à partir de la somme d'une moyenne fixe et d'une combinaison linéaire de ses états antérieurs affinée d'un bruit blanc. La composante MA affine le modèle précédent en modulant la valeur de la moyenne par les bruits blancs précédemment expérimentés et perturbant l'aspect stable du système. L'approche est particulièrement bien adaptée aux séries temporelles.
Ahmed *et al.* [52] ont ouvert la voie en introduisant cette approche dans le domaine du trafic à la fin des années 70. De multiples études ont suivi, affinant les variables d'entrées considérées et la composition du modèle (AR, AM, ARMA ou ARIMA) [53, 54].
Quelques évolutions du modèle ARIMA sont à souligner. La prise en considération des variations saisonnières du trafic via le *seasonal* ARIMA (SARIMA) aboutit à des modèles plus robustes que l'approche basique [55, 56]. De même, l'introduction d'informations spatialement connectées, tels que les états de trafic au voisinage de la section étudiée (*Spatio-Temporal* ARIMA) peut s'avérer efficace [57, 58, 59]. La version ARMAX intégre des variables explicatives exogènes telles que la météorologie [60]. Enfin, des approches fournissant des sorties multidimensionnelles sont également produites (VARMA, multivarite ARIMA) [61, 62, 63, 64]. Force est de constater que tout phénomène du flux de trafic doit être préalablement identifié et intégré au modèle.

Le **modèle espace-état** constitue une autre famille de méthodes paramétriques de prévision largement étudiée dans la littérature [65, 66, 67]. L'originalité de ces modèles résulte de la conjonction entre une équation d'état (modèle de trafic) et une équation de mesure (intégration de données réelles). Il est courant que l'équation d'état prenne la forme d'un système simple d'équations différentielles du premier ordre. Le filtre de Kalman est la plus commune des méthodes de résolution des modèles espace-état. Le filtre de Kalman se partage en deux phases conjointement liées : la prévision basée sur les états précédemment observés et les covariances estimées, puis la mise à jour confrontant l'état prédit à l'observation et ajustant la matrice de covariance. L'ajustement repose sur un système de pondérations attribuant le plus de poids

aux valeurs possédant l'incertitude la plus faible. Ce procédé de mise à jour possède l'avantage indéniable de répondre naturellement à l'aspect dynamique et saisonnier des séries temporelles en trafic. Tandis que le filtre de Kalman repose sur la linéarité des équations d'état et de mesure, sa version étendue procède à une linéarisation locale du problème permettant d'appliquer les équations classiques du filtre. Ce filtre de Kalman étendu [67] montre de bonnes performances, même face à des approches non paramétriques.

Ces modèles statistiques s'avèrent extrêmement efficaces pour prévoir les tendances usuelles du trafic [55], à condition d'être bien calés. En revanche, la durée de paramétrage est élevée et croît avec le nombre de dimensions du problème considéré. Le passage à l'échelle du réseau occasionne donc un surplus du temps d'estimation des paramètres. Du fait de leur structure figée (variance constante), ces modèles échouent à retranscrire certains aspects non-linéaires du trafic. La sensibilité de ces modèles face aux événements inhabituels, tels que l'apparition d'incidents, est plus conséquente que pour une grande part des méthodes non paramétriques. Yoon *et al.* [68] ont montré que les méthodes paramétriques (SARIMA) souffraient d'un décalage temporel plus important et d'une moins bonne qualité de prévision face à des événements non récurrents que certaines régressions non paramétriques (KNN). Des constats similaires sont observés en période congestionnée [69], notamment par comparaison à des méthodes à réseaux de neurones [70].

1.3.3.1.2 Méthodes par apprentissage, méthodes non paramétriques

Tirées de l'intelligence informatique, les méthodes développées prennent part au domaine de la classification au sein de l'apprentissage. Plus spécifiquement, elles appartiennent aux classifications supervisées (*supervised learning*), par opposition aux approches non supervisées (*clustering*). Toutes deux s'intéressent à la classification d'individus caractérisés par des variables, mais le procédé diffère [71].
La **classification non supervisée** classe directement un groupe d'individus en catégories homogènes à partir de leurs seules variables. Les sous-groupes construits répondent aux tendances et densités observées sur les variables caractérisant les individus. Pour la méthode des K-moyennes (*K-means*), l'objectif consiste à accentuer la distance inter-classe et réduire la distance intra-classe.
La classification **supervisée** s'appuie sur un set d'individus, dont la catégorie est connue, pour déterminer des règles de décision. Ces règles sont **apprises** depuis l'historique de labels associés à chaque individu et ses caractéristiques intrinsèques. Cette phase d'**apprentissage** permet le classement de tout nouvel individu.

Il existe des méthodes de prévision basées sur la classification non supervisée ou l'identification de motifs telles que celles développées par Rakha *et al.* [26, 72]. Adoptant le point de vue de l'usager, les auteurs s'intéressent au problème particulier de la prévision des temps de parcours sur des itinéraires complets et non plus des seules caractéristiques macroscopiques de trafic.

Dans le cadre de la prévision du trafic, les concepts de la classification supervisée sont étendus à la régression pour répondre à l'exigence de continuité du flux de trafic prévu. Dans de tels modèles, la valeur prédite tient lieu de catégorie. Les catégories doivent donc être continues, d'où la nécessité de procédés régressifs.

Concept de la prévision par apprentissage

Le terme d'apprentissage regroupe une grande variété de méthodes, seules les approches rencontrées dans la littérature du trafic sont exposées. Quelles que soient les méthodes employées, le procédé reste le même [71] :

1. **Définition de la fonction de prévision** $\widehat{Y} = f(\boldsymbol{X})$, fonction de modélisation de la transition d'un espace $\boldsymbol{X}$ vers un espace Y propre à la méthode employée. $\boldsymbol{X}$ est un individu caractérisé par des variables et $\widehat{Y}$ est le label proposé par le modèle f. En termes de prévision, $\boldsymbol{X}$ est l'ensemble des variables explicatives et Y la variable expliquée.

2. **Usage d'un critère de performance**. Il définit une fonction de coût L traduisant la qualité de la prévision fondée sur l'écart entre $\widehat{Y}$ et Y (erreur de prévision).

3. **Évaluation des hyperparamètres**. Les hyperparamètres sont évalués depuis différents jeux d'historiques (ensembles de couples $(\boldsymbol{X}, Y)$). Ces paramètres externes jouent un rôle de régularisation en vue d'aboutir à un modèle conservant une bonne capacité de généralisation. Ils forcent donc le modèle à considérer des hypothèses plus simples et plus aptes à s'adapter aux différentes situations. Un modèle collant trop aux données apprises (sur-apprentissage) n'est pas apte à traiter convenablement de nouveaux cas, tandis que le sous-apprentissage conduit à des erreurs de prévision conséquentes. Un compromis doit donc être trouvé. Cet aspect joue un rôle essentiel dans la distinction entre les approches de régression par apprentissage supervisé et les approches paramétriques classiques. L'usage des hyperparamètres explique en partie l'aptitude des approches par apprentissage à mieux faire face à des conditions inhabituelles.

4. **Apprendre les éventuels paramètres internes (ou cachés)** du modèle depuis un nouveau jeu de données. À partir d'un jeu de couples $(\boldsymbol{X}, Y)$ connus et fixés, la fonction de perte est optimisée pour coller au mieux aux données d'apprentissage. Les éventuels paramètres internes et propres à la méthode choisie sont donc calibrés.

Bien que toutes les approches par apprentissage disposent d'hyperparamètres, une distinction est envisageable suivant la façon d'appréhender l'apprentissage :

– les méthodes apprenant à décrire la structure latente de l'historique à l'aide de leurs paramètres cachés (internes) ;

– les méthodes apprenant à gérer les informations contenues au sein de l'historique.

Les premières apprennent longuement, mais ne requièrent plus l'usage de l'historique lors de toute nouvelle évaluation. Les secondes apprennent rapidement, mais nécessitent la connaissance de l'historique pour toute nouvelle prévision.

Méthodes apprenant la structure latente de l'historique

La famille des **réseaux de neurones artificiels (ANN)** connaît un franc-succès dans la littérature du trafic depuis la fin des années 90. Dougherty *et al.* [73] ont produit une revue de leurs premiers usages en trafic, complétée par Vlahogianni *et al.* [74] pour les implémentations les plus récentes.

Les réseaux de neurones sont inspirés de l'apprentissage biologique prenant place dans un cerveau à partir de neurones connectés. Le concept repose sur l'observation que les neurones, simples de par leur fonctionnement et structure, permettent, ensemble, des tâches complexes telles que la reconnaissance de motifs, la classification et l'anticipation [75]. La formulation usuelle des réseaux de neurones réside dans la création d'un réseau organisé en couches entièrement connectées dont les nœuds sont les neurones. Chaque neurone reçoit un ensemble de signaux, y applique une opération simple et suivant un seuil renvoie en sortie une information. Entre deux couches, les neurones sont reliés par des arcs pondérés transmettant l'information de sortie d'un neurone à l'entrée de ses successeurs.

Classiquement les réseaux de neurones sont organisés suivant trois couches : une couche d'entrée recevant les variables explicatives, une couche cachée et une couche de sortie proposant une estimation des variables expliquées. L'augmentation du nombre de couches intermédiaires conduit aux méthodes de type apprentissage profond (*deep learning*) [76] apparues récemment. Les pondérations des arcs liant tout couple de neurones et les fonctions des neurones cachés sont apprises et constituent les paramètres internes du système. La couche cachée peut contenir autant de neurones que nécessaire. Plus il y a de neurones cachés, plus le pouvoir de retranscription de motifs complexes est élevé, mais également, plus le phénomène de sur-apprentissage est constaté. Le nombre de couches, le nombre de neurones de la couche cachée et la pénalisation sont les principaux hyperparamètres rencontrés dans la littérature.

La sélection du nombre approprié de nœuds n'est pas aisée et aucun véritable procédé n'a été établi, ce qui conduit à des performances parfois hasardeuses. De plus, étant donné le nombre d'arcs conséquent, l'apprentissage des pondérations se révéle long et fastidieux, ce qui ne garantit pas d'adaptation en temps réel. Enfin, il est souvent reproché aux réseaux de neurones d'être une boîte noire : la structure pondérative étant apprise, complexe et cachée, leur interprétation est très peu aisée. Néanmoins, après quelques tâtonnements, les modèles actuellement construits sont aptes à fournir des performances convenables sur les autoroutes [77, 78, 79] comme les zones urbaines [80, 81, 82] et ce même en prévoyant plusieurs sections simultanément [83].

Moins fréquente, la famille des méthodes à **forêt aléatoire** (*Random Forest*) se révèle être l'une des approches les plus performantes à l'heure actuelle [84, 85, 86]. Elle a été introduite formellement en 2001 par Breiman et al. [87]. Le concept repose sur la construction d'une multitude de modèles de prévision non corrélés prenant la forme d'arbres de décisions, puis la production d'une sortie comme moyenne pondérée des prévisions. La prévision par arbres, inspirée de la méthode dite de segmentation [88], crée un graphe de partitionnement décrivant la répartition d'une population d'individus en groupes homogènes selon un ensemble de variables discriminantes. Chaque arbre tire aléatoirement ses variables discriminantes depuis

un ensemble de variables proposées. L'agencement du graphe et les variables discriminantes choisies constituent les paramètres internes. Les hyperparamètres sont le nombre d'arbres construits et de variables d'entrée tirées aléatoirement. Cette approche possède l'avantage indéniable d'être peu sensible au sur-apprentissage et ouvre des perspectives pour la pratique de prévision à partir de données incomplètes[84].

À la frontière entre approches paramétriques et non-paramétriques, les approches par **réseau bayésien** ont gagné en popularité ces dernières années [89, 90]. S'appuyant sur la théorie bayésienne, le concept repose sur un réseau représenté par un graphe direct acyclique. Les nœuds du graphe sont associés aux variables explicatives comme expliquées. Les directions des arcs s'interprètent comme les dépendances conditionnelles entre les variables ou nœuds. À chaque nœud explicatif est associée une distribution représentant l'ensemble de ses valeurs potentielles. Tout nœud explicatif recueille un état de trafic expérimenté sur la section ou son voisinage. Sa valeur est donc connue à tout instant. L'objectif de prévision vise alors à optimiser la probabilité du nœud expliqué (variable expliquée) sachant que les caractéristiques de trafic associées aux nœuds explicatifs sont connues. Suivant la formulation des hypothèses faites sur les distributions associées aux nœuds explicatifs, cette approche est tantôt perçue comme paramétrique (loi de distribution fixée) [89, 91], tantôt non-paramétrique (distribution apprise) [92, 90]. Leur concision, leur facilité d'interprétation et leur complexité temporelle en font des méthodes extrêmement prometteuses.

Méthodes à gestion de l'historique

Les méthodes de **régression non paramétrique** sont les plus simples et intuitives du domaine de l'apprentissage. La régression produite est néanmoins non-linéaire et repose pleinement sur l'information contenue dans la base de données de l'historique [93]. Ces méthodes visent à déduire la prévision (label ou variable expliquée) à partir de la moyenne pondérée des labels associés aux individus de l'historique d'apprentissage les plus proches de l'état courant (variables explicatives). Deux approches se distinguent par la façon d'appréhender la notion de similarité de l'individu courant avec ceux de l'historique :

- La méthode **KNN** des K plus proches voisins [94, 95, 96, 97] consiste en la comparaison de l'individu courant à ses K plus proches voisins de l'historique d'apprentissage. La similarité est généralement estimée par la distance euclidienne. La prévision du label (variable expliquée) résulte de la moyenne pondérée des valeurs prises par les K labels associés aux individus de l'historique. Des améliorations de la formulation de prévision telles que des pondérations particulières ou des distances autres qu'euclidiennes sont également rencontrées [98, 99]
  Malgré sa simplicité, cette approche s'est révélée plus performante que les méthodes ARIMA et réseaux de neurones [100, 101] sur des cas d'étude autoroutiers.
- La méthode de **régression à noyaux** [102, 103, 104] consiste en la comparaison de l'individu courant à l'ensemble des individus de l'historique d'apprentissage. La contribution de chaque individu de l'historique à la prévision est pondérée par une fonction à noyau, fonction adoptant diverses formes : gaussiennes, à base radiales (RBF) etc. [105].

Ces approches présentent de multiples avantages. Outre leur simplicité théorique et d'interprétation, leur temps d'apprentissage est quasiment nul, contrairement aux approches paramétriques et aux réseaux de neurones. Elles ne possèdent, en effet, pas de paramètres internes à caler. Les hyperparamètres usuels sont respectivement le nombre de voisins K et les paramètres du noyau. Deux bémols sont à souligner : la dépendance du temps d'estimation de la similarité à la taille de l'historique et l'imprécision potentielle face à des événements non récurrents pas recensés par l'historique [106, 37]. Néanmoins, ce second aspect est propre à toute méthode de prévision.

La famille des **méthodes à astuce du noyau** est apparue récemment dans la littérature de l'apprentissage [107, 108, 109]. Ces méthodes se distinguent des régressions non paramétriques précédemment introduites par l'usage qu'il est fait du noyau. Le concept repose sur l'idée qu'il existe un espace de dimension potentiellement infinie, dans lequel la relation entre les variables d'entrée et de sortie est modélisable par une régression linéaire. Le passage de l'espace d'entrée à l'espace en haute dimension est approché par une projection prenant la forme de noyaux. Cette projection correspond à l'astuce du noyau ou *Kernel Trick*. La régression proposée est dès lors non-linéaire.
L'approche la plus connue est la régression par machines à séparateur à vaste marge ("Support Vector Regression" - SVR) développée par Vapnik *et al.* [108]. Cette méthode vise à procéder à une régression non linéaire s'appuyant sur un sous-ensemble de l'historique d'apprentissage bien choisi. Les individus de ce sous-ensemble sont nommés vecteurs supports. La non-linéarité de la régression, sa bonne capacité de généralisation due à la minimisation du risque structurel et la limitation de l'historique aux seuls vecteurs supports en font une méthode de prévision particulièrement robuste [110, 111]. Une analyse succincte de la littérature révèle une réduction conséquente de l'erreur de prévision face à des méthodes simples [112], ou plus complexes telles que les réseaux de neurones artificiels [113] quand la taille de l'historique d'apprentissage est réduite, voire l'apprentissage profond [114]. La variété des cadres d'études, des horizons de prévision et des caractéristiques de trafic considérées garantit une très bonne résilience de cette approche quels que soient les cas étudiés. Enfin, une méthodologie en ligne, *i.e.* de prévision temps réel, implémentée par Castro-Neto *et al.* [115] s'est montrée particulièrement performante en comparaison à de multiples autres méthodes en conditions non-récurrentes.
Une approche alternative est la Régression Kernel Ridge basée sur une régression par les moindres carrés [116]. Méthode plus simple à implémenter que le SVR, elle n'en présente pas l'avantage de la réduction de l'historique.

Finalement, chacune des approches présente ses avantages et inconvénients, mais certains défauts des méthodes fondées sur l'exploitation des données historiques se traitent plus facilement. Les méthodes favorisant la structure sont certes aptes à répondre à toute situation et indépendantes de l'historique, mais toute modification conséquente de ce dernier ou simple mise à jour nécessite une longue évaluation pour saisir la nouvelle structure. Leur temps d'apprentissage étant conséquent, elles ne peuvent donc pas être mises à jour en ligne. Les versions à gestion d'historique présentent plus de souplesse quant à ce dernier point. De plus,

la rigidité des méthodes apprenant la structure s'avère problématique face à des scenarii de rupture ou mettant en œuvre une dynamique de trafic inhabituelle. Il est donc nécessaire de disposer d'un jeu de données complet pour évaluer la structure, quand les algorithmes gérant l'historique nécessitent une rapide mise à jour par apprentissage. Ces derniers ont de meilleures facultés d'adaptation pour faire face à l'occurrence d'événements non récurrents. En revanche, les méthodes apprenant l'historique sont naturellement plus à même de faire face à des données manquantes au sein de l'historique d'apprentissage.

1.3.3.1.3 Optimisation des méthodes de prévision

Que ce soit le nombre de paramètres du modèle, l'espace d'entrée ou un lissage des données, de nombreuses méthodes prédictives nécessitent des phases d'optimisation de l'espace d'entrée et des variables employées. Les hyperparamètres des approches à réseaux de neurones nécessitent souvent une optimisation en 2 étapes, le nombre de neurones de la couche cachée étant déterminant quant à la performance de la méthode. Les méthodes biomimétiques ou algorithmes génétiques [80, 81, 117, 118] sont régulièrement observées afin de parvenir à optimiser simultanément les pondérations et la structure. L'espace d'entrée (les variables explicatives) peut se voir modifié soit par transformation de l'espace via notamment la décomposition en ondelettes [119, 120, 121], soit par réduction du nombre de variables d'entrée grâce à une Analyse en Composantes Principales (ACP) par exemple [122].

Enfin, certaines approches de la littérature combinent des prévisions tirées de divers modèles [123, 124]. Ces approches sont qualifiées d'ensemblistes. Elles procèdent par un vote attribuant un modèle à un sous-espace tiré de l'espace d'entrée. Les plus connues se dénomment *bagging* ou *boosting*. Le concept repose sur le constat que les méthodes de machine learning possèdent chacune leur domaine spécifique d'application. Suivant la section considérée et le pas de temps de la journée, telle ou telle méthode est plus adaptée. L'approche ensembliste vise à trouver l'adéquation d'usage de différentes méthodes les unes avec les autres. Les forêts d'arbres décisionnels peuvent se voir classées parmi les approches ensemblistes du fait de la création et comparaison d'une multitude de modèles.

1.3.3.2 Approches spatio-temporelles

L'émergence de nouvelles technologies au sein des ITS aboutit à des recueils de données couvrant des zones plus étendues et continues. Les informations ne sont donc plus disponibles sur des sections isolées, mais sur un ensemble de liens interconnectés en réseau. La vision de la prévision du trafic comme une série temporelle associée à une unique section s'est modifiée. Avec l'introduction des séries spatio-temporelles, la section est désormais perçue comme appartenant à un système complexe et interconnecté. Les modèles de prévision les plus récents proposent désormais des prévisions sur la section intégrant en entrée des variables explicatives temporelles et spatiales telles que dépeintes en Figure 1.8. Il en a résulté la problématique de la délimitation du voisinage spatial à considérer pour prévoir une section spécifique. En effet, si certaines informations provenant de sections en amont ou en aval sont utiles, suivant

les dynamiques du trafic visées, le voisinage pertinent de la section d'étude varie. Toutes les sections n'apportent pas une information ayant un impact sur la prévision des caractéristiques de trafic de la section considérée.

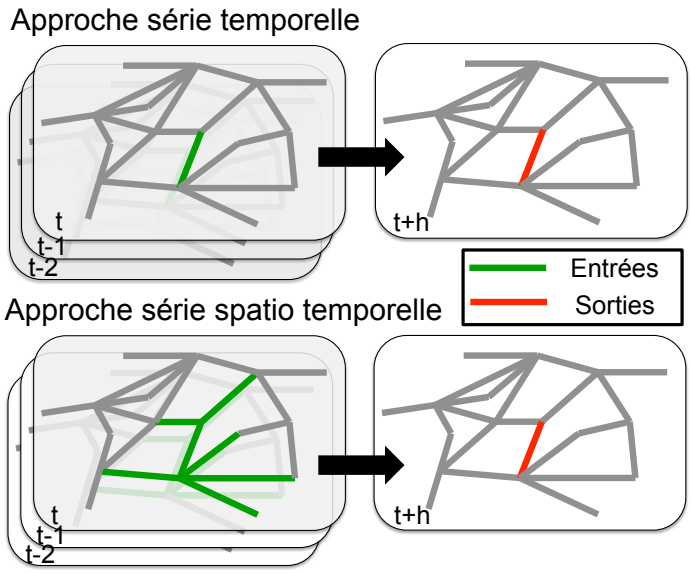


Fig. 1.8.: Évolution des variables d'entrée et de sortie de la prévision lors du passage d'une approche de type série temporelle à une approche par série spatio-temporelle.

Le voisinage spatio-temporel est intégré de diverses façons, s'accordant aux particularités des méthodes de prévision employées.
Les modèles statistiques de type STARIMA définissent un vecteur de pondération spatiale liant toutes les sections du réseau à la section d'étude [125]. La pondération pouvant être binaire et seulement délimiter le voisinage spatial. Cette pondération est indépendante du modèle et se détermine depuis des hypothèses d'écoulement du trafic ou des mesures de dépendance statistique telles que la corrélation. Les versions spatio-temporelles du VARIMA construisent directement des matrices de pondération spatiale [63, 126].
Les modèles non paramétriques de type réseaux de neurones ou forêts aléatoires [83] ne nécessitent pas d'adaptation particulière si ce n'est l'intégration de variables d'entrée suffisantes. En effet, ces approches sont en mesure d'apprendre les variables discriminantes utiles à l'optimisation de la fonction objectif associée au problème de prévision considéré. En contrepartie, les voisinages sélectionnés sont inconnus et non interprétables.
Les réseaux bayésiens intègrent les variables spatiales en leur associant des nœuds connectés au réseau de dépendances conditionnelles [90]. La construction du graphe bayésien et la sélection du voisinage obéissent à des considérations de l'écoulement physique du flux de trafic.
Les approches basées sur l'exploitation de l'historique incluent les informations spatiales comme des variables explicatives complémentaires [127]. Tout comme pour les approches paramétriques, une recherche du voisinage optimal est requise et estimée par une matrice de pondérations spatiales, de corrélations ou liée (*Graphical LASSO*) [127, 128, 129].

Le voisinage spatial est construit de diverses manières, depuis des considérations empiriques jusqu'à une analyse fine le faisant varier dynamiquement dans le temps et l'espace. Le voisinage se définit par 2 paramètres : son étendue spatiale et la contribution ou poids attribué aux sections voisines. Haworth [37] a analysé la définition du voisinage spatio-temporel suivant 5 catégories :

- **Voisinage fixé**. Pour toute section, les pondérations et le nombre de sections composant le voisinage sont prédéterminés, soit par des considérations empiriques [70, 65], soit par des estimations *a priori* [63, 129, 130], soit par apprentissage [81]. Cette méthodologie reste la plus commune.
- **Voisinage variant temporellement**. Chaque section possède un voisinage spatial aux dimensions préalablement fixées, mais les contributions de chacune des sections varient en fonction des périodes de la journée (pic du matin, nuit *etc*). Les sections du voisinage ont un impact variable au cours du temps [110].
- **Voisinage variant spatialement**. Pour chaque section, l'influence des sections du voisinage est fixée dans le temps, mais le nombre de sections le composant est variable [59, 78].
- **Voisinage variant spatio-temporellement**. Pour chaque section, la composition du voisinage et ses dimensions varient par période temporelle fixée [126, 66, 131].
- **Voisinage évoluant avec les conditions de trafic**. Pour chaque section, la structure du voisinage s'adapte en temps réel à la dynamique du trafic. Peu de méthodes sont en mesure d'y aboutir. Les méthodes s'en approchant ré-évaluent en temps réel la matrice de pondération spatiale depuis une évaluation du transfert de flux à chaque intersection du réseau [57] ou par une analyse locale des différences de vitesses [132].

Globalement les apports de l'introduction d'un voisinage spatial sont controversés. Certaines études [132, 130] ont montré que le STARIMA ne fournit pas forcément de meilleures performances que l'approche par série temporelle ARIMA. De tels résultats sont probablement dus à l'absence de prise en considération de la non-stationnarité du flux de trafic. Toute structure spatio-temporelle non dynamique ne semble pas apte à fournir les informations requises aux modèles ARIMA. Néanmoins, Min *et al.* [57] montrent, au prix d'efforts conséquents, une amélioration de l'erreur de prévision. Ces pertes de performances sont essentiellement constatées à partir de méthodes paramétriques. Elles font également l'objet du plus grand nombre d'études de voisinage. En effet, les approches non linéaires par apprentissage reportent en partie la tâche de sélection du voisinage et/ou l'aspect non-stationnaire sur la phase d'apprentissage et les métriques de la méthode de prévision. La non-linéarité des approches par apprentissage leur apporte plus de souplesse quant à la problématique de sélection du voisinage. L'introduction de la structure spatio-temporelle peut s'avérer intéressante néanmoins [37]. De tels constats encouragent, d'une part, à l'usage de méthodes non paramétriques pour s'attaquer aux enjeux actuels du trafic et d'autre part, à y intégrer des notions liées à la structure spatio-temporelle du réseau.

1.3.3.3 Unification des modèles à l'échelle d'un réseau : transition d'une approche de décomposition à une vision systémique

Ces dernières années, nombre d'études ont péché par une tendance à estimer la performance à travers le seul filtre de l'erreur de prévision et du temps d'exécution. Ce phénomène est probablement dû à la facilité de tester diverses approches tirées du domaine informatique et la volonté d'expérimenter les opportunités offertes par les nouvelles sources de données [74]. D'autres éléments essentiels ont été régulièrement négligés tels que le nombre de paramètres ou hyperparamètres à considérer, le temps d'apprentissage ou de calibrage des méthodes et la facilité de mise en œuvre à l'échelle du réseau.
Par le passé, l'analyse section par section se justifiait du fait de la quantité limitée de données et la disposition des capteurs contraignait l'étude à une section isolée ou un corridor. L'émergence de nouveaux moyens de mesure fournit, en tout point, l'ensemble des états de trafic sur un réseau de façon continue. Cet apport a conduit à remettre en question la perception de la prévision et est exploité via l'introduction de séries spatio-temporelles et l'analyse du voisinage. Cependant, le problème de prévision rencontré dans la littérature vise systématiquement à construire un modèle pour chacune des sections et parfois pour chaque période temporelle identifiée. La plupart des méthodes introduites précédemment sont appliquées en vue de prévoir sur une seule section (sortie unidimensionnelle). Or, l'objectif poursuivi vise à être en mesure de fournir une prévision de qualité à partir de modèles parcimonieux, efficace face à des événements non récurrents et facilement implémentable en temps réel pour toutes les sections d'un réseau.
L'opportunité de l'emploi de modèles **spatialement unifiés** en vue de prévoir sur un petit réseau n'a été que très peu exploitée [133]. Pourtant, par comparaison à la prévision de toute section par un modèle distinct, la possibilité de prévoir un ensemble de sections à partir d'un même modèle (*i.e.* méthode et paramètres communs) partagé augurerait un gain évident en termes de temps de calcul et de facilité d'implémentation en temps réel. Les gestionnaires d'infrastructures sont en effet confrontés à des réseaux entiers et non des sections isolées, ce qui leur impose d'estimer les états de trafic sur toute section. Le calibrage d'un modèle étant extrêmement chronophage et parfois compliqué, la possibilité de réduire drastiquement ce temps **en uniformisant les paramètres** est une opportunité à saisir.

Dans une certaine mesure, le questionnement peut se percevoir comme l'opportunité d'une transition d'une **approche de décomposition**, scindant le problème à l'échelle du réseau en des problèmes à l'échelle de la section, à une **approche systémique**, visant à analyser le fonctionnement du réseau dans sa globalité.

La problématique du **modèle unifié**, donc multidimensionnel, appliqué à un jeu cohérent de sections relève de deux verrous :

- partage des paramètres et/ou hyperparamètres calibrés pour toutes les sections ;
- prévision simultanée de toutes les sections.

La fonction objectif n'est donc plus d'optimiser la prévision sur une section isolée, mais bien sur l'ensemble des sections du réseau simultanément.

La littérature du trafic n'est pas totalement démunie de démarches s'en approchant. La sortie multidimensionnelle est une composante essentielle d'une méthode de prévision unifiée. Or, les méthodes de prévision de type VARIMA ou encore les réseaux de neurones offrent la possibilité de prévoir avec des variables en entrée et en sortie multivariées ou multidimensionnelles (*MultiInput-MultiOutput* ou MIMO).
La prévision à partir de variables d'entrée multivariées (MI) est usuelle par le traitement de séries spatio-temporelles et l'introduction du voisinage de la section étudiée. En revanche, la possibilité de sorties multidimensionnelles offre l'opportunité de prévoir la caractéristique de trafic de plusieurs sections simultanément.
L'usage de telles approches MIMO répond généralement à un besoin accessoire de facilité de traitement et de représentation, si bien que l'apport potentiel ne fait l'objet d'aucune analyse [134, 135, 136]. L'aspect multivarié est également envisagé en vue de prévoir simultanément plusieurs caractéristiques du trafic (débit, vitesse, densité) [128]. Une étude expérimentale de Chandra et Al-Drek [63] a montré l'absence de bénéfice de telles approches en les confrontant aux cas où une seule caractéristique du trafic est prédite.
Bien que produisant des résultats à l'échelle du réseau routier, les implémentations multivariées de réseaux de neurones [80, 83] ne proposent pas réellement de modèle unifié. En effet, les réseaux de neurones introduits s'adaptent simplement à l'exigence multi-dimensionnelle de la sortie en réclamant une quantité de neurones au sein de la couche cachée plus importante sous peine de performances nettement amoindries. La complexité temporelle et l'apprentissage en sont allongés. Une analyse fine et une intervention sur les neurones cachés serait nécessaire pour y aboutir. Néanmoins, le gain entre un réseau de neurones univariés et multivariés est possiblement conséquent, mais absolument pas documenté.
Certaines approches VARMA [126, 57] répondent à des tentatives délibérées d'aboutir à l'unification du modèle. Certes, les méthodes proposées sont en mesure de prédire à divers horizons depuis un seul et même modèle, mais elles correspondent en réalité à une juxtaposition de modèles aux sorties univariées. En effet, ces méthodes introduites de façon matricielle ont des sorties multivariées, mais les paramètres adoptent eux aussi une forme matricielle. Chaque section, voire période de la journée [126], requiert donc son propre jeu de paramètres nécessitant un calibrage individuel.
Enfin, une méthode unifiée est proposée par Lippi *et al.* [137], mais ces derniers ne traitent pas le problème de prévision du trafic. Ils s'intéressent à la classification entre régime fluide et congestionné.

Les premières véritables méthodes unifiées ont été développées simultanément au début de l'année 2015 par Dauwels *et al.* [133] et Laharotte *et al.* [138] à partir de méthodes simples de prévision : respectivement une régression linéaire généralisée et une régression non linéaire avec les K plus proches voisins.

1.4 Verrous scientifiques et démarche proposée

A l'issue de cet état de l'art, il résulte que les méthodes de prévision de type apprentissage s'avèrent souvent plus efficaces car plus souples pour traiter la prévision suite à l'occurrence d'événements non-récurrents ou les périodes de congestion usuelle à forte variabilité. Leur performance reste globalement équivalente aux modèles paramétriques sur les états usuels de trafic. Les approches modèles-centrées perdent de leur intérêt avec l'émergence des nouveaux Systèmes de Transport Intelligent, puisque des caractéristiques de trafic sont connues en tout point, sans nécessiter l'intervention de modèles physiques de propagation.
En outre, l'intégration de l'aspect spatio-temporel s'avère compliquée pour toute approche paramétrique, mais efficace. L'intégration de la structure spatio-temporelle est peu étudiée pour les méthodes par apprentissage.
Enfin, la prévision comme l'intégration de la structure spatio-temporelle du réseau sont perçues de façon localisée à l'échelle de la section. Le fonctionnement global du réseau n'est pratiquement pas analysé.

1.4.1 Positionnement de l'étude

Les principales évolutions et tendances constatées [74] au sein de la littérature du contrôle qualité des réseaux conduisent alors à identifier les verrous scientifiques suivants :

- **Exploration** des possibilités offertes par les **nouvelles technologies** et sources d'information associées, puis leur fusion ;
- **gestion**, traitement et **intégration des données manquantes** à l'issue des procédés de recueil ;
- sélection de modèles prédictifs aux paramètres **parcimonieux** et généralisables ;
- passage d'une prévision localisée à une **prévision unifiée** à l'échelle d'un réseau ;
- **intégration adéquate des relations spatio-temporelles** liant les composantes du réseau au modèle prédictif ;
- **adaptabilité du modèle prédictif** aux événements non-récurrents ;
- mesure du véritable **potentiel de l'apprentissage supervisé** ;
- identification et **qualification des conditions de performance** de toute méthode de prévision.

Face à ces verrous, le cadre de cette étude porte particulièrement sur l'opportunité de la transition d'approches localisées à des approches unifiées à l'échelle du réseau, sur l'analyse de fonctionnement de la structure spatio-temporelle du réseau et sur les potentiels des méthodes unifiées en terme de résilience. Les principaux objectifs de la thèse se résument à :

1. **construire un cadre prédictif** adapté au développement de méthodes multidimensionnelles à l'échelle de petits réseaux ;

2. **Adapter des méthodes de prévision** aux exigences de l'approche globale en s'inspirant de travaux du domaine informatique et les appliquer au trafic ;

3. Affiner l'**intégration de la structure spatio-temporelle** du réseau au sein du nouveau cadre prédictif ;

4. **analyser et comparer les performances** de ce procédé de passage à l'échelle du réseau au regard des approches usuelles (séries spatio-temporelles *etc*) ;

5. **estimer les apports et domaines de prédilection** de chacune des méthodes de prévision considérées, en particulier depuis des **méthodes d'apprentissage à gestion d'historique** ;

6. **étudier le potentiel** de la démarche développée à l'encontre des enjeux prédictifs actuels tels que la résilience face à l'apparition d'incidents, l'intégration de phénomènes externes (météo-sensibilité).

Le terme d'**approche globale** désignera par la suite l'approche adaptée au développement de méthodes unifiées. Elle s'opposera aux diverses approches alternatives existantes dénommées **approches locales**. La Figure 1.9 fournit une comparaison des variables d'entrée et de sortie de l'une et l'autre approche.

Les méthodes par apprentissage se révèlent être les plus robustes face à l'apparition de phénomènes non récurrents. Comme une perspective de l'étude vise à évaluer l'impact de tels phénomènes, la démarche prédictive adoptée sera préférentiellement adaptée pour les procédés de prévision de type classification. Dans la mesure où les méthodes de prévision à gestion d'historique [69] sont les plus souples en vue d'adaptation temps réel, l'approche globale s'y verra expérimentée.
Les deux évolutions des méthodologies de prévision tendent, en effet, à montrer que les approches non paramétriques sont plus à même de bien intégrer les aspects spatio-temporels de l'écoulement du flux de trafic et de faire face à des événements inhabituels. De plus, par leur procédé de gestion de l'historique, ces approches restent simples et adaptables dynamiquement. Elles semblent donc en mesure d'aborder les enjeux de la prévision court-terme du trafic soulignés par Vlahogianni *et al.* [74].
Différentes méthodes tirées des modèles de prévision à noyau serviront à confronter et confirmer les performances comparées entre approches locales et globales. Une contribution de cette thèse réside dans la proposition d'implémentation MIMO de ces méthodes de prévision et de leur application au trafic.

Bien que la gestion des données manquantes soit traitée, la faculté des méthodes à intégrer cet aspect ne sera pas analysée dans cette thèse. Par leurs procédés, les approches statistiques et à réseaux de neurones traitent plus naturellement de tels cas. Haworth [37, 103] a mené une analyse fine sur ce sujet en analysant les prises en considération potentielles depuis des approches gérant l'historique d'apprentissage.

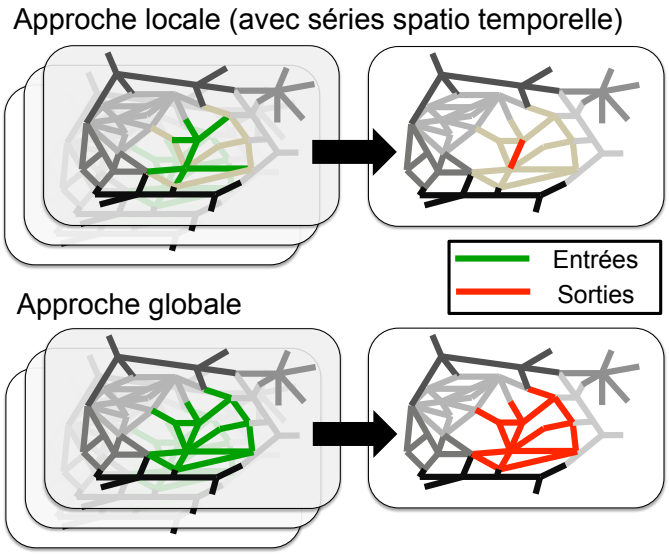


Fig. 1.9.: Évolution des variables d'entrée et de sortie de la prévision lors du passage d'une approche de type série spatio-temporelle (locale) à une approche globale.

La problématique de la méthode unifiée révèle bien évidemment celle de la délimitation du réseau à considérer. Cet aspect est extrêmement proche de la problématique rencontrée par Geroliminis *et al.* [39, 139] lors de la délimitation des zones d'estimation du MFD. Des études récentes ont également traité cette problématique de partage d'un réseau urbain en zones faisant sens [140, 141]. Cette thématique ne sera pas abordée. Les cas d'étude ont été spécialement sélectionnés pour répondre au mieux à une délimitation ayant un sens au regard des dynamiques étudiées.

1.4.2 Démarche adoptée

Partant des objectifs précédemment fixés, nous adoptons une démarche visant à analyser le potentiel d'une approche prédictive globale pour répondre aux enjeux actuels en matière de prévision du trafic. Focalisant l'étude sur les méthodes de prévision par apprentissage, l'agencement de la démarche est décrit par un procédé séquentiel à cinq étapes propre aux méthodes d'apprentissage et illustré en Figure 1.10.

Les deux premiers chapitres posent le cadre de la thèse et donc la structure du procédé séquentiel à cinq étapes dans son ensemble.
Les phases de détection et d'identification des individus sont assimilées au recueil de données trafic et leur mise en forme. Elles correspondent à la phase de gestion du réseau dépeinte en **Chapitre** 1.
Le **chapitre** 2 introduit la démarche prédictive globale spécifiquement adaptée aux méthodes de prévision par apprentissage supervisé. Il s'agit alors de décrire la phase de classification des individus aboutissant au label prédictif, puis de valider la classification sur un nouveau jeu de

données. Une comparaison entre approche locale et globale est produite depuis une méthode simple de prévision sur divers cas d'étude. L'analyse des performances de prévision correspond dès lors à la phase de validation.

Les deux chapitres suivant visent à affiner la définition du cadre prédictif en étendant le potentiel d'adaptation à d'autres approches (**chapitre** 3), puis en se focalisant sur la seule étape non véritablement traitée, à savoir la définition des caractéristiques via l'analyse de la structure spatio-temporelle du réseau (**chapitre** 4)
Le **chapitre** 3 valide les résultats issus de la classification prédictive par approche globale en proposant, respectivement, une analyse des performances de prévision de méthodes à noyaux sur diverses typologies de réseaux, puis une analyse comparative des performances face à des méthodes alternatives de l'état de l'art.
Le **chapitre** 4 affine les variables caractérisant l'individu à prévoir à travers une étude portant sur la dynamique spatio-temporelle du réseau et l'identification de ses sections critiques. L'objectif réside dans une meilleure intégration des relations spatio-temporelles liant les sections en vue d'assurer de meilleurs performances en toute situation de trafic.

Enfin, les **chapitres** 5 et 6 proposent un procédé d'intégration à l'approche globale de l'impact des événements non récurrents et la caractérisation d'un événement particulier, à savoir la météorologie. Ils interviennent sur le procédé par affinement de la définition des caractéristiques de l'individu à classer, via l'introduction de facteurs externes au système. La résilience du réseau face à de tels phénomènes se voit donc évaluée et des améliorations long-terme sont proposées.

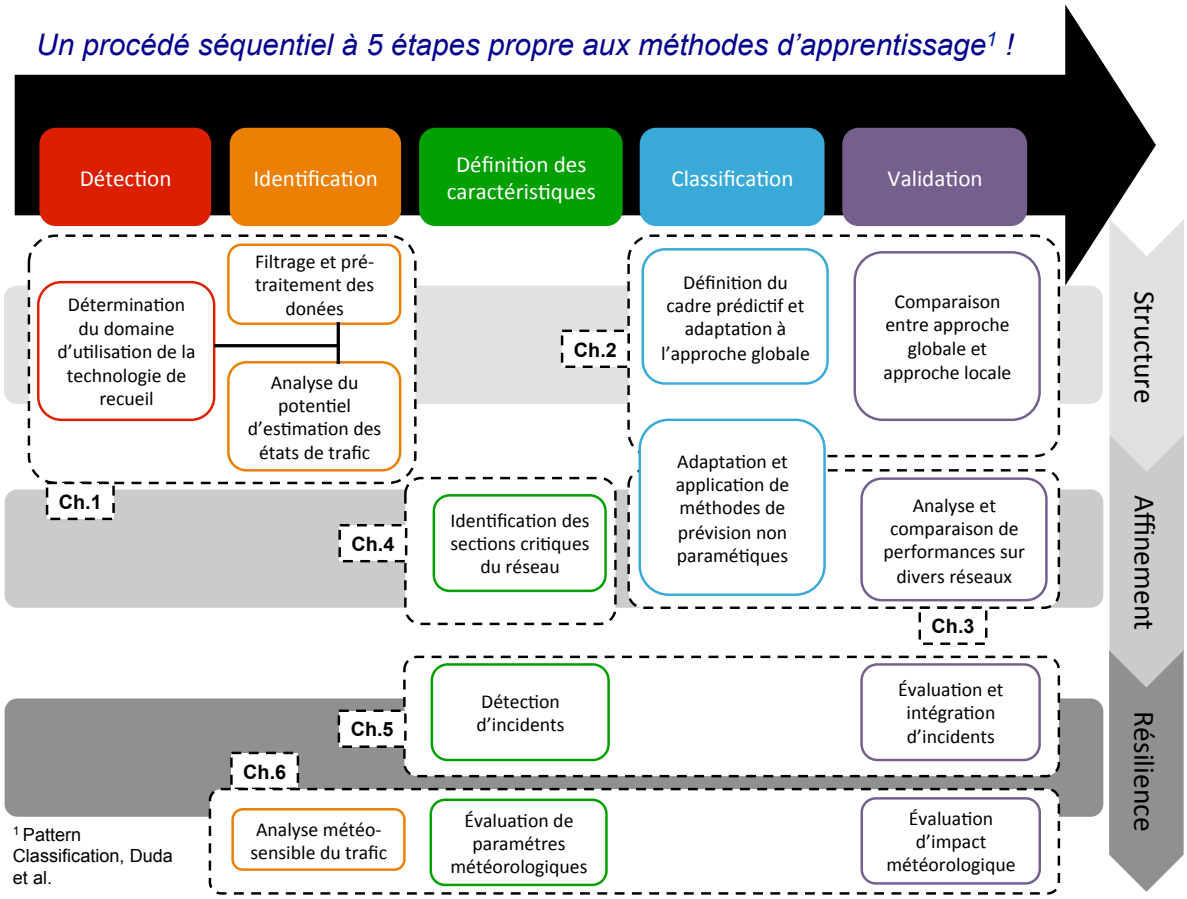


Fig. 1.10.: Illustration de la démarche méthodologique adoptée

2

Introduction d'un cadre prédictif à l'échelle du réseau

> *Les prévisions sont difficiles, surtout lorsqu'elles concernent l'avenir.*
>
> — **Pierre Dac**
> (Acteur)

Abstract :
This chapter aims to detail the traffic-forecasting framework able to tackle the network scale problems. Requirements for the transition from "local" to "global" approaches are exposed and some improvements of machine learning methods are included to the prediction framework. The performances of this network scale strategy are compared to local ones in similar conditions on 2 case studies. Both examples point out that a global approach requires significantly less time to perform as well as local strategy.

Ce chapitre vise à définir un cadre prédictif général permettant la mise en oeuvre en temps-réel de méthodes de prévision par apprentissage. Bien que l'approche prédictive à l'échelle du réseau soit totalement originale, le cadre prédictif reste celui d'un procédé usuel propre aux méthodes d'apprentissage supervisé.
Les méthodes de classification supervisée construisent un modèle de prévision f approximant le lien entre un individu en entrée, noté $\boldsymbol{X_t}$, et son label de sortie Y_{t+h}. La fonction régressive f est évaluée par la fonction de perte ou de coût $\mathcal{L}$ associée à l'écart entre les valeurs prévues $\widehat{Y_{t+h}}$ et les observations Y_{t+h}. La fonction de perte fait intervenir les paramètres internes de la méthode de prévision. Ces derniers sont déterminés par apprentissage sur un historique de données et définissent un modèle. Des hyperparamètres sont, néanmoins, intégrés en amont à la fonction de coût et viennent pondérer les paramètres internes en vue de garantir une bonne généralisation du modèle. Ces hyperparamètres nécessitent alors d'être optimisés. Usuellement, cette optimisation est opérée, hors-ligne, par apprentissages successifs depuis un historique (composé de T_{hist} individus) représentant les $\frac{2}{3}$ de la base de données disponible (voir Figures 2.1 et 2.7). Ceci assure une variété suffisante de cas rencontrés. Différents modèles sont construits et testés. Les meilleurs modèles et donc jeux d'hyperparamètres sont conservés pour construire le modèle de test. Le tiers restant (composé de T_{test} individus) est destiné à qualifier la capacité de généralisation du modèle produit.

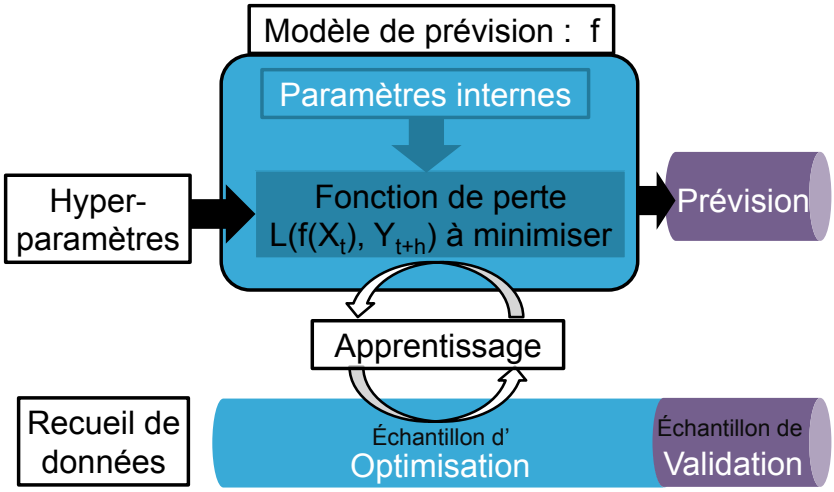


Fig. 2.1.: Procédé de prévision depuis une méthode par apprentissage supervisé.

Finalement, l'usage de méthodes de prévision par classification supervisée requiert un calibrage opérant en trois phases :

1. **Problématisation et construction du cadre prédictif** : définition des individus en entrée et en sortie, de la méthode de prévision employée et de sa fonction de perte associée ;
2. **Optimisation des hyperparamètres** : identification des hyperparamètres requis, puis évaluation de leurs valeurs optimales par analyse des performances de divers modèles appris sur des portions de l'historique d'apprentissage (phase de calibrage hors-ligne) ;
3. **Test de la capacité prédictive** sur un jeu de validation permettant de vérifier la capacité de généralisation de la méthode développée depuis les hyperparamètres optimaux. Après

une phase d'apprentissage sur un historique défini, le modèle est en mesure de prévoir les labels d'un nouveau jeu de données. Cette phase est opérationnelle en ligne ou en temps-réel et constitue la véritable phase de prévision expérimentée par le gestionnaire d'infrastructures.

Après une succincte introduction aux exigences de l'approche globale, ces trois phases structurent ce chapitre destiné à mettre en lumière les adaptations requises par cette approche dans un contexte d'apprentissage supervisé. Le cadre méthodologique proposé est général et adapté à toute méthode de prévision. Néanmoins, en guise d'illustration du procédé, les résultats obtenus à partir de la méthode des K plus proches voisins sont présentés tout au long de ce chapitre.

2.1 Exigences associées à la démarche globale

Le chapitre 1 a mis en évidence le potentiel du passage d'approches prédictives centrées sur une section à l'évaluation simultanée d'un groupe de sections formant un réseau. L'intérêt de la prise en considération des connexions spatio-temporelles liant les sections du réseau est également souligné.
La transition d'une méthode centrée sur la section à un modèle unifié à l'échelle du réseau vise essentiellement à réduire les temps de calcul et à exploiter la structure du réseau avec ses corrélations spatio-temporelles. À objectifs équivalents, la complexité des modèles proposés est simplifiée. En effet, contrairement aux approches usuelles, la pratique d'une prévision à l'échelle du réseau nécessite un nombre réduit de paramètres, indépendant de la quantité de sections considérées. Le temps de calage des paramètres est pratiquement divisé par le nombre de sections considérées. En outre, partageant les mêmes paramètres et un même modèle, toutes les sections du réseau peuvent être prévues simultanément, à condition que le modèle autorise des sorties multidimensionnelles.
L'intégration de l'aspect spatio-temporel répond au besoin de saisir en amont les informations se propageant via le flux de trafic. Si la quantité de véhicules en amont (demande) est élevée, il y a de fortes chances que ces véhicules se retrouvent également sur une section connexe en aval ultérieurement. De même, une congestion en aval occasionne une chute de vitesse, qui tend à remonter en amont au gré des arrivées de nouveaux véhicules. Ces phénomènes de propagation définissent la structure spatio-temporelle du réseau à intégrer en vue d'améliorer les prévisions de toute section.

La démarche globale proposée vise à répondre conjointement à ces deux enjeux. Les exigences s'exerçant sur le cadre prédictif sont de deux ordres :

- Adapter la méthode de classification supervisée pour **aboutir à une méthode de prévision spatialement unifiée**. Cette exigence se traduit par la nécessité de :
  - produire en sortie un **label multidimensionnel** associant une dimension à toute section du réseau considéré ;

– adapter la **fonction de coût à l'objectif multidimensionnel** : passage d'une optimisation par section à une optimisation du réseau aboutissant à une **identification commune des hyperparamètres**. Cette transition d'objectif local à global s'avère souvent la plus délicate. Les classifications supervisées ne sont que très peu adaptées à un objectif multidimensionnel, puisqu'originellement introduites pour ne distinguer que deux catégories distinctes.

Les exigences d'une méthode unifiée portent sur la nature et le fonctionnement en-soi de la méthode de prévision employée.

– Introduire **la structure spatio-temporelle** du réseau via une **sélection de variables pertinentes**.

L'introduction de l'approche globale dans une démarche de type apprentissage supervisé n'a finalement de véritable effet que sur la phase de problématisation et de construction du cadre prédictif. Les autres phases telles que l'évaluation des hyperparamètres optimaux, puis la validation sont inchangées et suivent des procédés pas ou peu altérés.

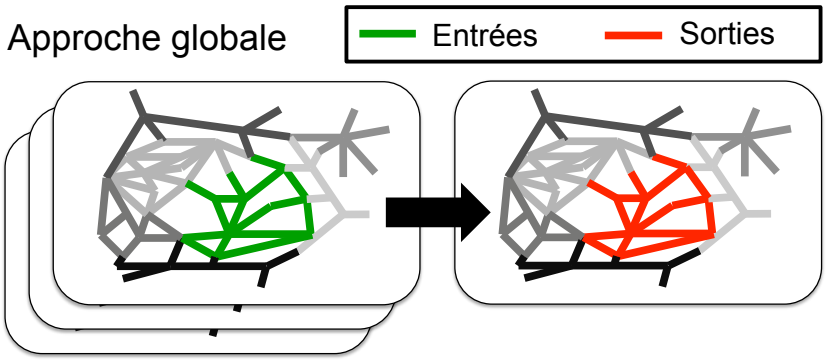


Fig. 2.2.: Illustration des dimensions des variables d'entrée et de sortie pour l'approche globale.

L'approche globale proposée trouve un équivalent dans la problématique de la reconnaissance d'images. L'individu à classer est l'ensemble des clichés du réseau aux instants antérieurs à t (t inclus). L'ensemble des clichés forme un petit film des événements récemment expérimentés. Le label construit par la méthode de prévision est le cliché cible. La caractéristique du trafic sur chaque section est alors assimilée aux pixels de l'image variant dans le temps. La prévision du trafic se ramène alors à l'estimation d'un cliché-cible par confrontation du petit film des événements récémment expérimentés par le réseau à l'historique des films observés par le passé. La confrontation aux événements passés est opérée par une mesure de similarité des clichés pour les approches à gestion d'historique ou par décomposition de ce petit film en fragments latents comparés à des seuils (motifs latents) pour les approches apprenant la structure.

Dans un premier temps, la structure spatio-temporelle du réseau est construite de façon naïve. Elle repose entièrement sur les facultés de la méthode de prévision et la mesure de similarité associée. Toutes les sections sont conservées et représentées en entrée de l'algorithme et il est supposé que ce dernier soit suffisamment souple pour ne conserver que les informations essentielles lors de l'estimation.

2.2 Prévision par apprentissage

Le cadre prédictif usuel est bien adapté à une mise en oeuvre temps-réel pour des prévisions unidimensionnelles à l'échelle de la section. Face aux exigences de l'approche globale, quelques adaptations du cadre méthodologique sont nécessaires.

Par la suite, Z_t^l désignera la caractéristique de trafic mesurée à l'instant t sur la section l. Cette notation permet de décrire l'aspect série temporelle des mesures de trafic. Pour toute section l, les mesures de trafic sont décrites par $Z^l = (Z_t^l)_{t\in[1,\mathrm{T}_{data}]}$, où T_{data} est le nombre de pas de temps constituant la base de données recueillies.

2.2.1 Cadre prédictif usuel

Dans le cas de la prévision de trafic à sortie unidimensionnelle, le procédé de prévision en temps-réel et à l'horizon h prend la forme du diagramme dépeint en Figure 2.3.

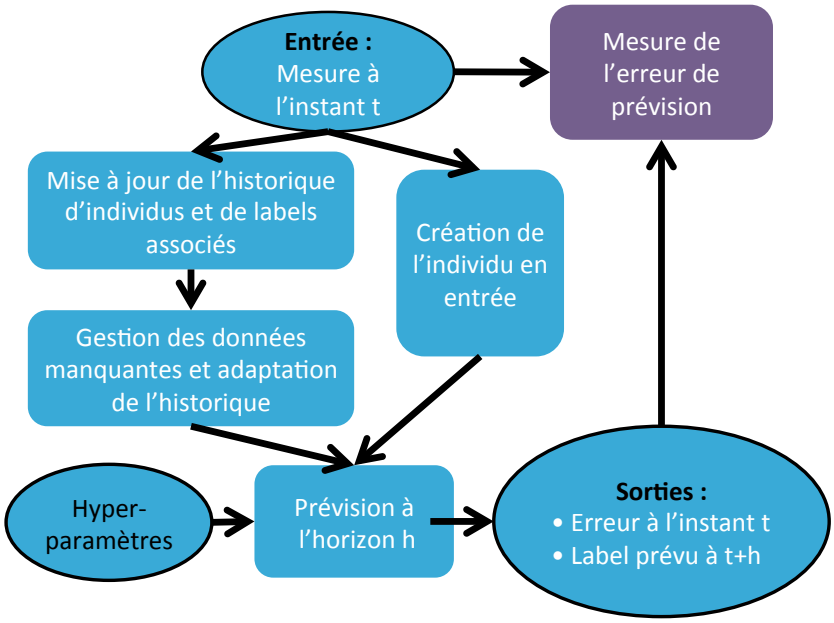


Fig. 2.3.: Cadre prédictif usuellement mis en oeuvre lors d'application en temps-réel (sortie unidimensionnelle).

Le cadre habituel vise à prévoir, pour une section spécifique l, le label ou variable expliquée Y_{t+h}^l depuis un ensemble de variables explicatives rassemblées dans un individu $\boldsymbol{X_t^l}$. L'individu $\boldsymbol{X_t^l}$ adopte la forme vectorielle suivante :

$$\boldsymbol{X_t^l} = \left(w_t^X Z_t^l, ..., w_{t-\mu}^X Z_{t-\mu}^l, w_t^X Z_t^{\Omega_1}, ..., w_{t-\mu}^X Z_{t-\mu}^{\Omega_{N_\Omega}}\right)$$

où N_Ω est le nombre de voisins, $(\Omega_1, ..., \Omega_{N_\Omega})$ les sections du voisinage, μ la longueur de mémoire temporelle conservée et $(w_t^X, ..., w_{t-\mu}^X)$ des pondérations, telles que $\sum_{j=0}^{\mu} w_{t-j}^X = 1$,

visant à favoriser les états les plus récents.
Dans le cadre spécifique de la prévision du trafic, les individus correspondent à un ensemble de caractéristiques macroscopiques du trafic (vitesse, débit) attachées à un instant temporel et une section, tandis que le label est usuellement une caractéristique de trafic ultérieurement expérimentée par la section étudiée. Une correspondance existe donc entre le vecteur d'état à l'entrée et le label de sortie. Chacune des variables composant le vecteur d'entrée correspond à un état de trafic à un instant donné et est également un label de sortie déjà expérimenté : $\forall t, \forall l, Z_t^l \equiv Y_t^l$.

Le procédé de prévision tel qu'exposé est destiné à prévoir en ligne, *i.e.* en temps-réel, dès qu'une mesure des états de trafic est disponible. Toute nouvelle mesure des états de trafic à l'instant t, Z_t^l détermine le label Y_t associé à l'individu $\boldsymbol{X_{t-h}}$. Trois actions sont alors accomplies :

1. Estimation de l'erreur entre la prévision $\widehat{Y_t}$ produite à l'instant $t-h$ et la nouvelle mesure Y_t ;

2. Intégration du couple $(\boldsymbol{X_{t-h}}, Y_t)$ à l'historique d'événements expérimentés ;

3. Construction du nouvel individu en entrée $\boldsymbol{X_t}$ à partir de la nouvelle mesure à l'instant t, Y_t, des mesures antérieures et, éventuellement, de celles des sections de son voisinage.

Suit alors une phase de gestion des valeurs manquantes au sein de l'historique et, éventuellement, de restriction de ce dernier pour une mise en oeuvre améliorée (voir section 2.2.4). La prévision de l'état de trafic $\widehat{Y_{t+h}}$ à l'horizon h est alors produite. Une première phase définit le modèle par apprentissage depuis l'historique prédéterminé, puis l'estimation du label-cible est obtenue. Ainsi, une prévision à tout horizon h est envisagée. Il est intéressant de remarquer que la prévision sur de longs horizons ne procède pas par prévisions successives, mais par évaluation directe. Le phénomène de propagation d'erreurs est donc limité.

2.2.2 Adaptation du cadre prédictif

Bien que le cadre prédictif associé aux méthodes par apprentissage soit suffisamment souple pour se prêter à l'usage de toute méthode non supervisée, quelques adaptations fines sont requises. La couleur bleue au sein de l'algorithme 1 met en valeur l'ensemble des éléments nécessitant modification.

Algorithme 1 Cadre prédictif temps-réel pour une approche globale.

▷ INITIALISATION (VALEURS CONNUES) :
Soit $\boldsymbol{Hyperparameters}$ Vecteur de paramètres propres à la méthode de prévision
Soit h Valeur de l'horizon de prévision
Soit $\underline{\boldsymbol{XHist_{t-h}}}$ Historique d'individus observés (jusqu'au pas de temps $t-1-h$ (temps-réel) ou sur une période spécifique (hors ligne)) - Matrice au format d'entrée : $t-h \times Card(\boldsymbol{X_t})$
Soit $\underline{\boldsymbol{YHist_t}}$ Historique d'individus observés (jusqu'au pas de temps $t-1$ (temps-réel) ou sur une période spécifique (hors ligne)) Matrice au format de sortie : $t-h \times Card(\boldsymbol{Y_t})$
Soit $\widehat{\boldsymbol{Y_t}}$ Résultat de prévision des états de trafic depuis l'instant $t-h$
Soit $\boldsymbol{X_{t-h}}$ Individu à l'instant $t-h$
▷ NOUVELLE MESURE À L'INSTANT t
$\boldsymbol{Y_t} \leftarrow$ Mesure(t)
▷ ESTIMATION DE L'ERREUR DE PRÉVISION :
$\boldsymbol{Error^t_{t-h}} \leftarrow$ Error($\boldsymbol{Y_t}, \widehat{\boldsymbol{Y_t}}$) Structure recueillant les diverses formes d'erreurs
▷ MISE À JOUR DES HISTORIQUES :
Intégration du couple $(\boldsymbol{X_{t-h}}, \boldsymbol{Y_t})$ aux historiques
$\underline{\boldsymbol{YHist_t}} \leftarrow \underline{\boldsymbol{YHist_t}} \cup \boldsymbol{Y_t}$
$\underline{\boldsymbol{XHist_{t-h}}} \leftarrow \underline{\boldsymbol{XHist_{t-h}}} \cup \boldsymbol{X_{t-h}}$
▷ CRÉATION DE L'INDIVIDU D'ENTRÉE
$\boldsymbol{X_t} \leftarrow$ BuildX($\boldsymbol{Y_t}, \underline{\boldsymbol{YHist_t}}$)
▷ GESTION DES DONNÉES MANQUANTES ET ADAPTATION DE L'HISTORIQUE :
$\boldsymbol{Bool_{NotMissingInHist}} \leftarrow$ IsNotMissing($\underline{\boldsymbol{XHist_{t-h}}}$)
$\boldsymbol{Bool_{InHist}} \leftarrow$ RestrictHist($\underline{\boldsymbol{XHist_{t-h}}}$)
$\underline{\boldsymbol{CutHistX}} \leftarrow \underline{\boldsymbol{XHist_{t-h}}}(\boldsymbol{Bool_{NotMissingInHist}} \ \& \ \boldsymbol{Bool_{InHist}}, :)$
$\underline{\boldsymbol{CutHistY}} \leftarrow \underline{\boldsymbol{YHist_t}}(\boldsymbol{Bool_{NotMissingInHist}} \ \& \ \boldsymbol{Bool_{InHist}}, :)$
▷ PRÉVISION :
$\widehat{\boldsymbol{Y_{t+h}}} \leftarrow$ Prediction $(\boldsymbol{X_t}, \underline{\boldsymbol{CutHistX}}, \underline{\boldsymbol{CutHistY}}, \boldsymbol{Hyperprarameters})$
▷ SORTIES :
$\boldsymbol{Error^t_{t-h}}$; $\widehat{\boldsymbol{Y_{t+h}}}$.

Cohérentes avec les exigences de l'approche globale préalablement identifiées, les modifications concernent principalement trois aspects :

- la dimensionnalité du label $\boldsymbol{Y_t}$ décrivant non plus l'état de trafic sur une section à l'instant t, mais sur l'ensemble des sections du réseau considéré. Son historique associé $\underline{\boldsymbol{YHist_t}}$ subit par conséquent les mêmes modifications.
- l'algorithme de prévision $Prediction$ devant s'adapter à la nouvelle dimensionnalité de la sortie, ainsi qu'à la nouvelle évaluation de sa fonction de coût. Cette dernière est adaptée pour estimer les paramètres internes durant la phase d'apprentissage à partir d'un critère multidimensionnel. Le chapitre 3 détaille les adaptations requises sur différentes méthodes de prévision.
- l'évaluation de l'erreur de prévision $Error$ n'est plus seulement évaluée localement pour une section, mais, globalement, à l'échelle du réseau. Ceci est une conséquence directe de l'évolution de la fonction de coût.

L'adaptation et la gestion de l'historique ne sont pas impactées par la transition à l'échelle globale. En revanche, elles contribuent pleinement à la mise en oeuvre en temps-réel du cadre prédictif.

2.2.3 Composantes fonctionnelles du cadre prédictif à adapter à l'approche globale

La transition d'une approche locale à une approche globale nécessite une modification de la dimensionnalité de la sortie, du procédé interne de la fonction de prévision et l'introduction de nouveaux échelons d'estimation de l'erreur de prévision.

2.2.3.1 Définition des variables d'entrée et de sortie

Le problème de prévision de l'approche globale vise à optimiser la fonction objectif à l'échelle du réseau, quand les approches locales optimisent la prévision d'une section isolée. Les entrées et sorties du modèle de prévision se voient adaptées en conséquence.

Le vecteur d'état en entrée $\boldsymbol{X_t}$ représente l'ensemble des sections du réseau considéré.

$$\boldsymbol{X_t} = \left(w_t^X Z_t^1, ..., w_{t-\mu}^X Z_{t-\mu}^1, w_t^X Z_t^2, ..., w_{t-\mu}^X Z_{t-\mu}^N\right)$$

où N est le nombre de sections composant le réseau, $\forall t, \forall i \in [1, N], Z_t^i$ l'état de trafic sur la section i à l'instant t et $(w_t^X, ..., w_{t-\mu}^X)$ un ensemble de pondérations. Par la suite, les pondérations adoptent la forme suivante : $\forall j \in [0, \mu], w_{t-j}^X = \dfrac{1}{j+1}$. Les états les plus récents se voient donc favorisés.

La vision série temporelle du trafic pousse à définir $\boldsymbol{X_t}$ depuis les états expérimentés à l'instant t, mais également aux μ instants antérieurs. Le paramètre μ est un hyperparamètre définissant la taille de mémoire optimale. Dans le cadre d'une introduction naïve de la structure spatio-temporelle, tous les événements expérimentés par le réseau jusqu'au μ pas de temps antérieurs sont conservés en entrée.
De même que pour l'approche locale, un lien existe entre les labels et les composantes du vecteur d'entrée. La fonction $BuildX$ de l'algorithme 1 fait référence à la construction de $\boldsymbol{X_t}$ depuis les mesures d'états de trafic ou labels antérieurs.

La prévision $\boldsymbol{Y_{t+h}}$ est un vecteur recueillant, pour chacune des N sections, l'état prévu par le modèle supervisé à partir de $\boldsymbol{X_t}$. $\boldsymbol{Y_{t+h}}$ s'écrit sous la forme vectorielle suivante :

$$\boldsymbol{Y_{t+h}} = (Y_{t+h}^1, ..., Y_{t+h}^N)$$

où $\forall i \in [1, N], Y_{t+h}^i = Z_{t+h}^i$ est une estimation de la caractéristique de trafic (vitesse, débit) expérimentée par la section i à l'horizon $t + h$.

Les historiques associés adoptent des formes liées aux dimensions des entrées et sorties. Ils sont dissociés dans l'algorithme 1, mais se composent en réalité d'un ensemble de T_{hist} couples individu $\times$ label : $(\boldsymbol{X_t}, \boldsymbol{Y_{t+h}})_{t\in[1,T_{hist}]}$ expérimentés par le réseau. La matrice associée à l'historique comparé à l'individu d'entrée est qualifiée de matrice $Individu \times Variables$. Chaque ligne correspond à un nouvel individu.

2.2.3.2 Méthode de prévision

2.2.3.2.1 Cas général

Dans la méthode de prévision, $Prediction$ de l'algorithme 1, l'adaptation la plus importante porte sur l'optimisation de sa fonction objectif ou perte, qui doit se voir adaptée aux exigences multidimensionnelles du passage à l'échelle. Une telle modification est exigée en vue de parvenir à un modèle unifié. Les hyperparamètres se voient alors optimisés pour toutes les dimensions de sortie et, donc, pour l'ensemble des sections simultanément. Un seul et même jeu d'hyperparamètres permet dès lors la prévision d'un ensemble cohérent de sections. Enfin, la mesure de similarité, traduite par une distance ou une fonction à noyaux pour les algorithmes à gestion d'historique, est adaptée aux dimensions des individus considérés.
Le passage d'une approche locale à globale est illustré depuis la méthode simple des K plus proches voisins (KNN). Des extensions plus complexes sont exposées en chapitre 3.

2.2.3.2.2 Cas de la méthode des K plus proches voisins

La méthode prédictive des k plus proches voisins (KNN) consiste en une régression non paramétrique. Simple et intuitive d'usage, elle est particulièrement employée en trafic lors de comparaisons entre approches paramétriques et non paramétriques [101, 55]. Elle s'avère souvent particulièrement efficace [142, 143, 93].
Cette méthode repose sur l'idée que le label associé à un nouvel individu dépendra de celui porté par son voisinage. S'appuyant sur une distance euclidienne, le voisinage, restreint aux K plus proches individus de l'historique, est exploité pour déterminer le label du nouvel individu à partir de leurs propres labels.

À l'usage, une telle méthode de régression ne possède pas de paramètres internes. Sa fonction de perte ne dépend que des hyperparamètres, dont le nombre de voisins directs K. Une fois les hyperparamètres définis, le modèle est parfaitement identifié, la phase d'apprentissage est donc immédiate. En revanche, la phase d'évaluation peut se révéler plus longue, puisque l'individu testé se voit comparé à l'intégralité de l'historique. L'adaptation de cette régression non paramétrique à l'approche globale est, dès lors, intuitive : seul le label de sortie est redimensionné.

La méthode KNN se prête bien à la comparaison au traitement d'images telle qu'illustrée en Figure 2.4. La technique du KNN multidimensionnel est appliquée pour prédire une image du réseau, lui-même composé des caractéristiques de trafic associées aux sections. L'individu en entrée de l'algorithme, est un petit film composé des $\mu + 1$ derniers clichés expérimentés par

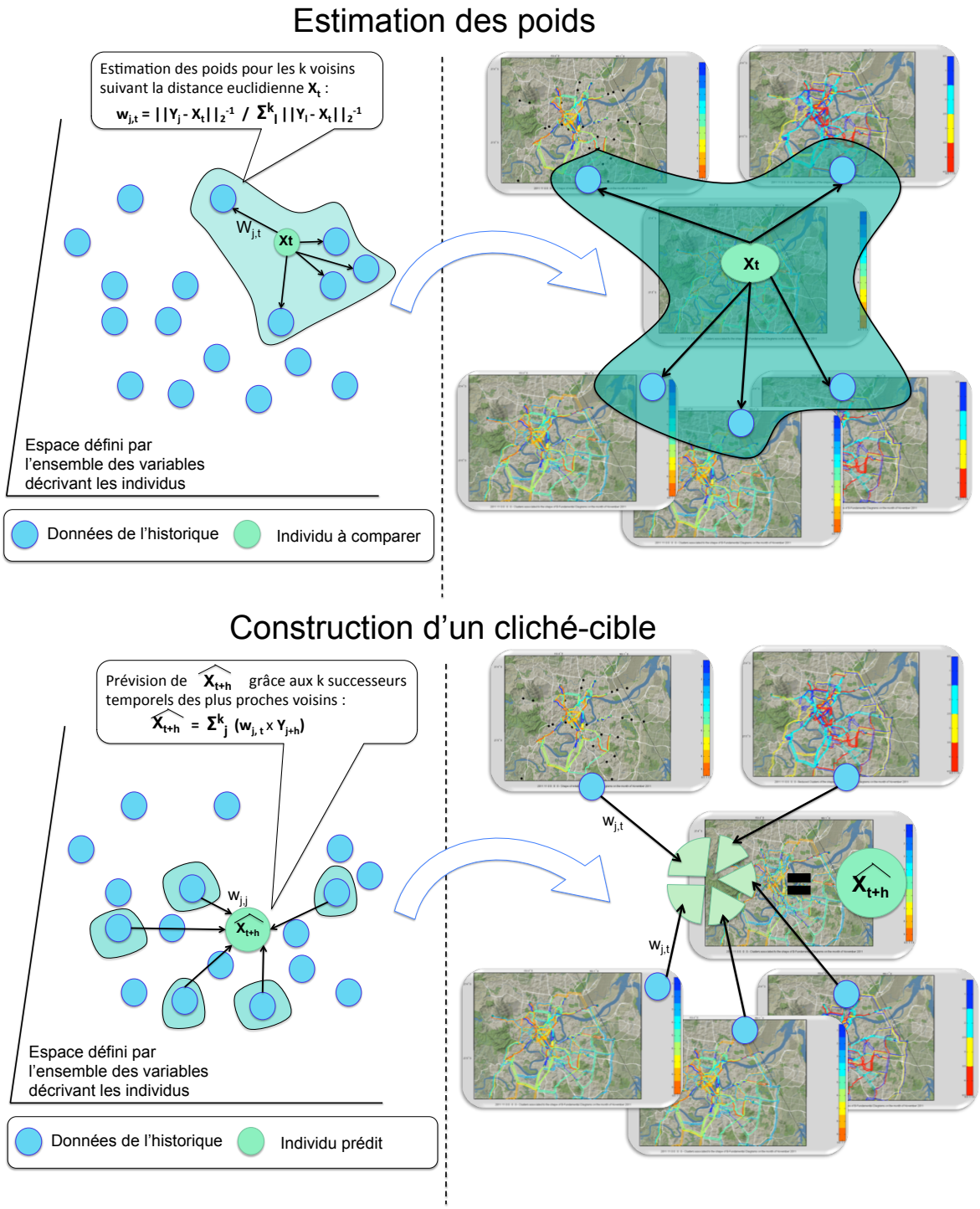


Fig. 2.4.: Application de la règle de décision des k plus proches voisins à un cliché du réseau (μ=0)

le réseau. Le petit film se voit confronté à ceux d'un historique de films du réseau d'ores et déjà connus. Les K films les plus proches de l'individu contribuent, à partir de leurs successeurs temporels à l'horizon h, à construire la suite du film évalué.

L'algorithme de régression par le KNN a été adapté afin de pondérer les K plus proches clichés du réseau issus de la base de données. Une moyenne pondérée des états des successeurs temporels immédiats des K plus proches voisins permet la génération des caractéristiques de trafic prédites aux horizons temporels $t+h$. Les différentes étapes du procédé sont les suivantes :

- **Identification des voisins** : la donnée d'entrée est un prototype caractérisé par le vecteur d'état $\boldsymbol{X_t}$, dont il faut prévoir l'état futur. L'historique de données est composé d'un ensemble de prototypes $(\boldsymbol{X_j})_{j\in[1,T_{Hist}]}$. Chacun des prototypes de l'historique $\boldsymbol{X_j}$ représente un cliché des états de trafic à un instant j. Aucune relation n'est imposée entre t et j.
- **Estimation des poids** : basé sur une distance euclidienne $||\boldsymbol{X_j} - \boldsymbol{X_t}||_2$, le poids des K plus proches prototypes est fixé de la manière suivante :

$$w_{i,t}^{KNN} = \frac{||\boldsymbol{X_i} - \boldsymbol{X_t}||_2^{-1}}{\sum_{j=1}^{K} ||\boldsymbol{X_j} - \boldsymbol{X_t}||_2^{-1}}, i \in 1, ..., K.$$

  Les individus de l'historique les plus proches du vecteur d'état courant contribuent le plus à la construction du label de sortie.

- **Détermination du profil cible** : les clichés des successeurs immédiats des K plus proches voisins sont utilisés pour construire le cliché cible $\widehat{\boldsymbol{Y_{t+h}}}$ résultant de leur moyenne pondérée.

$$\widehat{\boldsymbol{Y_{t+h}}} = \sum_{i=1}^{K} w_{i,t}^{KNN} \times \boldsymbol{Y_{i+h}}.$$

Cette approche multidimensionnelle du KNN est appelée **MKNN**. Deux versions du KNN apparaissent dans ce chapitre :

- une version originelle (**Direct**), pour laquelle le label prévu $\widehat{\boldsymbol{Y_{t+h}}}$ représente bien les caractéristiques de trafic expérimentées par le réseau ;
- une version indirecte (**Dev**), pour laquelle le label prévu $\widehat{\boldsymbol{Y_{t+h}}}$ correspond à l'évolution des caractéristiques de trafic sur le réseau. Le label $\widehat{\boldsymbol{Y_{t+h}}}$ se voit alors sommé à l'état courant $\boldsymbol{Y_t}$ pour aboutir à l'estimation des caractéristiques de trafic.

2.2.3.3 Erreur de prévision

L'erreur de prévision mesure l'écart entre le label prévu et le label effectif. Dans cette perspective, elle est extrêmement liée, dans le domaine du trafic, à la fonction de coût définie pour toute approche par apprentissage supervisé. Elle dépend donc du choix des hyperparamètres.

Dans la littérature du trafic, la performance de prévision est évaluée à travers les mesures de l'erreur. Des indicateurs variés ont été développés, mais s'appuient tous sur la différence entre prévision et observation :

$$\forall t, \forall l \in [1, N], \quad Error_t^{l,t+h} = Y_{t+h}^l - \widehat{Y_{t+h}^l}.$$

Suivant les objectifs, cette mesure est souvent agrégée en conséquence :

- à l'échelle locale de la **section** :
  - l'erreur absolue (AE) :
    $\forall t, \forall l \in [1, N], \quad AE_t^{l,t+h} = |Y_{t+h}^l - \widehat{Y_{t+h}^l}|$
  - l'erreur moyenne quadratique ($RMSE_{section}$) agrégeant le carré des erreurs sur une période d'estimation T :
    $\forall l \in [1, N], \quad RMSE_{section}^l = \sqrt{\frac{1}{T} \sum_{t=1}^{T_{test}} \|Y_{t+h}^l - \widehat{Y_{t+h}^l}\|^2}$
- à l'échelle du **réseau** :
  - l'erreur moyenne quadratique ($RMSE_{TimeStep}$) agrégeant le carré des erreurs sur l'ensemble des N sections du réseau à l'instant t :
    $\forall t, \quad RMSE_{TimeStep}^{t+h} = \sqrt{\frac{1}{N} \sum_{l=1}^{N} \|Y_{t+h}^l - \widehat{Y_{t+h}^l}\|^2}$
  - La moyenne des erreurs quadratiques moyennes ($RMSE_{global}$) :
    $RMSE_{global} = \sqrt{\frac{1}{T \times N} \sum_{t=1}^{T} \sum_{l=1}^{N} \|Y_{t+h}^l - \widehat{Y_{t+h}^l}\|^2}$

L'évaluation de la performance depuis l'erreur absolue de prévision est la plus répandue. Elle est bien adaptée pour identifier les sections produisant, dans l'absolu, l'erreur la plus importante. Cependant, il s'avère difficile de comparer les sections du réseau les unes par rapport aux autres sans souffrir d'un biais. Toutes les sections n'expérimentent pas la même intensité de flux de trafic, ne possèdent pas des capacités ou des réglementations de vitesse identiques. Les caractéristiques de trafic recueillies varient en intensité spatialement (d'une section à l'autre) et temporellement (d'une période de la journée à une autre).
Un indicateur d'erreur relative est introduit et pallie une telle difficulté. Il adopte les mêmes échelles d'évaluation que les erreurs absolues. On définit ainsi :

- à l'échelle locale de la **section** :
  - le pourcentage d'erreur absolue (APE) :
    $\forall t, \forall l \in [1, N], \quad APE_t^{l,t+h} = |\frac{Y_{t+h}^l - \widehat{Y_{t+h}^l}}{Y_{t+h}^l}|$
  - Le pourcentage moyen d'erreur ($MAPE_{section}$) agrégeant les erreurs relatives sur une période d'estimation T :
    $\forall l \in [1, N], \quad MAPE_{section}^l = \frac{1}{T} \sum_{t=1}^{T} |\frac{Y_{t+h}^l - \widehat{Y_{t+h}^l}}{Y_{t+h}^l}|$
- à l'échelle du **réseau** :
  - Le pourcentage d'erreur ($MAPE_{TimeStep}$) agrégeant les erreurs relatives sur l'ensemble des N sections du réseau à l'instant t :
    $\forall t, \quad MAPE_{TimeStep}^{t+h} = \frac{1}{N} \sum_{l=1}^{N} |\frac{Y_{t+h}^l - \widehat{Y_{t+h}^l}}{Y_{t+h}^l}|$
  - La moyenne des pourcentages d'erreur ($MAPE_{global}$) :
    $MAPE_{global} = \frac{1}{T \times N} \sum_{t=1}^{T} \sum_{l=1}^{N} |\frac{Y_{t+h}^l - \widehat{Y_{t+h}^l}}{Y_{t+h}^l}|$

Pour l'analyse à l'échelle du réseau, les deux indicateurs de performance décrits précédemment seront utilisés pour évaluer le bon fonctionnement du réseau. Le premier indicateur ($RMSE_{TimeStep}$ ou $MAPE_{TimeStep}$) fournit une estimation pour tout pas de temps de la qualité de prévision d'un cliché complet du réseau. Le second indicateur ($RMSE_{global}$ ou $MAPE_{global}$) évalue la performance agrégée sur les périodes de la journée et les sections. Cet indicateur décrit donc le comportement moyen de l'approche globale.

2.2.4 Composantes fonctionnelles classiques du cadre prédictif

2.2.4.1 Gestion des données manquantes

Lors de l'introduction de la phase de supervision du réseau et, plus spécifiquement, des cas d'étude (section 1.2.3), il a été montré qu'un enjeu majeur de l'exploitation des sources de données résidait dans le traitement des informations manquantes. Face aux causes multiples d'absence de données (détecteur défaillant, phénomène météorologique etc), toute source d'information est touchée. Cet enjeu se révèle crucial dans le cadre des méthodes de prévision données-dépendantes (*data-driven*). L'aspect série temporelle du flux de trafic est, en effet, directement affecté. Or, la continuité temporelle des caractéristiques de trafic est nécessaire :

- pour garantir l'existence de couples Individus $\times$ Label ;
- pour construire des individus d'entrée (ou variables explicatives) complets.

L'absence d'une seule information quant aux caractéristiques de trafic d'un individu d'entrée empêche sa définition et limite, en principe, la possibilité de prévoir les caractéristiques de trafic à venir. Le phénomène de données manquantes est particulièrement gênant lors de la construction de l'historique d'apprentissage.

2.2.4.1.1 Complétion des données dans la littérature

Afin de pallier une telle difficulté, la littérature du trafic est riche en propositions. La plupart des méthodologies s'adressent à un traitement hors-ligne, *i.e.* applicable uniquement en dehors de la fonction prédictive, mais quelques unes sont d'ores et déjà développées pour intervenir en ligne [103, 144]. Les approches rencontrées visent à reconstruire les caractéristiques de trafic manquantes depuis les états antérieurs et/ou les phénomènes se produisant au voisinage. Quatre méthodologies se distinguent de par leur degré de complexité :

1. la méthode de **complétion temporelle** : les données manquantes sont complétées à partir des événements déjà expérimentés par la seule section [145, 122, 146]. Seul l'aspect série temporelle est exploité. Des régressions simples, des splines ou des lissages exponentiels sont les approches les plus basiques [147]. Des versions plus évoluées font intervenir des méthodes de classification supervisée [146], voire des analyses en composantes principales probabilistes (Probabiliste ACP) [122].
2. la méthode de **complétion spatio-temporelle** : les données manquantes sont complétées depuis les événements expérimentés par la section et son voisinage [148, 149].

3. la méthode à **complétions multiples** : la valeur manquante est estimée de multiples fois tantôt via une méthode aléatoire [150], tantôt via diverses approches de classification [151], la moyenne pondérée par un degré de fiabilité aboutit à la valeur cible. La variance des multiples valeurs proposées permet d'évaluer le degré de confiance de toute valeur complétée.
4. la méthode de **complétion intégrée** à la fonction de prévision : par modification de l'espace d'entrée ou des variables explicatives [152, 103] (décomposition en signaux, ACP bayésienne *etc*), par intégration d'un réseau bayésien [153, 148] ou, pour le cas de réseaux de neurones, par intégration d'une couche de neurones de prétraitement [154, 144], telles sont les complétions intégrées aujourd'hui rencontrées.

L'approche de complétion par série temporelle se révéle très performante lors de courtes périodes d'absence de données. Pour de plus longues périodes, en revanche, les approches spatio-temporelles ou intégrées sont recommandées.

2.2.4.1.2 Procédé de traitement des données manquantes adopté

Le procédé de traitement des données manquantes reste dans la lignée des approches à complétion temporelle. Le procédé est véritablement créé en vue de pallier les difficultés induites par des individus incomplets. Pour tout couple Individu $\times$ Label $(\boldsymbol{X_t}, Y_{t+h})$, deux configurations sont envisageables :

- Soit l'individu et/ou le label sont vides de toute information, auquel cas la prévision n'a pas de sens et n'est pas produite depuis ce couple. **Le couple n'est pas intégré à l'historique d'apprentissage et est supprimé**.
- Soit les informations de trafic sont complètes ou partielles, auquel cas, l'individu et ou le label subissent **un procédé de complétion**.

Le résultat de ce test aboutit à la variable $Bool_{NotMissingInHist}$ résultant de la fonction $IsNotMissing$ dans l'algorithme 1. Il est à remarquer que le test élimine directement toute période sans information de trafic excédant $\mu + 1$ pas de temps.

Le **procédé de complétion** des données manquantes est accompli en deux phases :

1. **Substitution d'une valeur nominale à la valeur manquante**. Des profils nominaux de caractéristiques de trafic sont construits depuis un ensemble de valeurs issues de l'historique d'apprentissage. A chaque typologie de journée, est associé un profil typique de 24 heures obtenu en moyennant les caractéristiques de trafic à chaque pas de temps. Sept typologies de jour de semaine sont distinguées : cinq profils distincts du lundi au vendredi pour les jours ouvrés, un profil pour le Week-End et un dernier profil pour les congés. Ce partage en typologies a été validé statistiquement sur les données Bluetooth de Brisbane [155]. Toute valeur manquante est remplacée par celle du profil nominal de 24h associé au même pas de temps et même typologie de jour. Le procédé est illustré en Figure 2.5.
2. **Intégration de la valeur de substitution au sein de la série temporelle** par lissage.
   Le lissage est opéré par un filtre de type moyenne glissante pondérée défini par la relation

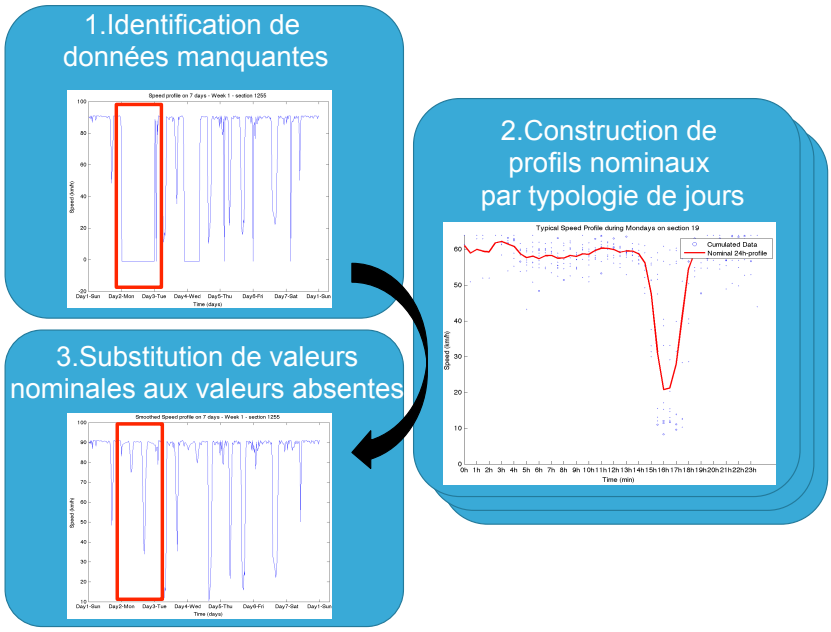
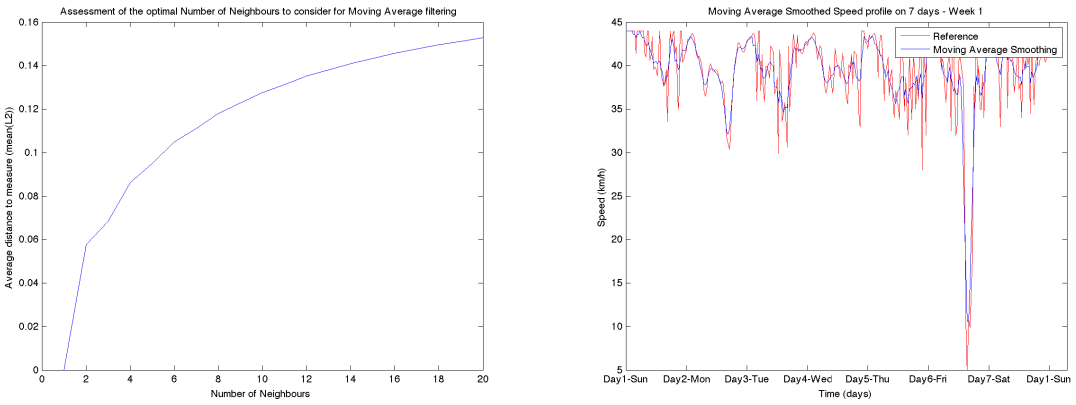


Fig. 2.5.: Illustration du procédé de complétion des données manquantes via la construction de profils quotidiens par typologie de jours de semaines (Application : données vitesses depuis FCD, Pont de Cheviré, 1ère semaine de septembre 2013).

suivante pour toute valeur manquante à l'instant t sur la section l :
$\tilde{Z}_t^l = \sum_{i=t-\frac{\mathrm{T}}{2}}^{t+\frac{\mathrm{T}}{2}} w_i^{smooth} \times Z_i^l$, où Z_i^l est l'état de trafic observé ou nominal à l'instant i sur la section l. La pondération est la suivante : $\forall i \neq t, w_i^{smooth} = \frac{1}{\mathrm{T}+1}$ et $w_i^{smooth} = \frac{2}{\mathrm{T}+1}$ quand $i = t$.
La longueur de fenêtre T du lissage est obtenue en analysant l'erreur de type $RMSE_{global}$ produite par application du filtre à un ensemble de pas de temps renseignés sur les sections du réseau. La valeur optimale est considérée par recherche d'un coude au sein de la courbe d'erreur. Un exemple de résultat est exposé en Figure 2.6 depuis des données FCD de Nantes sur une section du Pont de Cheviré. L'estimation sur un mois de données FCD sur la zone de Nantes montre une erreur globale d'environ 4% pour toutes les sections et pas de temps considérés. La fenêtre optimale est évaluée à $\mathrm{T} = 9$ pas de temps.

Les deux composantes du procédé de complétion jouent un rôle bien distinct. La création de profils nominaux quotidiens suivant des typologies prédéfinies constitue une méthode efficace et rapide pour proposer une caractéristique de trafic cohérente avec les observations habituelles. Cependant, la substitution d'une valeur nominale typique en lieu et place d'une valeur absente efface toute variabilité liée à l'évolution temporelle à court-terme. L'aspect série temporelle est restitué par la pratique du lissage.

Le traitement des données manquantes aboutit à l'obtention de couples Individu $\times$ Label bien définis et donc à une base de données épurée.



(a) (b)

Fig. 2.6.: Illustration du lissage des valeurs se substituant aux données manquantes : a) Courbe d'erreur $RMSE_{global}$ en fonction de la longueur de fenêtre T b) Illustration du lissage des vitesses sur la 1ère semaine de septembre 2013 sur une section du Pont de Cheviré.

2.2.4.2 Gestion de la taille de l'historique

Habituellement, l'implémentation d'une classification supervisée opère hors-ligne depuis une base de donnée complètement connue. Les 2/3 de la base de données servent d'historique d'apprentissage fixe pour la classification du tiers restant (voir Figure 2.7). Le modèle apprend une unique fois, puis classifie tout nouvel individu en entrée. Les performances du tiers restant de la base de données estiment le potentiel de généralisation de la méthode de classification.

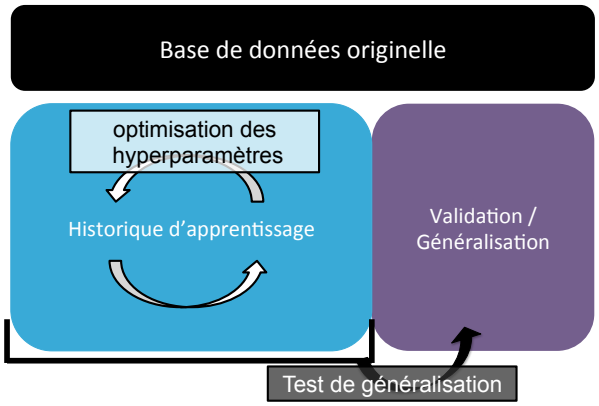


Fig. 2.7.: Gestion usuelle de la base de données dans le cadre de l'apprentissage supervisé.

S'adaptant aux exigences temps-réel du trafic, le cadre prédictif proposé autorise la mise à jour en temps-réel de l'historique d'apprentissage. La fonction $RestrictHist$ traduit cette possibilité de gestion de l'historique d'apprentissage. Les modifications de l'historique sont souples et peuvent adopter divers objectifs :

– **agrandir l'historique** d'apprentissage originel, fixé aux deux tiers de la base de données, par mise à jour en temps-réel : l'objectif est alors d'augmenter la variabilité des clichés ;

– **faire glisser l'historique** en ne conservant que les informations les plus récentes en vue de bien saisir les variations saisonnières ou de réduire le temps d'estimation ;

– **réduire l'historique** à des événements ou phénomènes préalablement identifiés et caractérisés.

L'analyse de l'état de l'art rapporte un défaut des méthodes d'apprentissage fondées sur l'historique, à savoir la forte dépendance aux informations contenues dans l'historique et l'impact de sa taille en termes de temps de calcul. En effet, pour que ces méthodes soient pertinentes, l'historique doit pouvoir contenir le moins de données possibles, mais les informations les plus variées. Les comportements des usagers sur un réseau sont souvent récurrents et des typologies de flux de trafic sont identifiables. Or, il n'est pas utile de disposer d'un grand nombre d'informations pour saisir ces typologies. De plus, ces dernières tendent à varier par saisonnalité.
L'adaptation pour application temps-réel des méthodes de prévision supervisée nous conduit alors à définir un historique d'apprentissage glissant. Ainsi, seuls les derniers événements rencontrés sur le réseau sur une période M sont pris en considération pour construire le modèle lors de la phase d'apprentissage. À chaque nouvel individu, l'historique d'apprentissage est modifié et l'apprentissage des paramètres internes évalué une nouvelle fois.
Afin de déterminer la taille de la fenêtre glissante M optimale de l'historique d'apprentissage, une étude portant sur l'évolution de l'erreur de prévision globale ($MAPE_{global}$) en fonction de la longueur de fenêtre a été menée sur le Pont de Cheviré à partir des données vitesses issues de FCD. Les résultats sont fournis pour 6 horizons de prévision de 30 min à 3h en Figure 2.8. Il en ressort qu'un historique de 30 jours suffit à saisir les principales typologies et sources de variabilité du flux de trafic. Les possibles extensions de l'historique d'apprentissage en vue d'intégrer des clichés du réseau apportant de la variabilité sont envisagées en section 5.4.

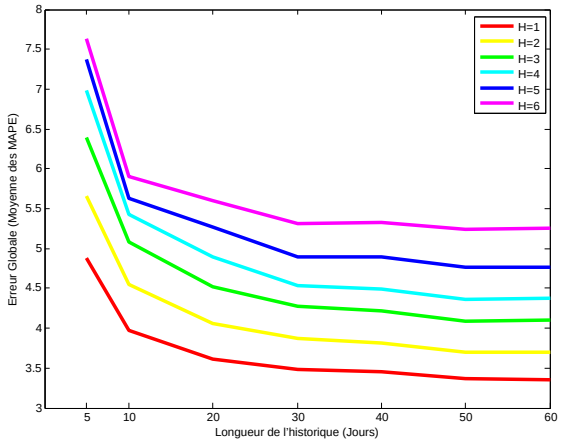


Fig. 2.8.: Evaluation de la taille d'historique glissant optimale à partir des performances de prévision de la méthode KNN sur la zone du Pont de Cheviré (Mois de novembre 2013). Chaque horizon de prévision h est réprésenté par une couleur. Le pas de temps considéré est de 30 min.

Il est à noter que la mise en ligne de la prévision ne se traduit pas obligatoirement par un nouvel apprentissage du modèle à chaque évaluation d'individus en entrée. Suivant la méthode employée, une légère modification de l'historique ne modifie pas les paramètres internes de la méthode. Au moyen d'une légère adaptation, le SVR a ainsi pu être implémenté en ligne par Castro *et al.* [115]. Les méthodes de régression non paramétriques n'ont pas de paramètres internes, si bien que le temps d'apprentissage n'en est pas altéré.

Toutes les méthodes de prévision testées par la suite adoptent un tel historique d'apprentissage glissant de $M = 30$ jours. Ce choix pour l'hyperparamètre M a été vérifié sur une méthode alternative (SVR) dans un travail encadré sur ce sujet [156].

Bien que l'historique d'apprentissage soit modifié en temps-réel, les données testées lors de la phase de validation restent inchangées. Il s'agit toujours du dernier tiers de la base d'information originelle. Sans le moindre conflit, les deux premiers tiers de la base de donnée sont maintenus pour l'optimisation des hyperparamètres.

2.3 Mise en oeuvre de la stratégie d'optimisation des hyperparamètres

2.3.1 Recherche du compromis entre représentativité et généralisation

Les hyperparamètres sont présents dans toute approche de classification supervisée et visent à assurer la capacité de généralisation des modèles appris. Ils sont intégrés au sein de la fonction de perte ou de risque empirique à minimiser dans le seul but de modérer les régles de décisions construites par le modèle appris. Leur rôle vise à intégrer la notion de bruit et de représentativité des individus de l'historique d'apprentissage. Apprenant les règles de décisions depuis des données réelles, l'espace des individus à caractériser n'est pas uniformément couvert et des biais de mesures sont possibles. Les hyperparamètres évitent aux modèles de tomber dans des optima locaux ou de trop généraliser. Un modèle de régression apte à reproduire exactement les valeurs de son ensemble d'apprentissage risque de fournir des prévisions très éloignées de la réalité. Ceci caractérise ce qui est communément appelé le sur-apprentissage. A contrario, un modèle en sous-apprentissage généralise trop les tendances, élimine la variabilité et accroît le biais. Ces phénomènes sont illustrés en Figure 2.9.

Le problème d'optimisation des hyperparamètres se résume dès lors à trouver le meilleur compromis entre capacité de reproduire des situations historiques et aptitude de généralisation sur de nouvelles situations. La complexité du problème est illustrée par la Figure 2.10 montrant l'évolution de l'erreur de prévision en fonction de la quantité d'informations apportées. La complexité du modèle traduit ainsi le choix des valeurs des hyperparamètres. Plus le modèle s'appuie sur une grande quantité d'informations, plus l'erreur de prévision des informations

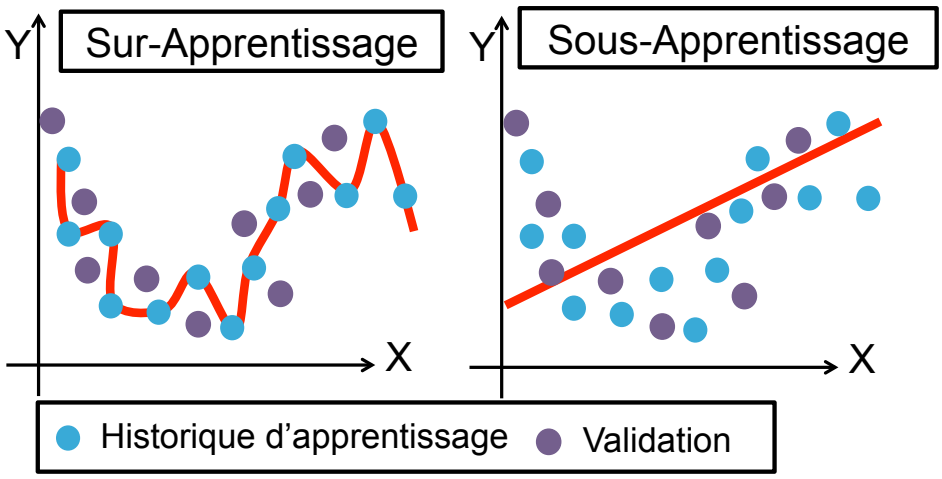


Fig. 2.9.: Illustration des phénomènes de sur - et sous-apprentissage.

de l'historique d'apprentissage diminue. Au-delà d'un certain seuil, l'erreur de prévision de nouvelles données tend à augmenter. Tout l'enjeu du calage des hyperparamètres consiste à déterminer un tel seuil malgré l'ignorance de la courbe d'erreur de validation. Il sagit du dilemne biais-variance.

2.3.2 Procédés d'optimisation des hyperparamètres

Le problème de l'optimisation des hyperparamètres est traité via diverses stratégies. Elles ont toutes en commun d'utiliser les données portées par l'historique d'apprentissage, composé de T_{hist} individus et usuellement composé de deux tiers des données originelles.
Ne pouvant disposer de la courbe de validation (voir Figure 2.10), mais ayant accès à l'intégralité de l'historique d'apprentissage, l'idée consiste à apprendre itérativement des modèles de prévision sur des sous-ensembles de l'historique d'apprentissage, puis à les tester sur un sous-ensemble distinct. Pour chaque sous-ensemble, les différentes valeurs d'hyperparamètres sont explorées. Une cartographie de la performance dépendant des valeurs prises par les hyperparamètres est alors produite, dont sont déduits les hyperparamètres optimaux pour le sous-ensemble de l'historique considéré. Une moyenne des meilleurs hyperparamètres obtenus est alors produite.

Cette approche met en oeuvre deux phases :

1. une **phase d'échantillonnage de l'historique d'apprentissage** originel en deux groupes : l'un destiné à servir à l'apprentissage et l'autre à la validation. Suivant la façon de choisir les sous-ensembles, différentes stratégies d'échantillonnage ont vu le jour, dont beaucoup se regroupent autour du principe de **validation croisée** ;
2. une **phase d'optimisation des hyperparamètres au regard du sous-historique** d'apprentissage considéré et du groupe de validation. Divers modèles liés aux valeurs prises par les hyperparamètres doivent être testés afin de cartographier la performance de

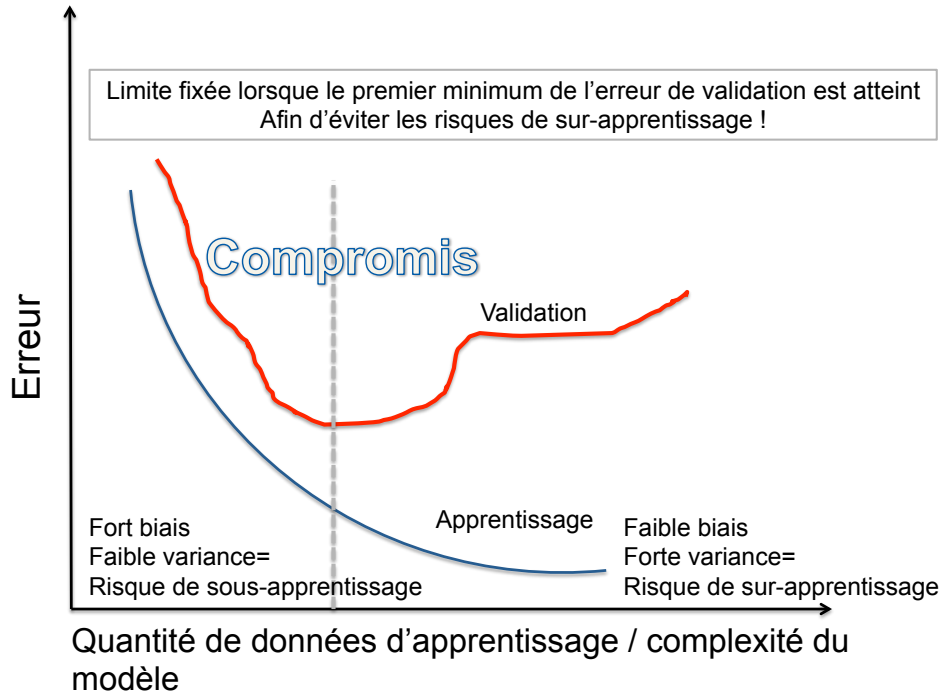


Fig. 2.10.: Illustration de la recherche d'un compromis entre généralisation et représentativité, *i.e.* du compromis biais - variance.

prévision et d'identifier les jeux d'hyperparamètres optimaux pour le sous-ensemble. Différentes stratégies de maillage de la cartographie de performance ou de recherche par heuristiques du jeu d'hyperparamètres optimaux sont employées.

2.3.2.1 Stratégies d'échantillonnage et validation croisée

Le choix des stratégies d'échantillonnage et le nombre d'itérations d'échantillonnage de l'historique originel se distinguent suivant deux catégories :

- **validation croisée exhaustive**, adoptant une stratégie de partage de l'historique en jeu d'apprentissage et de validation. Pour un apprentissage de longueur p, l'ensemble des p parmi T_{hist} combinaisons de jeux d'apprentissage est testé. Les $C^{p}_{\mathrm{T}_{hist}}$ jeux d'hyperparamètres optimaux sont moyennés pour obtenir le jeu final. Ces stratégies sont nommées *Leave-p-out cross-validation*. Elles présentent le principal désavantage d'être extrêmement longues, mais sont principalement recommandées quand l'historique est de taille faible.
- **validation croisée non exhaustive**. Elle regroupe des approximations de la stratégie exhaustive. On distingue deux catégories :
  - une approche fondée sur un sous-échantillonage aléatoire : à chaque itération, l'historique originel est partagé par tirages aléatoires en deux groupes de taille aléatoire. Son principal désavantage réside dans l'absence de garantie que tout individu apparaisse dans le jeu de validation.

– la méthode de validation croisée à m-échantillons (*m-fold cross-validation*) : l'historique originel est partitionné en m échantillons aléatoires de taille similaire. Alternativement, chacun des m échantillons est adopté comme jeu de validation, tandis que les $m-1$ restant intègrent le jeu d'apprentissage. Le jeu d'hyperparamètres final résulte de la moyenne de m jeux optimaux obtenus. Cette dernière version est préconisée lorsque l'historique est de grande taille. Elle est également plus rapide par construction.

Dans un premier temps, le choix de la stratégie d'optimisation s'est porté sur une validation croisée à m-champs. Cette dernière s'est révélée inadaptée à cause du caractère série temporelle de la base de données. Des résultats pour le moins surprenants apparaissaient lors de la phase d'optimisation. La difficulté provient de la dépendance temporelle des individus au sein de la base de données. La *m-fold cross-validation*, comme les autres approches introduites, pousse à prédire des portions du passé à partir d'événements futurs. Le problème est particulièrement marqué aux bornes des sous-groupes considérés. L'introduction d'un tirage aléatoire pour la construction des m sous-ensembles accroît le phénomène. Une approche, récemment développée par Hyndman [157] et spécifiquement adaptée pour le traitement de séries temporelles, a finalement été adoptée. Cette approche est nommée *time-series cross-validation* et consiste à progressivement agrandir le jeu d'apprentissage en testant sur l'individu associé au pas de temps suivant. La comparaison du procédé entre l'approche de validation croisée à m-échantillons et celle de validation-croisée pour séries temporelles est illustré en Figure 2.11.



Fig. 2.11.: Confrontation du procédé de validation croisée à m-échantillons à celle spécialement adaptée aux séries temporelles.

Étant donné sa prise en considération de la chronologie des événements lors de l'échantillonnage, l'approche par *time-series cross-validation* se révèle plus fiable et stable que les approches usuelles [157] au regard de l'erreur de prévision.

2.3.2.2 Choix d'une stratégie d'optimisation

Une fois le jeu d'apprentissage fixé, la phase de recherche du jeu d'hyperparamètres optimaux est enclenchée. Chaque jeu d'hyperparamètres correspond à un modèle. La performance de

chacun des modèles testés est associée au jeu d'hyperparamètres aboutissant à une cartographie de l'erreur de prévision. L'objectif est d'identifier la zone de la cartographie minimisant l'erreur de prévision et donc le jeu d'hyperparamètres associés. Il s'agit d'un véritable problème d'optimisation traité depuis des outils adaptés. Deux approches sont communément distinguées dans la littérature d'apprentisage :

- les **approches exhaustives** : il s'agit pour l'essentiel de l'**analyse par grille** (*Grid Analysis*). Cette approche simple et intuitive consiste à cartographier intégralement une zone préalablement spécifiée de l'espace des hyperparamètres en testant l'ensemble des jeux possibles associés à la zone. La cartographie établie, sa lecture aboutit naturellement à l'identification du jeu optimal de paramètres. La théorie du coude recherchant l'inflexion de la courbe d'erreur est employée en absence de minimum atteignable. Il est requis de fixer les bornes de la zone étudiée. De plus, une discrétisation des valeurs prises par les hyperparamètres se révèle utile en cas de valeurs continues. Cette approche pêche principalement par le temps nécessaire pour bâtir l'intégralité de la cartographie, d'autant que ce temps dépend de la discrétisation de l'espace. Elle souffre du fléau de la dimension (*Curse of Dimensionnality*).
- les **approches non exhaustives** : ces algorithmes d'optimisation visent à aboutir de façon itérative à une solution optimale ou s'en approchant. L'objectif consiste à limiter le nombre d'itérations aboutissant au critère de fin de l'algorithme et donc au jeu d'hyperparamètres optimal. Diverses stratégies sont observées :
  - **l'optimisation bayésienne** : cette stratégie consiste à définir de façon aléatoire ou par expertise un ensemble de modèles (ou points de la cartographie), puis à leur attribuer une confiance correspondant à la probabilité *a priori* que l'optimum soit à leur proximité. Après évaluation des performances de chacun des modèles, les probabilités *a priori* sont mises à jour. L'inférence des probabilités *a posteriori* permet alors d'identifier les prochaines zones à étudier et de restreindre progressivement la zone d'analyse.
  - **l'optimisation par recherche aléatoire** : cette technique d'optimisation repose sur une première évaluation aléatoire d'un modèle-pivot (jeu d'hyperparamètres) , puis sa comparaison par rapport à celle d'un modèle aléatoirement choisi dans son entourage. Si ce dernier se révèle meilleur, alors il devient le modèle-pivot. Dans le cas contraire, un autre modèle est estimé. Cette approche peut se révéler particulièrement longue et peu efficace suivant l'allure de la courbe de performance.
  - **l'optimisation par gradient** : ce procédé d'optimisation se rapporte au concept porté par la méthode de Newton itérée. Dans le cas d'une descente de gradient, cas bien adapté dans le contexte, un premier jeu d'hyperparamètres est choisi aléatoirement, puis les suivants sont sélectionnés par analyse de la courbure (ou gradient) de la fonction de performance. Suivant la direction d'une diminution de l'erreur, le jeu de paramètres voisins est considéré. Par itération du procédé, le modèle testé tend à s'approcher de l'optimum.

– l'optimisation **heuristique par biomimétisme** : ces méthodes sont réservées à la résolution approchée de problèmes difficiles. Elles comportent plusieurs approches s'appuyant sur des phénomènes physiques, biologiques ou socio-psychologiques. On recense ainsi les algorithmes de colonies de fourmis, les algorithmes génétiques et évolutionnistes *etc*. Leur principal avantage réside dans la réinjection régulière d'une composante aléatoire en vue d'éviter les minima locaux.

Ces approches sont plus ou moins sensibles aux optima locaux, ce qui les distinguent et orientent vers des usages spécifiques. Quand la courbe de performance est sujette à une forte volatilité, les méthodes à gradient tendent sous performer, là où des méthodes heuristiques ou aléatoires sont plus performantes.

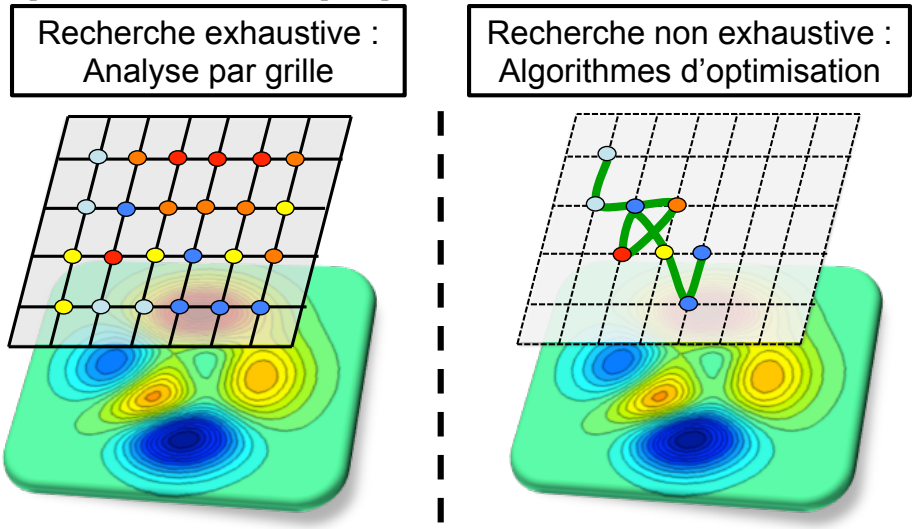


Fig. 2.12.: Illustration des deux principales approches d'optimisation : lecture exhaustive de la courbe (surface) de perforamnce versus recherche algorithmique de l'optimum.

Pour des questions de simplicité et de facilité d'usage, une analyse par grille est employée pour toutes les méthodes prédictives ne nécessitant pas plus de trois hyperparamètres discrets. L'analyse par grille permet une représentation exhaustive de la courbe de performance et facilite l'identification des hyperparamètres optimaux. Ceci est d'autant plus vrai que l'erreur de prévision peut ne jamais connaître de minimum, auquel cas la règle du coude (*i.e.* identification du fort changement de courbure) est utilisée.
Pour toute approche de prévision excédant trois hyperparamètres, l'optimisation est opérée en deux phases : (1) une **analyse par grille** ciblant la délimitation de l'espace à étudier ; (2) la mise en œuvre d'un **algorithme d'optimisation par colonies de fourmis** (ACO). Ce dernier processus vise à fournir un jeu d'hyperparamètres précis et non discret. L'optimisation par les colonies de fourmis n'est pas directement appliquée car souvent longue à converger.

Le choix de l'**algorithme ACO** [1] est lié aux objectifs d'usage à venir. Ce dernier a été observé dans la littérature du trafic avec les régressions à vastes marges (SVR) [109, 111]. Initialement

1. L'algorithme ACO implémenté dans notre travail repose sur la librairie MIDACO 4.0 : `http://www.midaco-solver.com`.

proposé par Dorigo [158], le processus de la méthode est basé sur le principe que chaque fourmi (agent) est susceptible d'atteindre le plus court chemin entre sa fourmilière et la nourriture. En envoyant les fourmis dans diverses directions aléatoires, différents chemins sont parcourus vers la source de nourriture. Sur leur passage, les fourmis déposent des phéromones détectables par leurs congénères, mais volatiles avec le temps. Lorsque la source de nourriture est trouvée, chaque fourmi se rend à la source de nourriture et en ramène au nid, renforçant au passage le taux de phéromones. Explorant de façon aléatoire d'autres chemins, le plus court chemin reliant la nourriture au nid est ainsi identifié. Bien que chaque agent ait un comportement isolé, l'attitude collective conduit à l'optimum global.
Dans le cas de la recherche du modèle optimal, chaque jeu d'hyperparamètres constitue un chemin potentiel vers la nourriture. La distance parcourue par ce chemin est l'erreur de prévision. Depuis ces informations, une probabilité d'usage du chemin est établie et dépend du taux de phéromone. À chaque itération, de nouvelles fourmis sont générées et le taux de phéromone de chaque chemin et donc sa probabilité d'usage sont mises à jour. Les nouvelles fourmis adoptent, suivant les probabilités prédéfinies, un itinéraire vers la nourriture. Par itérations successives, le comportement moyen des agents aboutit au modèle optimum.

La mise en oeuvre algorithmique du procédé d'optimisation complet est présentée en Annexe A. L'optimisation des hyperparamètres étant accomplie, le cadre prédictif est calé pour test sur le jeu de validation, en adoptant une mise en situation temps-réel.

2.4 Tests sur deux cas d'étude

Les données recueillies depuis les deux sources distinctes introduites en section 1.2.3 sont analysées. Le réseau de Brisbane (Australie) et celui du Pont de Cheviré (Nantes, France) sont tous deux modélisés et le cadre prédictif y est développé à partir de la méthode KNN.
Pour ces deux cas d'étude, trois mois de données sont à disposition. Les deux premiers mois permettent l'optimisation des hyperparamètres de l'approche (K, μ et M), tandis que le dernier mois évalue la capacité de généralisation des modèles. M a été fixé précédemment pour toute méthode de prévision à 30 jours. K et μ sont optimisés à partir des stratégies décrites en section 2.3.2 depuis les deux premiers mois de données.

Les approches locales et globales sont comparées à partir de la méthode de prévision KNN. Cependant, toutes les méthodes de prévison adoptées partagent le cadre prédictif suivant :

- **l'approche globale** reçoit en entrée $(\mu + 1)$ clichés successifs des caractéristiques de trafic sur l'intégralité du réseau. Elle fournit à la prévision simultanée des caractéristiques de trafic sur l'ensemble des sections.
- **l'approche locale** reçoit en entrée les caractéristiques de trafic de la section l étudiée et de son voisinage immédiat (voisinage directement connexe) sur $(\mu + 1)$ périodes consécutives. Seules les sections partageant un nœud commun avec la section d'étude sont intégrées aux variables d'entrée. Il en résulte la prévision sur la seule section l.

Pour chaque modèle de prévision créé, une phase d'optimisation visant à fixer les hyperparamètres est requise. L'approche globale ne nécessite donc qu'une optimisation par horizon de prévision h, soit N_h recherches de jeux d'hyperparamètres. En revanche, les approches locales ont besoin de définir un modèle pour chacune des N sections du réseau, soit $N \times N_h$ jeux d'hyperparamètres optimaux. Le nombre de sections composant les réseaux étudiés est conséquent, de même que le temps d'estimation des hyperparamètres. En pratique, seul un nombre restreint de sections (environ 10% du réseau) est concerné par une optimisation des hyperparamètres. Ces sections bien choisies ont été préalablement identifiées comme critique à tout instant sur le réseau (voir chapitre 4). Les hyperparamètres des autres sections sont obtenus depuis un algorithme de classification de type KNN appliqué via une distance prédéfinie entre sections. Suivant les caractéristiques des hyperparamètres considérés, la distance prend la forme d'une distance physique (proche voisinage en termes de connexité), d'une similarité entre les vitesses moyennes expérimentées *etc*. Le temps de construction des modèles prédictifs pour toute section du réseau est ainsi réduit à un temps admissible. Seuls $N_{critique} \times N_h$ modèles sont alors réellement à optimiser pour toute approche locale.

Afin de disposer d'une **référence de prévision**, une méthode simple est introduite. Le procédé consiste à utiliser l'état de trafic observé à l'instant t comme prévision à l'instant $t+h$: $\widehat{\boldsymbol{Y_{t+h}}} = \boldsymbol{Y_t}$. Il est alors supposé que le trafic n'évolue pas à l'horizon considéré. Cette méthode, qui s'apparente à un processus de marche aléatoire, est qualifiée de **Naïve** par la suite.
Outre les formulations par approche locale ou globale, la méthode KNN est développée sous ses formes directes (**Direct**) et indirectes (**DEV**) précédemment décrites.

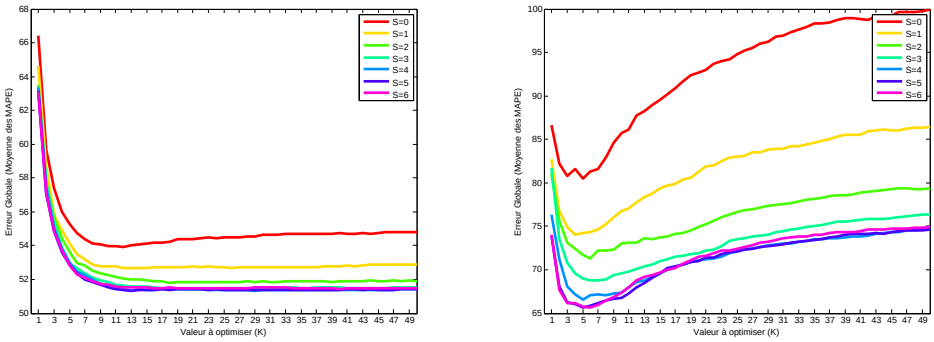
2.4.1 Prévision sur le réseau de Brisbane à partir de données Bluetooth

Les données Bluetooth à disposition s'étalent sur trois mois : janvier et février 2012, puis octobre 2012. Pour chacune des 466 sections du réseau, le débit d'usage de la section et la vitesse du flot (*i.e.* vitesse moyenne spatiale) sont estimés avec un pas de 1 heure. Les horizons de prévision h s'étalent de 1h à 3h.

2.4.1.1 Optimisation des hyperparamètres

Le cadre prédictif appliqué via la méthode KNN ne nécessite que deux hyperparamètres. Une analyse par grille faisant varier K dans l'intervalle $[1, 50]$ et la mémoire temporelle μ entre 0 et 6 pas de temps est pratiquée. Le jeu optimal d'hyperparamètres est lu directement sur le graphique par recherche du minimum ou du coude. La Figure 2.13 illustre l'allure des courbes de performances obtenues depuis l'approche globale (MKNN) appliquée au débit de trafic. Il est constaté que l'erreur de prévision est déjà très élevée. Les variations saisonnières du taux de pénétration (ou taille de l'échantillon de trafic) sont certainement à mettre en cause.

L'ensemble des hyperparamètres considérés pour construire les modèles de prévision aux trois horizons envisagés est exposé dans la table 2.1. Aucune tendance ne semble véritablement se



(a) (b)

Fig. 2.13.: Représentation de la cartographie de performances depuis une analyse par grille appliquée à l'approche globale MKNN sur le réseau de Brisbane : a) Horizon de prévision de 1h ; b) Horizon de prévision de 3h. Evolution de la performance de prévision en fonction des hyperparamètres K et μ.

distinguer d'un horizon de prévision à un autre. Il apparaît néanmoins que la mémoire μ requise croît avec l'horizon de prévision. Peu de différences sont observées suivant les caractéristiques de trafic (débit, vitesse).

MKNN		**Hyperparamètre**	**Horizon** h		
			1h	**2h**	**3h**
Variables à prévoir	**Débit** (Q_d)	K_Q	9	8	8
		μ_Q	3	4	4
	Vitesse (V_L)	K_V	8	9	9
		μ_V	2	1	3

Tab. 2.1.: Valeurs optimales des hyperparamètres obtenues pour application du MKNN au réseau de Brisbane.

Le premier apport de l'approche globale s'illustre véritablement à travers les résultats présentés dans le tableau 2.2. En effet, le temps de recherche du modèle optimum via une analyse par grille est sensiblement réduit en adoptant une approche globale. Un facteur 7 est relevé. Néanmoins, le temps nécessaire à l'optimisation par approche locale n'est pas linéairement lié à celui requis pour optimiser une approche globale. En effet, l'estimation des hyperparamètres est pratiquée sur une quarantaine de sections critiques, tandis que le facteur multiplicatif est limité à 7. Le temps d'estimation d'une approche globale est donc plus élevé que celui d'une approche locale sur une unique section.

Une fois les hyperparamètres déterminés pour toute approche, les performances de prévision sont comparées sur le mois d'octobre 2012.

1. Pour toute évaluation de performances temporelles effectuée au cours de cette thèse, le processeur disposait des caractéristiques suivantes : Intel Xeon CPU E3-1225 V2 - 3GHz - RAM : 8Go

Approche	Caractéristique de trafic	Horizon h			Temps approximatif[1]
		1h	2h	3h	
globale (MKNN)	*Débit* (Q_d)	4 609s	4 567s	4 548s	**1,1h**
	Vitesse (V_L)	3 649s	3 635s	3 645s	
locale (KNN)	*Débit* (Q_d)	27 006s	26 703s	26 994s	**7,4h**
	Vitesse (V_L)	24 633s	24 526s	24 583s	

Tab. 2.2.: Comparaison des temps d'optimisation des hyperparamètres depuis la méthode KNN : confrontation approche locale et approche globale lors d'une analyse par grille.

2.4.1.2 Performances de prévision

La performance de prévision est estimée à partir du temps d'estimation requis, son adéquation avec les objectifs temps-réel et les diverses formes d'erreur de prévision.

Le tableau 2.3 fournit une estimation du temps requis pour prévoir un mois de données suivant les approches considérées. Une fois encore, l'approche globale propose une durée de prévision nettement inférieure à celle requise par l'approche locale. L'usage de l'approche globale réduit, non seulement, la durée de la phase d'optimisation, mais également celle de la prévision en temps-réel. Les deux approches sont compatibles avec les objectifs de prévision à 1h, mais une réduction du pas de temps considéré ou une croissance de la taille du réseau se révèlerait problématique pour l'approche locale.

Approche	Caractéristique de trafic	Horizon h			Temps approximatif
		1h	2h	3h	
globale (MKNN)	*Débit* (Q_d)	14,9s	21,5s	21,2	**10-20s**
	Vitesse (V_L)	9,7s	12,3s	14,8s	
locale[2] (KNN)	*Débit* (Q_d)	2 854s	2 907s	2 915s	**48mn**
	Vitesse (V_L)	2 634s	2 663s	2 822s	

Tab. 2.3.: Comparaison des temps de prévision par la méthode KNN pour 1 mois de données : confrontation de l'approche locale à l'approche globale.

L'analyse de la Figure 2.14 met en évidence la difficulté de prévision des caractéristiques de trafic issues de Bluetooth, voire de leur imprédictabilité. L'approche naïve aboutit à des erreurs de prévision extrêmement élevées. Si, sur de courtes périodes, limitées à un mois, les informations recueillies contiennent suffisamment d'information pour retrouver les principales composantes du trafic (voir section 1.2.3), la variabilité du taux de pénétration sur plusieurs jours consécutifs s'avère problématique. Les débits partiels recueillis par Bluetooth varient suivant les sections considérées et de façon saisonnière, introduisant une forte non-stationnarité non expliquée par les variables de trafic en entrée. Seuls des facteurs externes et méconnus tels que le comportement d'activation ou non du Bluetooth sont susceptibles d'expliquer de telles variations. Il est donc naturel de constater une forte erreur de prévision des débits Bluetooth telle qu'exposée en Figure 2.14 a). Le débit partiel reste pertinent en tant qu'indicateur de

2. Résultats obtenus depuis la ParallelToolBox de Matlab.

fiabilité des mesures de vitesses. En revanche, il est plus surprenant d'observer une erreur globale ($MAPE_{global}$) conséquente pour les vitesses (cf Figure 2.14 b)). Une analyse des séries temporelles de vitesses par section souligne effectivement d'importantes non-linéarités sur certaines périodes de janvier et février 2012.

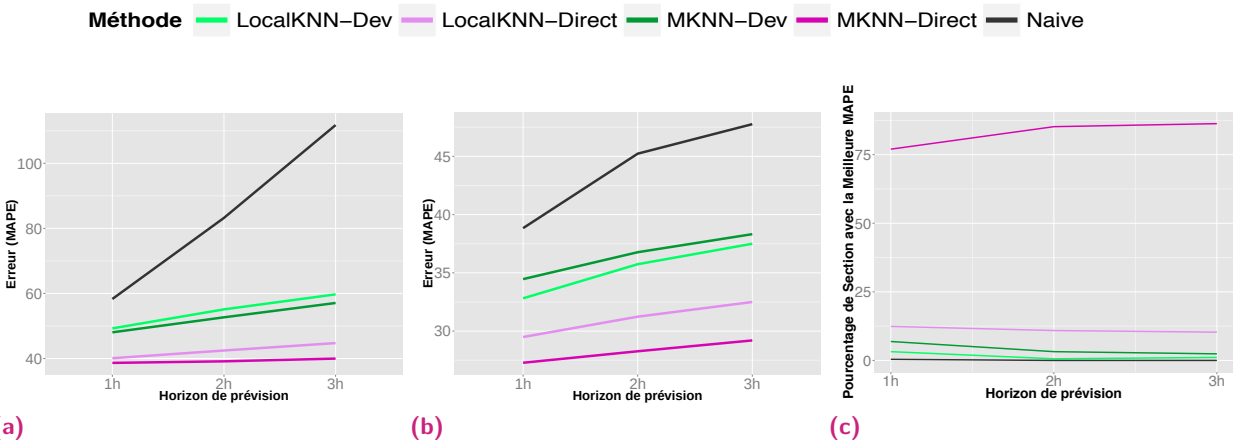


Fig. 2.14.: Confrontation des performances à l'échelle du réseau ($MAPE_{global}$) sur le mois de validation : a) Performance de prévision du débit Bluetooth ; b) Performance de prévision de la vitesse ; c) Identification des méthodes produisant le moins d'erreur ($MAPE_{section}$) en pourcentage du nombre de sections pour divers horizons.

La difficulté de prévision du jeu considéré n'empêche pas la confrontation des méthodes de prévision en analysant le comportement et la réponse de chacune des méthodes. Les résultats exposés en Figure 2.14 soulignent la performance des approches globales face aux approches locales développées. L'optimisation des hyperparamètres sur un nombre restreint de sections, puis leur généralisation entraînent une incertitude quant à l'optimalité des modèles employés pour prévoir chacune des sections par approche locale. Cependant, cette incertitude résulte de la difficulté à optimiser toutes les sections dans un temps raisonnable et est donc imputable à l'approche adoptée. Force est de constater que l'approche globale aboutit de façon globale ($MAPE_{global}$) à des performances au moins équivalentes à celles de l'approche locale. La Figure 2.14 c) montre même que quel que soit l'horizon de prévision plus de 80% des sections sont mieux prévues par l'approche globale directe (MKNN-Direct). Du fait de l'instabilité du taux de pénétration, l'approche directe est plus performante que celle estimant la variation de la caractéristique de trafic.

L'analyse de l'erreur de prévision par période de la journée est fournie en Figure 2.15. Les pics de trafic recueillent les principales périodes de congestion et sont habituellement difficiles à prévoir. Il est dès lors pertinent de constater que l'approche globale semble apte à saisir les interactions spatio-temporelles du réseau et se révèle la plus performante durant le pic du matin. Peu de distinctions sont réellement observables entre les périodes sensibles (pic du matin et du soir) et les périodes nocturnes. Les variations du taux de pénétration sont susceptibles d'expliquer un tel phénomène. Le milieu de journée semble expérimenter tout de même peu d'erreur de prévision. La différentiation des performances entre les périodes de pic de trafic et

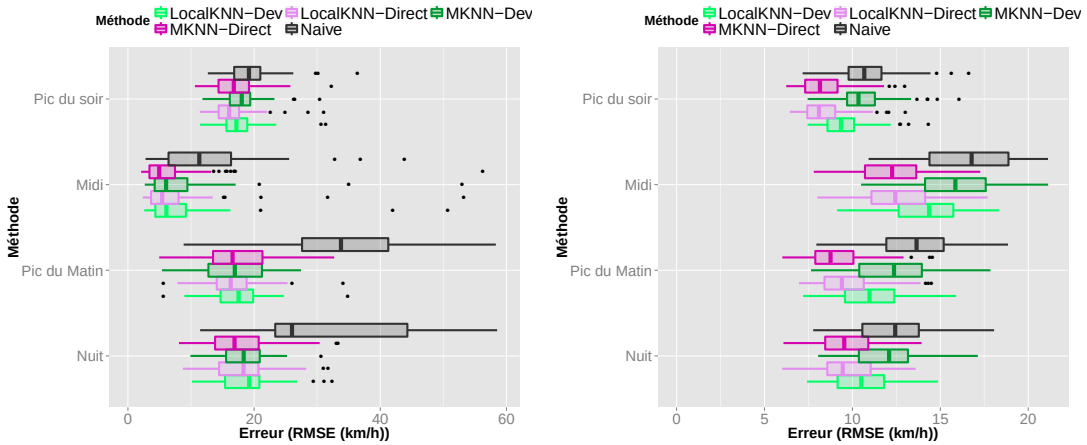


Fig. 2.15.: Distribution de l'erreur de prévision ($RMSE_{TimeStep}$) en fonction des périodes de la journée : a) Performance de prévision du débit Bluetooth à 1h ; b) Performance de prévision de la vitesse à 1h.

les périodes fluides, usuellement observée, ne l'est pas sur ce cas d'étude. Plus que les capacités des méthodes, l'irrégularité de certaines plages de données est sans doute à mettre en cause.

Une analyse plus fine de l'erreur de prévision en fonction des caractéristiques des sections considérées est proposée en Figure 2.16. L'erreur absolue de prévision par KNN est peu dépendante du débit moyen traversant une section, mais son erreur relative l'est en revanche fortement (Figure 2.16 a)). La méthode naïve est moins marquée par une telle différence suivant le débit moyen expérimenté par les sections. En effet, le gain d'usage d'une méthode évoluée est plus conséquent sur les sections à fort débit Bluetooth et donc les sections les mieux caractérisées. Ceci est encourageant dans les perspectives d'application à venir du Bluetooth.

Enfin, la Figure 2.16 b) montre qu'aucune distinction de performance n'est observée suivant les vitesses expérimentées par les sections et donc suivant les typologies d'infrastructures (urbaine, autoroutière etc). La Figure 2.16 c) vient affiner ce constat en précisant qu'il existe en revanche une dépendance à l'écart-type des vitesses observées sur la section. Il est en effet attendu que les sections soumises à d'importantes variations de vitesses soient plus difficiles à prévoir que les autres (volatilité).

L'analyse des sections les plus fiables du réseau de Brisbane (*i.e.* des sections caractérisées par des débits Bluetooth plus conséquents) s'avère riche en informations. Outre la meilleure qualité des résultats de prévision ($MAPE_{section}$) illustrée par la Figure 2.16 a), les phénomènes perturbant le réseau sont également plus facilement observables. La figure 2.17 montre la propagation d'un événement non récurrent sur deux sections proches du réseau. Le phénomène inhabituel est marqué par une chute de vitesse et une hausse du débit sur le début de la nuit. De telles caractéristiques suggèrent un événement important de type incident ou travaux

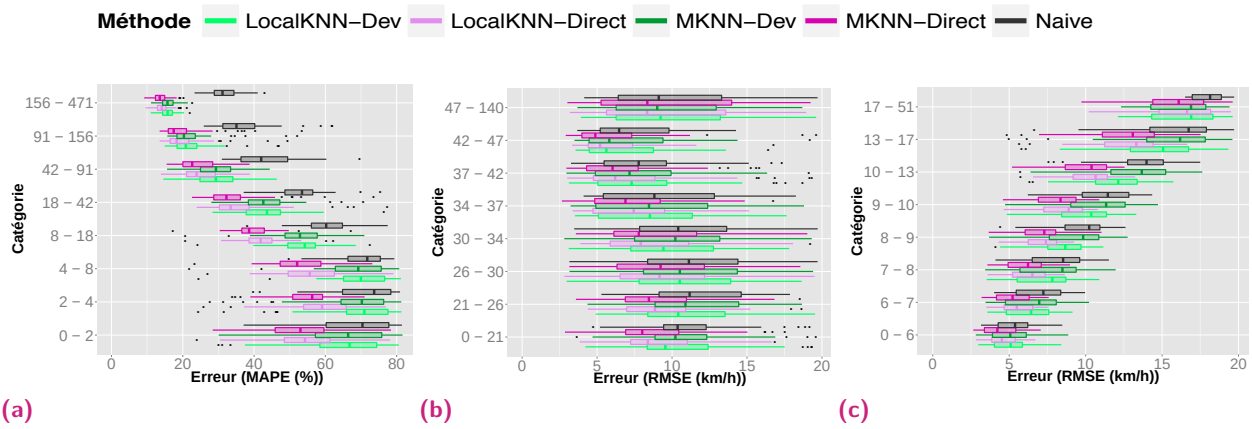


Fig. 2.16.: Distribution de l'erreur de prévision en fonction des caractérisitiques des sections : a) Performance de prévision ($MAPE_{section}$ du débit Bluetooth à 1h en fonction du débit moyen de la section ; b) Performance de prévision ($RMSE_{section}$) de la vitesse à 1h en fonction de la vitesse moyenne de la section c) Performance de prévision ($RMSE_{section}$) de la vitesse à 1h en fonction de l'écart-type des vitesses expérimentées par la section.

impactant la capacité de la section en aval (24-25). Il est intéressant d'observer la propagation de la congestion de l'aval vers l'amont. Visiblement plus contraignante que la section amont, la diminution de vitesse est observée dès 22h sur la section (24-25) et se propage en moins d'une heure à la section amont (16-24). La demande sur la section amont augmente rapidement dès 22h, tandis que la section aval expérimente la hausse de débit à 23h seulement. Force est de constater que l'approche globale directe reproduit mieux la structure spatio-temporelle que les autres approches. Sa prévision des débits à 22h suit de façon étonnante la courbe expérimentée par les sections. La faculté à intégrer la propagation des perturbations du trafic à travers le réseau est également observée grâce à la prévision adéquate de la hausse de débit.

Le cas d'étude de Brisbane suffit à mettre en évidence la bonne capacité de prévision de l'approche globale depuis des méthodes de prévision simples. Néanmoins, les limites actuelles de l'information portée par la technologie Bluetooth se font également sentir. La forte variabilité du taux de pénétration est problématique et ne permet pas d'envisager la mise en oeuvre de prévision à partir de ces seules informations. Leur potentiel est en revanche avéré comme l'ont illustré les analyses fines menées sur le comportement des méthodes de prévision.

2.4.2 Prévision sur le réseau de Nantes à partir des données FCD

Le même procédé de prévision par la méthode KNN a été mis en place sur le réseau du Pont de Cheviré à Nantes. Le réseau est composé de 177 sections du périphérique nantais et de leurs voies d'accès urbaines. Trois mois de données ont été mis à disposition par notre partenaire MEDIAMOBILE : septembre, octobre et novembre 2013. Les vitesses du flot sont recueillies avec un pas de 30 minutes. Les horizons de prévision s'étendent de 30 minutes à 3 heures.

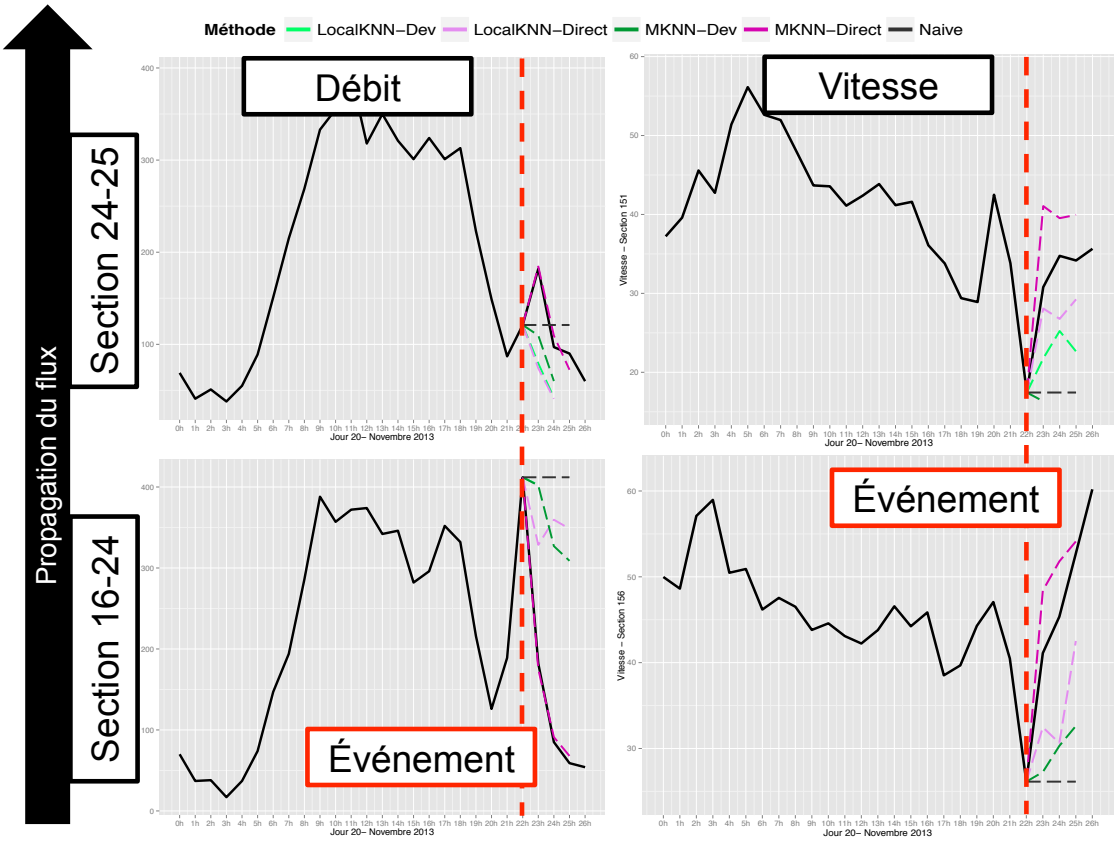


Fig. 2.17.: Analyse focalisée sur la propagation d'une congestion non récurrente sur le réseau de Brisbane.

Les deux premiers mois de données sont utilisés pour l'optimisation des hyperparamètres pratiquée par analyse par grille. Le nombre de voisins K varie de 1 à 50 et la mémoire considérée peut s'étendre de 0 à 6 pas de temps. De même que pour le cas du réseau de Brisbane, les hyperparamètres optimaux prennent des valeurs comprises entre 7 et 12 proches voisins et requièrent des mémoires entre 3 et 6 pas de temps suivant les horizons considérés.

Les performances globales de l'erreur de prévision sont données par la Figure 2.18 et confirment les conclusions obtenues sur le cas de Brisbane. Les approches globales observent des erreurs de prévision voisines, voire plus faibles que celles obtenues par application des approches locales associées. Un motif intéressant réside dans la transition à l'horizon 1h30 entre l'approche indirecte et l'approche directe. Franchi ce cap de 1h30, l'approche directe se révèle plus performante. Un tel phénomène trouve son origine dans la modification des dynamiques de trafic observées. En-dessous de l'horizon 1h30, la dynamique est celle du très court-terme. Au-delà, les méthodes de prévision doivent être aptes à saisir les tendances. L'identification de la meilleure méthode pour chacune des sections et chaque horizon étudié (Figure 2.18 b)) abonde dans le sens des constats précédents : changement de dynamique au passage de l'horizon 1h30 et bonnes performances des approches globales. Les valeurs prises par l'erreur de prévision globale sont, en revanche, sensiblement plus satisfaisantes que celles observées

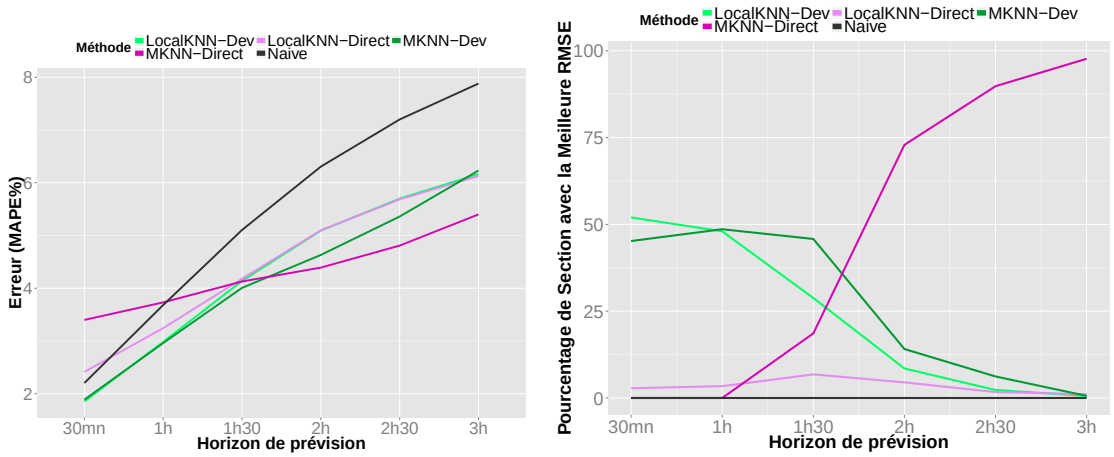


Fig. 2.18.: Comparaison des performances des diverses méthodes et approches de prévision par KNN considérées : a) Performance de prévision globale ($MAPE_{global}$) à l'échelle du réseau b) Identification des méthodes prédisant le mieux les sections (en pourcentage du nombre total de sections) pour tout horizon h de prévision

sur le réseau de Brisbane. Suivant l'horizon de prévision, l'indicateur global varie entre 1,5 et 6%. De telles performances sont qualifiées de très bonnes.

Une analyse plus fine de l'erreur a été menée par période de la journée et suivant les caractéristiques habituelles de trafic des sections du Pont de Cheviré. Les résultats à l'horizon de 30 minutes sont exposés en Figure 2.19. Une nouvelle fois, les observations concordent avec celles du cas australien. L'erreur de prévision de l'approche globale MKNN-Dev est plus faible ou équivalente que toute autre approche durant les pics du matin. En revanche, à cet horizon court-terme, la prévision KNN directe se montre peu performante, même vis-à-vis de l'approche naïve. La comparaison des performances suivant l'écart-type des vitesses observées sur les sections confirme que la prévision est plus aisée, lorsque les vitesses sont peu variables. Néanmoins, au regard des performances de la méthode naïve, les réductions d'erreur liées à l'usage des méthodes KNN sont plus conséquentes sur les sections, dont les variations de vitesse sont intermédiaires. Aucune distinction n'est envisageable par analyse suivant la vitesse moyenne.

Ces premiers résultats sur le cas nantais concordent et entérinent les constats dégagés sur le cas australien avec les données Bluetooth. D'autres méthodes et contextes d'étude sont dès lors envisageables.

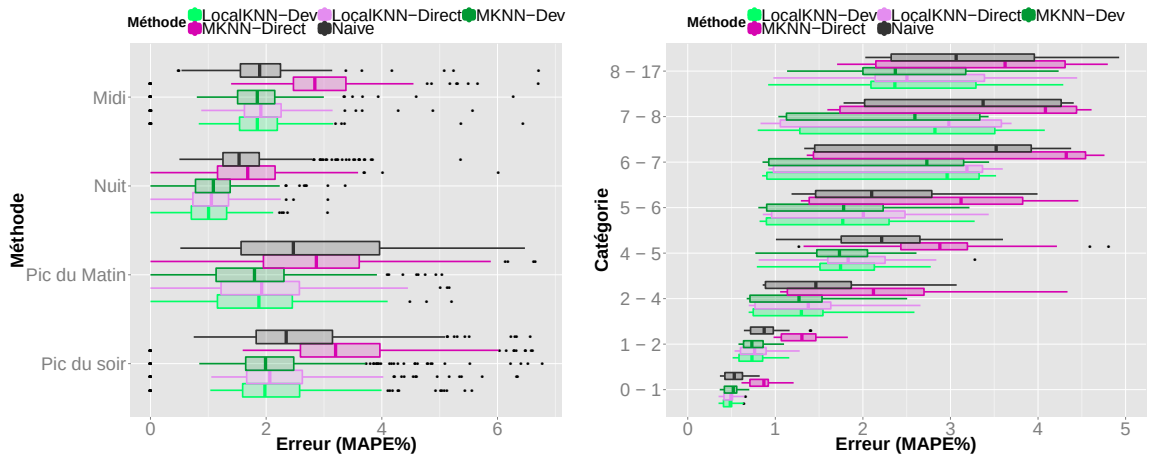


Fig. 2.19.: Analyse des distibutions de l'erreur de prévision par les diverses méthodes KNN sur la zone de Cheviré : a) Comparatif des performances par période de la journée ; b) Comparatif des performances en fonction de la catégorie d'écart-type caractérisant la section.

2.5 Conclusions et perspectives

► Synthèse

Ce chapitre décrit en détails la démarche scientifique adoptée pour toute la suite de la thèse. Le procédé de prévision, depuis des méthodes d'apprentissage, y est généralisé pour toute méthodologie et adapté pour mise en oeuvre en temps-réel de l'approche globale. Les différentes étapes conduisant à la construction du modèle sont décrites et illustrées à partir de l'approche des K plus proches voisins. Les premiers résultats sont encourageants, dans la mesure où, outre le gain de temps, les approches globales retiennent suffisamment d'informations de l'agencement spatio-temporel du réseau pour améliorer occasionnellement les performances de prévision. L'usage d'un modèle unifié apparaît dès lors extrêmement pertinent dans les cadres d'étude présentés.

► Objectifs à venir

Les étapes restantes consistent à :

- étendre l'analyse à des méthodes de prévision nécessitant une adaptation plus complexe et susceptibles de fournir de meilleures performances ;
- évaluer les apports de l'introduction d'une structure spatio-temporelle plus fine et moins naïve ;
- approfondir l'analyse d'événements non récurrents se propageant sur le réseau et mieux intégrer leur prise en considération.

Ces trois points font l'objet des études menées dans les chapitres suivants.

Face aux performances de prévision obtenues à partir des données Bluetooth, il apparaît clairement que ces dernières ne sont pas, en l'état, suffisantes à la mise en œuvre de prévision en temps-réel. Partant de là, l'analyse des données Bluetooth et donc l'étude de Brisbane ne sont pas exploités par la suite. Les imprécisions dues aux variations du taux de pénétration sont conséquentes et effacent les distinctions entre méthodologies de prévision. L'intérêt d'une exploitation plus approfondie de ces données est très limité.
Seules les données traceurs et les informations vitesses associées seront exploitées sur les divers réseaux identifiés de la région nantaise durant la suite de cette étude.

3

Adaptation de méthodes à noyaux et analyse comparative

> *The best way to predict the future is to create it.*
>
> — **Abraham Lincoln**
> (USA president)

Abstract :
This chapter introduces the modifications of Kernel based prediction methods required by the brand new traffic-forecasting framework. The first adaptation of a multivariate SVR method to traffic engineering is presented. The predictions resulting from the Kernel based methodologies enable to validate the good performances of the global approach with respect to the local one. Furthermore the spatial and temporal specificities of each method are highlighted in order to identify their best application cases. The identification of such typologies draws some perspectives to develop forecasting process.

Après l'introduction au cadre prédictif et les performances encourageantes de l'approche globale appliquée aux cas d'études simples dans le chapitre 2, l'objectif poursuivit désormais est l'adaptation de méthodes de prévision plus complexes, nécessitant une modification méthodologique plus conséquente.
Ce chapitre vise à exposer les modifications requises sur des modèles de prévision de la famille des méthodes à noyaux, à évaluer les apports de l'approche globale sur de telles méthodes et à aborder les domaines de prédilection de chacune d'elles. Les méthodes à noyaux sont particulièrement bien adaptées au cadre de l'approche globale et aux enjeux soulevés par l'état de l'art, à savoir la facilité d'adaptation à un nouvel historique pour faire face à des événements non récurrents.
Après présentation et adaptation à l'approche globale des méthodologies introduites, les performances de prévision sont estimées sur le réseau du Pont de Cheviré, ainsi que sur la zone urbaine de Nantes.

3.1 Adaptation des méthodes à noyaux aux exigences de l'approche globale

Les méthodes à base de noyaux se sont distinguées par leur souplesse et leur faculté d'adaptation. Les approches employées dans la littérature du trafic se différencient par la nature des fonctions à noyaux utilisées. Les régressions non paramétriques s'appuient sur ces noyaux pour couvrir par inférence les zones de l'espace peu renseignées, tandis que les approches utilisant l'astuce du noyau (*Kernel Trick*) les exploitent comme projecteurs en haute dimension. À pratique distincte, procédé divergent et donc processus d'adaptation à l'approche globale différent.
Outre une introduction générale à chacune de ces deux méthodologies, les modifications liées à l'implémentation d'une approche globale sont exposées et discutées au cours de cette section. Comme décrit dans le chapitre 2, l'approche gobale requiert deux adaptations majeures :

- la dimension du label de sortie, qui doit être un vecteur pour retranscrire l'ensemble des caractéristiques de trafic à l'échelle du réseau ;
- les objectifs de la fonction de perte doivent être redéfinis pour répondre à la prévision multidimensionnelle requise par le passage à l'échelle du réseau.

3.1.1 Régression non paramétrique : méthodes à noyaux

3.1.1.1 Méthodes de régression à noyaux

La méthode de régression non paramétrique à noyaux (*Kernel*) est proche, conceptuellement, de la méthode KNN introduite antérieurement (section 2.2.3). Moins populaire que son homologue dans la littérature du trafic, elle suscite néanmoins un certain intérêt dans la littérature récente

de la prévision du trafic [103]. Elle joue un rôle d'intermédiaire entre l'approche extrêmement simple du KNN et celles plus évoluées du SVR [103, 104, 105].
L'accès à de grandes bases de données complètes et, de plus en plus, exhaustives n'est pas étrangère à cette popularité. Cette méthodologie prédictive s'appuie effectivement sur l'intégralité de l'historique de données pour prévoir le label de tout nouvel individu. L'espace d'entrée est supposé entièrement couvert grâce au jeu d'apprentissage et aux noyaux associant à tout individu connu une zone d'influence. Tout nouvel individu se positionne au sein de l'espace d'entrée, puis prend place dans les zones d'influence des individus issus du jeu d'apprentissage. L'influence de tout individu issu de l'historique sur la sortie est évaluée par le noyau. Le label attribué au nouvel individu résulte de la moyenne des labels de l'intégralité de l'historique pondérée par les valeurs prises par les fonctions à noyaux.
Il existe plusieurs formes de noyaux. Le plus populaire est le **noyau gaussien**, appartenant au groupe des noyaux **RBF** (*Radial Basis Function*). Dans le cas d'individus caractérisés par n variables, il prend la forme suivante :

$$\forall(\boldsymbol{x_1}, \boldsymbol{x_2}) \in (\Re^n)^2, \;\; \mathcal{K}(\boldsymbol{x_1}, \boldsymbol{x_2}) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{(\|\boldsymbol{x_1} - \boldsymbol{x_2}\|_2)^2}{2\gamma}} .$$

où γ est la fenêtre du noyau. La zone d'influence correspond alors à une nappe de densité de probabilité associée à chaque individu de l'historique. Les pondérations, associées à tout membre de l'historique, sont définies par la probabilité que le nouvel individu adopte le label associé. La moyenne des labels pondérée par la probabilité d'apparition aboutit à la prévision. Cette moyenne pondérée rappelle le fonctionnement de l'évaluation par KNN. Une analogie reliant ces deux méthodologies est proposée en Figure 3.1.

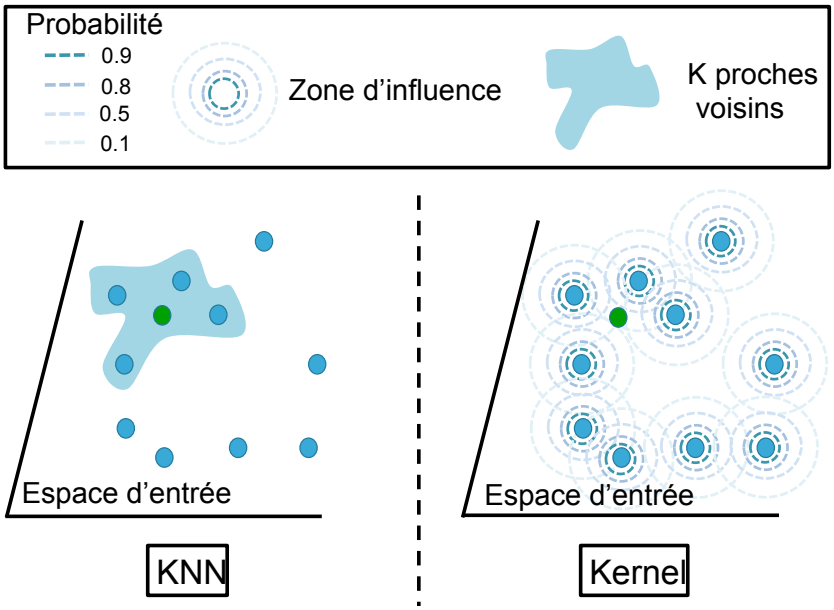


Fig. 3.1.: Analogie de fonctionnement entre la régression non paramétrique des K plus proches voisins (KNN) et celle à fonctions noyaux (Kernel de type RBF).

3.1.1.2 Formulation pour l'approche globale

De façon similaire aux régressions des K plus proches voisins, la régression non paramétrique à noyaux connaît une phase d'apprentissage immédiate. Seuls des hyperparamètres définissent la forme du modèle, aucun paramètre interne n'est à caler. La fonction de perte à minimiser ne dépend que du seul hyperparamètre, à savoir la fenêtre du noyau : γ. Dans le cas des noyaux gaussiens, la fenêtre du noyau est assimilée à une largeur de bande. Cet hyperparamètre module la largeur des zones d'influence de chaque individu. Il se révèle très dépendant de la diversité des individus de l'historique. Si l'espace d'entrée est bien couvert par l'historique, la fenêtre du noyau pourra être fine, tandis qu'en cas de disparités, une bande large permet de mieux couvrir l'espace. Les seules véritables modifications dues à l'application d'une approche globale se manifestent à travers la dimension de sortie du label $\widehat{\boldsymbol{Y_{t+h}}}$ et la représentation de l'intégralité du réseau en entrée. Ceci conduit à déterminer une unique fenêtre du noyau pour toute section du réseau. Les individus en entrée de l'approche globale restent les mêmes que ceux introduits en chapitre 2, à savoir des ensembles de clichés successifs du réseau.

L'algorithme de régression à noyaux a été adapté à l'approche globale en proposant des labels de sortie multidimensionnels. Lors du traitement de tout nouveau cliché du réseau $\boldsymbol{X_t}$, son positionnement dans l'espace d'entrée est estimé par rapport aux individus de l'historique $(\boldsymbol{X_j})_{j\in[1,\mathrm{T}_{hist}]}$. Le processus se déroule en deux phases :

- **Estimation des poids associés à tous les individus de l'historique** : la pondération est directement estimée par la fonction à noyaux jouant le rôle de la distance euclidienne dans l'approche des K plus proches voisins. La différence notable est que la régression à noyaux fait intervenir l'ensemble des individus de l'historique, quand KNN se restreint à ses k proches voisins. L'expression des pondérations est la suivante :

$$\forall j \in [1, \mathrm{T}_{hist}],\ \ w_{j,t}^{kernel} = \frac{\mathcal{K}(\boldsymbol{X_t}, \boldsymbol{X_j})}{\frac{1}{\mathrm{T}_{hist}} \sum_{i=1}^{\mathrm{T}_{hist}} (\mathcal{K}(\boldsymbol{X_i}, \boldsymbol{X_t}))}$$

- **Détermination du profil cible** : le cliché prévu résulte de la moyenne pondérée des clichés $\boldsymbol{Y_{j+h}}$, successeurs temporels à l'horizon h des individus de l'historique. La formulation mathématique est la suivante :

$$\widehat{\boldsymbol{Y_{t+h}}} = \sum_{j=1}^{\mathrm{T}_{hist}} w_{j,t}^{kernel} \times \boldsymbol{Y_{j+h}}$$

Cette approche multidimensionnelle de la régression à noyaux est nommée pour toute la suite **MKernel** par opposition aux approches locales : **LocalKernel**.

3.1.1.3 Optimisation des hyperparamètres

L'introduction de la méthode **MKernel** au sein du cadre prédictif conduit à identifier trois hyperparamètres : la fenêtre du noyau γ, la longueur de mémoire μ et la longueur de fenêtre

de l'historique glissant M. M étant supposé fixé à 30 jours pour toute méthode, seuls γ et μ sont réellement à évaluer. Ces deux hyperparamètres sont, dans un premier temps, ciblés par analyse par grille, et lecture du coude sur la courbe d'erreur ($MAPE_{global}$) tel qu'illustré en Figure 3.2 b) et c). La fenêtre du noyau γ étant continue, une seconde phase met en oeuvre l'algorithme d'optimisation par les fourmis (ACO) pour optimiser le seul hyperparamètre γ. Le développement de l'algorithme ACO est effectué sur une zone-cible de l'espace des hyperparamètres directement bornée par les résultats de l'analyse par grille. Pour tout horizon de prévision et toute longueur de mémoire μ, les meilleures valeurs de γ sont approchées par l'algorithme des fourmis. Les meilleurs couples d'hyperparamètres (γ, s) sont représentés par les points noirs en Figure 3.2 a). Pour cette méthodologie encore peu complexe, l'optimisation par les colonies de fourmis fournit une vérification des résultats obtenus par lecture graphique et un test de leur sensibilité. De plus, le procédé de sélection des hyperparamètres est ainsi totalement automatisé.

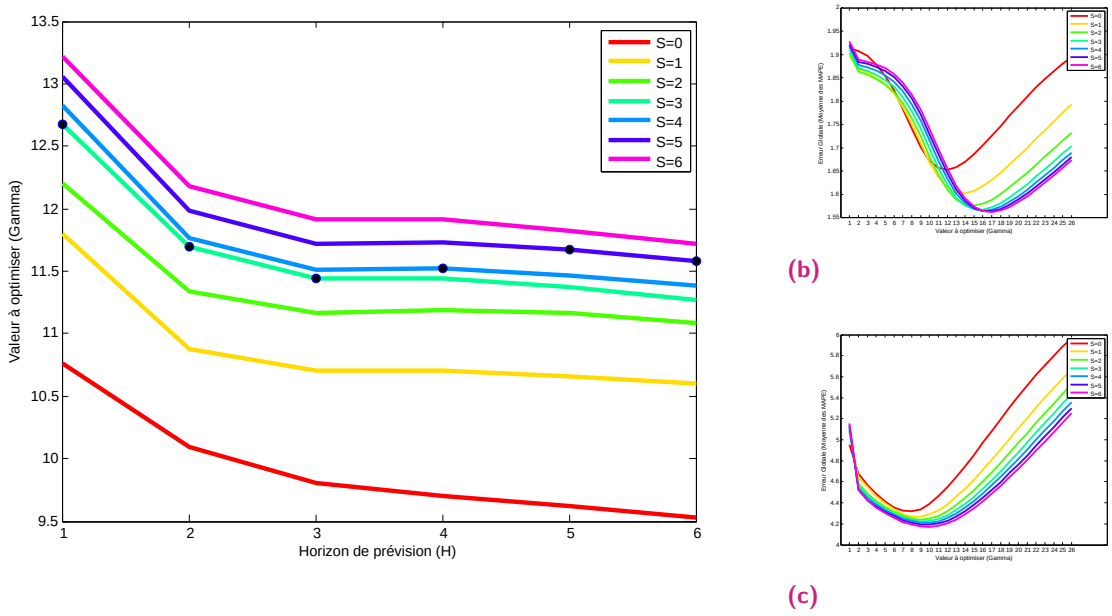


Fig. 3.2.: Illustration des différentes étapes du processus d'optimisation des hyperparamètres pour la méthode MKernel sur le Pont de Cheviré : a) Valeurs Optimales de γ en fonction de l'horizon de prévision et de la longueur de mémoire μ ; b) Représentation en coupes de la surface d'erreur obtenue par analyse par grille à l'horizon 30 mn ; c) Représentation en coupes de la surface d'erreur obtenue par analyse par grille à l'horizon 3h.

Pour l'**approche locale**, l'optimisation des hyperparamètres, calant les modèles propres à chacune des sections, est effectuée sur un sous-ensemble de sections bien choisies. Le même procédé d'optimisation en deux phases leur est appliqué. Les hyperparamètres optimaux sont propagés aux sections restantes du réseau via un algorithe KNN et la définition d'une mesure de similarité. Dans le cas du Kernel à noyau RBF, la fenêtre du noyau γ est dépendante de la variabilité des vitesses. La proximité physique et l'écart-type des vitesses expérimentées

par les sections sont employés pour mesurer les distances inter-sections, puis ré-affecter les hyperparamètres à toute section.

3.1.2 Régression par machines à Vecteurs Supports (SVR)

La méthode SVR, plus récente que les Réseaux de Neurones Artificiels (ANN), connaît un vrai succès aussi bien dans la littérature des transports que dans divers autres domaines applicatifs. Le SVR nécessite un historique plus restreint que les approches apprenant la structure de l'historique pour aboutir à des performances équivalentes [113]. De plus, ce dernier se prête aisément à une implémentation en temps-réel, sans nécessiter systématiquement un nouvel apprentissage du modèle [115]. Sa rapidité d'apprentissage lui fournit une souplesse non négligeable vis-à-vis de phénomènes non-récurrents [115, 114]. Le SVR est, à l'heure actuelle, sans conteste la méthode la plus performante parmi celles à gestion de l'historique.

3.1.2.1 Méthode à Vecteurs Supports usuelle

La méthode de Régression par machines à Vecteurs Supports (SVR), développée au milieu des années 90 [107, 109, 108], est une méthode de régression directement tirée des Séparateurs à Vaste Marge (SVM), introduits par Vapnik *et al.* [159] dans le courant des années 60.

3.1.2.1.1 Séparateurs à vaste marge (SVM)

Les **Séparateurs à Vaste Marge** ou Machines à Vecteurs Supports (**SVM**) consistent en une méthode de classification supervisée d'individus en deux catégories. Ils reposent sur deux idées maîtresses illustrées en Figure 3.3 :

- l'hypothèse qu'il existe un hyperplan en haute dimension permettant la séparation linéaire entre les deux groupes d'individus. L'accès à cet espace de haute dimension, voire de dimension infinie, est proposé par la fonction ϕ et approché par des projecteurs prenant la forme de **fonctions à noyaux**.
- une grande variété d'hyperplans pouvant effectivement séparer le jeu d'individus, l'hyperplan **maximisant la marge** (i.e. la distance entre l'hyperplan et l'ensemble des individus) est considéré comme le séparateur optimal et assurant la meilleure généralisation.

D'un point de vue mathématique, la **première idée** adopte la formulation suivante :

$$\forall \boldsymbol{x} \in \Re^n, \quad f(\boldsymbol{x}) = \boldsymbol{W}^{\mathrm{T}} \phi(\boldsymbol{x}) + b \tag{3.1}$$

où ($\boldsymbol{W}$, b) sont les caractéristiques du vecteur orthogonal à l'hyperplan affine séparateur. Cet hyperplan est donc caractérisé par l'ensemble des vecteurs $\phi(\boldsymbol{x})$ orthogonaux à $\boldsymbol{W}$ et vérifiant donc : $\{\boldsymbol{x} \in \Re^n;\ f(\boldsymbol{x}) = 0\}$. Il est dès lors supposé que le signe strict de $f(\boldsymbol{x})$ détermine la classe de l'individu $\boldsymbol{x}$.

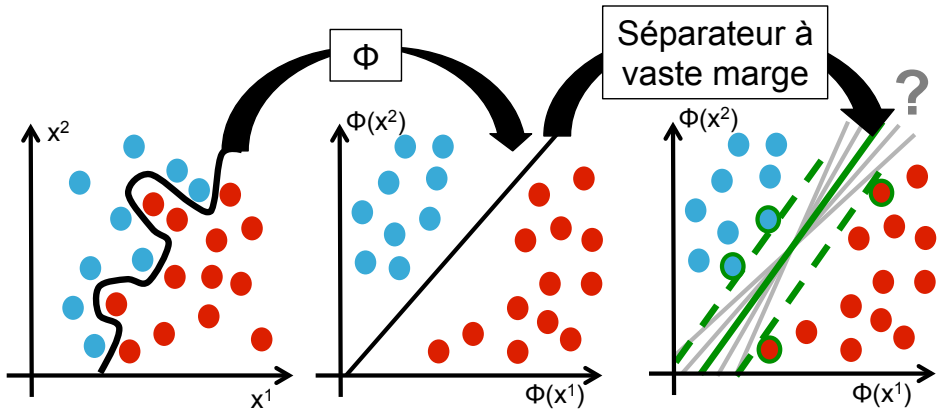


Fig. 3.3.: Illustration des deux idées directrices du processus de classification par la méthode SVM.

La **seconde notion**, dite de **marge maximale**, vise à déterminer l'unique hyperplan optimal. Il est défini par Vapnik comme étant le seul hyperplan maximisant la marge entre les échantillons d'apprentissage et l'hyperplan séparateur. La marge étant déterminée par la distance minimale entre les individus de l'historique d'apprentissage $\{\boldsymbol{x_k};\ k \in [1, \mathrm{T}_{hist}]\}$ projetés en haute dimension et l'hyperplan, le meilleur séparateur vérifie la formule mathématique suivante :

$$\arg\max_{\boldsymbol{W},b} \underbrace{\min_{k}\{\|\phi(\boldsymbol{x}) - \phi(\boldsymbol{x_k})\|;\ x \in \Re^n, \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x}) + b = 0\}}_{marge}$$

La distance d'un échantillon $\phi(\boldsymbol{x_k})$ à un hyperplan est donnée par sa projection orthogonale sur celui-ci : $\dfrac{\boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b}{\|\boldsymbol{W}\|};\ \forall k \in [1, \mathrm{T}_{hist}]$. La recherche de l'hyperplan séparateur de marge maximale adopte alors la forme :

$$\arg\max_{\boldsymbol{W},b}\{\frac{1}{\|\boldsymbol{W}\|} \quad \underbrace{\min_{k}(\boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b)}_{Projection\ des\ Vecteurs\ Supports} \quad \}$$

Les individus de l'historique à la frontière de la marge, *i.e.* ceux atteignant la distance minimale à l'hyperplan, sont qualifiés de **vecteurs supports**. Afin de faciliter l'optimisation, les paramètres de l'hyperplan $(\boldsymbol{W}, b)$ sont normalisés de telle façon que la projection orthogonale des vecteurs supports est égale à plus ou moins 1. Cette mise à l'échelle conduit alors à maximiser $\dfrac{1}{\|\boldsymbol{W}\|}$, ce qui est couramment traduit par le problème d'optimisation de la fonction de perte $\mathcal{L}(\boldsymbol{W}, b)$:

$$\min_{\boldsymbol{W},b}\ \mathcal{L}(\boldsymbol{W}, b) = \min_{\boldsymbol{W},b}\ \frac{1}{2}\|\boldsymbol{W}\|^2 \tag{3.2}$$

sous les contraintes : $(|\boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b| \geq 1);\ \forall k \in [1, \mathrm{T}_{hist}]$.

En général, trouver un hyperplan séparateur en haute dimension s'avère peu aisé et les contraintes sont souvent trop strictes, d'autant que les échantillons d'apprentissage sont susceptibles d'être mal étiquetés à la frontière. En 1995, Vapnik *et al.* [160] ont introduit la

notion de **marge souple**, permettant de tolérer quelques mauvais classements, *i.e.* autoriser quelques individus au sein de la marge. La technique consiste à chercher un hyperplan minimisant le nombre d'erreurs via l'introduction des **variables ressorts** ξ_k relâchant les contraintes, mais pénalisées par un **coût** C. La minimisation de la fonction de perte évolue alors selon la formulation suivante :

$$\min_{\boldsymbol{W},b,C} \ (\mathcal{L}(\boldsymbol{W},b,C)) = \min_{\boldsymbol{W},b,C} \ \frac{1}{2}\|\boldsymbol{W}\|^2 + C\sum_{k=1}^{\mathrm{T}_{hist}} \xi_k \tag{3.3}$$

sous les contraintes : $(|\boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b| \geq 1 - \xi_k); \ \ k \in [1, \mathrm{T}_{hist}] \ \ ; \ \ \xi_k \geq 0 \ \ et \ \ C > 0.$

Résoudre le problème d'estimation des coefficients du modèle $\boldsymbol{W}$ et b revient donc à résoudre le problème d'optimisation visant à minimiser la fonction de perte $\mathcal{L}(\boldsymbol{W}, b, C)$. $\boldsymbol{W}$ et b sont les **paramètres internes** du modèle de classification supervisée. C est un **hyperparamètre** visant à équilibrer le nombre d'exceptions à la règle de décision. Le choix de la règle de décision et la notion de variables ressorts sont illustrés par la figure 3.4 dans l'espace à haute dimension construit par ϕ.

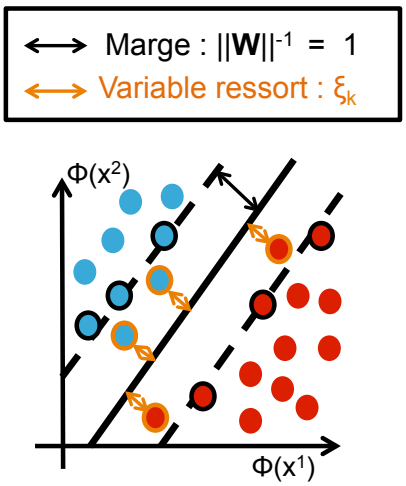


Fig. 3.4.: Illustration du procédé de recherche de l'hyperplan optimal avec relâche des contraintes liées à la marge maximale dans l'espace à haute dimensions résultant de ϕ.

Les machines à vecteurs supports ont été déclinées sous diverses formes, permettant de traiter des cas multi-classes et, depuis 1995, de procéder à de la régression non-linéaire [107].

3.1.2.1.2 De la classification SVM aux Régressions par machines à Vecteurs Supports (SVR)

Les Régressions par machines à Vecteurs Supports s'appuient sur un jeu de vecteurs de l'ensemble d'apprentissage ($\{(\boldsymbol{x_k}, y_k) \in \Re^n \times \Re; \ \ k \in [1, \mathrm{T}_{hist}]\}$) pour déterminer l'hyperplan de régression optimal en haute dimension. La marge ne consiste plus en la distance minimale séparant les échantillons de l'hyperplan séparateur, mais en la distance maximale. Les vecteurs supports, situés en marge, sont les plus éloignés de l'hyperplan de régression. Le problème de régression est donc inversé par rapport à celui de classification (voir Figure 3.5) : l'objectif

consiste à trouver l'hyperplan minimisant l'écart aux données d'apprentissage en s'autorisant une erreur maximale ϵ, tandis que la classification vise à trouver l'hyperplan le plus éloigné de toutes les données. Le cas de la régression consiste à définir un ϵ-tube autour de la régression contenant toutes les valeurs de l'historique d'apprentissage. Cette contrainte est relâchée par l'introduction d'une ϵ-fonction de perte insensible [1] : ζ_ϵ définissant les variables ressorts autorisées. La formulation du problème d'optimisation associé au procédé de régression adopte la même forme que les équations 3.1 et 3.3 du SVM :

$$\forall(\boldsymbol{x}, y) \in \Re^n \times \Re, \;\; \widehat{y} = f(\boldsymbol{x}) = \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x}) + b \tag{3.4}$$

dont les paramètres ($\boldsymbol{W}$, b) de la droite de régression sont à optimiser à partir de la fonction de perte :

$$\min_{\boldsymbol{W},b,C,\epsilon} \; (\mathcal{L}(\boldsymbol{W}, b, C, \epsilon)) = \min_{\boldsymbol{W},b,C,\epsilon} \; \frac{1}{2}\|\boldsymbol{W}\|^2 + C \sum_{k=1}^{\mathrm{T}_{hist}} (\xi_k^+ + \xi_k^-) \tag{3.5}$$

sous les contraintes :

$$\begin{aligned} \forall k \in [1, \mathrm{T}_{hist}]; \;\; (y_k - \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) - b &\leq \epsilon + \xi_k^+) \\ \forall k \in [1, \mathrm{T}_{hist}]; \;\; (\boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b - y_k &\leq \epsilon + \xi_k^-) \\ \xi_k^+; \xi_k^- &\geq 0 \end{aligned}$$

avec $C > 0$ et la ϵ-fonction de perte insensible définie par :

$$\zeta_\epsilon(\boldsymbol{x_k}) = |\xi_k^\pm| \;\; \rightarrow \begin{cases} 0 & si \;\; |y_k - \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) - b| < \epsilon \\ |y_k - \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) - b| - \epsilon & sinon \end{cases}$$

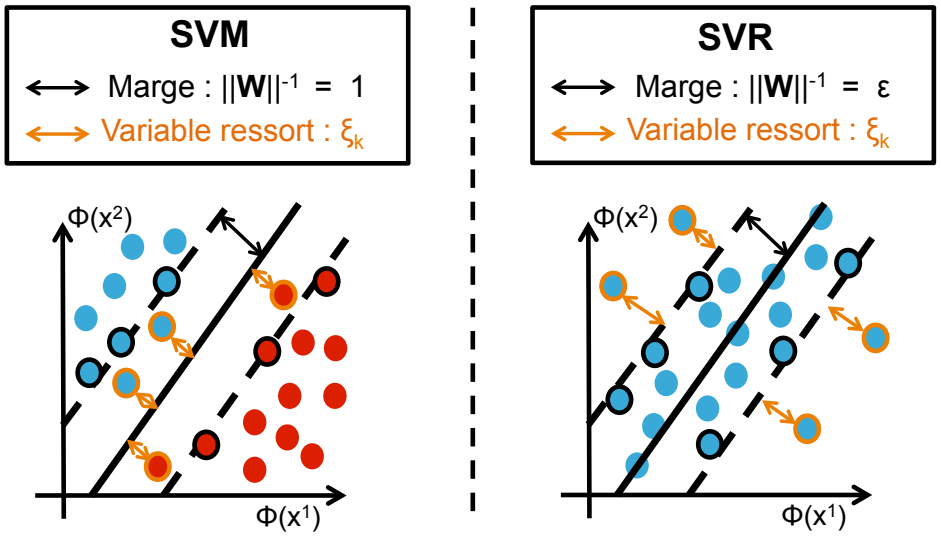


Fig. 3.5.: Illustration de la transition d'un problème de classfication SVM à un procédé de régression SVR (Représentation dans l'espace à haute dimension).

1. perte sans coût, tant que l'erreur est inférieure à ϵ

Il s'avère que ce problème d'optimisation est plus facilement et plus directement résolu depuis sa formulation duale. Cette dernière fait intervenir les multiplicateurs de Lagrange α_k^+, α_k^-, η_k^+ et η_k^- permettant l'intégration des contraintes au sein de la fonction de perte. La formulation duale à optimiser revient à minimiser :

$$\begin{aligned} D_{ual}(\boldsymbol{W}, b, C, \epsilon, \alpha_k^+, \alpha_k^-, \eta_k^+, \eta_k^+) = \frac{1}{2}\|\boldsymbol{W}\|^2 + C\sum_{k=1}^{\mathrm{T}_{hist}}(\xi_k^+ + \xi_k^-) \\ - \sum_{k=1}^{\mathrm{T}_{hist}}(\eta_k^+\xi_k^+ + \eta_k^-\xi_k^-) - \sum_{k=1}^{\mathrm{T}_{hist}}\alpha_k^+(\epsilon + \xi_k^+ - y_k + \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) + b) \\ - \sum_{k=1}^{\mathrm{T}_{hist}}\alpha_k^-(\epsilon + \xi_k^- + y_k - \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) - b) \end{aligned} \quad (3.6)$$

sous les contraintes : $\forall k \in [1, \mathrm{T}_{hist}];\ \ \alpha_k^+, \alpha_k^-, \eta_k^+, \eta_k^- \geq 0$.
Le but étant de minimiser la fonction de perte, les dérivées partielles de la forme duale D sont annulées et les conditions supplémentaires de Karush-Kuhn-Tuker appliquées :

$$\partial_b D_{ual} = \sum_{k=1}^{\mathrm{T}_{hist}}(\alpha_k^- - \alpha_k^+) = 0 \quad (3.7)$$

$$\partial_{\boldsymbol{W}} D_{ual} = \boldsymbol{W} - \sum_{k=1}^{\mathrm{T}_{hist}}(\alpha_k^- - \alpha_k^+)\phi(\boldsymbol{x_k}) = 0 \quad (3.8)$$

$$\partial_{\xi_k^\pm} D_{ual} = C - \alpha_k^\pm - \eta_k^\pm = 0 \quad (3.9)$$

$$min(\eta_k^\pm;\ \xi_k^\pm) = 0 \quad (3.10)$$

$$min(\alpha_k^\pm;\ |\epsilon + \xi_k^\pm - y_k + \boldsymbol{W}^{\mathrm{T}}\phi(\boldsymbol{x_k}) - b|) = 0 \quad (3.11)$$

Les équations supplémentaires de Kuhn-Tuker (équations 3.10 et 3.11) identifient **les vecteurs supports**, puisque les multiplicateurs lagrangiens ne sont non nuls que pour ces seuls échantillons de l'historique. Les dérivées partielles (équations 3.7, 3.8 et 3.9) sont intégrées à la formulation duale d'optimisation (équation 3.6) pour aboutir à l'expression quadratique :

$$\begin{aligned} \max_{\alpha_k^\pm} \ -\frac{1}{2}\sum_{i,k=1}^{\mathrm{T}_{hist}}(\alpha_i^+ - \alpha_i^-)(\alpha_k^+ - \alpha_k^-)\phi(\boldsymbol{x_i})^t\phi(\boldsymbol{x_k}) - \epsilon\sum_{k=1}^{\mathrm{T}_{hist}}(\alpha_k^+ + \alpha_k^-) \\ + \sum_{k=1}^{\mathrm{T}_{hist}} y_k(\alpha_k^+ - \alpha_k^-) \end{aligned} \quad (3.12)$$

sous les contraintes : $\sum_{k=1}^{\mathrm{T}_{hist}}(\alpha_k^+ - \alpha_k^-) = 0$ et $\alpha_k^\pm \in [0, C]$. L'annulation des dérivées partielles de cette expression par rapport aux multiplicateurs lagrangiens aboutit directement au modèle optimal.
La **fonction à noyaux** $\mathcal{K}$, telle que $\mathcal{K}(\boldsymbol{x_i}, \boldsymbol{x_k}) \equiv \phi(\boldsymbol{x_i})^{\mathrm{T}}\phi(\boldsymbol{x_k}));\ \ \forall (i,k) \in [1, \mathrm{T}_{hist}]^2$, se substitue au produit scalaire indéfini en haute dimension ou extrêmement coûteux en temps de calcul.

Cette dernière formulation liée aux conditions d'annulation de la dérivée conduit à réécrire l'équation 3.8 pour obtenir l'expression du vecteur directeur $\boldsymbol{W}$:

$$\boldsymbol{W} = \sum_{k=1}^{\mathrm{T}_{hist}} (\alpha_k^+ - \alpha_k^-)\phi(\boldsymbol{x_k}) \tag{3.13}$$

L'expression du **biais** systématique b est évaluée depuis les équations tirées des hypothèses de Karush-Kuhn-Tucker (équations 3.10 et 3.11).

La résolution habituelle de la régression à vecteurs supports est donc directement atteignable depuis sa formulation duale. L'expression du **modèle de régression** est obtenue, pour tout nouvel individu $\boldsymbol{x}$, en fonction des multiplicateurs de Lagrange optimisés lors de la phase d'apprentissage :

$$\forall \boldsymbol{x} \in \Re^n;\ \widehat{y} = f(\boldsymbol{x}) = \sum_{k=1}^{\mathrm{T}_{hist}} (\alpha_k^+ - \alpha_k^-)\mathcal{K}(\boldsymbol{x_k}, \boldsymbol{x}) + b$$

Les paramètres lagrangiens $\alpha_k^{\pm}$ étant nuls pour tout élément de l'historique distinct des **vecteurs supports**, seuls ces derniers interviennent réellement dans l'expression finale.
L'apprentissage détermine les **paramètres internes** du modèle de régression à vecteurs supports, à savoir les caractéristiques de la droite de régression en haute dimension : $\boldsymbol{W}$ et b. Le procédé de prévision SVR possède dès lors **trois hyperparamètres** : la largeur de marge ϵ, le coût dû à l'assouplissement de la marge C et l'hyperparamètre associé au choix du noyau. Divers choix de fonctions à noyaux sont possibles (linéaires, polynomiaux, gaussiens), mais, dans notre contexte, le noyau gaussien (RBF) est le plus adapté et est donc sélectionné. L'ultime hyperparamètre est donc la fenêtre du noyau : γ.

Le processus de régression, tel que décrit, correspond au procédé employé par l'**approche locale** de prévision appliquée à chacune des sections l du réseau. Les hyperparamètres μ de la taille de mémoire, liée à la construction de l'individu d'entrée $\boldsymbol{X_t^l}$, et M la taille d'historique, préalablement fixée à 30 jours, s'ajoutent aux trois précédents lors de la phase d'optimisation. Cette dernière repose sur le procédé en deux étapes introduit en section 2.3.2. Le processus s'appuie sur une délimitation de la zone de recherche par analyse par grille, puis le lancement de l'algorithme des colonies de fourmis pour optimisation. Comme pour les approches locales précédemment introduites, seul un sous-ensemble de sections est optimisé. Les meilleurs hyperparamètres obtenus sur le petit jeu de sections sont étendus au réseau [156].

L'extension de ce procédé de prévision à l'approche globale requiert des adaptations fines, mais conséquentes.

3.1.2.2 Adaptation de la Régression SVR à l'approche globale

Les exigences de l'approche globale conduisent à introduire l'aspect multidimensionnel au sein de la Régression par machines à Vecteurs Supports, ce qui se traduit par :

– **la modification des dimensions** des variables d'entrée ($\boldsymbol{X_t}$), de sortie ($\boldsymbol{Y_{t+h}}$) et des paramètres internes ($\underline{\boldsymbol{W}}, \boldsymbol{b}$). La variable d'entrée $\boldsymbol{X_t}$ représente les caractéristiques de trafic expérimentées aux instants t et antérieurs sur l'ensemble des N sections du réseau considéré. Elle prend une forme vectorielle, tout comme désormais le label de sortie $\boldsymbol{Y_{t+h}}$ de dimension N. Les paramètres internes se déclinent respectivement sous forme matricielle pour le coefficient directeur $\underline{\boldsymbol{W}} = (\boldsymbol{W^1}, ..., \boldsymbol{W^N})^{\mathrm{T}}$, où chaque composante est un vecteur directeur en haute dimension, et sous forme vectorielle pour le biais $\boldsymbol{b} = (b^1, ..., b^N)^{\mathrm{T}}$. La formulation du **modèle de prévision** à l'horizon h répond alors à l'équation 3.14.

$$\forall(\boldsymbol{X_t}, \boldsymbol{Y_{t+h}}),\ \ \widehat{\boldsymbol{Y_{t+h}}} = f(\boldsymbol{X_t}) = \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_t}) + \boldsymbol{b} \tag{3.14}$$

– **l'adaptation du problème d'optimisation** aux nouvelles dimensions. Le changement de dimension du coefficient directeur $\underline{\boldsymbol{W}}$ et l'évaluation de la distance à la marge via les variables ressort imposent des métriques distinctes. La ϵ-fonction de perte insensible aboutissant aux variables ressort n'est plus évaluée selon une seule dimension, mais selon les N dimensions de sortie. Elle s'exprime en comparant l'erreur $RMSE_{TimeStep}$ du résultat de régression à la largeur de marge ϵ [161, 162] :

$$\zeta_\epsilon(\boldsymbol{X_k}) = |\xi_k| = \begin{cases} 0 & si\ \ (||\boldsymbol{Y}_{k+h} - \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_k}) - \boldsymbol{b}||_2)^2 < \epsilon \\ (||\boldsymbol{Y}_{k+h} - \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_k}) - \boldsymbol{b}||_2)^2 - \epsilon & sinon \end{cases}$$

La fonction de perte à minimiser intègre dès lors ces nouvelles métriques et est rectifiée en conséquence :

$$\min_{\underline{\boldsymbol{W}}, \boldsymbol{b}, C, \epsilon} \ (\mathcal{L}(\underline{\boldsymbol{W}}, \boldsymbol{b}, C, \epsilon)) = \min_{\underline{\boldsymbol{W}}, \boldsymbol{b}, C, \epsilon} \ \sum_{j=1}^{N} \|\boldsymbol{W^j}\|^2 + C \sum_{k=1}^{\mathrm{T}_{hist}} \xi_k \tag{3.15}$$

sous les contraintes :

$$\begin{aligned} \forall k \in [1, \mathrm{T}_{hist}];\ \ \|\boldsymbol{Y_{k+h}} - \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_k}) - \boldsymbol{b}\|_2 &\leq \epsilon + \xi_k \\ C > 0;\ \xi_k &\geq 0 \end{aligned}$$

Cette formulation multidimensionnelle du problème de prévision autorise toujours l'intégration des contraintes au sein de la fonction de perte via l'usage des **multiplicateurs de Lagrange** (équation 3.16).

$$\begin{aligned} D_{ual}(\underline{\boldsymbol{W}}, \boldsymbol{b}, C, \epsilon, \alpha_k, \eta_k) = &\sum_{j=1}^{N} \|\boldsymbol{W^j}\|^2 + C \sum_{k=1}^{\mathrm{T}_{hist}} \xi_k \\ &- \sum_{k=1}^{\mathrm{T}_{hist}} \eta_k \xi_k - \sum_{k=1}^{\mathrm{T}_{hist}} \alpha_k (\epsilon + \xi_k - (\|\boldsymbol{Y_{k+h}} - \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_k}) - \boldsymbol{b}\|_2)^2) \end{aligned} \tag{3.16}$$

sous les contraintes : $\forall k \in [1, \mathrm{T}_{hist}];\ \ \alpha_k, \eta_k \geq 0$.

Cependant, du fait des nouvelles métriques, la minimisation par annulation des dérivées partielles ne permet plus, ni l'identification d'une forme quadratique (équation 3.12), ni la résolution exacte précédemment opérée. En effet, les solutions au problème proposées par le théorème de Karush-Kuhn-Tuker [163] aboutissent aux présentes équations :

$$\partial_{\boldsymbol{b^j}} D_{ual} = \boldsymbol{\alpha}(\boldsymbol{Y^j} - \underline{\boldsymbol{\Phi}}\boldsymbol{W^j} - \mathbf{1}b^j) = 0 \quad (\forall j \in [1, N]) \tag{3.17}$$

$$\partial_{\boldsymbol{W^j}} D_{ual} = 2\boldsymbol{W^j} - 2\underline{\boldsymbol{\Phi^t}}\underline{\boldsymbol{D_\alpha}}(\boldsymbol{Y^j} - \underline{\boldsymbol{\Phi}}\boldsymbol{W^j} - \mathbf{1}b^j) = 0 \quad (\forall j \in [1, N]) \tag{3.18}$$

$$\partial_{\xi_k} D_{ual} = C - \alpha_k - \eta_k = 0 \quad (\forall k \in [1, \mathrm{T}_{hist}]) \tag{3.19}$$

$$min\Big(\eta_k;\ \xi_k\Big) = 0 \quad (\forall k \in [1, \mathrm{T}_{hist}]) \tag{3.20}$$

$$min\Big(\alpha_k;\ \epsilon + \xi_k - (\|\boldsymbol{Y_{k+h}} - \underline{\boldsymbol{W}}^{\mathrm{T}}\phi(\boldsymbol{X_k}) - \boldsymbol{b}\|_2)^2\Big) = 0 \quad (\forall k \in [1, \mathrm{T}_{hist}]) \tag{3.21}$$

avec $\alpha_k;\ \eta_k \geq 0$ et les notations : $\underline{\boldsymbol{\Phi}} = (\phi(X_1), ..., \phi(X_{\mathrm{T}_{hist}}))^t$; $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha\mathrm{T}_{hist})^t$; $(\underline{\boldsymbol{D_\alpha}})_{p,q} = \alpha_p\delta(p-q)$ et $\boldsymbol{Y^j} = (Y^j_{1+h}, ..., Y^j_{\mathrm{T}_{hist}+h})^t$ vecteur des labels de l'historique pour la section j.

Les équations 3.17 et 3.18 ne sont plus directement exploitables pour aboutir à l'expression finale du modèle de prévision. Dans le domaine informatique, Perez-Cruz *et al.* [161, 164, 162] ont étudié à diverses reprises les difficultés induites par ce passage à une régression MIMO (multiples entrées et sorties) et ont posé les bases d'un algorithme approchant le modèle optimal. Ils s'appuient sur un processus itératif dénommé IRWLS (*Iterated Re-Weighted Least Squares*).
Ce même procédé est appliqué à notre problématique. Dans une première phase, $\underline{\boldsymbol{W}}$ et $\boldsymbol{b}$ sont évalués en supposant les α_k connus. La seconde étape ré-évalue les multiplicateurs lagrangiens depuis le nouveau modèle prédéterminé. L'algorithme est itéré jusqu'à ce que le gain sur la fonction de perte soit en-dessous d'un seuil prédéterminé. La mise en oeuvre de l'algorithme exploite une reformulation matricielle des équations 3.17 et 3.18 à l'origine des difficultés de passage aux N dimensions de sortie :

$$\begin{bmatrix} \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\underline{\boldsymbol{D_\alpha}}\underline{\boldsymbol{\Phi}} + \underline{\boldsymbol{I}} & \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{\mathrm{T}}\underline{\boldsymbol{\Phi}} & \boldsymbol{\alpha}^{\mathrm{T}}\underline{\mathbf{1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{W^j} \\ b^j \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{\Phi}}^{\mathrm{T}}\underline{\boldsymbol{D_\alpha}}\boldsymbol{Y^j} \\ \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y^j} \end{bmatrix} \tag{3.22}$$

Le processus vise alors à résoudre de façon itérative le système matriciel 3.22 en faisant graduellement évoluer les multiplicateurs lagrangiens en vue de réduire la fonction de coût :

1. **Initialisation** : en fixant $\underline{\boldsymbol{W}}_0$ et $\boldsymbol{b}_0$ à des valeurs nulles, les α^1_k et ξ^1_k sont initialisés.
2. **Évaluation du modèle à l'étape** $i+1$:
   - **Résolution du système 3.22** proposant un nouveau modèle ($\underline{\boldsymbol{W}}_p$, $\boldsymbol{b}_p$) et évaluation de la perte associée : $\mathcal{L}(\underline{\boldsymbol{W}}_p, \boldsymbol{b}_p, \alpha^{i+1}_k)$.
   - **Confrontation des fonctions de perte** : antérieure $\mathcal{L}(\underline{\boldsymbol{W}}_i, \boldsymbol{b}_i, \alpha^i_k)$ & nouvelle $\mathcal{L}(\underline{\boldsymbol{W}}_p, \boldsymbol{b}_p, \alpha^{i+1}_k)$.

– En cas de *décroissance de la perte*, le nouveau modèle est conservé : $(\underline{\boldsymbol{W}}_{i+1}, \boldsymbol{b}_{i+1}) = (\underline{\boldsymbol{W}}_p, \boldsymbol{b}_p)$;

– En cas de *croissance de la perte*, le nouveau modèle est raffiné suivant une direction descendante définie par un algorithme itératif de retour sur trace (*backtracking*) : $\begin{bmatrix} \boldsymbol{W_{i+1}} \\ \boldsymbol{b_{i+1}} \end{bmatrix} = (1-\nu^\kappa) \begin{bmatrix} \boldsymbol{W_i} \\ \boldsymbol{b_i} \end{bmatrix} + \nu^\kappa \begin{bmatrix} \boldsymbol{W_p} \\ \boldsymbol{b_p} \end{bmatrix}$ où $\nu \in]0,1[$ et κ est le facteur, borné, d'itération de l'algorithme *backtracking*.

– **Évaluation du gain** sur la fonction de perte et comparaison au **critère de fin** de l'algorithme

3. **Mise à jour des multiplicateurs de Lagrange** et indirectement des vecteurs supports : estimation de α_k^{i+2} et ξ_k^{i+2}, puis réitération par passage à l'étape $i+2$.

L'exécution de cet algorithme n'est possible qu'à condition de connaître la fonction non-linéraire de transition en haute dimension $\phi(.)$, ce qui n'est pas le cas dans notre contexte. De même que pour le cas usuel du SVR, les fonctions à noyaux sont introduites pour approcher le produit vectoriel : $\mathcal{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) \equiv \phi(\boldsymbol{x_i})^{\mathrm{T}} \phi(\boldsymbol{x_j}))$. Le théorème *Representer* [165] est alors exploité et permet d'exprimer le coefficient directeur $\underline{\boldsymbol{W}}$ comme une combinaison linéaire des échantillons du jeu d'apprentissage : $\boldsymbol{W}^j = \underline{\boldsymbol{\Phi}}^{\mathrm{T}} \boldsymbol{\rho}^j$. Moyennant un passage à l'inverse [2], le système matriciel 3.22 adopte une forme basée sur des produits scalaires d'éléments de l'historique en haute dimension :

$$\begin{bmatrix} \underline{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}} + \underline{\boldsymbol{D}_{\boldsymbol{\alpha}}^{-1}} & \mathbf{1} \\ \boldsymbol{\alpha}^{\mathrm{T}} \underline{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}} & \boldsymbol{\alpha}^{\mathrm{T}} \underline{\mathbf{1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}^j \\ b^j \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}^j \\ \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{Y}^j \end{bmatrix} \tag{3.23}$$

Cette nouvelle forme matricielle est alors exprimée à partir des fonctions à noyaux en les substituant aux produits scalaires $\underline{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}}$. Le **modèle optimal de régression** est approché par application de l'algorithme IRWLS lors de toute phase d'apprentissage.

Le procédé, tel que décrit, définit la version multidimensionnelle de la régression à vecteurs supports, baptisée **MSVR** pour *MultiOutput-SVR*. Bien que le modèle MSVR proposé soit largement inspiré des travaux de Perez-Cruz *et al.* [162], nous avons développé une implémentation de l'algorithme favorisant sa convergence et disponible en annexe B.

3.1.2.3 Optimisation des hyperparamètres du MSVR

L'intégration de la méthode MSVR au sein de la démarche prédictive de type globale est opérée sans altération du cadre prédictif. L'usage d'un algorihme d'approximation des meilleurs paramètres internes $(\underline{\boldsymbol{W}}, \boldsymbol{b})$ ne transforme pas la fonction de perte. Aucun hyperparamètre n'est donc ajouté à la version habituelle admettant une sortie unidimensionnelle du SVR. Les hyperparamètres à optimiser se composent donc des deux hyperparamètres propres au cadre prédictif, à savoir la longueur de mémoire μ et la largeur d'historique glissant M, ainsi que ceux propres à la méthode de prévision : la largeur de marge ϵ, le coût C et la fenêtre du noyau γ. La fenêtre du noyau γ est liée au choix de noyaux gaussiens. La largeur d'historique glissant

2. Multiplication de tous les termes par : $(\underline{\boldsymbol{D}_{\boldsymbol{\alpha}}}^{\mathrm{T}} \underline{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}} \underline{\boldsymbol{D}_{\boldsymbol{\alpha}}})^{-1} \underline{\boldsymbol{D}_{\boldsymbol{\alpha}}}^{\mathrm{T}} \underline{\boldsymbol{\Phi}}$

est préalablement fixée à 30 jours. Quatre hyperparamètres, dont deux continus, restent donc à caler.

Le processus d'optimisation résulte des deux phases préalablement introduites en section 2.3.2 :

1. la **restriction de la zone d'exploration** des valeurs potentielles des hyperparamètres grâce à une analyse par grille ;
2. l'optimisation des hyperparamètres par **application de l'algorithme à heuristique** des colonies de fourmis sur la zone d'exploration identifiée.

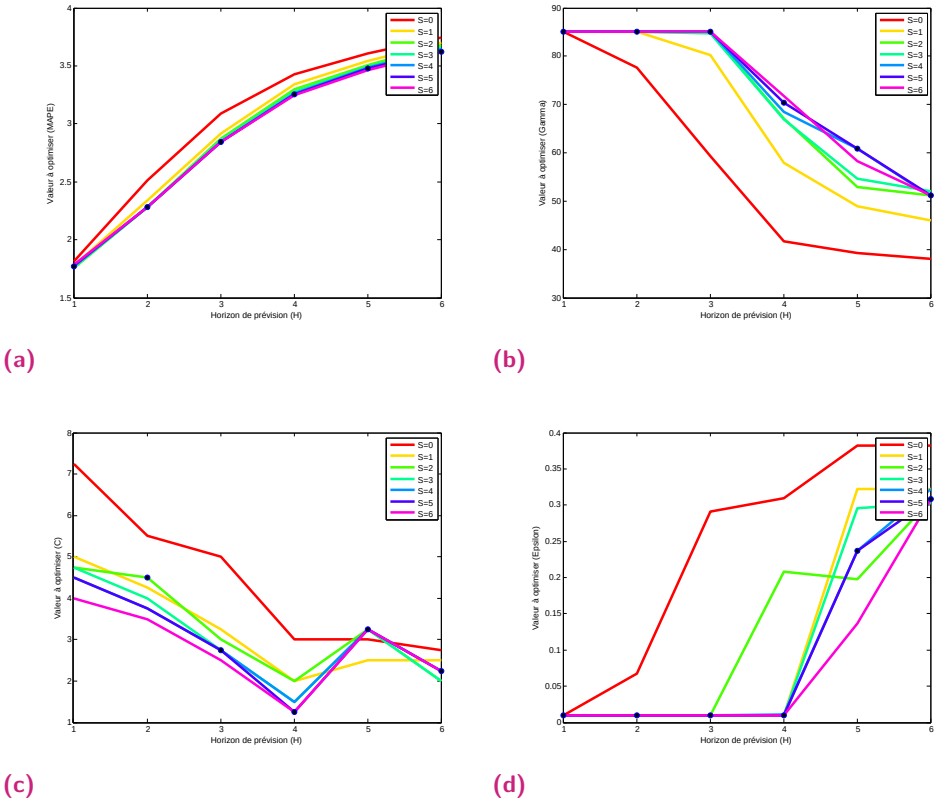


Fig. 3.6.: Illustration des résultats du procédé ACO d'optimisation des hyperparamètres pour la méthode MSVR sur le Pont de Cheviré en fonction de l'horizon de prévison h et de la mémoire μ : a) Erreur optimale exprimée par la $MAPE_{global}$; b) Valeurs optimales de γ ; c) Valeurs optimales de coût C ; d) Valeurs optimales de marges ϵ.

Afin de réduire le temps de recherche et la complexité temporelle, l'optimisation par les colonies de fourmis est, en réalité, pratiquée pour chacune des mémoires μ disponibles. La méthode dite de *Time Series Cross-Validation* est bien toujours utilisée pour assurer la bonne généralisation du modèle à venir. Les hyperparamètres sélectionnés par minimisation de l'erreur pour tout horizon de prévision sont représentés par des points noirs sur les Figures 3.6 b), c), et d). La Figure 3.6 a) fournit un relevé de l'erreur optimale de prévision recensée pour tout horizon de prévision. L'erreur croît avec l'horizon de prévision, phénomène naturel dû à l'impossibilité de connaître les événements intervenant durant la période h.

Une fois encore, la seconde phase du procédé de sélection des hyperparamètres est complètement automatisée grâce à une descente de gradient.

3.2 Analyse comparative des performances des méthodes à noyaux sur la zone de Cheviré

Les sections 3.2 et 3.3 opposent les performances de l'ensemble des méthodes préalablement introduites. Les méthodes naïve, KNN, Kernel et SVR déclinées dans leurs versions locales et globales sont évaluées, comparées et caractérisées en fonction de leurs spécificités suivant les typologies de sections ou de périodes temporelles.
Cette section fait état des performances relevées à partir des relevés de vitesse du flot recensés depuis les données traceurs sur la zone du Pont de Cheviré à Nantes. $N = 177$ sections sont ainsi prises en considération pour un historique de données composé de $\tau_{hist} = 61$ jours avec un pas de temps de 30 minutes. L'erreur de prévision est présentée sur la période de validation composée des 30 jours du mois de novembre 2013.

3.2.1 Performances en termes de temps d'estimation en temps-réel

Il est essentiel de rapporter les gains en termes d'erreur de prévision aux pertes de complexité temporelle. En effet, plus les modèles développés sont complexes et font appel à un grand nombre de paramètres, plus l'effort de calibrage, puis d'estimation en temps-réel est conséquent. Un bon compromis doit être trouvé pour parvenir à des modèles parcimonieux. Le développement des approches globales travaillant à l'échelle du réseau vise justement à proposer des modèles plus parcimonieux que ceux proposés par les approches locales.
La durée d'optimisation des hyperparamètres a déjà été évaluée sur le cas de Brisbane à partir de la méthode KNN. Sur la zone du Pont de Cheviré, les approches globales requièrent entre 7h pour les régressions de type MKNN et MKernel à trois jours pour la méthode plus complexe du MSVR, tandis que les approches locales nécessitent *a minima* une semaine pour les approches les plus aisées. Ces durées d'optimisation des hyperparamètres n'ont pas de véritable impact sur le potentiel de mise en œuvre en temps-réel des méthodes. Il s'agit de procédés agissant hors-ligne et visant seulement à calibrer les modèles à venir. Néanmoins, un temps conséquent de mise à jour du calibrage peut être préjudiciable du fait de l'occupation de mémoire vive.

Le temps d'estimation en temps-réel sur le réseau du Pont de Cheviré est présenté dans le tableau 3.1 pour l'intégralité du mois de novembre 2013, soit 1440 pas de temps. Quelle que soit l'approche locale ou globale considérée, la méthode à vecteurs supports nécessite un temps d'estimation nettement supérieur aux approches plus simples de régression non paramétrique.

1. Les approches locales sont estimées en parallélisant les estimations sur chaque section.
2. L'approche locale du SVR a été testée sous un logiciel distinct de Matlab : R et sur un ordinateur différent, mais avec des performances équivalentes. Seul l'ordre de grandeur du temps d'évaluation a donc un sens.

Approche	Méthode	Horizon h						Ordre de grandeur
		30mn	**1h**	**1h30**	**2h**	**2h30**	**3h**	
globale	Naïve	0.4s	0.3s	0.03s	0.03s	0.3s	0.3s	**0.2s**
	MKNN	33s	44,7s	44,9s	44,7s	44,7s	44,7s	**43s**
	MKernel	54s	55s	55s	61s	66s	65s	**1mn**
	MSVR	2 915s	2 888s	2 776s	2 646s	3 494s	3 594s	**51mn**
locale[1]	KNN	1 943s	1 921s	1 920s	1 885s	2 071s	2 238s	**33mn**
	Kernel	4 740s	4 915s	4978s	4 771s	4 817s	4 740s	**80mn**
	SVR [2]	NC	NC	NC	NC	NC	NC	**∼50h**

Tab. 3.1.: Comparaison des temps d'apprentissage et de prévision en fonction de la méthode de prévision pour le mois de novembre 2013 (1440 évaluations) sur le Pont de Cheviré : confrontation de l'approche locale à l'approche globale.

Cette différence de durée s'explique par l'absence de réelle phase d'apprentissage pour les méthodes de régression KNN ou Kernel.
La seconde observation importante réside dans la différence de temps de calcul requis entre les approches locales et globales. Un facteur 70 environ est constaté. Ce facteur ne correspond pas à $N = 177$ fois le temps d'estimation de l'approche globale pour deux raisons : les approches locales sont estimées en parallèle et la durée d'évaluation d'une section par approche locale est légèrement inférieure à celle du réseau par approche globale. Les temps indiqués étant ceux pour évaluer un mois entier, toutes les méthodes présentées, quel que soit l'horizon et l'approche adoptée, sont en mesure de fournir des prévisions en temps-réel pour des horizons au-delà de 30 minutes. L'approche locale du SVR nécessite néanmoins pour le petit réseau du Pont de Cheviré quelques minutes d'estimation. Toute réduction de l'horizon ou augmentation de la taille du réseau considéré pourraient se révéler problématique à terme.
Enfin, peu de variabilité de la durée de prévision est observée d'un horizon de prévision à un autre. L'estimation directe à l'horizon de prévision, sans cumul d'évaluations en est la cause.

3.2.2 Performances globales en fonction de l'horizon de prévision

Un premier aperçu des performances de prévision des 7 méthodes considérées est proposé dans la Figure 3.7. Les moyennes des indicateurs usuels d'erreur sont comparées en Figure 3.7 a) et b), tandis que leurs distributions spatiales et temporelles à l'horizon 1h30 sont représentées, respectivement en Figure 3.7 c) et d).

Quel que soit l'horizon de prévision, la méthode **naïve**, servant de base de comparaison (*benchmark*), est surpassée. Les méthodes proposées sont donc pertinentes.
Seule la méthode de régression à noyaux (Kernel) se distingue, pour toute approche, par des performances de moindre qualité que ses homologues, phénomène accentué aux horizons élevés. L'approche **LocalKernel** atteint même des niveaux d'erreur équivalents à la méthode naïve. La forte sensibilité de la régression **LocalKernel** à l'hyperparamètre dû au

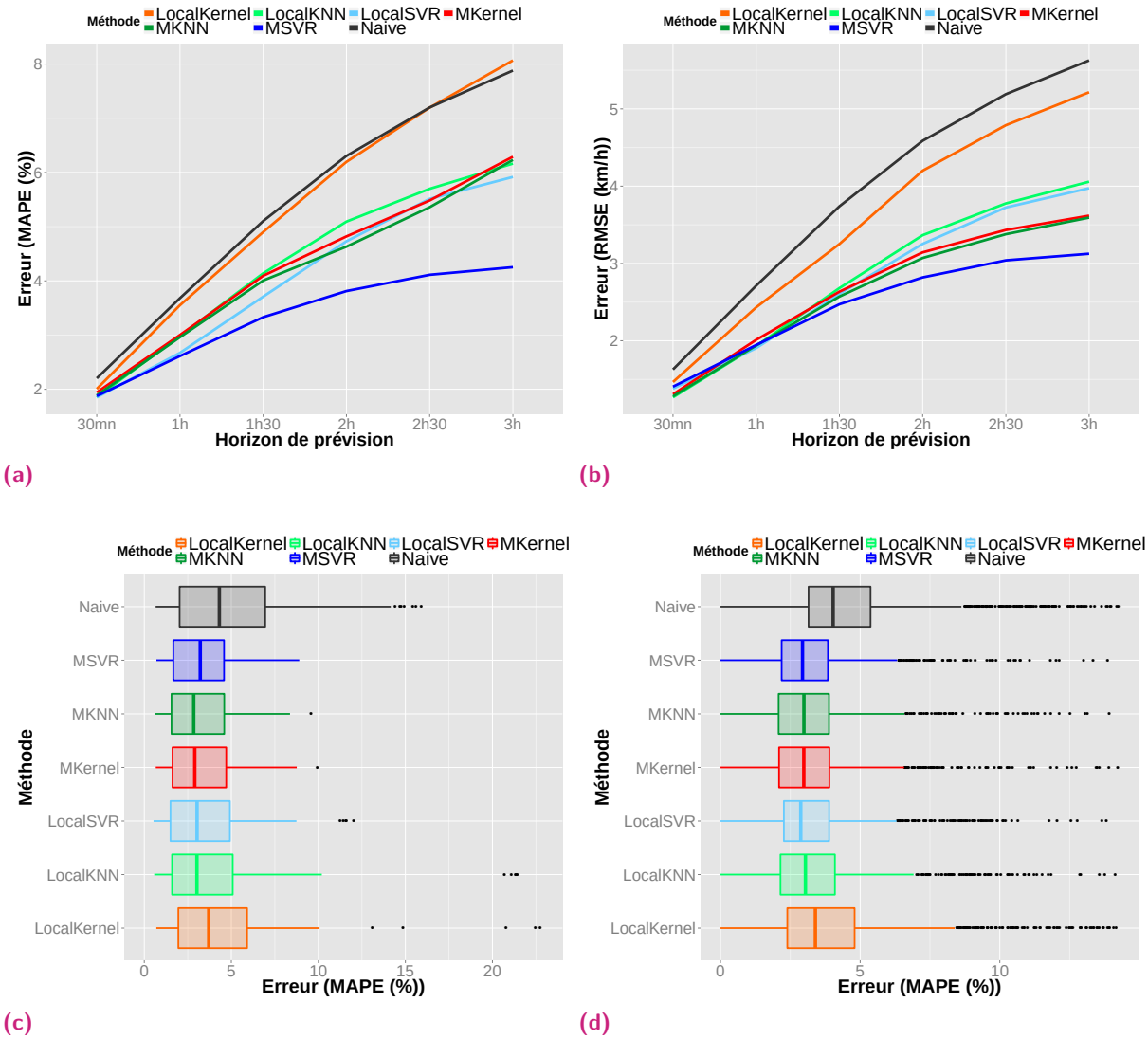


Fig. 3.7.: Performances globales de prévision des 7 méthodes testées suivant l'horizon de prévision : a) Erreur relative globale ($MAPE_{global}$) b) Erreur globale ($RMSE_{global}$) c) Distribution des erreurs par section ($MAPE_{section}$) à l'horizon 1h30 d) Distribution des erreurs par pas de temps ($MAPE_{TimeStep}$)à l'horizon 1h30.

noyau (fenêtre du noyau γ) constitue une cause possible. L'allure des distributions d'erreur par section (Figure 3.7 c)) ne montre pas de différence significative entre **LocalKernel** et **LocalKNN**, si bien que le procédé de réaffectation des hyperparamètres ne peut-être accusé, mais constitue également une piste à envisager. Bien que la version globale (**MKernel**) soit plus efficace que certaines approches locales, elle reste la moins performante des approches globales.

Globalement, les performances de prévision sont relativement équivalentes avec un léger avantage aux approches globales. Peu de différences sont observables d'une méthode ou approche à une autre. **L'analyse sur un vaste champ de méthodes conforte donc l'intérêt de**

l'approche globale en termes de performances de prévision. Aucune perte de performance n'est constatée.
Avec l'horizon de prévision, les écarts de performance se creusent. **Au-delà de 1h30, l'approche globale MSVR se distingue** et tend à borner son erreur globale en suivant une croissance logarithmique. Cependant, **sur de plus courts horizons, les méthodes les plus simples restent très efficaces** et le gain en termes d'erreur de prévision est extrêmement faible. Même l'approche naïve se révèle globalement performante à 30 minutes (environ 2 % d'erreur). Les faibles variations de vitesse du flux sur de courts horizons expliquent de telles performances. Des procédés plus évolués assurent en revanche une bonne robustesse sur le long terme.
Néanmoins, ces indicateurs globaux cachent une réalité à l'échelle locale parfois bien différente, comme l'illustre la répartition des erreurs par section ou pas de temps de l'approche naïve (Figure 3.7 c) et d)). Certaines sections connaissent des erreurs relatives moyennes supérieures à 10 % pour cette méthode. De plus, l'intérêt de développer des méthodologies plus complexes de prévision réside dans l'opportunité de mieux répondre à des phénomènes peu récurrents ou à forte variabilité. Une analyse plus fine des erreurs est proposée par la suite.

3.2.3 Analyse des performances suivant les périodes de la journée

Suivant la période de la journée considérée, la difficulté de prévision est variable, ainsi que le coût de toute erreur de prévision. Malheureusement, les périodes de la journée appelant le plus grand intérêt sont également les plus difficiles à prévoir. Il s'agit en effet des périodes de congestion récurrente connaissant une forte variabilité des comportements des usagers et des vitesses sur de courtes périodes : les pics de demandes liés aux déplacements pendulaires. La Figure 3.8 e) montre que la plupart des méthodes développées ne font pas exception, bien que **l'approche globale est nettement moins sensible aux pics de trafic**. À court terme, les erreurs les plus élevées sont observées durant les phases de forte activité de la journée. On constate entre 1 et 3 km/h de différentiel d'erreur à l'horizon 60 minutes entre les périodes nocturnes ou à mi-journées et celles de forte activité.
Étant donné le sens de circulation considéré, le pic du matin est particulièrement marqué au sein des données. Ce pic est le plus pertinent à analyser car il se retrouve sur l'erreur de prévision. Si la Figure 3.8 a) donne peu d'éléments permettant de distinguer les méthodes de prévision, **la croissance de l'horizon de prévision met en évidence une meilleure robustesse des approches globales lors de ce pic du matin**. Cette tendance est confirmée par la Figure 3.8 c) et d) montrant que les méthodes MSVR et MKNN, puis seulement LocalSVR réduisent la disparité d'erreur suivant les pas de temps dès l'horizon 1h30. L'analyse par section suivant les périodes de la journée (Figure 3.9 c)) abonde en ce sens. **L'usage des approches globales réduit donc la variabilité de l'erreur par section comme par période temporelle lorsque l'horizon croît.**

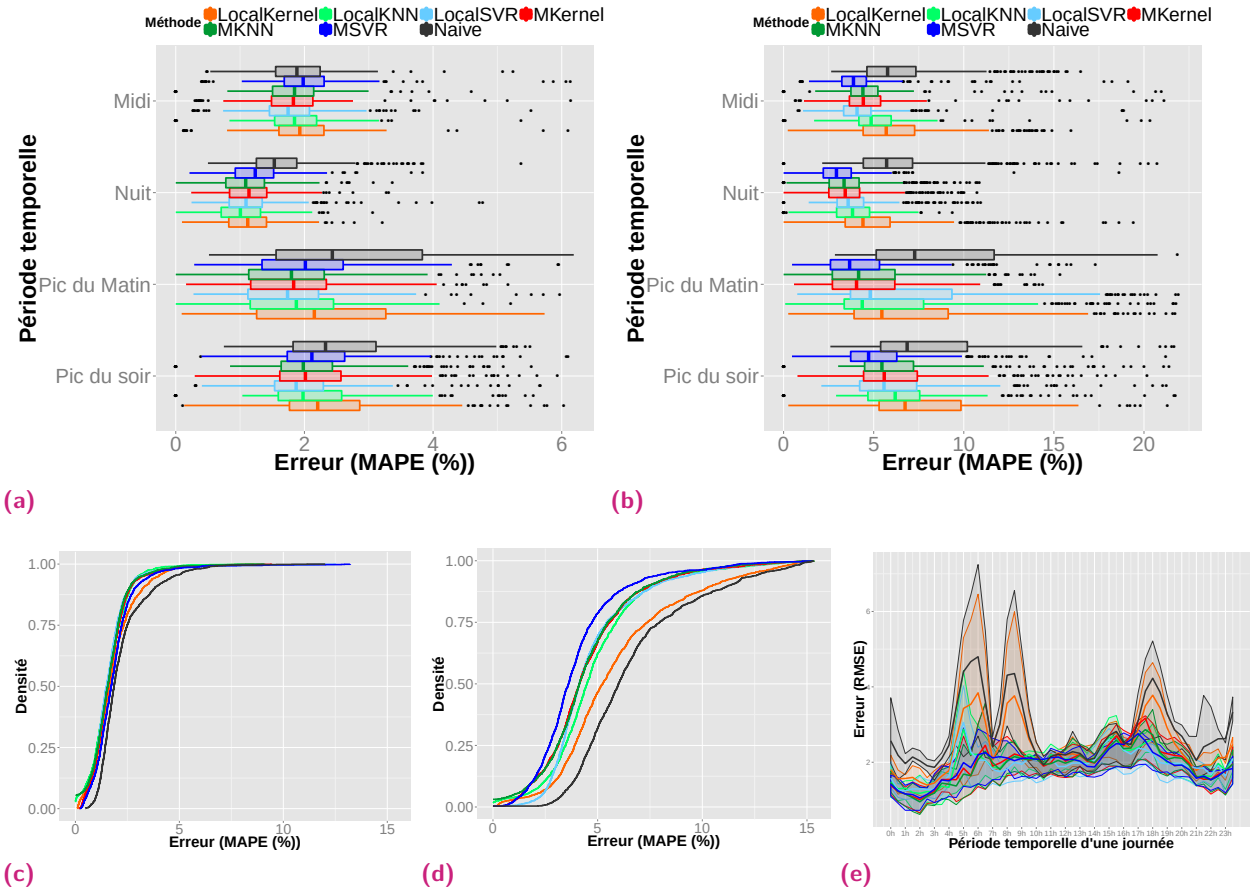


Fig. 3.8.: Analyse locale et par pas de temps ($MAPE_{TimeStep}$) des erreurs de prévision : a) Distribution des erreurs suivant la période de la journée considérée à l'horizon 30mn b) Distribution des erreurs suivant la période de la journée considérée à l'horizon 3h c) Distribution cumulée des erreurs de prévision par pas de temps à l'horizon 30mn d) Distribution cumulée des erreurs de prévision par pas de temps à l'horizon 3h e) Courbe moyenne et premier et dernier quartiles de l'erreur de prévision ($RMSE_{TimeStep}$) en fonction de l'heure de la journée (horizon 60mn).

3.2.4 Analyse des performances suivant les sections considérées

Les performances de prévision des diverses méthodes sont également susceptibles de se distinguer suivant les sections ou zones du réseau évaluées. En effet, les zones fortement maillées, les bretelles d'accès ou de sorties constituent autant de zones sensibles du réseau difficiles à prévoir. Il est dès lors tout à fait possible que des méthodes soient plus aptes à prévoir certaines sections que d'autres, d'autant que des approches locales et globales sont comparées. L'intuition laisse penser qu'une approche locale est plus apte à gérer les difficultés de prévision d'une section spécifique, puisque sa fonction objectif vise à minimiser la seule erreur de prévision de cette section. Pourtant, les distributions d'erreur, exposées en Figure 3.9 a) et b), mettent en évidence que les approches globales MKNN et MSVR produisent en moyenne légèrement moins d'erreur sur l'ensemble des sections. L'analyse des distributions durant le pic du matin (Figure 3.9 c)) montre que l'erreur n'excède 12 % sur aucune section

pour MSVR et MKNN, quand les moins bonnes sections des approches locales atteignent 20 %. En revanche, sur toute autre période de la journée, les distributions sont similaires.

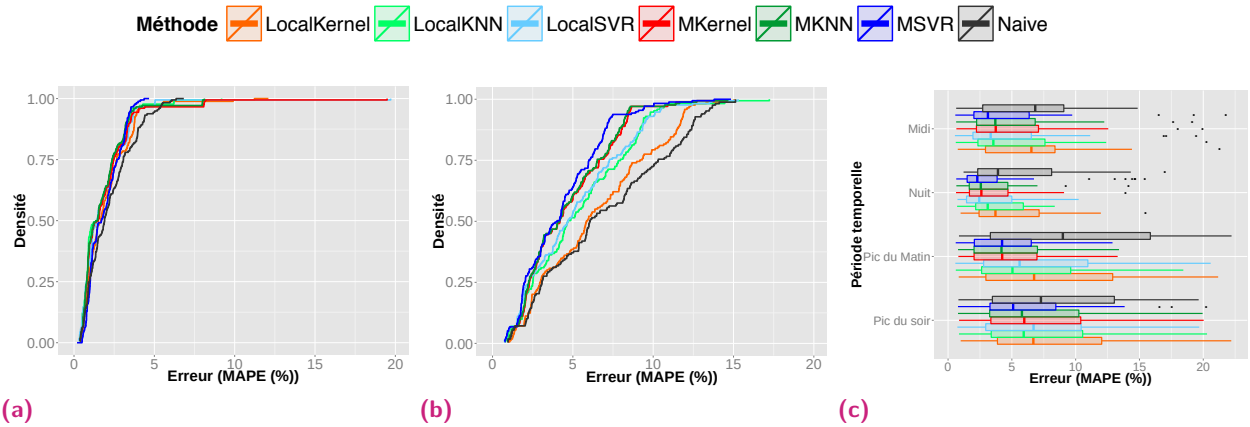


Fig. 3.9.: Analyse locale et par section ($MAPE_{section}$) des erreurs de prévision : a) Distribution cumulée des erreurs moyennes expérimentées par les sections à l'horizon 30 mn b) Distribution cumulée des erreurs moyennes expérimentées par les sections à l'horizon 3h c) Distribution des erreurs relatives moyennes expérimentées par les sections suivant les périodes de la journée à l'horizon 3h.

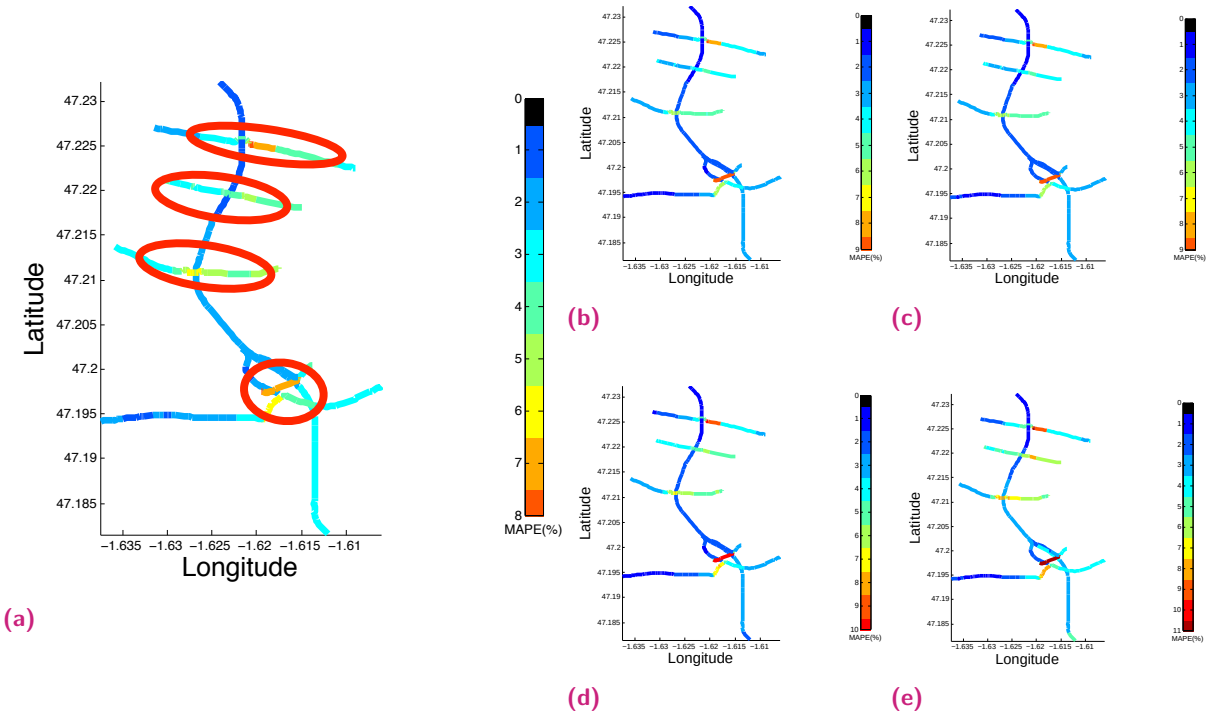


Fig. 3.10.: Analyse de la répartition spatiale de l'erreur de prévision ($MAPE_{section}$) : a) Résultats de la méthode MSVR b) Résultats de la méthode MKNN c) Résultats de la méthode MKernel d) Résultats de la méthode LocalKNN e) Résultats de la méthode LocalKernel.

En vue d'approfondir l'analyse, la distribution spatiale de l'erreur relative moyenne par section du réseau est représentée dans la Figure 3.10.

Il y est effectivement constaté que **les mêmes portions du réseau posent problème aux diverses méthodes testées et ce quelle que soit l'approche considérée.** Les zones identifiées sont les principales branches d'accès au Nord, ainsi que l'échangeur (ou apparenté) au Sud au niveau de la Loire. Ces portions expérimentent également les plus faibles vitesses et la plus forte variabilité. On remarque que la branche d'accès Est, la plus au Nord, est marquée par une erreur plus élevée que les autres. Cette branche conduit directement aux principaux axes du centre-ville nantais. L'unique autre portion atteignant des erreurs aussi élevées est la branche Est de l'échangeur Sud reliant la rive opposée de la zone urbaine.
Cependant, l'intensité d'erreur varie en fonction du processus de prévision adopté. Ainsi, il paraît plus aisé aux approches locales d'anticiper les événements se produisant sur l'axe principal Sud-Nord du périphérique nantais, tandis que les approches globales expérimentent légèrement moins d'erreurs ($\sim$ 1%) sur les branches d'accès.

Contredisant l'intuition, les approches globales ne favorisent pas un sous-ensemble de sections, malgré l'usage d'une fonction objectif minimisant l'erreur globale à l'échelle du réseau. Les corrélations spatio-temporelles liant les sections sont sans doute à l'origine d'un tel phénomène. Les sections étant connectées les unes aux autres, il est difficile de négliger une section sans occasionner une propagation d'erreur se ressentant à l'échelle globale. La fonction objectif globale est donc indirectement contrainte par la structure spatio-temporelle du réseau, maintenant une qualité de prévision spatialement bien répartie.
L'une ou l'autre approche possède sa propre zone de confort en termes géographiques, mais aucune disparité spatiale de l'erreur de prévision n'est constatée sur les approches globales.

3.2.5 Typologies spatiales et temporelles des meilleures méthodes

Nous proposons, dans cette partie, de mettre en concurrence les sept méthodes de prévision considérées et d'identifier celles parvenant à prévoir la plus grande quantité de sections ou de pas de temps suivant les horizons ou les périodes de la journée. L'idée sous-jacente est d'identifier les domaines spatiaux et temporels de prédilection de chacune des méthodes.

Outre la prédominance des principales méthodes MKNN, LocalSVR et MSVR suivant les horizons de prévision, la caractérisation spatio-temporelle proposée par les Figures 3.11 et 3.12 met en avant les spécificités de chacune des méthodes testées. Tandis que la Figure 3.11 confirme les constats quant aux meilleures méthodes de prévision, la Figure 3.12 qualifie les performances au regard de la dynamique du trafic.
À court-terme, l'axe principal du périphérique nantais est mieux couvert durant les périodes de fortes affluences par MKNN, tandis que LocalSVR est performant de nuit. MKNN remporte également les principales rampes ou branches d'accès à l'exception de la zone difficile à prévoir au Nord-Est.
À plus long terme, LocalSVR prédomine sur l'axe principal du périphérique nantais durant la nuit, mais les périodes de pics de trafic sont principalement couvertes par les approches

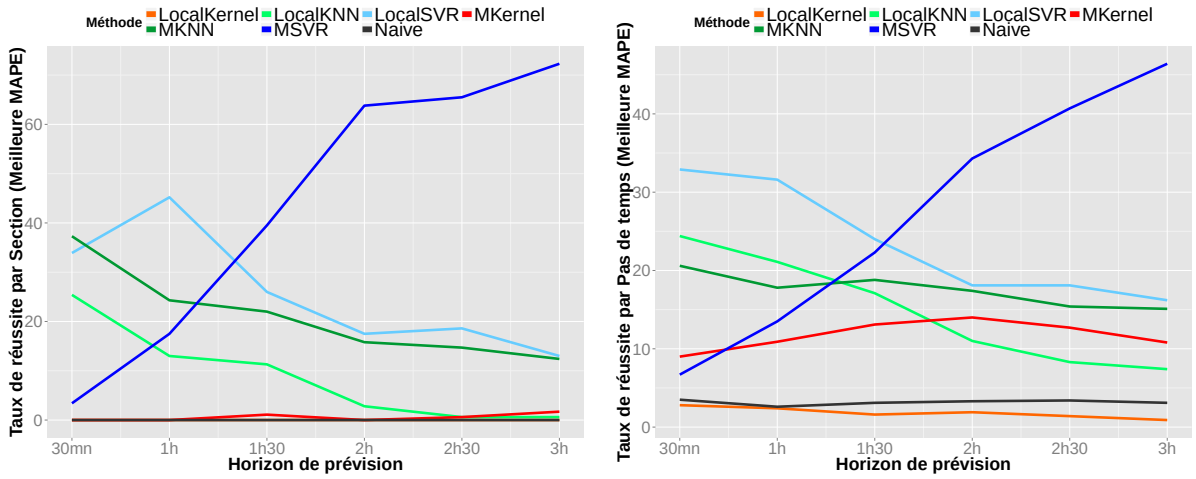


Fig. 3.11.: Mise en concurrence des sept méthodes de prévision : a) Évaluation du pourcentage de sections remportées par chaque méthode en fonction de l'horizon de prévison b) Évaluation du pourcentage de pas de temps remportés par chaque méthode en fonction de l'horizon de prévison.

globales. La méthode MSVR couvre l'essentiel des branches d'accès et quelques portions de l'axe principal.
Les approches globales présentent, finalement, une meilleure robustesse à l'horizon de prévision et sont plus aptes à anticiper l'état de trafic des sections sur les branches d'accès.

La mise en relation de ces observations avec la Figure 3.10 exposant la distribution spatiale des erreurs par méthode se révèle riche, en particulier sur les zones qualifiées de difficiles à prévoir. La branche d'accès Nord-Est donnant sur le centre-ville est, à court-terme, alternativement mieux prévue par LocalKNN et LocalSVR durant les périodes d'affluence, puis par MKernel la nuit. À plus long terme, MSVR et MKernel présentent les meilleures performances. L'échangeur Sud au niveau de la Loire est lui aussi très disputé suivant les sections considérées et la période de la journée. **Ces zones difficiles à prévoir exacerbent les propriétés de chacune des méthodologies et les mettent en avant.** Ainsi, la méthode MKernel, qui se révèle, de manière générale, moins performante que les autres approches globales, est en réalité extrêmement performante sur certaines sections compliquées.

Ce procédé de mise en concurrence des méthodes de prévision n'est pas sans rappeler le **concept des méthodes ensemblistes**. Ces dernières évaluent divers modèles extrêmement simples et parcimonieux pour en construire un nouveau composé des différents modèles s'activant suivant des critères bien définis. L'approche globale ayant été développée dans une perspective de simplification et d'unification spatiale des modèles de prévision, l'étape suivante, naturelle, consisterait à proposer un modèle unique composé d'un ensemble d'approches locales

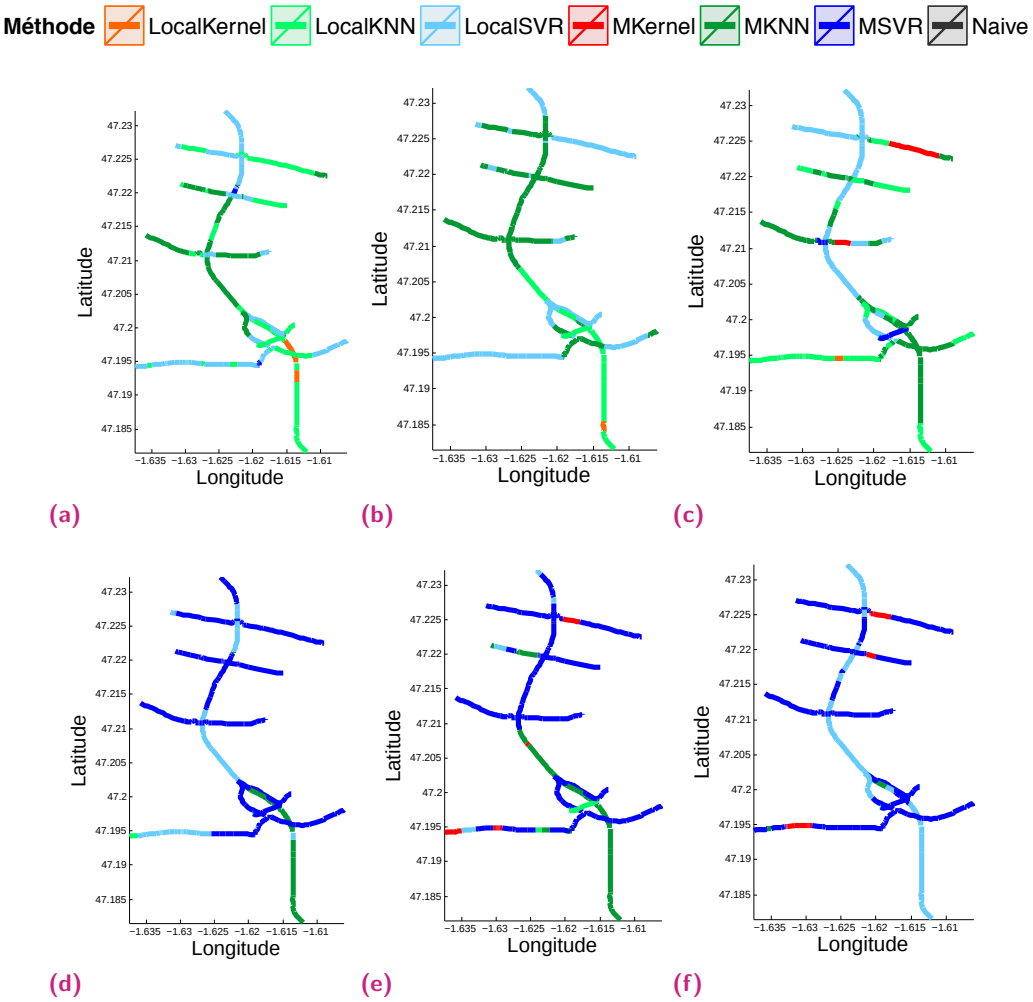


Fig. 3.12.: Représentation spatiale des meilleures méthodes en fonction de la période de la journée : a) Évaluation durant le pic du soir à l'horizon 30mn b) Évaluation durant le pic du matin à l'horizon 30mn c) Évaluation durant la nuit à l'horizon 30mn d) Évaluation durant le pic du soir à l'horizon 3h e) Évaluation durant le pic du matin à l'horizon 3h f) Évaluation durant la nuit à l'horizon 3h. La $MAPE_{section}$ est considérée pour comparer les performances.

s'activant à tour de rôle. Notre effort a posé les bases d'une analyse pouvant conduire à un tel processus.

3.3 Analyse comparative pour le réseau très maillé d'une zone urbaine (centre-ville nantais)

Après l'introduction du cas d'étude du Pont de Cheviré, réseau semi-urbain peu maillé et liant portions périphériques à grande vitesse et branches d'accès, il est proposé d'évaluer le passage à grande échelle de l'approche globale sur le réseau très maillé du centre-ville nantais. Outre le maillage resserré du réseau, les feux tricolores, stops ou rond-points promettent une dynamique

de trafic bien particulière. Cette nouvelle zone d'étude constitue l'occasion de confirmer les performances de l'approche globale. La zone urbaine de Nantes est composée de 796 liens, sur lesquels les vitesses moyennes spatiales sont recueillies avec un pas de temps de 30 minutes entre septembre et novembre 2013. Les performances de prévision sont évaluées à partir des 1440 pas de temps du mois de novembre 2013.

3.3.1 Évolution de la performance quant à la complexité temporelle

En ce qui concerne la complexité temporelle, le nombre conséquent de sections composant le réseau urbain signifie une forte croissance de la durée d'optimisation des hyperparamètres et du temps d'estimation de toute nouvelle prévision pour les approches locales. La dépendance des approches globales à la taille du réseau étant indirecte, les temps de calibrage et d'estimation sont moins altérés.
En termes de durée d'optimisation, entre deux et quatre jours sont nécessaires pour les approches globales, quand 1 mois et demi est requis pour estimer les meilleurs hyperparamètres d'un sous-ensemble de sections représentant ∼10 % du réseau pour l'approche locale KNN. Étant donné les performances médiocres de la méthode LocalKernel sur le cas de Cheviré pour le temps d'optimisation requis, cette dernière n'est pas estimée sur la zone urbaine de Nantes.

Zone	Approche	Méthode	Durée totale	Durée par pas de temps
Urbaine (796 sections)	**globale**	MKNN	3,5mn	0,2s
		MKernel	3,2mn	0,2s
		MSVR	86mn	3,6s
	locale[3]	KNN	∼14h	∼35s
		SVR	∼10 jours	∼10mn
Cheviré (177 sections)	**globale**	KNN	43s	$\ll$ 1s
	locale	KNN	33mn	1,4s

Tab. 3.2.: Table comparative des ordres de grandeur des temps d'apprentissage et de prévision en fonction de la méthode de prévision pour le mois de novembre 2013 (1440 évaluations) sur la zone urbaine de Nantes : confrontation de l'approche locale à l'approche globale.

Les durées totales et moyennes par pas de temps sont introduites en table 3.2 pour chacune des méthodes de prévision considérées. La forte croissance de la durée d'évaluation requise est observée sur les approches locales, atteignant, pour LocalSVR, des durées pratiquement inacceptables pour une application en temps-réel. Néanmoins, le passage à grande échelle avec un réseau très maillé est supporté par toutes les méthodes et approches. Il est à souligner que l'approche globale n'excède pas quelques secondes d'estimation pour la méthode la plus complexe (SVR) considérée.
Les approches les plus simples KNN et Kernel effectuent un apprentissage immédiat et, une

3. Les approches locales sont évaluées par des calculs parallèles. Seule la méthode LocalKNN est significative, car les autres approches locales ont été évaluées sur un ordinateur distinct.

nouvelle fois, présentent des temps d'exécution extrêmement courts au regard des performances temporelles du SVR. Le temps d'exécution du MSVR reste toutefois parfaitement acceptable.

3.3.2 Évaluation des performances vis-à-vis de l'erreur de prévision

3.3.2.1 Analyse de l'erreur globale sur la période de validation

Pour ce qui est de l'erreur de prévision, les performances de toutes les méthodes sont moindres sur ce réseau aux multiples contraintes (maillage resserré, feux tricolores *etc*). Quel que soit l'horizon de prévision, au moins 3% d'erreur relative moyenne ($MAPE_{global}$) supplémentaires sont constatés. La confrontation des Figures 3.7 c) et 3.13 b) met même en évidence un différentiel de 5 % à partir de l'horizon 1h30. Les erreurs relatives moyennes s'étendent ainsi de 4 % pour un horizon court-terme de 30 minutes à 9-10% pour une prévision à 3h. Les dynamiques de trafic distinctes de celles du Pont de Cheviré mettent à rude épreuve les méthodes de prévision quelles qu'elles soient.
Les tendances identifiées sur le réseau du Pont de Cheviré se trouvent exacerbées sur la zone urbaine de Nantes, comme en témoigne la Figure 3.13. Ainsi, à court-terme, les performances de la méthode MSVR sont comparables à celles de la méthode naïve de référence. MSVR rejoint ses concurrentes dès l'horizon 1h30 et produit à 3h une erreur au moins 20 % inférieure à toute autre approche. MKNN, MKernel et LocalSVR fournissent suivant l'indicateur global des performances équivalentes, tandis que localKNN s'en distingue lorsque l'horizon de prévision excède 1h30. Finalement, seule l'approche globale à vecteurs supports apporte une variabilité de comportement identifiable à l'échelle globale d'erreur.

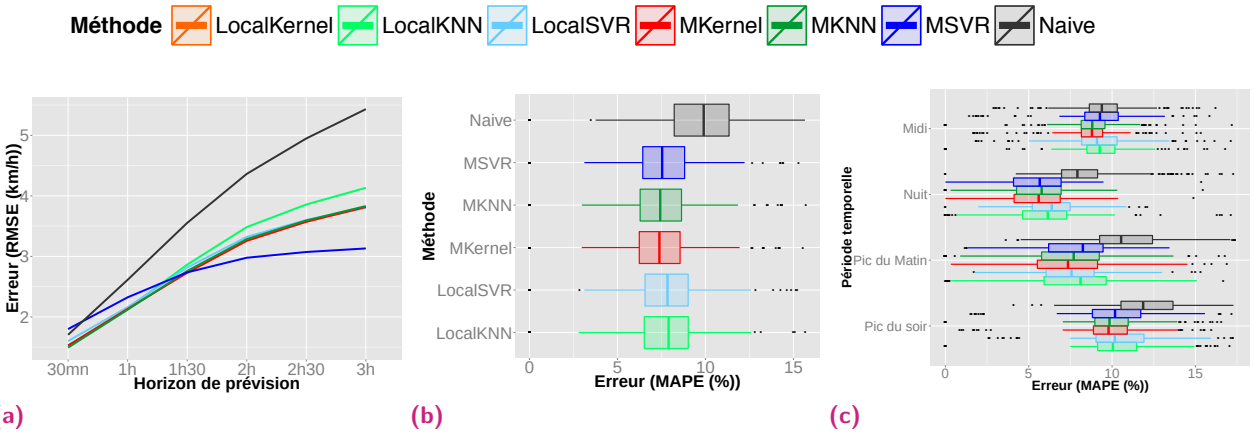


Fig. 3.13.: Performances globales de prévision des six méthodes testées suivant l'horizon de prévision sur la zone urbaine de Nantes : a) Performance en termes d'erreur globale ($RMSE_{global}$) b) Distribution des erreurs par section ($MAPE_{section}$) à l'horizon 1h30 c) Distribution des erreurs par pas de temps ($MAPE_{TimeStep}$) à l'horizon 1h30.

Sur cette zone urbaine, aucune méthode ne néglige des sections plus qu'une autre (voir Figure 3.13 b)), tandis que les comportements varient légèrement suivant les périodes de la journée.

Les pics de trafic du matin sont quelque peu mieux prévus par les approches globales MKNN et MKernel et l'étendue de leur variabilité d'erreur est moindre au milieu de la journée (Figure 3.13 c)).

3.3.2.2 Analyse spatiale et temporelle

Une analyse plus fine, se focalisant sur les performances suivant les périodes temporelles et les zones du réseau considérées, met en avant trois points :

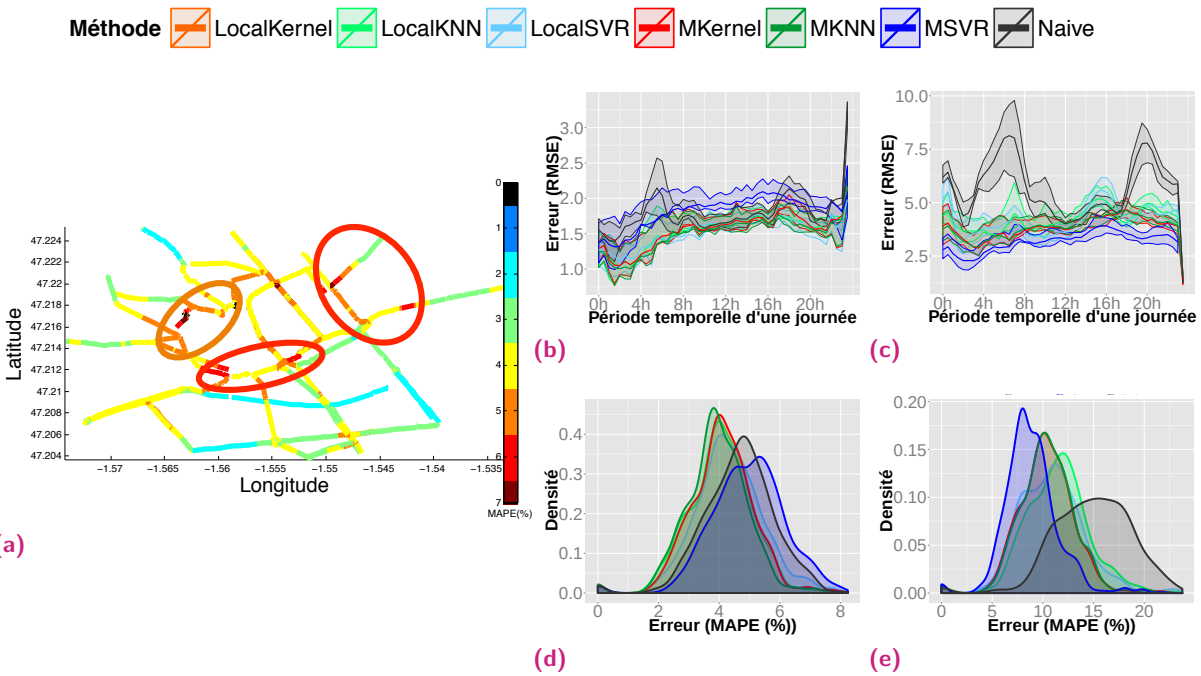


Fig. 3.14.: Analyse locale ($MAPE_{TimeStep} et MAPE_{section}$) des erreurs de prévision : a) Répartition spatiale de l'erreur de prévision par section ($MAPE_{section}$) à l'horizon 30mn pour la méthode MKNN b) Courbe moyenne et premier et dernier quartiles de l'erreur de prévision en fonction de l'heure de la journée (horizon 30mn) c) Courbe moyenne et premier et dernier quartiles de l'erreur de prévision en fonction de l'heure de la journée (horizon 3h) d) Distribution des erreurs de prévision ($MAPE_{section}$) pour l'ensemble des sections du réseau à l'horizon 30mn e)Distribution des erreurs de prévision ($MAPE_{section}$) pour l'ensemble des sections du réseau à l'horizon 3h.

– Bien que les périodes de forte affluence de la journée soient, en principe, plus difficiles à prévoir, **peu de dépendances par rapport aux moments de la journée** sont observées sur les Figures 3.14 b) et c) ainsi que 3.13 c). Seules les performances de la méthode naïve sont encore marquées par ces pics de demande de trafic. L'origine d'un tel phénomène reste inconnue, d'autant que toutes les approches sont concernées à l'exception de la méthode naïve. Pourtant, la variation de l'erreur de prévision de l'approche naïve suivant les heures considérées témoigne, de par sa construction, de l'existence de variabilité durant les phases de pics. L'existence de dynamiques spécifiques à la zone urbaine ne suffit donc pas à expliquer cette indépendance.

– À l'exception de la méthode naïve, **peu de disparité dans le traitement des sections entre les méthodes** est observée (Figure 3.14 d) et e)). Les formes des distributions sont similaires, à l'exception de celles de la méthode naïve. Les seules véritables différences proviennent de performances globales de prévision distinctes.

– De même que pour le cas de Cheviré, peu de distinctions entre les approches ou méthodes de prévision sont observées en termes de répartition spatiale de l'erreur. La figure 3.14 a) constitue dès lors un bon représentant de la répartition de l'erreur court-terme sur les diverses zones du réseau urbain. Outre les sections en bordure interne [4] du réseau, deux zones se distinguent par leur maigre performance (entourées en rouge) : les axes à l'Est de la zone urbaine, conduisant au Pont de Bellevue, et ceux longeant la rive Nord au Sud de la zone urbaine dense. Finalement, **les zones les plus difficiles à prévoir sont celles du centre-ville dense combinant feux tricolores, zones piétonnes et rues plus étroites à proximité.** Tous ces facteurs introduisent une variabilité compliquée à saisir. Beaucoup de ceux-ci sont influencés ou même liés aux comportements des personnes (piétons, recherche de stationnement etc) et donc à des facteurs extérieurs au trafic impossibles à déterminer à l'avance.

3.3.2.3 Identification des typologies spatio-temporelles des meilleures méthodes

Pour affiner cette analyse comparative en zone urbaine dense, les performances de prévision sont comparées sur toute section et par période de la journée. Les résultats sont exposés en Figure 3.15.

Contrairement au cas de la zone du Pont de Cheviré, peu de différences sont observées d'une période temporelle à une autre de la journée. À court-terme, MKNN remporte la grande majorité des sections, tandis que MSVR prédomine largement lorsque l'horizon excède 1h30. Néanmoins, **les sections réputées plus difficiles ou plus faciles à prévoir sont disputées**. On remarque ainsi que les 2 axes longeant la rive Sud ne dépassant pas 2-3 % d'erreur à l'horizon 30 minutes (Figure 3.14 a)) sont partagées suivant les périodes considérées par LocalSVR, localKNN et MKernel. Le phénomène est accentué durant la phase du pic du matin et, à l'horizon 3h, n'est présent que durant cette période avec la prédominance du MKernel.
Le même partage est observé sur la zone des sections en bordure interne de réseau apparaissant en orange sur la Figure 3.14 a) : ces liens du réseau s'avèrent mieux prévus par les approches locales et MKernel. Pour ces sections en bordure de réseau, la spécificité des approches locales, parfaitement conçues pour saisir l'information requise localement, conduit naturellement à de meilleures performances. Le fait que le réseau soit tronqué en son centre cache une partie importante de l'information et éventuellement des sections critiques pour le réseau, ce défaut d'information est mieux contourné par l'approche locale. Ceci souligne **une limite de l'approche globale, à savoir le choix du réseau et sa délimitation**. Il est néanmoins à

4. Zone entourée en orange. Seuls les grands axes du centre-ville nantais sont considérés comme partie intégrante du réseau. Les petites rues ou rues commerçantes du centre-ville ne sont pas représentées, ce qui constitue une perte d'information et limite les performances des sections en bordure.

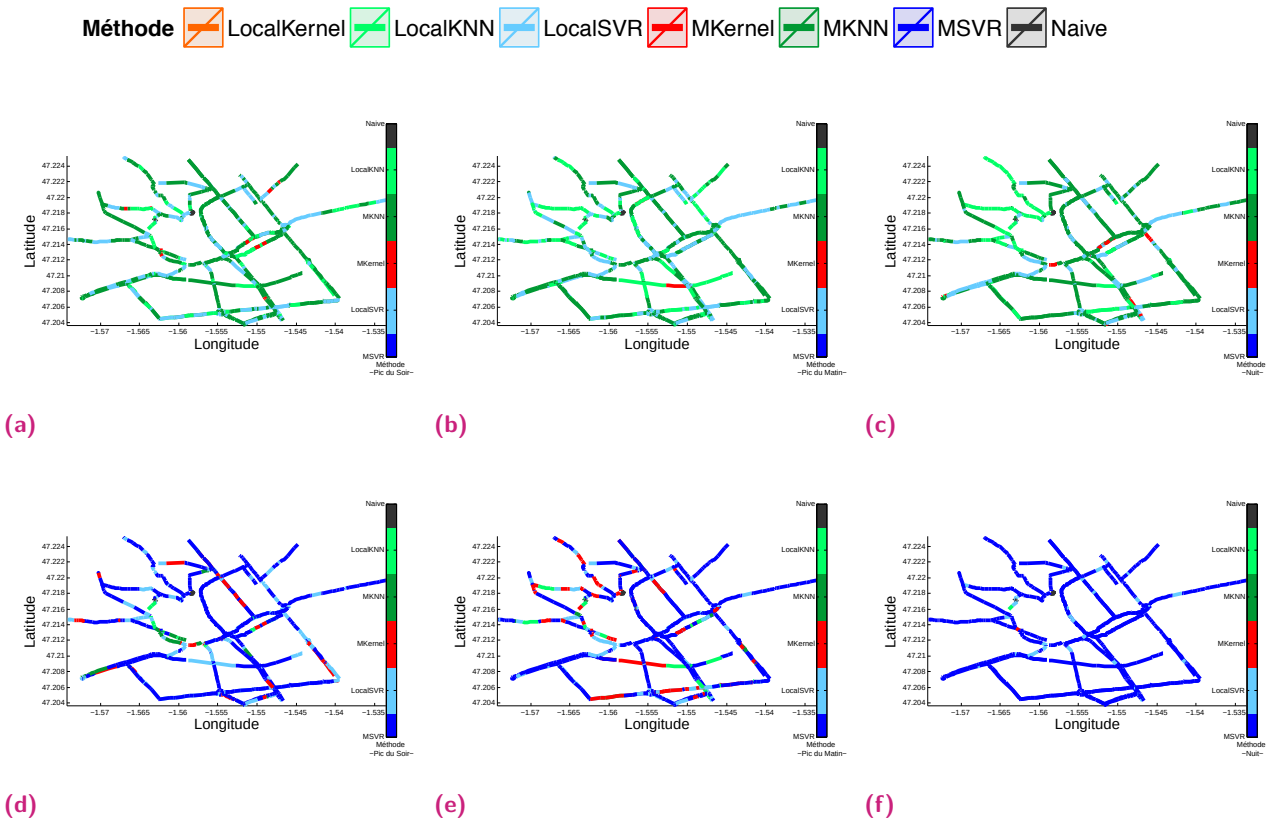


Fig. 3.15.: Représentation spatiale des meilleures méthodes en fonction de la période de la journée sur la zone urbaine de Nantes : a) Évaluation durant le pic du soir à l'horizon 30mn b) Évaluation durant le pic du matin à l'horizon 30mn c) Évaluation durant la nuit à l'horizon 30mn d) Évaluation durant le pic du soir à l'horizon 3h e) Évaluation durant le pic du matin à l'horizon 3h f) Évaluation durant la nuit à l'horizon 3h. La $MAPE_{section}$ est considérée pour comparer les performances.

noter que, de manière générale, MKernel concurrence directement les approches locales sur ce terrain, comme en témoigne son omni-présence sur ces mêmes sections. MKernel semble effectivement être moins soumise ou contrainte par ces conditions difficiles, sans doute grâce à son noyau et la richesse de l'historique complet considéré.

Les deux zones identifiées difficiles à prévoir, en rouge sur la Figure 3.14 a), sont souvent remportées par les méthodes majoritaires de l'horizon de prévision et LocalSVR. L'approche locale à Vecteurs Supports se révèle particulièrement efficace pour anticiper les états de trafic à tout instant d'une section à l'Est expérimentant une erreur à 30 minutes supérieure à 6%. Pour de rares cas, des approches globales alternatives sont plus performantes.

Finalement, **les cas difficiles sont mieux traités par les approches globales MKNN et MSVR suivant les horizons de prévision et l'approche locale du SVR, tandis que les approches locales et MKernel semblent plus en mesure de prévoir les zones mal représentées du réseau urbain (maillons manquant du réseau, relation spatio-temporelles partielles *etc*).** En ce qui concerne les sections posant moins de difficultés, les différentes approches globales prédominent sans contestation.

L'étude de cette nouvelle zone nantaise propose une dynamique urbaine caractérisée par un maillage et une taille distincte du réseau. Les constats formulés sur la zone semi-urbaine du Pont de Cheviré sont confirmés : **prédominance des méthodes MKNN et localSVR à court-terme, puis progressivement du MSVR lorsque l'horizon croît.** Les performances globales de prévision restent extrêmement proches quel que soit le processus de prévision choisi, mais des divergences à l'échelle plus locale sont constatées. Les sections préalablement identifiées comme difficiles sont particulièrement disputées par les méthodes de prévision, ce qui confirme l'intérêt d'une mise en œuvre ensembliste du procédé de prévision. Enfin, une limite de l'approche globale a été mise en évidence : un mauvais découpage du réseau conduit à de moindres performances de cette approche sur les sections en bordure de réseau. L'approche globale, de par son principe de construction, se révèle sensible à la qualité de découpage du réseau.

Avant de véritablement conclure, toutes ces performances des méthodes à gestion d'historique sont à mettre au regard d'autres méthodes de prévision de la littérature.

3.4 Confrontation à des méthodes alternatives de l'état de l'art

Afin de remettre les performances de prévision introduites préalablement dans un contexte plus général, cette section présente la confrontation sur la zone du Pont de Cheviré d'une des meilleures méthodes exposées, MSVR, à d'autres procédés de prévision tirés de la littérature. Ces approches alternatives ont été testées [166] dans un esprit de mise en concurrence des méthodes. Les conditions d'expérimentation les plus similaires possibles ont été maintenues, mais quelques modifications sont néanmoins à noter. Les méthodes implémentées intègrent des informations extérieures au trafic telles que l'heure et la typologie de jour. L'ajout de telles informations est à double tranchant : il permet, certes, de restreindre la zone de recherche de l'information similaire au sein des historiques tout en limitant la variabilité de l'erreur pour des périodes habituelles, mais, lors de phénomènes moins récurrents, bride la variabilité des réponses apportées par l'algorithme de prévision et endigue tout lien éventuel entre deux situations n'appartenant pas aux mêmes périodes temporelles. Les effets d'une telle mesure seront donc pris en considération durant l'analyse des performances de prévision.

3.4.1 Introduction succincte aux méthodes de l'état de l'art testées

Les trois méthodes testées ont été choisies pour représenter les différentes familles de méthodes de prévision non encore évaluées, à savoir les méthodes paramétriques, les méthodes apprenant la structure de l'historique et les approches ensemblistes.

3.4.1.1 Méthode paramétrique simple

La méthode paramétrique sélectionnée est la plus simple existant dans la littérature, à savoir la régression linéaire. La formulation de la prévision est la suivante :

$$\forall t, \quad \widehat{y_{t+h}} = f(\boldsymbol{x_t}) = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x_t} + b$$

Il est dès lors supposé qu'une relation linéaire suffit à décrire la relation entre les individus en entrée et le label de sortie correspondant à l'état de trafic prévu. Une approche **locale** a été adoptée pour construire le modèle testé. L'introduction de l'information temporelle implique donc la création d'un modèle par typologie de jour, période horaire et section. En vue de conserver suffisamment d'informations pertinentes, seules trois typologies de jour sont distinguées : jours d'activité, week-end, vacances. L'optimisation des paramètres ($\boldsymbol{W}$, b) procède par la détermination de la droite des moindres carrés, donc la minimisaiton de l'erreur $RMSE$. La résolution est immédiate en annulant les dérivées partielles de la formule prédictive.

3.4.1.2 Deux approches ensemblistes apprenant la structure de l'historique

Les deux approches ensemblistes proposées sont les forêts d'arbres décisionnels (*Random Forest*) [87] et les *Gradient Tree Boosting* [167]. Bien qu'apparus très récemment, ces procédés de prévision connaissent un franc succès de par leur simplicité et leur efficacité [85, 86, 167].
Ces deux méthodes reposent sur l'apprentissage par arbres de décision et, en ce sens, **apprennent la structure de l'historique**. Chaque arbre (*CART*) se construit en subdivisant l'espace d'entrée. Chaque subdivision marque le partage d'une branche en 2 sous-branches de l'arbre. Les labels de sortie ou prévisions sont les feuilles des arbres. Ainsi, toute subdivision de l'espace d'entrée est associé à un unique label de sortie homogène sur toute la zone. La construction de l'arbre est dictée par les couples Individu $\times$ Label du jeu d'apprentissage. L'arbre ou modèle de prévision explicite donc la structure décisionnelle aboutissant à toute prévision. À profondeur croissante de l'arbre, ces derniers s'avèrent extrêmement efficaces pour identifer les motifs spécifiques de l'historique d'apprentissage. Malheureusement, cette particularité les conduit souvent au sur-apprentissage.
Les méthodes proposées font toutes deux appel à une sélection ou combinaison d'une grande quantité de modèles construits depuis des arbres de décision pour aboutir à un unique modèle performant. Elles appartiennent donc également aux **approches ensemblistes**. Ces dernières permettent de pallier le sur-apprentissage occasionné par la profondeur des arbres de décision. Les deux approches introduites se distinguent néanmoins par la construction et la mise en relation des modèles en concurrence.

3.4.1.2.1 Forêts d'arbres décisionnels

La méthode de type *Random Forest* évalue en parallèle un nombre fixé J de modèles basés sur des arbres décisionnels, puis détermine un unique modèle en considérant la moyenne

(éventuellement pondérée) des J modèles construits. L'expression finale adopte la forme suivante :

$$\forall t,\ \ \widehat{y_{t+h}} = F_J(\boldsymbol{x_t}) = \sum_{j=1}^{J}(w_j^{RF} f_j(\boldsymbol{x_t})),$$

où f_j est un arbre de décision et w_j^{RF} la pondération associée au modèle j pouvant dépendre de sa performance sur le jeu d'apprentissage.
Afin de garantir la diversité des J modèles obtenus, ces derniers sont, d'une part, évalués sur un sous-ensemble du jeu d'apprentissage composé de B couples Individu $\times$ Label (*bagging* itéré), et, d'autre part, le nombre de variables des individus $\boldsymbol{x}$ est restreint à q. Le jeu de permutations sur les variables occasionné par la restriction à un maximum de q composantes pour l'individu d'entrée joue un rôle de sélection de variables. Les meilleurs modèles sont alors favorisés dans la formulation finale.
Trois hyperparamètres sont identifiés : le nombre J d'arbres, la longueur des sous-ensembles construits par *bagging* B et le nombre maximum de variables permises q. Leur optimisation suit le procédé proposé par *skicit* (Python) et opère sur les deux mois d'apprentissage habituels (septembre et octobre 2013).
L'implémentation Python étant extrêmement permissive, l'algorithme du *Random Forest* a été décliné sous trois approches : une approche équivalente à l'approche **globale** car MIMO ; une approche équivalente à l'approche **locale** telle qu'introduite préalablement (*i.e.* les états de trafic sur une section et son voisinage en entrée pour prévoir celui d'une seule section) ; une approche dite **unique**, où seuls les états de trafic de la section sont considérés.

3.4.1.2.2 *Gradient Tree Boosting*

Contrairement aux forêts d'arbres décisionnels, les modèles simples de prévision ne sont plus évalués en parallèle, mais sont construits itérativement jusqu'à obtention d'un meilleur modèle. La formulation finale du modèle prédictif suit l'expression suivante :

$$\forall t,\ \ \widehat{y_{t+h}} = F_J(\boldsymbol{x_t}) = \sum_{j=1}^{J}(w_j^{GTB} f_j(\boldsymbol{x_t})) + w_0^{GTB},$$

où f_j est un arbre de décision et w_j^{GTB} un paramètre interne à optimiser.
Le concept consiste à compléter itérativement le modèle F_{j-1} par une fonction de prévision simple f_j décrivant la différence entre le label observé et celui prévu par le modèle à l'étape antérieure. L'ajout de la nouvelle fonction permet de minimiser l'erreur de prévision sur le jeu d'apprentissage considéré.
D'un point de vue algorithmique, suite à la phase d'initialisation recherchant la constante w_0^{GTB} minimisant la fonction de perte ($F_0(\boldsymbol{x_t}) = \arg\min_{w^{GTB}} \sum_{k=1}^{T_{hist}} \mathcal{L}(y_k, w^{GTB})$), le processus opère, à chaque nouvelle itération j, en 4 étapes :

1. Évaluation du gradient (dérivées partielles) de la fonction de perte $\mathcal{L}(y_k, F_{j-1}(\boldsymbol{x_k}))$ par rapport aux $F_{j-1}(\boldsymbol{x_k})$ en vue de déterminer la direction descendante pour minimiser la fonction de perte.

2. Construction de la fonction simple f_j à partir d'un arbre de décision proposant une partition de l'espace d'entrée. L'arbre est appris sur un jeu d'apprentissage composé des individus $\boldsymbol{x_k}$ et de la direction descendante évaluée à l'étape 1.

3. Estimation du facteur multiplicateur w_j^{GTB} par minimisation de la perte : $w_j^{GTB} = \arg\min_{w^{GTB}} \sum_{k=1}^{T_{hist}} \mathcal{L}(y_k, F_{j-1}(\boldsymbol{x_k}) + w^{GTB} f_j(\boldsymbol{x_k}))$.

4. Mise à jour du modèle : $F_j(\boldsymbol{x_k}) = F_{j-1}(\boldsymbol{x_k}) + w_j^{GTB} f_j(\boldsymbol{x_k})$

Outre le nombre d'itérations J, de multiples autres hyperparamètres sont souvent proposés pour assurer la meilleure généralisation du modèle final. Délimitation de la quantité de variables d'entrée caractérisant l'individu, délimitation du taux d'apprentissage, pénalisation de la complexité sont autant d'hyperparamètres traduisant des objectifs de généralisation. Six hyperparamètres ont ainsi été pris en considération. Leur optimisation procède directement de la fonction sklearn.ensemble.GradientBoostingRegressor proposée par *Skicit*. Seule l'approche **locale** est proposée pour cet algorithme ensembliste de prévision.

3.4.2 Analyse comparative des performances

Les trois méthodes de prévision et leurs déclinaisons sous différentes approches sont comparées à la méthode à vecteurs supports ayant présenté de très bonnes performances sur ce réseau de Cheviré durant le mois de novembre 2013. Les performances en termes d'erreur de prévision sont présentées en Figure 3.16.

Globalement, quels que soient les horizons de prévision, les performances en termes d'erreur de prévision sont très similaires, aussi bien en termes d'allure que d'ordre de grandeur de l'erreur. Seule RFglobal expérimente des erreurs bien supérieures à ses homologues à court-terme. Une telle distinction est le fruit du passage à l'échelle multi-dimensionnelle pour une méthodologie apprenant la structure de l'historique. Les mêmes durées d'optimisation et d'apprentissage ont été appliquées aux approches RFlocale, RFunique et RFglobale. Or, la fonction objectif à l'échelle du réseau est plus complexe que celle à l'échelle de la section, du fait des variabilités multidimensionnelles. La structure latente associée à l'historique est donc elle aussi plus complexe et difficile à saisir. Un temps d'apprentissage plus long et des arbres plus profonds sont donc requis pour atteindre des performances équivalentes. Ceci illustre parfaitement un défaut régulièrement formulé à l'encontre des méthodes apprenant la structure de l'historique telles que les Réseaux de Neurones Artificiels : la quantité d'informations requises dans l'historique, donc sa longeur et la durée d'apprentissage nécessaire pour atteindre des performances adequates sont souvent extrêmement conséquents.
Cependant, en termes de performance globale, les méthodes GBM et RFlocal sont les plus efficaces à tout horizon de prévision et surpassent MSVR. Néanmoins, ces bonnes performances globales sont à considérer au regard des informations horaires intégrées et donc à nuancer par une analyse plus locale.

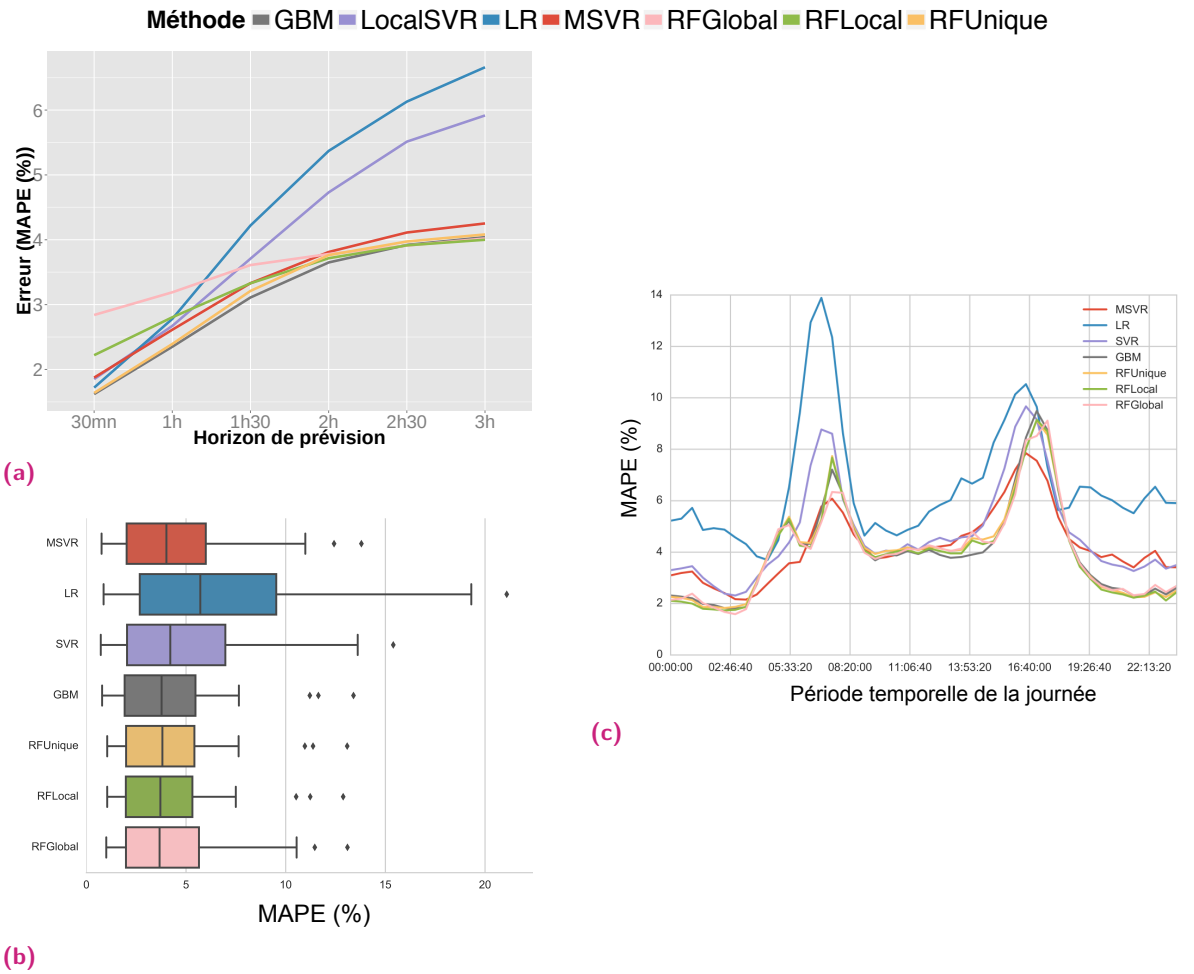


Fig. 3.16.: Performances globales de prévision des méthodes de l'état de l'art testées suivant l'horizon de prévision sur le Pont de Cheviré : a) Performance en termes d'erreur relative globale ($MAPE_{global}$) b) Distribution des erreurs par section ($MAPE_{section}$) à l'horizon 3h c) Distribution des erreurs par pas de temps ($MAPE_{TimeStep}$) à l'horizon 3h.

À très court-terme ($\leq$1h), la méthode paramétrique de régression linéaire (LR) est particulièrement performante à l'échelle globale d'analyse. Ceci montre qu'à très court-terme, les méthodes de prévision les plus simples restent les plus efficaces pour prévoir la plupart des sections et pas de temps d'une journée. Cette observation est à modérer par l'analyse plus locale de l'erreur de cette méthode montrant une forte disparité de l'erreur d'une section à l'autre et une forte dépendance aux périodes d'affluence de la journée.

L'analyse locale de l'erreur par section, proposée en Figure 3.16 b), met en évidence que les méthodes basées sur des arbres décisionnels traitent toutes les sections du réseau sur un même pied d'égalité, quand des disparités plus importantes sont observées sur les autres méthodes. Le fait que GBM comme RF soient concernés par cette observation suggère que l'origine d'un tel comportement réside dans l'usage d'arbres décisionnels.

Enfin, les individus en entrée des méthodes alternatives (GBM, RF, LR) intègrent les facteurs horaires et les typologies de jours, ce qui favorise la prévision des périodes usuelles. Néanmoins, la Figure 3.16 c) met clairement en évidence que la méthode MSVR est plus efficace durant les périodes de fortes activité de la journée. Seules les périodes nocturnes, présentant moins d'intérêt, sont moins bien anticipées par la méthodologie à vecteurs supports. L'introduction des paramètres horaires ne permet finalement que de réduire artificiellement l'erreur de prévision globale présentée en Figure 3.16 a). Les meilleures performances globales des approches à base d'arbres décisionnels sont en partie dues à l'introduction de facteurs externes. De plus, les phénomènes inhabituels étant peu récurrents, de façon générale, l'introduction de ces facteurs s'avère bénéfique, mais permet seulement de réduire l'erreur sur des périodes, dont l'intérêt est limité.

3.5 Conclusions et perspectives

► Synthèse

Au cours de ce chapitre, un large spectre de méthodes de prévision, adaptées aux approches locales comme globales, est comparé sur des réseaux aux dynamiques variées. Les principaux apports se partagent suivant trois points :

- **adaptation de méthodologies classiques de prévision du trafic aux exigences de l'approche globale en s'inspirant de méthodes du domaine informatique, puis première application au trafic.** Les méthodes MKernel et MSVR sont ainsi adaptées, puis testées sur les cas d'étude nantais.
- **confirmation sur de nouvelles méthodologies des performances de l'approche globale au regard de l'approche locale.** Outre un gain de temps non négligeable en termes d'exécution en temps-réel, les erreurs de prévision des approches globales sont au moins équivalentes à leurs homologues locales. Les méthodes MKNN et LocalSVR se distinguent sur le court terme, tandis que MSVR est très efficace dès l'horizon 1h30. De façon surprenante, la disparité spatiale de traitement des sections n'est pas davantage marquée pour les approches globales que locales. Les mêmes zones difficiles de prévision occasionnent les mêmes ordres de grandeur d'erreur.
- **mise en évidence de domaines de prédilection propres à chaque méthode de prévision.** L'analyse spatiale et temporelle des meilleures méthodes révèle que les zones difficiles de prévision sont particulièrement disputées et exacerbent les spécificités de chacun des modèles. Des typologies d'usage des méthodes sont mises en évidence. Ainsi, les méthodes à noyaux expérimentent la même difficulté à prévoir toute section, mais, du même coup, sont performantes sur les zones difficiles.

Finalement, les approches globales évaluées se révèlent particulièrement efficaces au point de suppléer les approches locales sur la grande majorité des sections des réseaux et de rester très performantes sur les zones les plus difficiles. Elles permettent donc, au moins, une restriction

non négligeable du temps d'estimation pour la plupart des sections et peuvent se voir complétées par des approches locales pour traiter des cas spécifiques.

▶ Limites

Les bons résultats de l'approche globale, obtenus depuis les procédés à gestion d'historique, doivent être confirmés sur des méthodologies apprenant la structure de l'historique et des méthodes paramétriques. La principale difficulté réside dans l'adaptation de ces processus aux exigences de l'approche globale.

Les processus de type **Réseaux de Neurones Artificiels** se prêtent facilement à une extension multidimensionnelle des sorties du modèle de prévision, mais, au prix d'une croissance non négligeable du nombre de paramètres internes et donc de la durée d'apprentissage. Ce phénomène est partagé par les procédés à gestion d'historique et illustré en section 3.4.2 par les moins bonnes performances de la version *MIMO* du *Random Forest* face à sa version locale pour des durées d'optimisation égales.

Les **approches paramétriques** existantes et visant à l'unification spatiale du modèle consistent en une superposition de modèles et donc de paramètres internes associés. Ces méthodologies ne peuvent donc être parcimonieuses, à moins d'introduire un ou plusieurs hyperparamètres servant de facteurs multiplicateurs de matrices décrivant les disparités spatiales.

▶ Perspectives

L'extension de l'analyse concurrentielle des méthodes de prévision constitue une perspective de taille. Étant donné que l'approche globale conduit à un gain en temps d'estimation non négligeable et requiert une faible quantité d'hyperparamètres, il est tout à fait envisageable d'évaluer en parallèle différentes méthodologies de prévision par l'approche globale, puis de les activer à tour de rôle sur les sections. L'activation prendrait la forme d'un système probabiliste par votes ou d'un réseau bayésien liant le choix méthodologique aux conditions expérimentées par le réseau, aux risques de rupture (incidents *etc*), à la période de la journée et à la section concernée. Un tel processus se rapprocherait des **méthodes ensemblistes**. L'intérêt réside dans l'opportunité d'améliorer durablement les performances de prévision en tirant profit des avantages de chacune des méthodologies.

4

Sélection de sections critiques sur le réseau

Les mots nous éclairent, ils nous montrent la voie ...

— BlackSad
Tome 5 : Amarillo (Juan Diaz Canales)

Abstract :
This chapter intends to improve the prediction accuracy by understanding and integrating the spatio-temporal interactions between network links within the forecasting framework. A time dependent process of critical links identification is developed from local configurations. It is assumed that the local behaviour of the links with its neighbourhood reflects the instantaneous network dynamic and that some behaviour typologies, e.g. to create or propagate congestion, are matching with criticality of links for the network operations. Then the critical links identification is used as a ranking indicator for variable selection. It leads to a reduction of the time complexity by 3 and a more accurate prediction especially in case of unusual events.

Affiner la prévision et anticiper les propagations d'éléments perturbateurs en saisissant les interactions spatiales et temporelles entre sections, tels sont les objectifs des approches locales intégrant en variables d'entrée les états de trafic au voisinage de la section d'étude. L'ajout des informations des sections voisines est essentiel pour connaître en amont l'existence de phénomènes pouvant affecter la section analysée à l'horizon de prévision considéré. Une meilleure robustesse de la prévision est alors attendue. Néanmoins, pour être efficace, la délimitation du voisinage doit pouvoir évoluer dynamiquement et ne pas s'étendre indéfiniment (voir section 1.3.3). Un compromis entre extension du voisinage et quantité de variables d'entrée doit être trouvé. En effet, la croissance du nombre de variables en entrée conduit à séparer inéluctablement les individus au point d'effacer toute similarité entre eux. Cette notion, introduite par Bellman, porte le nom de fléau de la dimension (*Curse of Dimensionnality*). La communauté du trafic s'est ainsi heurtée à la difficulté de **délimiter le voisinage optimal**, notamment, du fait de son évolution suivant la dynamique du trafic [37]. Cette difficulté résulte de deux enjeux de taille, qui se font face :

- introduire de nouvelles variables d'entrée permettant de **saisir les interactions entre sections en vue d'améliorer la robustesse** de la méthode de prévision ;
- **limiter le fléau de la dimension** et le bruitage de l'information essentielle en restreignant la quantité de variables d'entrée.

Le passage à l'approche globale induit une légère modification de la problématique et des enjeux. La fonction objectif n'est plus d'optimiser les performances d'une section, mais celles du réseau considéré. Il ne s'agit dès lors plus de saisir les interactions entre une section spécifique et son voisinage, mais de **saisir la structure spatio-temporelle régulant les flux de trafic sur le réseau** en **limitant la quantité de variables en entrée**. Si l'approche globale, telle qu'introduite en chapitre 2, renseigne naïvement la structure spatio-temporelle en introduisant les informations relatives à l'ensemble des sections du réseau, les conséquences en termes de complexité temporelle et de fléau de la dimension sont totalement négligées. Une caractérisation plus fine de la structure spatio-temporelle du réseau, conservant la qualité de prévision lors d'événements non récurrents, réduisant le temps de calcul et les variables d'entrée, constitue tout l'enjeu de ce chapitre.

L'hypothèse fondamentale, sur laquelle repose ce chapitre, s'inspire des procédés de reconnaissance de visages pour lesquels seuls certains traits (sections) caractéristiques du visage (réseau) sont supposés permettre de saisir l'essentiel de l'image (états de trafic sur le réseau). Ainsi, il est supposé qu'à un instant t, **seul un nombre restreint de sections suffit pour retrouver l'intégralité des états de trafic sur le réseau**. Ces sections particulières seront appelées **les sections critiques** et diffèrent en fonction de la dynamique du réseau. Cette hypothèse trouve son origine dans l'observation que, sur une période donnée, le comportement des sections vis-à-vis du fonctionnement du réseau varie. À un moment donné, certaines sections ont plus tendance à propager la congestion, à la catalyser ou à en créer. Tout lien du réseau peut être perçu à travers l'évolution de son comportement affectant spatialement et temporellement le fonctionnement du réseau. Les sections, disposant du comportement le plus contraignant à un instant, sont également celles portant le plus d'informations quant à son fonctionnement, donc

les plus critiques.
Finalement, de même que pour les approches locales, la problématique poursuivie pour l'approche globale révèle deux enjeux suivant le point de vue adopté, auxquels deux démarches sont associées :

– Introduire la structure spatio-temporelle du réseau afin de **parvenir à une bonne robustesse durant les phases de congestion récurrentes comme non-récurrentes**. Pour y parvenir, il s'agit de faire ressortir, depuis des motifs latents du comportement global du réseau, le jeu de **sections critiques** et leur évolution en fonction de la dynamique du trafic. La classification des sections suivant un critère de criticité est supposée retranscrire alors l'influence des sections sur le fonctionnement du réseau. L'évaluation à différentes périodes temporelles transpose l'aspect évolutif de la criticité.
– **Éliminer les informations redondantes** ou introduisant un bruit de fond, afin d'aboutir à une reconnaissance robuste de motifs latents. Endiguer le fléau de la dimension requiert alors **une réduction du nombre de variables** en entrée.

La problématique introduite pourrait être généralisée au domaine du traitement de réseaux de grande envergure tels que les réseaux d'énergie, internet *etc*. Ces derniers font également face à des dynamiques variées de débit et de transfert d'information à travers un réseau composé de sections. Leur enjeu réside dans la difficulté à aboutir à de bonnes performances depuis des données spatio-temporelles complexes. La démarche de recherche des sections critiques introduite pourra dès lors être étendue à ces domaines.

Suite à une introduction affinant le contexte, ce chapitre propose un procédé de sélection de variables à partir d'un critère de qualification de la criticité des sections du réseau. L'originalité réside dans l'évolution temporelle des variables sélectionnées en fonction du degré d'appartenance au critère de chaque section. Le processus s'agence suivant deux phases. S'appuyant sur un historique de données, le procédé évalue les dynamiques locales du réseau, puis applique une méthode de classification issue de la catégorisation de corpus de documents pour ordonner les sections et, à terme, affiner l'intégration de la structure spatio-temporelle. Le procédé est évalué, dans diverses versions, sur le Pont de Cheviré et la zone urbaine de Nantes.

4.1 Positionnement des démarches envisagées au sein de la littérature

4.1.1 Traitement de la démarche d'identification des sections critiques dans la littérature

Au sein de la littérature du trafic et, par extension, celle du traitement de grands réseaux, le problème d'identification dynamique des sections critiques tel que dépeint est totalement original, au même titre que l'introduction de l'approche globale.

Néanmoins, la thématique s'en approchant le plus réside dans l'estimation statique de la vulnérabilité d'un réseau (*network disruption*). Ces études de vulnérabilité constituent un sous-domaine en soi de l'analyse des réseaux et s'avèrent nécessaires pour estimer la sensibilité des réseaux en cas de scenarii catastrophiques (ouragans, tremblements de terre, ruptures de barrage *etc*). L'objectif est de parvenir à identifier les arcs les plus sensibles et ceux aptes à évacuer le plus grand nombre de personnes. Un point de vue de pure analyse de graphes est adopté. Les sections sont alors supposées pleinement actives ou totalement inactives. Les études menées par Bell font office de référence dans ce domaine [168], mais ne se prêtent pas à la finesse requise par la dynamique du réseau. Cependant, des études plus récentes [169] s'intéressent à des scenarii n'ayant qu'un impact partiel sur la section concernée et identifient, puis classifient les arcs les plus critiques pour évaluer la robustesse du réseau [170, 171]. L'ordre des sections critiques qualifie le fonctionnement du réseau en saisissant ses composantes ayant la plus forte influence sur des cas testés.
La démarche proposée reprend cette notion de classification des sections en vue de saisir le fonctionnement du réseau, mais vise à y introduire l'aspect évolutif et dynamique de sa structure spatio-temporelle.

4.1.2 Traitement de la démarche de réduction de la dimensionnalité dans la littérature

Au sein de la littérature informatique, le problème de la dimensionnalité implique de construire des individus compacts en réduisant ou modifiant les dimensions d'entrée [172, 173, 174] tout en trouvant un compromis entre pertinence et redondance de l'information [175]. Cet objectif est ainsi partagé par divers domaines d'application (catégorisation de textes [176] ou faciale [177]), traité par différentes méthodes [178, 179] et présente un intérêt grandissant dans le domaine de l'apprentissage automatique dû à la croissance des dimensionalités [175], aux contraintes de budget [180] ou de sécurité [181].

La modification des variables d'entrée (*Feature Extraction*) étant peu envisageable du fait de la multidimensionnalité de la fonction objectif de l'approche globale et de la difficulté d'interprétation, seule la sélection de variables (*Feature Selection*) est considérée.
La sélection de variables consiste à réduire le nombre de dimensionalités en entrée d'algorithme en cherchant un sous-ensemble de variables d'entrée optimisant un critère. Le critère consiste en une mesure de séparabilité ou une mesure de l'erreur de prévision ou de classification. Suivant la façon d'appréhender les variables composant l'individu d'entrée, deux approches sont identifiées :

– **le "classement de Variables"** (*Variable Ranking*), appelé approche univariée dans le domaine. Il consiste à supposer chacune des variables indépendantes les unes des autres et à évaluer l'information utile portée par chacune séparément. En fonction du score de chacune, les variables sont ordonnées et sélectionnées. Le procédé, extrêmement généraliste, a beaucoup été appliqué au cas de classification, mais l'hypothèse d'indépendance

des variables est rarement vérifiée. De plus, il a été montré en 2003 par Guyon *et al.* [173] qu'un jeu de variables, jugées indivuellement inutiles par un procédé de classement, pouvait, en réalité, révéler un potentiel d'explication pertinent [1]. Cette approche univariée occupe une place de plus en plus restreinte dans la littérature.
Elle est particulièrement représentée par **les méthodes dites *Filter***. Elles interviennent comme une phase de prétraitement et mesurent la proximité entre le label prédit et chacune des variables. Les mesures habituelles sont l'information mutuelle, les coefficients de corrélation de Pearson *etc*. Ces méthodes sont rapides d'exécution et indépendantes des critères de prévision, mais seules les relations statistiques sont saisies.

– **la "sélection de sous-ensembles"** (*Subset Selection*) ou approche multivariée. Cette approche propose de tester la pertinence de jeux de variables d'entrée en les évaluant suivant un critère prédéfini. Deux classes d'algorithme sont rencontrées dans la littérature et se distinguent par le procédé d'application :

  – **les méthodes dites *Wrapper***. Elles s'appuient sur un modèle de prévision pour évaluer le score du sous-ensemble de variables testé. Tout nouveau jeu de variables est utilisé pour apprendre un modèle, dont les performances de prévision sont estimées. Le jeu de variables le plus performant est conservé. Les méthodes Wrapper sont spécifiques à un modèle et fournissent donc les meilleures variables propices à l'évaluation de cet unique modèle. Les procédés les plus communs [182] sont :

    – les procédés exhaustifs : le procédé le plus naïf (*All Subset Model - ASM*) consiste à tester l'ensemble des combinaisons possibles de variables d'entrée et d'identifier le sous-ensemble minimisant l'erreur. Cette recherche exhaustive s'avère régulièrement irréalisable puisque le nombre de sous-ensembles à tester croît exponentiellement avec le nombre total de variables.

    – les procédés aléatoires et/ou graduels : le processus consiste à faire évoluer itérativement un jeu de variables de départ pour atteindre le jeu optimal (*greedy hill climbing*) [183, 184]. Les méthodes usuelles sont la "recherche séquentielle" (*Sequential Search*) et la *Forward Selection*, débutant sans aucune variable et ajoutant itérativement une variable si minimisation du résidu observée, et la *Backward Elimination*, procédant, à l'inverse par élimination des variables peu pertinentes. Les versions les plus récentes minimisent des critères alternatifs tels que le critère d'Information de Akaike (AIC) ou le critère d'information bayésien (BIC).

    – les procédés à métaheuristiques : souvent bio-inspirés, les principaux procédés d'optimisation s'y retrouvent : optimisation par essaim particulaire (*Particle Swarm Optimization*) [185], optimisation par colonies de fourmis (ACO) [186] ou encore algorithmes génétiques [187]. Pour le cas des algorithmes génétiques, un chromosome constitue une représentation binaire des variables sélectionnées. Différents jeux de chromosomes sont croisés génération après génération jusqu'à

1. *It has been illustrated "the usefulness of selecting subsets of variables that together have good predictive power, as opposed to ranking variables according to their individual predictive power."*

obtention de la meilleure combinaison. Les autres procédés proposent la même représentation binaire de la combinaison de variables considérées.

– **les méthodes dites *Embedded***. Ces procédés se distinguent des méthodes *Wrapper* par l'intégration du procédé de sélection au sein d'un modèle simple de prévision. Le modèle de prévision simple intègre dans sa fonction objectif le facteur de minimisation du nombre de variables considérées. La plus populaire de ces méthodes [188, 182] est le *LASSO* (*Least Absolute Shrinkage and Selection Operator*), dont la fonction de perte à minimiser est : $\min_W(\sum_{k=1}^{T_{hist}}(y_k - b - \boldsymbol{W}^T\boldsymbol{x_k})^2 + \lambda N_1(\boldsymbol{W})$. Un coût λ pour toute composante non nulle du coefficient directeur est ajouté au terme des moindres carrés. Différentes évolutions sont rencontrées dans la littérature : *Elastic Net* [184], *Graphical LASSO* [189] *etc*. Certains étendent même le LASSO à la prévision multi-tâches [190, 191], mais peu d'adaptations sont constatées pour les cas MIMO.

Les procédés de type *Filter* ne sont pas applicables dans notre cadre d'étude puisque les variables sont considérées spatio-temporellement dépendantes. Les approches classiques de type *Wrapper* et *Embedded* négligent les spécificités propres au trafic lors de la phase de recherche des variables les plus adaptées. Ainsi, bien que ces procédés soient en mesure de minimiser l'erreur de prévision, aucun n'est apte à assurer la robustesse sur le long terme des variables sélectionnées, ni à s'adapter dynamiquement aux contraintes exercées sur le réseau. La démarche adoptée vise à proposer une méthode de sélection de variables répondant à cet objectif de robustesse en s'appuyant sur l'évolution dynamique de la criticité des sections liée au fonctionnement du réseau.

4.2 Identification de sections critiques pour la sélection de variables

Pour répondre à l'évolution dynamique des contraintes exercées sur le réseau, le jeu de variables à conserver doit être adaptable en fonction des périodes temporelles. La mise en œuvre de méthodes usuelles de sélection de variables est donc limitée. Seuls les procédés de sélection de type *Variable Ranking* bénéficient d'une souplesse suffisante en permettant de définir un critère de classement évaluable à tout moment. La démarche adoptée consiste dès lors à trouver une unique définition à la notion de criticité, puis à classer les sections en fonction du degré d'appartenance à cet unique critère durant la période d'évaluation. Les $\hat{n}$ meilleures sections du moment sont utilisées pour construire les variables d'entrée.

Cependant, la notion de criticité est particulièrement ardue à appréhender. Diverses approches sont envisageables. Néanmoins, l'**hypothèse fondamentale** suppose que la criticité des sections à un instant donné provient du comportement local des sections. L'**aspect critique des sections** serait ainsi marqué à l'échelle locale et résulterait du **comportement de chacun des liens avec son voisinage** durant une période donnée. L'influence exercée sur le fonctionnement du réseau

par une section est supposée être localement marquée par les tendances de son comportement. Détériorer les états de trafic ou propager la congestion pourraient être des indicateurs de criticité. Il reste alors à déterminer le type de comportement (critère) correspondant le mieux à l'aspect critique. À cette fin, un ensemble de critères potentiels sont bâtis par encodage en configurations des comportements locaux des sections. Les critères sont ensuite exploités, en parallèle, comme procédé de sélection de variables et mis en concurrence au regard de l'erreur de prévision et du nombre de variables requises. Le critère optimisant l'erreur de prévision à partir de la plus petite quantité de variables d'entrée est celui correspondant à un comportement critique des sections. Suivant les recommandations [2] de Guyon *et al.* [173], la méthode de prévision utilisée pour choisir le critère de criticité se doit d'être simple : MKNN est utilisée. Une fois le critère de criticité fixé, la sélection de variables est opérable depuis toute autre méthode de prévision et évaluable pour toute période temporelle.

Finalement, la démarche de sélection de variables s'appuie sur la recherche des sections critiques du réseau et permet leur identification en 2 étapes :

- **la détermination du critère de criticité** se partageant en 2 phases :
  - **une phase d'encodage dynamique des sections.** Elle consiste en une analyse locale du comportement des sections aboutissant à leur caractérisation pour toute période temporelle par une distribution sur l'ensemble des configurations d'état possibles. L'analyse locale s'appuie sur les corrélations liant une section à ses voisines en aval comme en amont.
  - **une phase de définition du critère de criticité**. Des motifs latents associés à des dynamiques locales spécifiques sont dévoilés grâce à un procédé génératif probabiliste de classification. Ensuite, la prévision est estimée depuis les différents jeux de variables d'entrée tirés du classement propre à chacune des typologies latentes. Les performances sont comparées pour identifier le motif le plus critique.
- **la sélection de variables** de type *Variable Ranking* s'appuyant sur la classification dynamique proposée par le critère de criticité.

Une telle méthodologie liant corrélation spatio-temporelle et modèle génératif probabiliste se rapproche d'un procédé efficace appliqué à l'étiquetage en ligne d'images [192]. De plus, l'introduction de concepts physiques de propagation du flux est déjà rencontrée dans la littérature du trafic : Li *et al.* [193] détectent les incidents à partir de variations de vitesse en amont liées par corrélation aux valeurs expérimentées. Néanmoins, personne n'encode les variations de vitesse afin d'en faire ressortir des motifs latents, ni n'utilise de tels motifs pour caractériser le comportement du réseau. Les contributions de cette étude sont donc de 2 ordres : l'identification des sections critiques caractérisant le mieux la dynamique du réseau et la proposition d'un procédé d'ordonnancement dynamique des sections pour sélection de variables.

2. "*It seems reasonnable to use a wrapper (or embedded method) with a linear predictor as a filter and then train a more complex non-linear predictor on the resulting variables*"

Afin de faciliter la compréhension, la démarche est illustrée sur un cas simple composé de $N_C = 8$ configurations d'état sur la zone du Pont de Cheviré.

4.2.1 Phase d'encodage par décomposition physique du système : les configurations locales

Cette partie propose de caractériser les sections critiques par leur comportement local, en analysant plus spécifiquement leur tendance à propager, catalyser et étendre la congestion. Dans ce but, les configurations locales d'état sont introduites comme un moyen de saisir simultanément les variations d'états de la section l au regard de celles observées chez ses pères et fils. Comme l'illustre la Figure 4.1, les pères (respectivement fils) se composent des sections directement atteignables en aval (respectivement en amont) de la section l. Il est donc supposé que les sections critiques se distinguent par des motifs spécifiques au sein des relations spatio-temporelles les liant à leur voisinage direct et caractérisant l'étendue de leur zone d'influence.

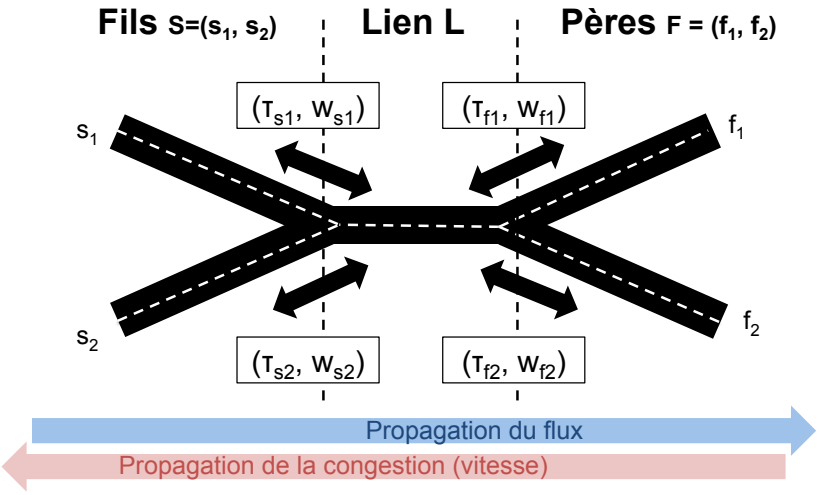


Fig. 4.1.: Représentation du fonctionnement local autour d'une section : identification des pères et fils.

4.2.1.1 Analyse du système local

Le problème est perçu sous l'angle général de la propagation d'un signal à travers un ensemble de liens interconnectés et dépendants. Du fait de la propagation du flux d'information, il est supposé que l'état Z_t^l sur le lien l est influencé par ses pères ($\{f, f \in [1, F]\}$, où F est le nombre de liens directement en aval de l) et possède un impact sur ses S fils ($\{s, s \in [1, S]\}$) avec des décalages temporels τ à estimer. Au-delà des retards temporels mesurés pour chacun des pères et fils, une pondération w^{config} est associée à tout père (respectivement fils) et représente leur contribution à l'information transmise (respectivement redistribuée). La mesure optimale des

paramètres τ et w^{config} propres à chacune des sections du voisinage nécessite la maximisation de la corrélation croisée entre la section l et ses pères ou fils définie par :

$$Z_t^l \star Z_{t-\tau}^f = \int Z_t^l (Z_{t-\tau}^f)^* dt$$

pour tout père f avec un décalage négatif τ et, respectivement, par :

$$Z_t^l \star Z_{t+\tau}^s = \int Z_t^l (Z_{t+\tau}^s)^* dt$$

pour tout fils s avec un décalage temporel positif τ.
La corrélation croisée estime la puissance de la relation statistique liant les états expérimentés par le lien l et ceux associés à un père ou fils avec un retard τ. Les valeurs optimales pour tout père ou fils sont donc obtenues en faisant varier la valeur du retard τ en vue de maximiser la corrélation. Les expressions des paramètres optimaux prennent les formes suivantes :

$$\forall f \in [1, F], \quad \begin{cases} \tau_f & = & \arg\max_\tau (Z_t^l \star Z_{t-\tau}^f) \\ w_f^{config} & = & \dfrac{Z_t^l \star Z_{t-\tau_f}^f}{\sum_{i=1}^{F} Z_t^l \star Z_{t-\tau_i}^i} \end{cases}$$

et

$$\forall s \in [1, S], \quad \begin{cases} \tau_s & = & \arg\max_\tau (Z_t^l \star Z_{t+\tau}^s) \\ w_s^{config} & = & \dfrac{Z_t^l \star Z_{t+\tau_s}^s}{\sum_{i=1}^{S} Z_t^l \star Z_{t+\tau_i}^i} \end{cases}$$

Bien que les contributions et retards temporels puissent être variables lors de transitions d'état congestionné à fluide, ces paramètres sont supposés stationnaires dans cette étude.

4.2.1.2 Encodage des configurations d'état

Résultant de l'analyse par corrélation croisée, la modélisation locale des relations liant une section l et son voisinage immédiat est perçue suivant trois étapes transparaissant en Figure 4.2 :

1. le lien l reçoit, à l'instant t, un signal résultant de l'ensemble des signaux émis par ses pères aux instants antérieurs $t - \tau_f$;
2. le signal reçu par le lien l est légèrement altéré par les contraintes exercées sur la section ;
3. le lien l redistibue un signal à chacun de ses S fils avec un retard τ_s et contribue à leur signal avec un poids w_s^{config}.

Cette modélisation pose les bases de la détermination de configurations locales. L'objectif d'une configuration locale vise à proposer une représentation agrégée de l'évolution du signal expérimenté à l'instant t par la section l en la mettant au regard de celui reçu par ses sections

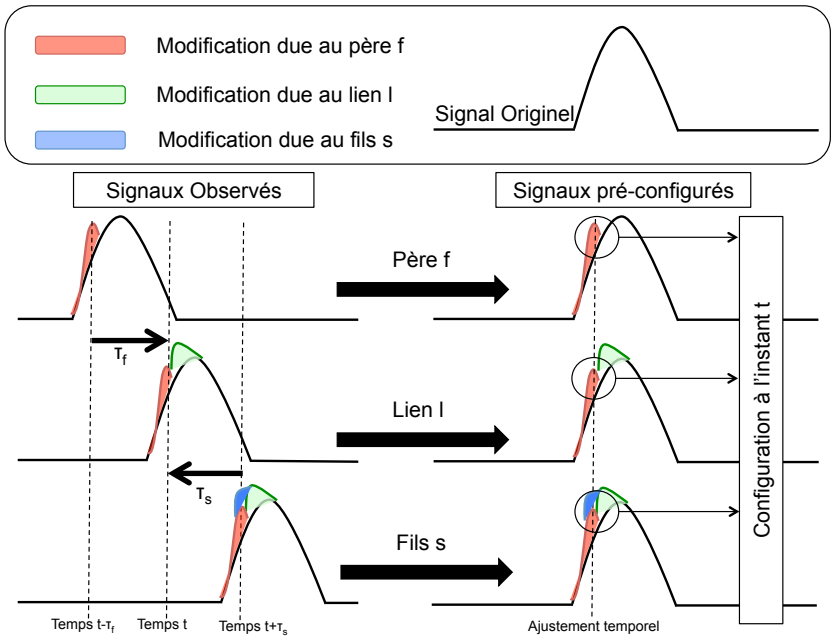


Fig. 4.2.: Illustration de la modélisation locale aboutissant au concept de configuration d'état.

voisines. La mesure simultanée de l'évolution du signal se voit alors caractérisée à tout instant par le triplet $\underline{\boldsymbol{\delta_l}} = (\boldsymbol{\delta_{Fathers}}, \boldsymbol{\delta_{Link}}, \boldsymbol{\delta_{Sons}})$ défini par :

$$\begin{cases} \boldsymbol{\delta_{Fathers}}(t) &= \sum_{f=1}^{F} w_f^{config}(Z^f_{t-\tau_f} - Z^f_{t-\tau_f-1}) \\ \boldsymbol{\delta_{Link}}(t) &= (Z^l_t - Z^l_{t-1}) \\ \boldsymbol{\delta_{Sons}}(t) &= \sum_{s=1}^{S} w_s^{config}(Z^s_{t+\tau_s} - Z^s_{t+\tau_s-1}) \end{cases} \tag{4.1}$$

À tout instant t, il est désormais possible de déterminer si la variation de signal observée sur la section l résulte des pères et si elle transmet ces variations à ses fils. L'encodage des configurations d'état consiste alors en une phase de discrétisation de la variable $\underline{\boldsymbol{\delta_l}}$ en r parties. $N_C = r^3$ configurations potentielles d'état en découlent. Cette discrétisation vise à réduire le triplet à la seule dimension des configurations et à fournir une interprétation manipulable et simplifiée des phénomènes se produisant. Pour des raisons de simplicité, le cas $r = 2$ faisant intervenir une discrétisation binaire est introduit dans un premier temps. La discrétisation de $\underline{\boldsymbol{\delta_l}}$ en un triplet binaire $\underline{\boldsymbol{\Delta_l}}$ identifie les variations négatives du signal marquant une diminution de la variable d'intérêt et donc une détérioration des conditions de propagation :

$$\begin{cases} \boldsymbol{\Delta_{Fathers}}(t) &= 1 \quad si\ \boldsymbol{\delta_{Fathers}}(t) < 0, \quad 0\ sinon \\ \boldsymbol{\Delta_{Link}}(t) &= 1 \quad si\ \boldsymbol{\delta_{Link}}(t) < 0, \quad 0\ sinon \\ \boldsymbol{\Delta_{Sons}}(t) &= 1 \quad si\ \boldsymbol{\delta_{Sons}}(t) < 0, \quad 0\ sinon \end{cases} \tag{4.2}$$

À tout instant t, la configuration de la section l, notée $c_{t,l}$ est totalement définie par le triplet $\underline{\boldsymbol{\Delta_l}}(t) = c$, où $c \in [1, N_C]$ est un indice de configuration. Comme illustré en Figure 4.3, huit

($N_C = 2^3$) configurations potentielles d'état résultent de la discrétisation binaire.
L'ultime étape consiste à regrouper ces configurations expérimentées par la section l par période de temps tp afin de saisir la dynamique l'animant sur cette même période. En effet, la dynamique et l'agencement des configurations expérimentées par la section durant une période est plus à même de caractériser son comportement. L'apparition d'une configuration à un instant donné peut n'être que passagère et requiert la connaissance des états futurs de ses fils. Elle ne peut, à elle seule, caractériser le comportement de la section. En vue d'assurer une bonne caractérisation des motifs, le jeu d'apprentissage est partagé en TP périodes temporelles homogènes. La durée et le mode de sélection des périodes temporelles dépendent des dynamiques du réseau à découvrir et intègrent des connaissances spécifiques au domaine. Le comportement local d'une section l durant la période temporelle tp est alors caractérisé par $\mathbf{\Phi}_{tp,l}$, la distribution de probabilité sur les N_C configurations possibles. Ces distributions sur les configurations d'état alimentent l'algorithme de classification visant à faire ressortir des motifs latents de distribution correspondant à des comportements typiques. Les performances associées à chacun de ces motifs sont alors testées afin de reconnaître la typologie associée aux sections critiques. Outre ce rôle de définition du critère de criticité, ces distributions sur les configurations sont appliquées pour déterminer les sections critiques pour toute nouvelle période de temps considérée.

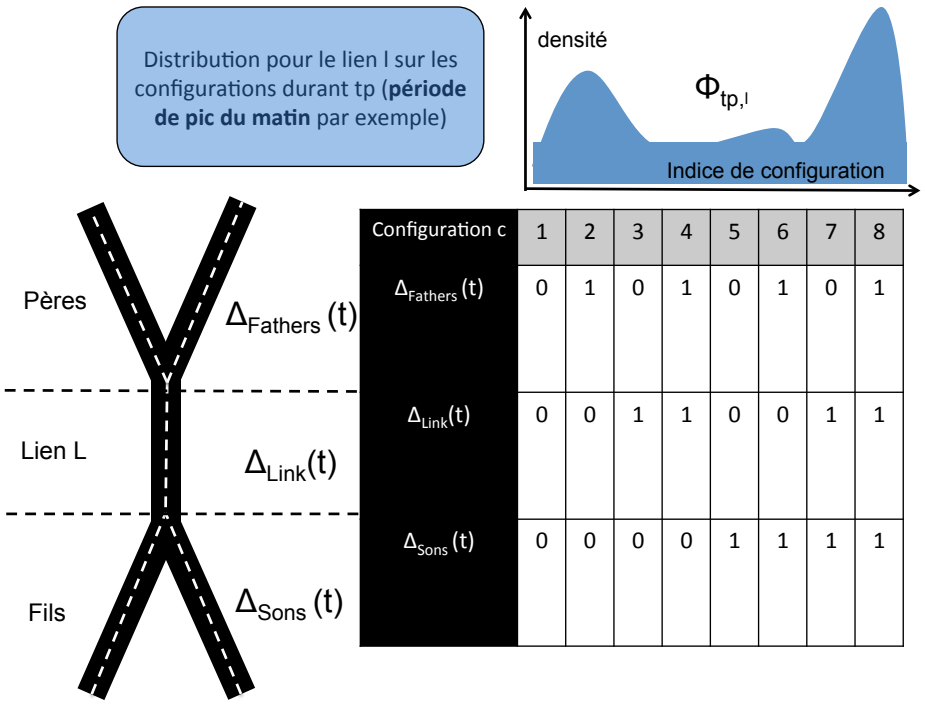


Fig. 4.3.: Illustration du procédé d'encodage et de distribution de configurations.

L'application de la modélisation locale, puis de l'encodage au domaine spécifique du trafic est dépendante de l'indicateur de trafic considéré :

– En tant que signal, le flux de véhicules se propage de l'amont vers l'aval, donc des fils vers les pères.

– Comme un signal se réfléchissant sur un obstacle, la congestion apparaît quand la demande locale est plus élevée que l'offre et se propage vers l'amont : des pères vers les fils. La vitesse, traduisant l'effet de la congestion, se propage elle aussi d'aval en amont.

La caractéristique de trafic connue étant la vitesse moyenne spatiale, le signal considéré se propage donc de l'aval vers l'amont. Les pères expérimentent les variations de signal avant les fils.
En vue d'assurer une bonne définition des distributions considérées et simplifier l'appréhension du concept, les périodes de temps doivent être suffisamment étendues pour assurer la significativité des distributions. Un découpage en $TP = 4$ périodes régulant une journée (nuit, pic du matin, midi et pic du soir) est envisageable. Seules les dynamiques de ces périodes temporelles seraient alors dépeintes.

4.2.2 Phase d'identification et de tri des motifs latents à partir de l'Allocation de Dirichlet Latente (LDA)

L'objectif est de parvenir à faire ressortir différentes typologies latentes parmi les distributions de configurations d'état expérimentées par les sections sur diverses périodes de la journée. Chacune de ces typologies correspond à un comportement spécifique de la section, dont l'un est supposé être associé à celui des sections critiques.
Cette terminologie de recherche de motifs latents n'est pas sans rappeler celle des modèles génératifs probabilistes. Le choix d'appliquer un procédé de classification issu d'un tel domaine a donc été opéré. La méthode de l'Allocation de Dirichlet Latente, usuellement appliquée à la classification de corpus de textes, est envisagée. Bien qu'elles aient été développées originellement dans le cadre de l'analyse de documents, les applications originales du LDA et, plus généralement, de modèles génératifs probabilistes ne sont pas nouvelles dans la littérature du transport. Farrahi et Gatica-Perez [194] ont tenté de mettre en évidence des habitudes de conduite depuis des données de type véhicules traceurs (téléphones mobiles), Huynh *et al.* [195] ont extrait les habitudes quotidiennes de personnes depuis des capteurs embarqués (bracelets) et Niebles *et al.* [196] ont analysé des trajectoires, puis modélisé des zones thématiques à partir de courts extraits vidéo. Le modèle LDA se rencontre, plus spécifiquement, dans le cadre de données massives recueillies par des BSS (*Bicycle Sharing Systems*) [197, 198] en vue d'identifier des comportements spatio-temporels sous-jacents tels que les couples Origine - Destination (O-D) correspondant au pic du matin ou les stations à usage de loisirs. Cette même exploitation du LDA se retrouve dans l'analyse de profils de demande par O-D depuis des données Bluetooth [22].
Suite au succès de ces démarches, le procédé génératif probabiliste LDA est naturellement exploité en vue de faire ressortir les thèmes latents au sein de distributions de configurations. Après une introduction à cette méthodologie de classification non supervisée, l'algorithme d'identification des sections critiques est décrit.

4.2.2.1 Le LDA, un modèle génératif probabiliste

Le modèle LDA, introduit par Blei [199] en 2003, appartient à la famille des modèles génératifs probabilistes permettant de faire ressortir les principales tendances ou thèmes d'un ensemble de ressources ou documents. Les documents généralement étudiés dans la littérature sont des images ou des textes. L'application de tels modèles vise alors à distinguer les différents éléments d'une image (ciel, terre, objet par exemple) ou faire ressortir les sujets traités par un texte.
L'architecture de ces modèles est décrite dans les cours de Bach et Obozinski [200] ou les écrits de Westerveld *et al*. [201], dont sont librement inspirées les informations fournies dans cette section.

Les modèles génératifs probabilistes consistent, pour l'essentiel, à reproduire le procédé aboutissant à un objet complexe (texte / image) à partir d'éléments unitaires (mots / pixels) structurés par des motifs latents, dont l'objectif est de donner du sens à l'ensemble. Ainsi, les objets complexes sont représentés comme des combinaisons d'éléments unitaires, dont les structures et agencements sont modélisés par des typologies simplifiées. Ces motifs latents sont appris, par calibrage du modèle, en tentant de reconstruire des objets complexes connus. Indirectement, la reconstruction des objets complexes depuis les thèmes latents permet une caractérisation des motifs latents par les éléments unitaires et une classification des objets complexes étudiés.

Le procédé des modèles génératifs probabilistes se partage en 4 étapes :

1. **Détermination d'un modèle probabiliste liant la structure complexe aux éléments unitaires** : identification des composants unitaires, des variables latentes et de l'objet complexe, puis mise en équation probabiliste des relations les animant afin de reconstruire l'objet complexe. Des paramètres régulant les lois de comportement de chacune des variables sont introduits. Il en résulte un modèle probabiliste pouvant décrire tout objet complexe.
2. **Optimisation de l'adéquation entre l'objet complexe et le modèle** pour un nombre de thèmes fixé : les paramètres du modèle sont définis en maximisant la probabilité de reproduire les objets complexes à classer.
3. **Délimitation du degré de précision des variables latentes** : évaluation de l'entropie du modèle pour identification du nombre optimum de thèmes latents.
4. **Exploitation du modèle** pour génération de nouveaux objets complexes ou classification non supervisée des éléments unitaires de l'objet complexe.

Bien qu'ils adoptent un même processus, les modèles probabilistes génératifs se distinguent par le nombre de niveaux au sein de leur structure. Les modèles les plus communs sont à deux niveaux : les éléments unitaires et la structure complexe. Le LDA est un modèle à trois niveaux faisant appel à des objets complexes intermédiaires. Afin de faciliter la compréhension du fonctionnement du modèle LDA, les approches à deux niveaux sont introduites puis généralisées et étendues pour décrire les modèles à trois niveaux.

4.2.2.1.1 Illustration sur un modèle génératif à deux niveaux

Les modèles génératifs à deux niveaux présentent l'avantage de modéliser directement les liens unissant l'objet complexe à ses composants unitaires, sans nécessiter l'intégration d'étapes intermédiaires complexes.
L'objectif ultime étant d'introduire le modèle LDA de classification de corpus de textes, le fonctionnement du modèle à deux niveaux est illustré depuis un cas de reconnaissance d'un document texte par un modèle génératif de langage. Il s'agit de faire ressortir les principaux sujets traités par le document. L'hypothèse formulée est que l'ensemble des mots employés pour construire le texte appartiennent à divers champs lexicaux latents à dévoiler. Les champs lexicaux latents se définissent alors depuis des probabilités d'apparition des mots appartenant à un vocabulaire donné.
Les différentes variables structurant la modélisation de l'objet complexe sont identifiées en Table 4.1. Le modèle créé vise à reproduire le document texte. Il se fonde alors sur les hypothèses suivantes :

- Le nombre K_{lda} de thèmes latents ou champs lexicaux caractérisant la structure du modèle est supposé connu, mais est à optimiser par minimisation de la perplexité.
- Tout document est considéré comme un "sac" d'éléments unitaires. L'ordre des mots est supposé sans importance. Toute permutation d'éléments unitaires composant le document le laisse inchangé.

Éléments théoriques	**Notation**	**Interprétation**
Élément unitaire	m	Mot du vocabulaire
Ensemble d'éléments unitaires	Voc $\forall m, \exists i \in [1, N_{Voc}], m \equiv Voc^i$ $\forall j \in [1, N_{Voc}], Voc^{i,j} = \begin{cases} 1 & si \quad j = i \\ 0 & si \quad j \neq i \end{cases}$	Vocabulaire (N_{Voc} mots)
Variable latente	$\boldsymbol{Q} = \{q_1, ..., q_{K_{lda}}\}$	Champ lexical
Objet complexe	$\boldsymbol{D} = \{m_1, ..., m_{N_d}\}$ où $\forall i \in [1, N_d], m_i \in Voc$	Document texte faisant apparaître N_d mots possiblement itérés

Tab. 4.1.: Identification des variables mises en jeu pour la construction d'un objet complexe de type texte.

Étape 1 : modélisation des relations entre les variables

Le but est de construire un modèle apte à reproduire la structure, qui lie le système complexe $\boldsymbol{D}$ à ses éléments unitaires m en s'appuyant sur une variable latente $\boldsymbol{Q}$ fixant les principales tendances de la structure. Les thèmes latents servant d'intermédiaires, seules les relations entre la variable latente $\boldsymbol{Q}$ et les documents $\boldsymbol{D}$, puis $\boldsymbol{Q}$ et les éléments unitaires m sont à modéliser.
Le procédé de génération du document $\boldsymbol{D} = \{m_1, ..., m_{N_d}\}$ officie en deux phases :

1. **Définition des K_{lda} thèmes latents** : cette étape vise à donner un sens aux thèmes issus de la variable latente $\boldsymbol{Q}$. Chaque thème représentant un champ lexical, une distribution sur l'ensemble des mots du vocabulaire Voc leur est associée : $\underline{\boldsymbol{\Lambda}}$. Cette

variable latente est mathématiquement définie par : $\underline{\mathbf{\Lambda}} = (\Lambda_j^{(k)})_{k\in[1,K_{lda}],j\in[1,N_{Voc}]} = (p(m_j|q_k))_{k\in[1,K_{lda}],j\in[1,N_{Voc}]}$.
De plus, il est supposé que les distributions sur le vocabulaire Voc associées à tout thème latent sont générées depuis une loi de Dirichlet, soit : $\forall k \in [1, K_{lda}],\ \Lambda^{(k)} \sim \mathcal{D}(\boldsymbol{\beta}^{\boldsymbol{lda}})$, où $\boldsymbol{\beta}^{\boldsymbol{lda}}$ est un vecteur de paramètres de longueur N_{Voc}. La forme de la loi de Dirichlet est dictée par le paramètre $\boldsymbol{\beta}^{\boldsymbol{lda}}$ et régule les distributions sur les éléments unitaires envisageables, donc l'ensemble des thèmes possibles. Les mots ayant la plus forte probabilité d'apparaître dans un thème donné appartiennent au même champ lexical.

2. **Tissage des relations liant les thèmes latents aux variables observables** :

   a) **Association du document $\boldsymbol{D}$ à sa proportion de thèmes latents** : le texte $\boldsymbol{D}$ est perçu comme un mélange de champs lexicaux. Il est nécessaire de caractériser une telle combinaison ou distribution sur les thèmes latents. À cette fin, un paramètre latent, nommé $\boldsymbol{\theta}$ est introduit et caractérise la proportion de chacun des motifs latents utilisés pour bâtir $\boldsymbol{D}$. Cette variable possède une composante par motif, soit K_{lda} composantes, et chacune d'elle est définie par : $\forall k \in [1, K_{lda}],\ \ \theta_k = p(q_k|\boldsymbol{\theta})$.

   b) **Tirage des N_d éléments unitaires constituant le document** : la contribution de chacun des thèmes au texte D ayant été fixée par θ, la probabilité d'apparition d'un mot au sein du document dépend encore de sa contribution ($\boldsymbol{\Lambda}^{(k)}$) à chacun des champs lexicaux q_k. Pour tout élément unitaire $m_j, j \in [1, N_d]$ du document D, deux étapes de sélection sont requises :

      i. **Choix d'un motif** q en supposant que ce choix suit une **loi multinomiale** $q \sim \mathcal{M}(1, \boldsymbol{\theta})$. La sélection du thème est donc guidée par la proportion $\boldsymbol{\theta}$ des thèmes au sein du document généré.

      ii. **Choix d'un élément unitaire** m depuis la distribution sur le vocabulaire Voc associée à q_k : $\Lambda^{(k)}$. La sélection est supoposée suivre une **loi multinomiale** : $m \sim \mathcal{M}(1, \Lambda^{(k)})$.

Comme il en est d'usage dans le domaine de la catégorisation de textes, une représentation en assiettes de ce modèle à deux niveaux est proposée en Figure 4.4.
Le processus met en évidence **deux paramètres** régulant le modèle : $\boldsymbol{\beta}^{\boldsymbol{lda}}$ pour la forme des thèmes potentiels et $\boldsymbol{\theta}$ pour la proportion d'apparition du thème au sein du document. Un calibrage est donc requis afin de reproduire au mieux le document désiré.

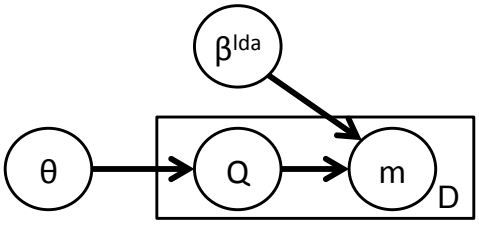


Fig. 4.4.: Représentation d'un modèle génératif probabiliste à deux niveaux.

Étape 2 : détermination des paramètres du modèle

L'objectif du modèle est de reproduire D avec le plus grand degré de fiabilité. La problématique se traduit par la maximisation de la probabilité de reproduire ce jeu d'éléments unitaires sachant que l'on dispose de la répartition sur les thèmes θ et de thèmes tirés d'une distribution de Dirichlet de paramètre $\boldsymbol{\beta^{lda}}$. Il s'agit donc de maximiser $p(D|\boldsymbol{\theta}, \boldsymbol{\beta^{lda}})$ aussi appelée vraisemblance du modèle. Mathématiquement, ce problème d'optimisation se traduit par :

$$(\boldsymbol{\theta_{ML}}, \boldsymbol{\beta^{lda}_{ML}}) = \arg\max_{\boldsymbol{\theta}, \boldsymbol{\beta^{lda}}} \sum_{m \in D} \log(p(m|\boldsymbol{\theta}, \boldsymbol{\beta^{lda}})) \tag{4.3}$$

L'optimisation de cette équation se révèle être difficile à résoudre analytiquement. Sa résolution approchée par des algorithmes récursifs (*Expectation-Maximisation* (EM) [202, 203], échantillonnage de Gibbs) est fréquente dans la littérature.

Étape 3 : optimisation du nombre de thèmes latents

Cette étape consiste à déterminer le nombre K_{lda} optimum de motifs nécessaires pour modéliser avec un degré de précision suffisant la structure de l'objet complexe (document texte). Cette capacité est évaluée par des techniques variées, dont la plupart se rapportent de près ou de loin à la mesure de l'entropie du système (la perplexité par exemple). Divers scénarii variant par le nombre de motifs considérés sont comparés. La valeur K_{lda} associée au scénario minimisant la mesure se rapportant à l'entropie est alors considéré pour la classification ou génération de nouveaux documents.

Étape 4 : classification des éléments unitaires

Une des grandes forces de ces modèles génératifs probabilistes, outre le fait qu'ils permettent de générer de nouveaux documents suivant un modèle prédéfini, est qu'ils permettent de classer à la fois les élément unitaires et le document ou sac d'éléments unitaires. Une fois le nombre de motifs latents optimal déterminé et le modèle calibré, la représentation du document texte D par le modèle est optimale. Deux sorties indirectes du modèle sont alors exploitables : $\underline{\boldsymbol{\Lambda}}$ et $\boldsymbol{\theta}$. La variable $\underline{\boldsymbol{\Lambda}}$ aboutit à la classification des éléments unitaires par thèmes en s'appuyant sur la probabilité de tirer chacun des mots en fonction du thème latent. $\boldsymbol{\theta}$ qualifie la répartition des sujets au sein du document texte, mais, plus précisément, exprime la probabilité de tirer le motif $k, k \in [1, K_{lda}]$ sachant qu'on dispose du document D. L'extension naturelle consisterait à classer un ensemble de documents en les regroupant en fonction des thèmes latents partagés. Ceci permet d'introduire la notion de modèles à trois niveaux, puisque la structure complexe serait le corpus de documents, ce qui conduit au modèle de l'Allocation de Dirichlet Latente (LDA).

4.2.2.1.2 L'Allocation de Dirichlet Latente, un procédé à trois niveaux

L'Allocation de Dirichlet Latente (LDA) introduit un nouveau niveau d'objet complexe, à savoir le corpus de textes. L'objectif n'est plus de simplement faire ressortir les principaux sujets d'un texte, mais d'identifier les thèmes ressortant d'un corpus de textes et de regrouper les textes suivant leurs thèmes majeurs. L'élément complexe, dont il faut comprendre la structure,

est le corpus de texte, tandis que tout document texte du corpus en est un élément unitaire. Cependant, les textes restent des éléments complexes générés par des objets unitaires : les mots. Ceci introduit une mise en abîme avec enchassement de deux modèles à deux niveaux. Le processus de modélisation reste globalement inchangé, mais subit des modifications en vue de s'adapter aux nouvelles contraintes. La principale modification repose sur la nécessité de modéliser les liens tissés entre le corpus de textes et chacun des textes, ce qui induit l'introduction de nouveaux paramètres à calibrer.
Les notations introduites pour le cas à deux niveaux restent inchangées (voir Table 4.1). Seul le corpus ou ensemble de documents est à définir. Il est noté E et caractérisé par N_E documents : $E = \{D_1, ..., D_{N_E}\}$.

Étape 1 : modélisation des relations entre les variables

La modélisation de chacun des documents texte reste inchangée. Seul le tissage des liens entre le corpus et chacun des documents texte est nécessaire. La construction d'un corpus de textes est perçue comme l'attribution, à chacun des documents le composant, d'une proportion de thèmes latents. Cette proportion est exprimée par $\boldsymbol{\theta_{D_d}}$ associée à un document $D_d, d \in [1, N_E]$ du corpus de textes. Il est supposé que chacune des variables $\boldsymbol{\theta_{D_d}}$ suit une loi de Dirichlet de paramètre $\boldsymbol{\alpha^{lda}}$: $\boldsymbol{\theta_{D_d}} \sim \mathcal{D}(\boldsymbol{\alpha^{lda}})$, où $\boldsymbol{\alpha^{lda}}$ est un vecteur de paramètres de longueur K_{lda}.
Le procédé de modélisation adopte alors la forme suivante :

1. **Choix d'une distribution de thèmes associée au corpus E considéré**. Il s'agit de fixer les paramètres du modèle, à savoir le paramètre de la loi régulant les proportions de thèmes au sein des documents $\boldsymbol{\alpha^{lda}}$ et celui de la loi régulant la forme des thèmes $\boldsymbol{\beta^{lda}}$. Il en découle les K_{lda} distributions sur les mots du vocabulaire Voc générées depuis la distribution de Dirichlet de paramètre $\boldsymbol{\beta^{lda}}$: $\forall k \in [1, K_{lda}],\ \boldsymbol{\Lambda^{(k)}} \sim \mathcal{D}(\boldsymbol{\beta^{lda}})$.
2. **Génération d'un document** $\forall d \in [1, N_E],\ D_d = \{m_1, ..., m_{N_d}\}$ du corpus E :
   a) Choix d'une proportion de thèmes pour le document D_d : $\boldsymbol{\theta^d} \sim \boldsymbol{\alpha^{lda}}$.
   b) Sélection des N_d mots m en fonction d'un des thèmes du document D_d :
      i. Choix d'un thème $q_k, k \in [1, K_{lda}]$ depuis la distribution de thèmes latents associée au document : $\boldsymbol{\theta^d}$ en suivant une distribution multinomiale : $q \sim \mathcal{M}(1, \boldsymbol{\theta^d})$.
      ii. Choix d'un mot m depuis la distribution sur le vocabulaire Voc associée au thème q_k suivant une distribution multinomiale : $m \sim \mathcal{M}(1, \Lambda^{(k)})$.

La représentation en assiette du modèle évolue donc légèrement avec l'intégration du paramètre $\boldsymbol{\alpha^{lda}}$ permettant de modéliser les distributions des thèmes sur l'ensemble des textes du corpus. La Figure 4.5 présente en rouge les évolutions dues au passage à une modélisation à trois niveaux.

Étape 2 : détermination des paramètres du modèle

L'objectif consiste désormais à maximiser la vraisemblance du corpus de textes considéré au regard du modèle construit. Il s'agit donc de calibrer convenablement les paramètres $\boldsymbol{\alpha^{lda}}$ et $\boldsymbol{\beta^{lda}}$ pour reconstruire au mieux l'intégralité du corpus de textes E, soit maximiser la

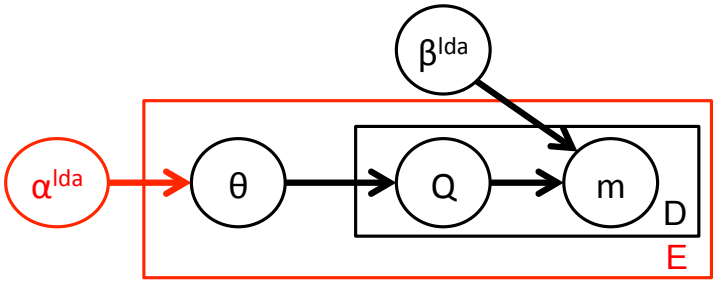


Fig. 4.5.: Représentation d'un modèle génératif probabiliste à trois niveaux : le LDA.

probabilité : $p(E|\boldsymbol{\alpha}^{\boldsymbol{lda}}, \boldsymbol{\beta}^{\boldsymbol{lda}})$. L'expression de la vraisemblance logarithmique étant préférée, l'équation prend la forme suivante :

$$(\boldsymbol{\alpha}^{\boldsymbol{lda}}_{\boldsymbol{ML}}, \boldsymbol{\beta}^{\boldsymbol{lda}}_{\boldsymbol{ML}}) = \arg \max_{\boldsymbol{\alpha}^{\boldsymbol{lda}}, \boldsymbol{\beta}^{\boldsymbol{lda}}} \sum_{d=1}^{N_E} \log(p(D_d|\boldsymbol{\alpha}^{\boldsymbol{lda}}, \boldsymbol{\beta}^{\boldsymbol{lda}})) \tag{4.4}$$

De même que pour les modèles à deux niveaux, la résolution analytique du problème n'est pas envisageable. L'algorithme d'approximation par Espérance - Maximisation (E-M) est donc mis en œuvre, mais nécessite quelques adaptations. Le grand nombre de paramètres latents dus à l'enchâssement de deux modèles à deux niveaux induit des difficultés liées aux couplages entre les paramètres $\boldsymbol{\theta}$ et $\boldsymbol{Q}$ tous eux liés à $\boldsymbol{\beta}^{\boldsymbol{lda}}$. Ce couplage est brisé par une méthode d'inférence variationnelle consistant à substituer des paramètres latents indépendants aux paramètres problématiques. L'optimisation des paramètres du modèle sur le corpus E est donc obtenue depuis un algorithme Variationnel d'Espérance - Maximisation (VEM).

Étape 3 : optimisation du nombre de thèmes latents

Le nombre optimal de thèmes latents est évalué grâce à un indicateur propre à la classification de textes, à savoir la perplexité. Elle est algébriquement équivalente à l'inverse de la moyenne géométrique de la probabilité des mots m. Une perplexité plus faible indique donc une meilleure capacité du modèle à saisir l'ensemble des thèmes généraux sous-jacents, promettant une meilleure aptitude à reproduire la réalité et donc pertinence des thèmes latents. De par son expression générale [199], le perplexité est proche de l'entropie du système :

$$perplexité(E) = \exp(-\frac{\sum_{d=1}^{N_E} log(p(D_d))}{\sum_{d=1}^{N_E} N_d})$$

On détermine le nombre optimum K_{lda} de sujets sous-jacents au sein du corpus de documents par l'étude de la courbe représentant la perplexité en fonction du nombre de thèmes latents considérés.

Étape 4 : Classification

Depuis le modèle constuit et optimisé pour reproduire le corpus de textes E, les principales entités intéressantes en termes de classification de documents sont :

– la variable latente $\underline{\mathbf{\Lambda}}$, dans la mesure où elle représente la façon dont sont construits les thèmes. Ainsi, pour tout thème k ($k \in [1, K_{lda}]$), $\mathbf{\Lambda^{(k)}} = (p(m_i|q_k))_{i\in[1,v]}$ caractérise la distribution du thème k sur les mots du vocabulaire.

– la variable latente θ^d associée au document D_d et caractérisant sa contribution à chacun des K_{lda} thèmes : $\theta^d = p(q_k|D_d)$.

Ces deux variables permettent de définir les thèmes, puis de classer les documents en fonction de leur contribution à chacun des thèmes.

Le procédé LDA est particulièrement efficace pour faire ressortir les principaux sujets abordés au sein d'un corpus de textes. La section suivante montre comment ce procédé de classification non supervisé peut être adapté au trafic pour faire apparaître les principaux comportements des sections du réseau par période de temps.

4.2.2.2 Sélection des sections critiques à partir du LDA

La phase d'analyse à l'échelle locale aboutit à un encodage sous forme de configurations d'état des variations du signal de trafic se propageant sur une section et son voisinage. L'analyse par pas de temps fournit, pour chaque période temporelle et toute section du réseau, une distribution sur les configurations potentielles. L'objectif consiste désormais à faire ressortir de ces comportements des informations quant à un objet complexe : la structure spatio-temporelle du réseau. Il est supposé que l'une de ces informations permet l'identification des sections critiques du réseau et peut être exploitée pour création d'un critère de sélection de variables.

Le procédé de caractérisation du critère de criticité, décrit dans l'algorithme 2, prend en entrée les distributions sur les configurations $\mathbf{\Phi_{tp,l}}$ **(phase d'initialisation)** associées à toute section l durant le moment tp et tirées du jeu d'apprentissage. **La première étape** du procédé consiste alors à déterminer les K_{lda} thèmes ou motifs latents émergeant de la classification non supervisée de l'Allocation de Dirichlet Latente (LDA). Pour chaque période temporelle tp, les variables les plus représentatives associées à chacun des K_{lda} thèmes sont testées en entrée de l'algorithme de prévision **(étape 2)**. Le motif latent recueillant la plus faible erreur avec un minimum de variables est identifié au thème des sections critiques **(étape 3)**.

Algorithme 2 Un procédé d'identification des sections critiques basé sur le LDA

▷ ENTRÉES :

X (taille $T_{hist} \times N$) matrice de séries temporelles, TemporalPeriod (taille $T_{hist} \times TP$) matrice booléenne indicatrice de l'appartenance des pas de temps aux TP périodes de temps, r nombre de catégories discrètes, K_{lda} nombre de thèmes latents , $\underline{\boldsymbol{\delta_l}} = (\boldsymbol{\delta_{Fathers}}, \boldsymbol{\delta_{Link}}, \boldsymbol{\delta_{Sons}})$ $\forall l \in [1, N]$.

▷ FONCTIONS PRÉDÉFINIES :

Discrete (eq 4.2), IdxConfig (Fig. 4.3), Freq, LDA, SortRows, AssessPerf, ElbowMethod

▷ INITIALISATION : ***encodage en configurations*** *voir section 4.2.1*

for $l = 1$ **to** N **do**

$\underline{\boldsymbol{\Delta_l}}$ = [Discrete($\boldsymbol{\delta_{Fathers}}$, r), Discrete($\boldsymbol{\delta_{Link}}$, r), Discrete($\boldsymbol{\delta_{Sons}}$, r)] Discrétisation

$\underline{\boldsymbol{C}}(., l)$ = IdxConfig($\underline{\boldsymbol{\Delta_l}}$) C (matrice $T \times N$) recueille toute configuration à tout moment

$N_c = r^3$

for $tp = 1$ **to** TP **do**

$\underline{\boldsymbol{Config}}_{tp,l}$ = $\underline{\boldsymbol{C}}$(TemporalPeriod(., tp), l)^{T} Recueil de configurations observées durant tp sur l

$\underline{\boldsymbol{\Phi}}((tp-1) \times N + l, .))$ = Freq($\underline{\boldsymbol{Config}}_{tp,l}$, 1 :$N_c$) Φ, matrice $(N * TP) \times N_c$ de distributions

end for

end for

▷ ÉTAPE 1 : ***détermination des motifs latents*** *voir section 4.2.2.2.1*

[$\underline{\boldsymbol{\Lambda}}$, $\underline{\boldsymbol{\theta}}$] = LDA(K_{lda}, $\underline{\boldsymbol{\Phi}}$) Λ est une matrice de taille $K_{lda} \times N_C$ & θ une matrice de taille $(N * TP) \times K_{lda}$

▷ ÉTAPE 2 : ***sélection du motif critique*** *voir section 4.2.2.2.2*

for $k = 1$ **to** K_{lda} **do**

- **Tri des variables suivant leur contribution à tout motif** k

for $tp = 1$ **to** TP **do**

$\boldsymbol{IdxL}_{tp} = (tp-1) \times N + [1 : N]^{\mathrm{T}}$ Indice des N sections du réseau pour la période tp

$\underline{\boldsymbol{RankVariable}}_k$ = [$\underline{\boldsymbol{\theta}}(IdxL_{tp}$, k), $[1 : N]^{\mathrm{T}}$] Association des $p(k|tp, l)$ aux N sections

$\underline{\boldsymbol{RankVariable}}_k$ = SortRows ($\underline{\boldsymbol{RankVariable}}_k$, 1) Tri par probabilité décroissante

$\underline{\boldsymbol{Indicator}}_k([1 : N]^{\mathrm{T}}$, tp) = $\underline{\boldsymbol{RankVariable}}_k([1 : N]^{\mathrm{T}}$, 2) Index ordonné des sections

end for

- **Évaluation des performances de prévision associées au motifs latents**

for $n = 1$ **to** N **do**

Application d'une stratégie 'Best First'

for t = 1 **to** T_{hist} **do**

$\underline{\boldsymbol{input}} = \underline{\boldsymbol{X}}(t, Indicator_k($ $[1 : n]^{\mathrm{T}}$, find(TemporalPeriod(t,.)))) Restriction des entrées

end for

$\underline{\boldsymbol{Error}}(k, n)$ = AssessPerf(input, ...) Évaluation des performances de prévision par MKNN

end for

end for

- **Identification du motif associé aux sections critiques**

$[\hat{k}, \dot{n}]$ = find(min(Error))

▷ ÉTAPE 3 : ***sélection du nombre optimal de variables*** *voir section 4.2.2.2.3*

$[\hat{n}]$ = ElbowMethod(Error($\hat{k}$, .))

À chaque période tp, les variables les plus représentatives de ce thème sont, à ce moment, sélectionnées pour prévoir les états futurs. L'évaluation de la prévision porte, spécifiquement, sur les états de trafic rencontrés par de telles sections.

4.2.2.2.1 Etape 1 : détermination des motifs latents

Cette première étape applique le procédé de l'Allocation de Dirichlet Latente (LDA) à chacune des distributions $\mathbf{\Phi_{tp,l}}$ tirées du jeu d'apprentissage. En tant que procédé génératif probabiliste, le LDA fournit deux sorties permettant, d'une part, de donner du sens aux thèmes identifiés, et, d'autre part, de qualifier et classer les distributions en fonction de leur contribution au thème. Ces motifs se révèlent être significatifs quant à la contribution locale des sections au comportement global du réseau.

Contrairement aux précédentes applications dans le transport, le procédé appliqué propose de revenir à un usage plus classique de la classification par LDA. Seules de légères adaptations sont requises : les configurations d'état se substituent aux mots et un document est perçu comme un sac de configurations d'état. Le LDA vise alors à faire ressortir les thèmes latents au sein d'un corpus de documents (sac de configurations) conduisant à une meilleure compréhension des relations entre le comportement du lien et la dynamique du réseau. L'analogie entre le LDA original et son application dans le contexte actuel se poursuit par :

- **Corpus de documents** : structure spatio-temporelle du réseau, traduite par l'ensemble des distributions sur les configurations expérimentées par les sections sur diverses périodes temporelles : $\mathbf{\Phi_{tp,l}},\ \forall tp \in [1, TP],\ \forall l \in [1, N]$;
- **Document** ou sac de mots : comportement local, traduit par $\mathbf{\Phi_{tp,l}}$, distribution sur l'ensemble des configurations (mots) possibles pour le lien l durant la période tp ;
- **Mot** : configuration locale $c \in [1, N_C]$. Les N_C configurations représentent l'ensemble des triplets distincts obtenus par $\underline{\mathbf{\Delta}_l}(t) = (\mathbf{\Delta_{Fathers}}(t), \mathbf{\Delta_{Link}}(t), \mathbf{\Delta_{Sons}}(t))$.

Suivant cette analogie, le procédé génératif se réécrit de la façon suivante :

1. Détermination des K_{lda} thèmes ou distributions nominales depuis une distribution de Dirichlet sur l'ensemble des configurations :

$$\mathbf{\Lambda^{(k)}} \sim \mathcal{D}(\boldsymbol{\beta}^{\mathbf{lda}}),\ \ \forall k \in [1, K_{lda}] \tag{4.5}$$

2. Tout sac de configurations locales $\mathbf{\Phi_{tp,l}}$ issu du corpus de documents ($[1, N \times TP]$) est supposé être généré par le processus à 2 étapes suivant :
   a) Choix d'une proportion de motifs pour construire la distribution sur les configurations locales $\mathbf{\Phi_{tp,l}}$: $\boldsymbol{\theta}^{\mathbf{tp,l}} \sim \mathcal{D}(\boldsymbol{\alpha}^{\mathbf{lda}})$.
   b) Sélection d'une configuration c contribuant à la distribution $\mathbf{\Phi_{tp,l}}$:
      i. Choix d'un motif $q_k, k \in [1, K_{lda}]$ depuis la proportion $\boldsymbol{\theta}^{\mathbf{tp,l}}$: $q_k \sim \mathcal{M}(1, \boldsymbol{\theta}^{\mathbf{tp,l}})$
      ii. Choix d'une configuration c au sein du $k^{ième}$ motif : $c \sim \mathcal{M}(1, \mathbf{\Lambda^{(k)}})$.

La modélisation de la structure spatio-temporelle introduite reprend strictement le procédé classique du LDA. Les étapes de calibrage des paramètres $\boldsymbol{\alpha}^{\boldsymbol{lda}}$ et $\boldsymbol{\beta}^{\boldsymbol{lda}}$, puis l'optimisation du nombre K_{lda} de motifs latents restent inchangées. Deux sorties du procédé sont particulièrement intéressantes à interpréter dans notre contexte de reconstruction de la structure spatio-temporelle du réseau :

- $\boldsymbol{\Lambda}_c^{(\boldsymbol{k})} \equiv p(c|k)$, la probabilité de tirer la configuration c sachant que l'on considère le $k^{ième}$ motif. Cette distribution discrète sur les configurations permet l'interprétation des thèmes latents. Plus une configuration a de fortes chances d'être tirée au sein d'un thème latent, plus elle contribue à définir le motif.
- $\boldsymbol{\theta}^{\boldsymbol{tp,l}}(\boldsymbol{k}) \equiv p(k|tp;l)$, la probabilité de tirer le $k^{ième}$ motif sachant que l'on considère la distribution $\boldsymbol{\Phi}_{\boldsymbol{tp,l}}$. Cette probabilité est interprétée comme la part de contribution du $k^{ième}$ motif latent à la distribution sur les configurations de la section l durant la période tp. Pour une période tp fixée, l'ensemble des $\boldsymbol{\theta}^{\boldsymbol{tp,l}}(\boldsymbol{k})$ associées à toute section est exploité pour aboutir au tri des sections suivant leur appartenance au $k^{ième}$ motif.

Ces deux informations sont exploitées, d'une part, pour donner du sens aux thèmes et, d'autre part, pour ordonner les sections et construire les potentiels jeux de variables à tester.

4.2.2.2.2 Etape 2 : identification du motif critique à partir d'une méthode de prévision globale

Chacun des K_{lda} motifs latents produits par le LDA caractérise une typologie de comportement local des sections, mais également une composante typique des traces laissées par la structure spatio-temporelle du réseau lors de son activation. L'usage du LDA s'attaque donc à l'aspect résilience aux événements non récurrents lié à la définition des sections critiques en proposant des jeux de variables ordonnés ayant un sens vis-à-vis de la structure spatio-temporelle du réseau.

En revanche, cette nouvelle étape répond purement à l'aspect sélection de variables de la problématique et évalue différents jeux de variables. Les jeux de variables proposés ne sont pas choisis aléatoirement, mais suivant les motifs latents représentant la structure du réseau. Pour toute période temporelle tp, le LDA est évalué, si bien que chacun des K_{lda} motifs latents dispose de la liste ordonnée par probabilité décroissante des N sections. La fonction $SortRows$ [3] de l'algorithme 2 assure le tri des sections pour chaque période temporelle et tout motif latent. Depuis ces listes et pour un nombre restreint à $n\,(n \in [1, N])$ variables d'entrée, tout motif latent propose un jeu de n sections évoluant dynamiquement en fonction des périodes de la journée. En faisant varier le nombre de sections sélectionnées (n), l'ensemble des jeux de variables sont testés comme entrées potentielles de l'algorithme de prévision (MKNN), puis leur erreur globale est évaluée. Cette phase correspond alors au travail effectué par la fonction $AssessPerf$. La métrique employée pour évaluer les performances de prévision est la $MAPE_{global}$.

L'identification du motif critique consiste alors à trouver le motif associé aux jeux de variables sélectionnées minimisant l'erreur globale de prévision et le nombre de variables. Il s'agit donc d'optimiser une fonction de perte liant l'erreur de prévision et le nombre de variables d'entrée. Une telle fonction d'optimisation n'a pas été construite, puisqu'une évaluation graphique visuelle

3. Au même titre que $find$ et min, $SortRows$ assure les fonctionnalités telles que décrites sous MatLab.

est rapide et efficace. En effet, le motif latent considéré comme critique tend à minimiser l'erreur de prévision quelle que soit la quantité de variables considérées.

4.2.2.2.3 Etape 3 : recherche du nombre optimum de sections

Cette ultime étape consiste à trouver le compromis entre minimisation de l'erreur globale et réduction du nombre de variables sélectionnées. Elle résulte directement du problème d'optimisation conduisant à l'identification du motif critique, puisqu'en découvrant le critère de criticité, la quantité optimale de variables $\hat{n}$ est également obtenue. Le procédé résulte d'une lecture graphique par identification du coude ($ElbowMethod$) de la courbe retraçant l'évolution des performances de prévision en fonction du nombre n de sections considérées.

Le critère de criticité est alors totalement défini par le motif latent $\hat{k}$. Le nombre optimal $\hat{n}$ de sections à introduire en entrée étant fixé, ainsi que le motif latent correspondant au sections critiques, **le procédé de sélection de variables est applicable pour pratiquer de la prévision en temps-réel**. Suivant l'appréhension du problème, différentes stratégies sont envisageables :

- **Une stratégie stationnaire** : pour toute nouvelle prévision, les variables d'entrée sont directement déterminées par l'appartenance du pas de temps à l'une des TP périodes considérées. Pour toute prévision depuis l'instant t appartenant à la période tp ($tp \in [1, TP]$), les variables d'entrée correspondent aux $\hat{n}$ premières sections de la période tp selon le motif critique $\hat{k}$.

- **Une stratégie dynamique** : les TP périodes permettant l'identification des sections critiques servent de bases à la construction de périodes glissantes. Les probabilités de contribution des N sections sur la période glissante considérée sont évaluées par application du modèle LDA construit sur les K_{lda} motifs et les TP périodes. La probabilité de contribution de chacune des N sections au motif critique $\hat{k}$ est alors connue et les $\hat{n}$ sections les plus critiques sont utilisées pour construire les variables d'entrée de l'algorithme de prévision. Pour tout pas de temps, le jeu de sections critiques est réévalué au regard des nouvelles distributions sur les configurations locales dynamiquement mises à jour.

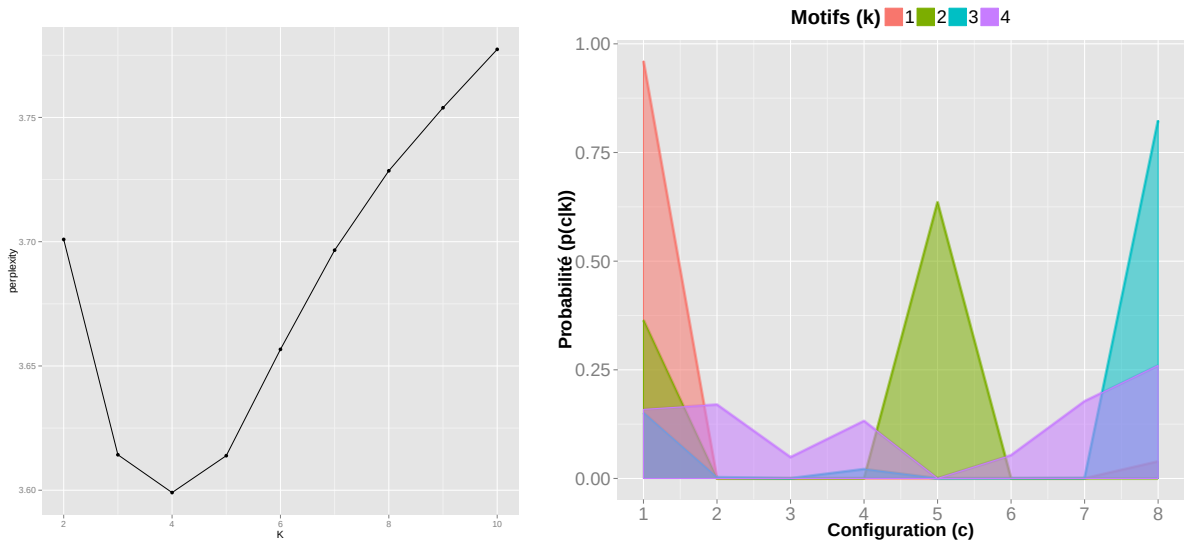
La détermination du motif critique fait systématiquement appel à une stratégie stationnaire lors de l'estimation de la performance de prévision. Seule la mise en oeuvre du procédé de sélection de variables, le motif critique étant connu, peut se voir traitée par diverses stratégies. Quelle que soit la stratégie adoptée, le critère de criticité, dont l'ordonnancement des sections évolue suivant la période tp considérée, fournit pour tout pas de temps un jeu de $\hat{n}$ sections aboutissant aux variables d'entrée.

4.2.2.3 Illustration de la phase de détermination du critère de criticité par le LDA sur un cas simple sur le réseau du Pont de Cheviré

Le procédé de détermination du critère de criticité est illustré depuis un cas simple sur le réseau du Pont de Cheviré à Nantes. La discrétisation adoptée est binaire et telle que décrite par l'équation 4.2. Huit configurations locales d'état sont donc envisageables à tout moment sur chacune des sections du réseau. Les périodes temporelles, considérées pour définir les motifs, se composent des $TP = 4$ périodes de la journée : nuit, pic du matin, midi et pic du soir. Les trois étapes du procédé d'identification du critère de criticité sont illustrées à partir de ce cas d'étude.

4.2.2.3.1 Etape 1 : détermination des motifs latents

L'algorithme du LDA tel que dépeint en section 4.2.2.2.3 est appliqué à un représentant de la structure spatio-temporelle composé du corpus de $TP \times N$ distributions [4] construites à partir de l'historique de données ($[1, T_{hist}]$, septembre-octobre 2013). L'optimisation du modèle LDA sur les données fournies aboutit à $K_{lda} = 4$ motifs, tels que représentés en Figure 4.6.



(a) (b)

Fig. 4.6.: Application du LDA au réseau de Cheviré (8 configurations et 4 périodes temporelles) : a) Identification du nombre optimal de motifs par la méthode du coude b)Représentation des $K_{lda} = 4$ distributions sur les configurations associées aux motifs latents.

Chacun de ces motifs est caractérisé par la distribution sur les configurations $\boldsymbol{\Lambda(k)}$ illustrée en Figure 4.6 b). À tout motif est associé pour chacune des $TP = 4$ périodes de la journée, une

4. Étant donné que les bordures du réseau font face à l'absence soit de père, soit de fils, ces derniers ont été encodés en fluide (0). Pour éviter de fausser l'évaluation des motifs latents, les cas en bordure de réseau sont exclus de cette phase.

liste ordonnée par probabilité $\boldsymbol{\theta}^{tp,l}$ décroissante des N sections du réseau. Ces listes ordonnées sont alors exploitées pour identifier le motif critique.

4.2.2.3.2 Etape 2 : identification du motif critique à partir d'une méthode de prévision globale

Chacune des listes ordonnées de sections associées aux motifs est testée en sélectionnant une quantité variable de sections en entrée de l'algorithme de prévision. Le motif critique correspond à celui minimisant l'erreur globale moyenne de prévision en nécessitant le moins de variables en entrée. La prévision est évaluée sur le mois de novembre 2013 en faisant varier le jeu de sections critiques suivant la période tp se rapportant à l'instant t d'estimation. L'évolution des performances de prévision ($MAPE_{global}$) en fonction du nombre de sections considérées comme critiques est exposée pour chacun des $K_{lda} = 4$ motifs en Figure 4.7 a) & b).
Le processus de sélection de variables s'appuie sur l'hypothèse fondamentale que les sections critiques suivent une définition à deux composantes : une composante dévoilée par les motifs du LDA supposés retranscrire la structure spatio-temporelle et assurer une bonne robustesse, puis une composante purement informatique de sélection de variables visant à réduire l'erreur globale de prévision. Afin d'affiner l'analyse et de vérifier l'hypothèse de résilience par intégration de la structrure spatio-temporelle, les Figures 4.7 c) & d) présentent les performances de prévision en conditions non récurrentes [5].
Les performances détaillées pour chaque horizon de prévision sont exposées dans le tableau 4.2.

À l'exception de l'horizon 30 minutes, le motif latent minimisant l'erreur moyenne globale de prévision est le 2ième motif représenté en vert sur les graphiques. Il est donc considéré pour tout horizon de prévision supérieur à 30 minutes comme le motif critique. D'après la Figure 4.6, ce motif est principalement généré par les configurations 5 (0, 0, 1) et 1 (0, 0, 0). Ceci signifie que 2 tendances principales définissent le comportement local des sections critiques : une tendance mineure (1), où les conditions de trafic sont non congestionnées, et une tendance majeure (5), où les conditions de trafic se dégradent sur les fils tandis que le père n'a rien transmis. Ainsi, d'après l'encodage adopté, le comportement local des sections critiques consisterait principalement à créer de la congestion sur ses sections en amont. Ceci suggère une contrainte forte dès l'entrée de la section l étudiée limitant son débit d'entrée et occasionnant des remontées de congestion sur les sections en amont. Cette interprétation correspond donc bien à un comportement susceptible de poser des difficultés de prévision à l'échelle du réseau.

L'analyse en conditions non récurrentes montre que quel que soit l'horizon de prévision, le motif 2 reste le plus performant en nécessitant moins de sections que les autres. Ceci tend à confirmer l'hypothèse de la relation liant la criticité des sections aux performances de prévision en conditions non récurrentes. L'importance du rôle joué par la classification LDA dans la saisie de la structure spatio-temporelle est ainsi soulignée.

5. voir chapitre 5 pour détails sur l'identification des conditions non récurrentes sur le mois de novembre 2013.

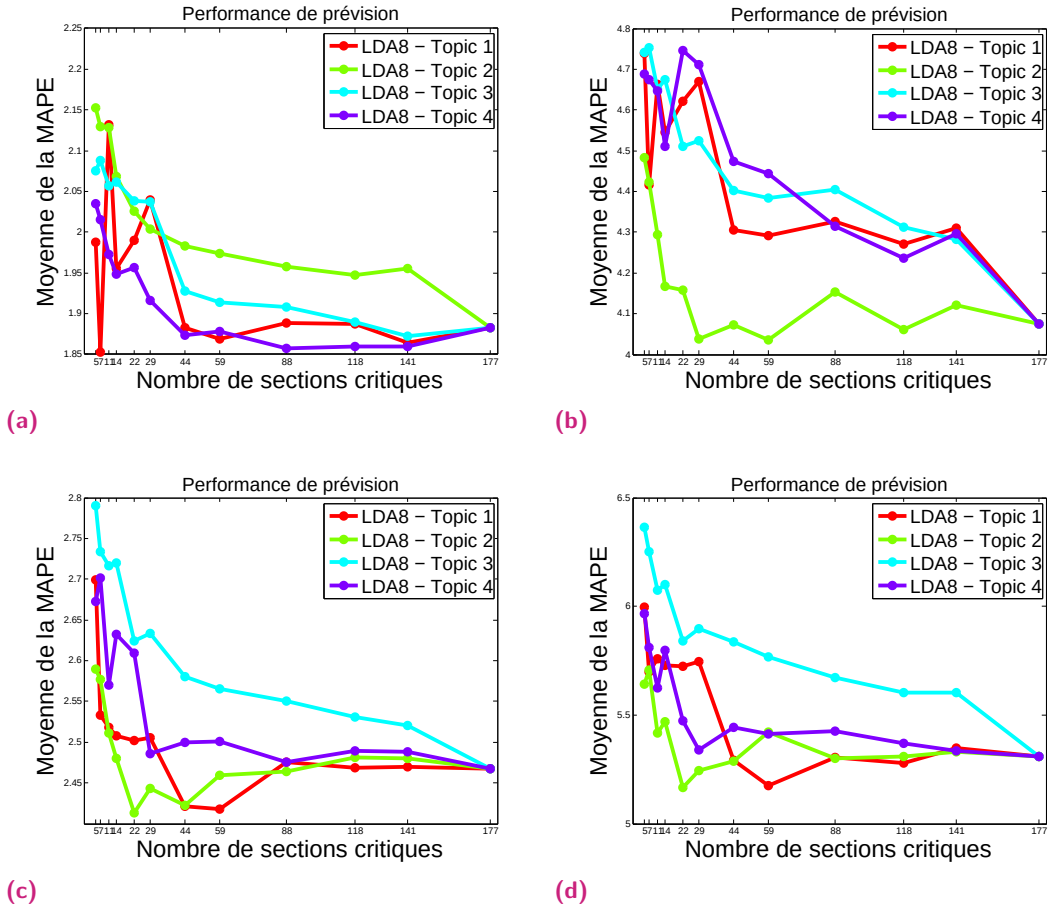


Fig. 4.7.: Évolution des performances associées aux jeux de variables produits par les $K_{lda} = 4$ motifs en fonction du nombre n de sections sélectionnées : a) à un horizon de 30 minutes sans distinction des conditions de trafic. b) à un horizon de 90 minutes sans distinction des conditions de trafic. c) à un horizon de 30 minutes en conditions non récurrentes. d) à un horizon de 90 minutes en conditions non récurrentes.

Seul l'horizon 30 minutes met en avant en conditions normales de trafic un motif autre que le 2ième. À cet horizon, le 4ième motif semble recueillir les meilleures performances globales et les plus stables en fonction du nombre de sections sélectionnées en entrée. Une telle modification de la criticité en fonction de l'horizon de prévision fait sens : les dynamiques de trafic saisies aux plus courts horizons de prévision diffèrent de celles estimées sur de plus longs horizons. Il est donc naturel que la définition des sections critiques en fasse autant. Cette observation est confirmée par l'instabilité du motif optimal, constatée à l'horizon 30 minutes, entre les cas récurrents et non récurrents. L'horizon 30 minutes jouerait un rôle intermédiaire entre les dynamiques très court terme (quelques minutes) et celles prenant place sur des horizons plus lointains. Néanmoins, seule une analyse sur de plus courts horizons de prévision permettrait d'entériner ou infirmer de telles hypothèses. En effet, si le motif 4 reste critique sur de très courts termes et améliore la robustesse en conditions non usuelles, alors l'horizon 30 minutes constitue bien un régime de transition entre des dynamiques de trafic distinctes.
Le motif 4 est influencé par trois configurations :

– la configuration 2 (1, 0, 0), où les pères dégradent les états de trafic, mais ne transmettent pas aux sections en amont ;

– la configuration 7 (0, 1, 1), où la section dégrade ses propres conditions de trafic et les transmet à ses fils ;

– la configuration 8 (1, 1, 1), où l'ensemble du bloc se dégrade simultanément, ce qui correspond à une congestion déjà bien installée.

Globalement, ces trois configurations soulignent la relation forte liant la section l à ses fils : l'état de trafic sur les sections critiques contrôle et détermine celui de ses fils. Un tel comportement correspond effectivement à des dynamiques de trafic très court-terme. Les sections critiques ne sont pas influencées par leurs parents, mais créent et propagent la congestion sur leurs fils. La moindre variation sur une section critique se répertorie immédiatement sur les sections en amont.
Finalement, la définition du motif critique à l'horizon 30 minutes reste difficile à trancher. A priori, la méthodologie développée conduit à sélectionner le 4ème motif ce qui est justifié également du point de vue des dynamiques à très court-terme, mais l'analyse sous conditions non récurrentes montre que le motif 2 est plus efficace. L'intérêt de l'identification des sections critiques est donc limité à cet horizon de 30 minutes. Certaines méthodes usuelles de sélection de variables devraient se montrer au moins aussi efficaces que le procédé proposé.

Horizon de prévision	H = 30 mn	H = 60 mn	H = 90 mn	H = 120 mn	H = 150 mn	H = 180 mn
$\hat{k}$	4 [1] 2 [2]	2	2	2	2	2
$\hat{n}$	48	30	35	35	32	31
Réduction du temps d'exécution	0.44	0.37	0.34	0.32	0.33	0.34

Tab. 4.2.: Amélioration de la complexité temporelle liée à la sélection de variables.

4.2.2.3.3 Etape 3 : recherche du nombre optimum de sections

L'identification du nombre $\hat{n}$ optimum de sections pour construire les variables d'entrée est déterminé par recherche du coude sur la courbe d'évolution de l'erreur de prévision en fonction du nombre n de sections pour le motif critique $\hat{k}$. Les valeurs optimales de n sont recueillies pour tout horizon de prévision dans le tableau 4.2. Tout comme le motif critique se modifie suivant l'horizon de prévision et les dynamiques de trafic, le nombre de sections requises évolue également. Force est de constater que l'horizon 30 minutes se démarque une nouvelle fois en réclamant une plus grande quantité de variables en entrée que les autres horizons.
En outre le tableau 4.2 montre également le gain en termes de complexité temporelle dû

1. En conditions de trafic usuelles.
2. en conditions de trafic non récurrentes.

à la restriction à $\hat{n}$ sections en entrée de l'algorithme de prévision, au lieu des N sections normalement introduites.

Le procédé appliqué au réseau du Pont de Cheviré fournit un critère de criticité pour tout horizon de prévision et chacune des 4 périodes de la journée. Ce critère est dès lors exploitable pour toute nouvelle prévision en sélectionnant les sections les plus critiques selon le motif $\hat{k} = 2$ (ou 4, suivant l'horizon de prévision) et la liste ordonnée des sections associée à la période concernée.

4.2.2.3.4 Interprétation géographique des sections critiques identifiées

Le processus fournit également un moyen de connaître l'évolution des sections critiques du réseau pour toute période de la journée, telle qu'illustrée en Figure 4.8.
Force est de constater que certaines sections critiques sont également celles identifiées comme difficiles à prévoir au chapitre 3 ou font partie de leur voisinage. L'échangeur Sud est particulièrement concerné par ce phénomène. Néanmoins, la forte erreur de prévision n'est pas la seule à expliquer l'intégration des sections de cet échangeur au sein des sections critiques. Elles contrôlent en effet les entrées et sorties sur le périphérique, ce qui implique que les informations portées par ces sections sont extrêmement impactantes. De même, les limites du réseau en entrée ou sortie du périphérique sont critiques pour toute période de la journée. Enfin, suivant la période concernée, certaines rampes d'accès ou portions urbaines sont considérées comme critique. L'axe principal du périphérique est parfaitement déterminé par ses branches d'alimentation et donc les informations à ses entrées et sorties. Comme pour un modèle de simulation, la connaissance aux bornes détermine en grande partie le comportement interne du réseau. Les sections qualifiées de critiques par le procédé d'identification ont effectivement un sens au regard de la structure du réseau. L'analogie avec la reconnaissance de visages paraît dès lors porter ses fruits.

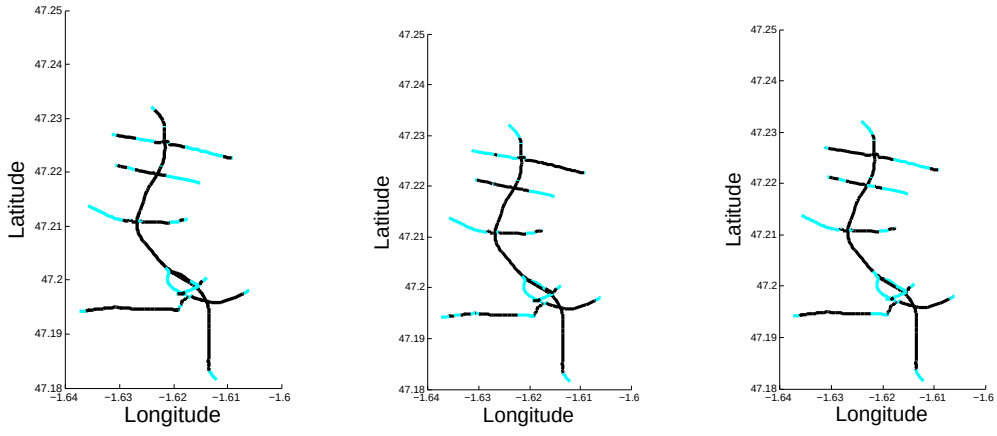


Fig. 4.8.: Identification de 44 sections critiques (cyan) suivant la période temporelle tp et le motif latent 2 (LDA à 8 configurations) : a) durant la nuit. b) durant le pic du matin. c) durant le pic du soir.

4.3 Évaluation du procédé de sélection de variables sur les réseaux nantais

Le processus exposé fournit une définition de la notion de section critique, puis l'exploite en vue de sélectionner les meilleures variables en entrée de l'algorithme de prévision. Cette méthodologie nécessite d'être adaptée au réseau et est sensible à la stratégie de tri des sections du réseau. Néanmoins, il s'agit de la méthodologie la plus performante et aboutie parmi celles envisagées lors de l'expérimentation. Certains de ces critères alternatifs envisagés sont exposés dans cette section. Tantôt issus d'hypothèses liées au trafic, tantôt adaptés de méthodes usuelles de sélection de variables, ces critères alternatifs sont testés, puis comparés au critère de criticité introduit précédemment.
Lors de cette confrontation de performances, la méthode précédemment introduite à partir d'une discrétisation binaire est notée LDA8. Les performances sont estimées en comparant l'évolution de l'erreur globale en fonction du nombre de sections à sélectionner pour construire les $\mu \times n$ variables d'entrée. Afin d'assurer la comparabilité des critères alternatifs, les hyper-paramètres de la fonction de prévision MKNN sont fixes et optimisés pour $n = N$ sections en entrée.
Ces critères alternatifs de sélection de variables sont proposés sur le réseau du Pont de Cheviré, mais également la zone urbaine nantaise. Les périodes temporelles sont inchangées et scindent en quatre phases homogènes toute journée.

4.3.1 Introduction de critères alternatifs de classement des sections

Cette section introduit quelques critères alternatifs testés, dont les performances sont les plus proches de celles expérimentées par application du procédé LDA8.

4.3.1.1 Critère M1

Ce critère suppose que les sections les plus critiques sont celles expérimentant les plus faibles vitesses sur la période temporelle tp. Le critère de criticité est alors défini par l'indicateur :

$$\forall tp \in [1, TP],\ \forall l \in [1, N],\ \theta^{M1}_{tp,l} = median_{tp}(\{Z^l_t, t \in tp\}) \tag{4.6}$$

Pour toute période tp, les sections du réseau, caractérisées par $\theta^{M1}_{tp,l}$, sont ordonnées par vitesse croissante. Les n premières sont alors utilisées pour construire le jeu de variables d'entrée.

4.3.1.2 Critère M1Bis

Ce critère suppose que les sections les plus critiques sont celles présentant la plus forte variabilité sur la période temporelle tp. Afin que la variabilité soit comparable d'une section à une autre,

les variations de vitesses réduites sont estimées : $\forall l \in [1, N],\ \delta_l(t) = \dfrac{Z_t^l - Z_{t-1}l}{\sigma_l}$. Pour toute période temporelle tp, la variation médiane de vitesse est estimée et exploitée comme variable indicatrice pour le tri des sections :

$$\forall tp \in [1, TP],\ \forall l \in [1, N],\ \theta_{tp,l}^{M1Bis} = median_{tp}(\{\delta_l(t), t \in tp\}) \tag{4.7}$$

Pour toute période tp, les sections sont ordonnées par variation décroissante, puis les n sections subissant les plus fortes variations sont conservées pour prévision.

4.3.1.3 Critère M2

Ce critère suppose que la criticité des sections resulte de la fréquence de détérioration de la vitesse expérimentée. Plus la vitesse tend à diminuer durant une période de la journée tp, plus la section est supposée critique pour le réseau. Ce critère est défini par :

$$\forall tp \in [1, TP],\ \forall l \in [1, N],\ \theta_{tp,l}^{M2} = freq_{tp}(\{\delta_l(t) \leq 0, t \in tp\}) \tag{4.8}$$

D'après ce critère, les sections critiques sont celles expérimentant le plus de congestions.

4.3.1.4 Critère M3

Ce critère reprend le procédé d'encodage en configurations locales associées à toute section pour tout instant. Tout comme chacun des K_{lda} motifs latents produits par le LDA apparaît comme un potentiel critère de criticité, toute configuration c est, ici, supposée possiblement dépeindre un comportement critique. L'hypothèse repose alors sur le principe de l'existence d'une configuration c vérifiant la règle suivante : plus une section expérimente cette configuration, plus elle est critique pour le réseau.
À toute configuration c, est associé un critère de criticité défini par la fréquence d'apparition de cette configuration durant la période tp sur la section l :

$$\forall tp \in [1, TP],\ \forall l \in [1, N]\ \forall c \in [1, C],\ \theta_{tp,l,c}^{M3} = \boldsymbol{\Phi_{tp,l}(c)} \tag{4.9}$$

Pour toute configuration (potentiel critère de criticité), les performances globales de prévision sont estimées. En principe, le critère sélectionné comme critique est associé à la configuration produisant la plus faible erreur globale de prévision.

Néanmoins, l'hypothèse formulée faisant porter l'aspect critique d'une section sur une unique configuration n'est pas vérifiée en pratique. En effet, l'analyse directe de l'erreur ne fournit aucune distinction significative entre plusieurs configurations.
Une origine statistique à ce problème est identifiée. Étant donné les dynamiques de trafic, il est rare qu'une section observe une même configuration sur l'intégralité d'une période temporelle tp. Plusieurs configurations co-habitent sur cette même période. Pour chacune, leur proportion

d'apparition caractérise le comportement de la section étudiée. Il n'est donc pas impossible que plusieurs configurations proposent un même ordonnancement des sections. Ceci est d'autant plus envisageable que l'apparition des configurations n'est pas équiprobable. Les cas d'absence de congestion ou de congestion générale sont susceptibles d'être les plus récurrents. D'autre part, la dynamique du trafic dicte également un agencement de l'apparition des configurations, si bien qu'une seule configuration ne suffit pas à définir l'aspect critique d'une section.

La méthodologie d'identification du critère de criticité par le LDA constitue une extension de ce procédé M3. Il est alors supposé que la seule proportion d'une configuration ne suffit pas à caractériser le comportement critique, mais bien l'agencement des proportions de configurations sur la période considérée. L'hypothèse de criticité ne porte plus sur la seule tendance à produire régulièrement une configuration, mais sur la dynamique des configurations expérimentées. En saisissant une telle dynamique par le LDA, l'unicité du motif critique est garantie par la sélection du nombre optimal de motifs latents K_{lda} et la quantité de critères potentiels à évaluer est réduite de r^3 à K_{lda}. Ce dernier point est particulièrement important surtout quand la discrétisation est affinée. Ainsi, en termes de performances, il est escompté que la criticité ressorte plus aisément d'une combinaison particulière de configurations que d'une configuration isolée.

À titre de comparaison, les performances produites par la 5ième configuration, contribuant le plus au motif $\hat{k} = 2$, sont proposées comme alternative.

4.3.1.5 Critère LDA64

Ce critère est construit en reprenant intégralement le processus d'encodage des configurations, puis de classification par le LDA. Seule la phase de discrétisation (fonction $Discrete$) est modifiée. Au lieu de considérer une discrétisation binaire des variations de vitesse, $r = 4$ catégories sont possibles, soit 64 configurations. L'équation 4.2 se réécrit de la façon suivante :

$$\forall i \in \{Fathers, Link, Sons\}, \quad \boldsymbol{\Delta_i}(t) = \begin{cases} 0 & si \quad \boldsymbol{\delta_i}(t) > 0.5 \\ 1 & si \quad 0 < \boldsymbol{\delta_i}(t) \leq 0.5 \\ 2 & si \quad -0.5 < \boldsymbol{\delta_i}(t) \leq 0 \\ 3 & si \quad \boldsymbol{\delta_i}(t) < -0.5 \end{cases} \tag{4.10}$$

Les 64 configurations en découlant caractérisent plus finement les variations de vitesses et sont susceptibles de produire des jeux de variables plus efficaces. La classification par LDA propose 7 motifs latents et l'évaluation des performances de prévision met en évidence que le 6ème motif caractérise la criticité des sections. Ce motif dépeint des typologies de comportements proches de celles exprimées par le motif 2 du LDA8.

4.3.1.6 Critère MLasso

Le critère de criticité est construit à partir d'une méthode *Embedded* de type graphical LASSO adaptée aux prévisions à sorties multidimensionnelles inspirée [6] de Friedman *et al.* [189]. Le critère de classement des sections est construit depuis la contribution de toute section à expliquer la variance des sorties du modèle pour la période temporelle tp :

$$\forall tp \in [1, TP],\ \forall l \in [1, N],\ \theta_{tp,l}^{MLasso} = mean(\{\underline{\boldsymbol{Lasso_{tp}(l, i)}}, i \in [1, N]\}) \tag{4.11}$$

où $Lasso_{tp}$ représente la matrice de co-variance, liant variables d'entrée et de sortie, issue du procédé du Lasso généralisé et adapté. Plus une section explique la variance globale de l'ensemble des sorties du modèle linéaire de prévision, plus elle est considérée comme critique et a de fortes chances d'être sélectionnée. Ce critère reproduit les sorties d'une méthode Lasso classique appliquée au problème et optimise donc les variables d'entrée en fonction des performances moyennes globales de prévision.

4.3.2 Comparaison des critères de criticité sur la zone du Pont de Cheviré

Afin d'appréhender les performances des divers critères de criticité ou de sélection de variables, l'évolution de l'erreur de prévision en fonction du nombre de sections sélectionnées par chacun des critères est évaluée en Figure 4.9. Les performances de prévision sont estimées sur le mois de novembre 2013 depuis un même modèle MKNN.

Les performances de prévision sont extrêmement proches pour tout critère de sélection de variables. Quel que soit le critère de criticité considéré, l'évolution de l'erreur en fonction du nombre n de sections en entrée de l'algorithme de prévision observe systématiquement une même tendance faisant apparaître un coude. En effet, pour de faibles valeurs de n, quelles que soient les informations fournies en entrée, elles contribuent malgré tout à expliquer la variance des sorties du modèle de prévision. Passé cette phase de forte diminution de l'erreur à tout ajout de variables, la composition du jeu de variables prend alors du sens et influe les performances. Une inflexion forte de la courbe (cuvette) est alors observée lorsque le jeu optimal de variables est atteint. Tout ajout de nouvelles variables est superflu et ne contribue aucunement à l'amélioration de la prévision, voire la déteriore. Ce premier constat confirme l'intuition qu'un nombre restreint de sections suffit à saisir l'essentiel du système et à limiter le fléau de la dimension et la complexité temporelle de l'algorithme. Au final, la comparaison des critères s'avère peu aisée, puisque toutes les courbes suivent une même tendance. Il s'agit dès lors d'analyser avec finesse les performances de prévision.

Les deux premiers graphiques (a) et b)) illustrent les performances en conditions récurrentes, tandis que les deux suivants (c) et d)) montrent la robustesse du critère de criticité face à

6. voir `https://publish.illinois.edu/xiaohuichen/code/graphical-lasso/`, pour codes MatLab employés.

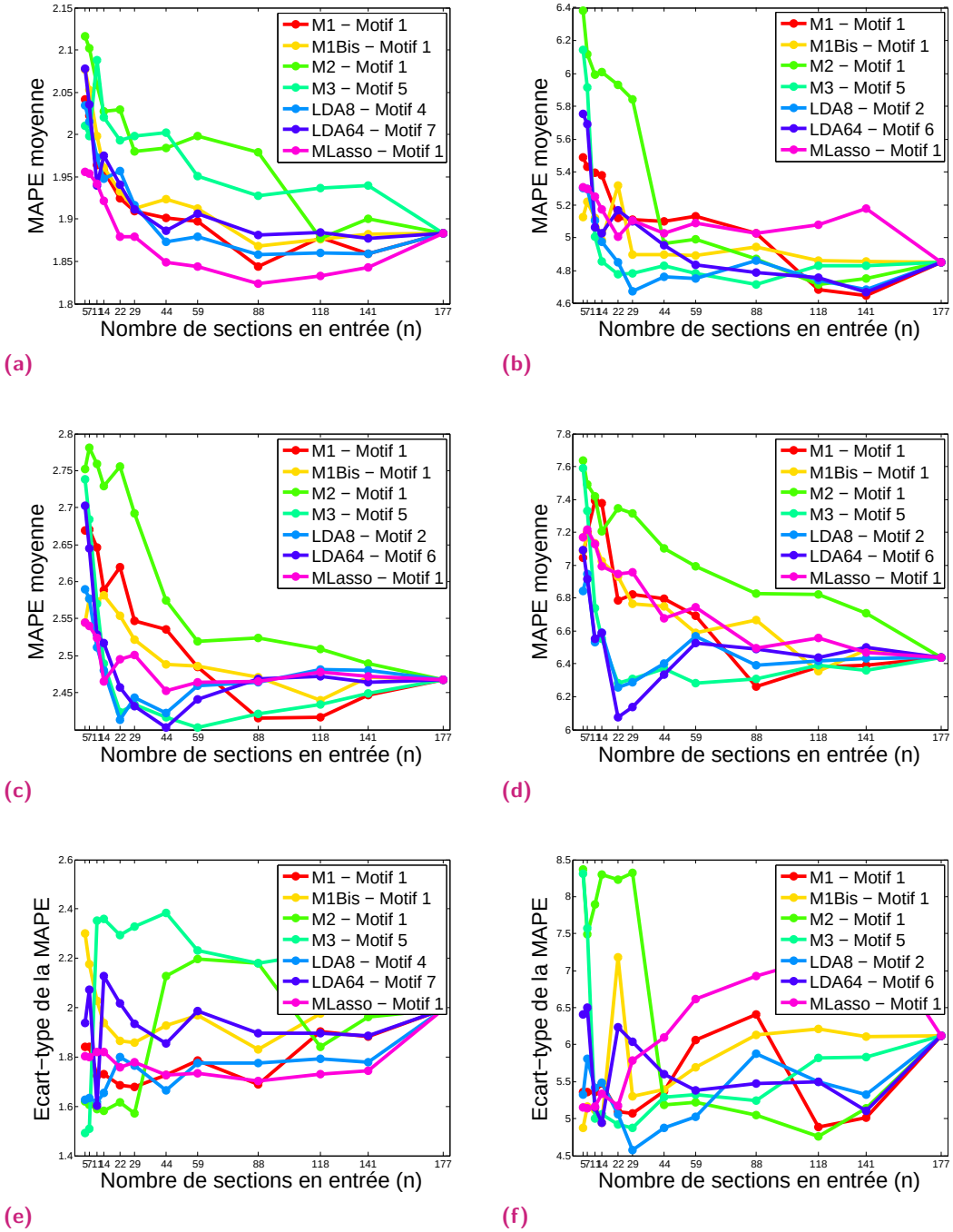


Fig. 4.9.: Évolution des performances associées aux divers critères testés en fonction du nombre n de sections sélectionnées : a) $MAPE_{global}$ à un horizon de 30 minutes sans distinction des conditions de trafic. b) $MAPE_{global}$ à un horizon de 2 heures sans distinction des conditions de trafic. c) $MAPE_{global}$ à un horizon de 30 minutes en conditions non récurrentes. d) $MAPE_{global}$ à un horizon de 2 heures en conditions non récurrentes. e) Évolution de la variance $MAPE_{section}$ pour un horizon de 30 minutes. f) Évolution de la variance de la $MAPE_{section}$ pour un horizon de 2 heures.

des conditions inhabituelles. L'analyse en conditions non récurrentes est particulièrement importante, puisque la robustesse constitue un objectif essentiel de cette étude.
À première vue, les méthodes s'appuyant sur le LDA (LDA8 & LDA64) sont parmi les plus performantes, surtout à horizon de prévision croissant pour LDA8.
Une différence de performances est clairement observée entre conditions récurrentes et non-récurrentes. La robustesse des méthodes de type LDA et M3 est plus conséquente que toute autre approche quel que soit l'horizon de prévision. Une telle observation confirme l'hypothèse que le comportement local des sections est apte à retranscrire, en partie au moins, la structure spatio-temporelle du réseau. On remarque que l'approche par LDA est légèrement plus robuste que la méthode M3. Le gain est dû à l'évaluation de distributions sur les configurations et non l'évaluation de configurations isolées.
Les performances de la méthode MLasso, sur de courts horizons de prévision (30 minutes), tendent à confirmer ces affirmations : l'erreur globale est minimisée par le Lasso en conditions habituelles, tandis qu'en conditions non récurrentes le Lasso est moins performant. La minimisation de l'erreur globale par le Lasso n'est plus effective dès que l'horizon excède 30 minutes. Il a été souligné que cet horizon de 30 minutes constitue possiblement un régime transitoire entre des dynamiques de trafic distinctes. Les méthodes LDA ont alors du mal à faire ressortir le motif critique, hésitation dont bénéficie le MLasso.
Exception faite de l'horizon 30 minutes en conditions habituelles, les méthodes LDA restent finalement légèrement plus performantes, tendance accentuée par analyse des conditions non récurrentes.

L'affinement du pas de discrétisation distinguant LDA8 et LDA64 montre une amélioration des performances de prévision uniquement sous conditions inhabituelles de trafic. Ainsi, mieux spécifier le comportement des sections, puis le motif de criticité assure une meilleure résilience au prix d'une diminution des performances globales de prévision. Un compromis entre finesse de la définition du thème critique et performances globales de prévision est donc à trouver.

Enfin, les deux derniers graphes (e) & f)) permettent l'analyse de la répartition des erreurs $MAPE_{section}$ en fonction du nombre n de sections en entrée pour chacun des critères. En termes d'ordre de grandeur, l'écart-type de l'erreur de prévision est approximativement le même quel que soit le critère. Néanmoins, une analyse plus fine révèle qu'au regard des autres critères, les méthodes LDA proposent moins de disparités d'erreur d'une section à l'autre, ce qui garantit une bonne prévision globale. Cette observation est également valide pour MLasso sur de courts horizons de prévision seulement. Il est à remarquer que le critère M3 produit certes de bonnes performances globales, mais accuse une disparité entre les sections plus importantes que l'approche LDA8. La diminution conjointe de l'erreur de prévision et de l'écart-type de l'erreur par section ($MAPE_section$) indique une amélioration pérenne des performances conjointement sur toutes les sections. Un tel phénomène est observé à l'horizon 30 minutes pour MLasso et LDA8.

4.3.3 Application à une zone urbaine étendue

En vue d'affiner l'analyse comparative des performances de la sélection de variables issue de l'identification des sections critiques, le centre-urbain de Nantes a été mis à contribution. Néanmoins, les informations quant aux phases récurrentes et non récurrentes n'étant pas disponibles pour ce réseau, l'interprétation y est nettement moins riche. En effet, le maillage intense de la zone urbaine et l'absence de corridors rend toute analyse manuelle des données trafic difficile et longue. L'usage de relevés d'informations exogènes est donc requis.
Le processus liant encodage en configurations et classement des sections selon des motifs latents par le LDA est appliqué à cette zone se composant de 796 sections pour les cas à 8 et 64 configurations.
Le critère LDA8 est construit une nouvelle fois essentiellement à partir du 2ème et du 4ème motif parmi les $K_{lda} = 4$ motifs latents identifiés. La proximité de performances entre ces deux motifs est plus importante que celle observée sur le cas du Pont de Chéviré. L'augmentation du nombre de sections considérées fait disparaître l'effet cuvette et lisse les courbes en allongeant la phase de convergence par croissance du nombre de sections n. En termes d'interprétation, ces deux motifs latents possèdent, à peu de choses près, la même signification que précédemment :

- le motif 2 est composé, à majorité, des configurations 4 (1, 1, 0) et 5 (0, 0, 1). La section critique n'expérimente donc pas les mêmes variations que ses fils. Une contrainte forte est exercée sur la section d'étude, si bien qu'une rupture de comportement existe entre la section l et ses fils. Les fils semblent être en opposition de phase avec la section étudiée. Étant donné le pas de temps de 30 minutes, il ne peut s'agir d'ondes de congestion. La présence de feux de signalisation et d'intersections pourrait expliquer une telle rupture de comportement. L'analyse de la Figure 4.11 révèle effectivement la présence de sections critiques à certaines intersections, mais reste difficile d'interprétation sans connaissance de la disposition des feux tricolores.

- le motif 4 est composé, comme précédemment, des configurations 2 (1, 0, 0) et 7 (0, 1, 1). Seule la configuration 8 (1, 1, 1) n'est pas retenue par ce motif. Le comportement de la section étudiée et de ses fils est fortement lié : la section contraint le signal et le transmet à ses fils.

Le développement du critère LDA64 aboutit à 8 motifs latents, dont le 7ème se distingue comme motif critique. Ce motif correspond à un ensemble de configurations, où la section étudiée détériore les conditions de trafic plus que ses pères et fils et tend à transmettre cette congestion à ses fils. On retrouve, en grande partie, le même schéma critique que pour le réseau du Pont de Cheviré. Il est intéressant de constater que la majorité des dynamiques de trafic, saisies à un pas de temps de 30 minutes, sont partagées par un réseau péri-urbain et un réseau urbain. Les comportements critiques ne diffèrent donc que légèrement d'un réseau à l'autre, rendant le critère de criticité et sa définition pratiquement génériques. Les motifs latents tirés du Pont de Cheviré sont éventuellement transposables à celui de la zone urbaine de Nantes.

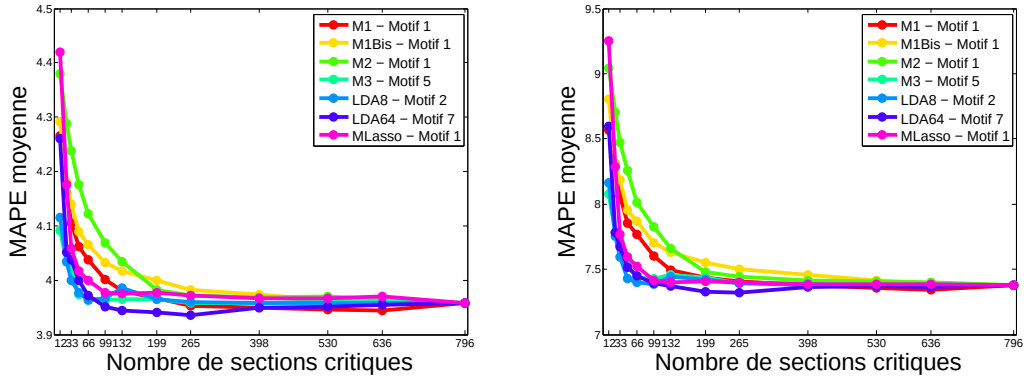


Fig. 4.10.: Évolution des performances associées aux divers critères testés en fonction du nombre n de sections sélectionnées sur la zone urbaine de Nantes : a) $MAPE_{global}$ à un horizon de 30 minutes sans distinction des conditions de trafic. b) $MAPE_{global}$ à un horizon de 90 minutes sans distinction des conditions de trafic.

La performance des critères de type LDA (LDA8 et LDA64), au regard des autres critères considérés, reste bonne, mais peu marquée. Sur cette zone urbaine, la confrontation des performances des divers critères est plus difficile d'interprétation, du fait du lissage par le grand nombre de sections composant le réseau. L'analyse de la Figure 4.10 montre néanmoins que le critère LDA64 est plus performant que toute autre méthode, mais atteint sa plus faible erreur tardivement. Le critère MLasso est le principal concurrent des critères de type LDA en proposant des performances équivalentes. Afin d'affiner l'analyse, un meilleur ciblage des zones de minimisation de l'erreur pour les faibles quantités de sections serait nécessaire. Finalement, en l'absence d'informations quant aux événements non récurrents, les conclusions sur cette aire d'étude restent limitées.

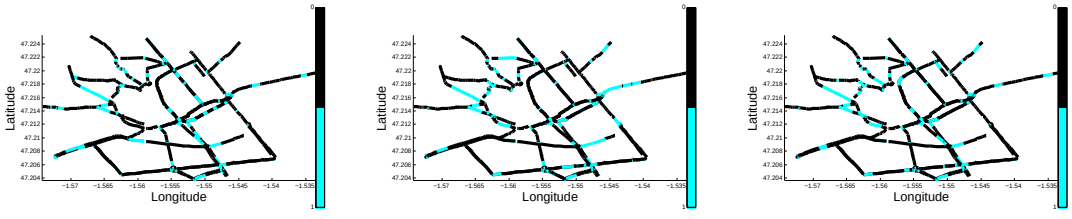


Fig. 4.11.: Identification de 132 sections critiques (cyan) suivant la période temporelle tp et le motif latent 2 (LDA à 8 configurations) sur la zone urbaine de Nantes : a) durant la nuit. b) durant le pic du matin. c) durant le pic du soir.

Les sections critiques identifiées par le 2ème motif tiré du LDA8 sont exposées en Figure 4.11. Certaines zones qualifiées de critiques correspondent à des zones identifiées au chapitre 3 comme difficile à prédire, notamment les zones proches du centre-ville et l'accès au Pont

de Bellevue à l'Ouest. Néanmoins, contrairement au cas du Pont de Cheviré, où les sections critiques visaient les bornes du réseau, la plupart des sections critiques tendent à couvrir l'étendue du réseau en se densifiant sur les zones se rapprochant du centre-ville.

Finalement, les comportements critiques du centre-ville urbain de Nantes ne sont pas tant différents de ceux observés sur la zone du Pont de Cheviré, mais adoptent une stratégie de positionnement géographique bien distincte. Le peu de distinction entre les comportements critiques d'un réseau à l'autre est à rapprocher des typologies urbaines des branches d'accès sur le réseau du Pont de Cheviré. Il est donc naturel que des dynamiques communes s'y développent.

4.3.4 Vers une application en temps-réel du procédé de sélection de variables

L'évolution des sections critiques est *a priori* permanente et dépend des conditions expérimentées par le réseau. Dans l'idéal, le critère de criticité devrait être en mesure de mettre à jour, à tout moment, l'ordre des sections du réseau. Il s'avère que le modèle LDA, stationnairement construit sur TP périodes temporelles, offre, en réalité, la possibilité de classer tout nouvel ensemble de N distributions sur les configurations (stratégie dynamique).
Cependant, les périodes temporelles doivent évoluer pour s'adapter à une fenêtre glissante, propre à un usage en temps-réel. Une nouvelle définition des périodes temporelles permettant la construction des distributions sur les configurations est requise. Cette étape s'avère particulièrement délicate : la fenêtre temporelle doit être suffisamment large pour que la distribution ait du sens, mais suffisamment fine pour caractériser la dynamique à l'instant présent.
La stratégie, envisagée pour définir de telles périodes temporelles et, par extension, les distributions sur les configurations, consiste à construire une **fenêtre temporelle artificielle** d'évaluation. Pour toute section l et tout instant t de prévision, la **distribution artificielle** associée $\mathbf{\Phi}_{t,l}$ se scinde en 2 composantes :

- **une composante stationnaire** correspondant à la distribution sur les configurations expérimentées au même instant de journées d'un historique fixé, soit 48 distributions pour un pas de 30 minutes ;
- **une composante dynamique** recueillant les configurations récemment [7] expérimentées par la section l.

La contribution à la distribution artificielle est de moitié pour chacune des composantes.
Le principal avantage de cette stratégie est que la définition du motif critique est possible depuis une stratégie stationnaire appliquée aux $TP = 48$ périodes constituant les composantes stationnaires. Le nombre de distributions à classer par le LDA est donc restreint. Le modèle de classification est construit rapidement et efficacement pour aboutir à la définition de motifs latents. L'application du cas à 8 configurations sur la zone du Pont de Cheviré parvient au

7. Les configurations les plus proches de l'instant t ne peuvent être intégrées, puisqu'elles font appel aux états de trafic futurs des fils.

même motif critique que ceux identifiés précédemment. Il est alors supposé que les distributions artificielles suivent les mêmes motifs que ceux identifiés par les distributions stationnaires. Cette hypothèse n'est malheureusement pas vérifiée en pratique lors de la mise en oeuvre en temps-réel. En effet, les modifications induites par les distributions artificielles occasionnent une modification des motifs critiques. Après vérification, il s'avère que le 2ème motif n'est plus le plus performant, mais celui correspondant aux configurations intégralement fluides. Une réévalution directement portée sur les distributions artificielles des motifs critiques confirme cette observation. L'introduction de périodes et distributions artificielles altère donc en profondeur les motifs latents du système et efface l'effet de la structure spatio-temporelle. **La définition de la criticité n'a alors plus de signification physique, au regard de ces périodes artificielles**. La caractérisation par fenêtre glissante artificielle n'est donc pas adaptée au problème. Des approches alternatives doivent être envisagées.

4.4 Conclusions et perspectives

▶ Synthèse

L'objectif poursuivi par ce chapitre est ambitieux car à double enjeux. Il s'agit, d'une part, de procéder à une sélection de variables visant à passer outre le fléau de la dimension pour réduire erreur de prévision globale et complexité temporelle et, d'autre part, de maintenir une bonne robustesse de la prévision face à des événements peu récurrents en saisissant finement la structure spatio-temporelle du réseau. La recherche d'un compromis liant ces deux aspects est requise. Le processus adopté aboutit à deux principales contributions :

- l'**identification des sections critiques** du réseau, de leur agencement et de leur évolution en fonction des périodes de la journée pour caractérisation de la structure spatio-temporelle du réseau ;
- la proposition d'**une méthode de sélection de variables** de type *Variable Ranking* apte à s'adapter dynamiquement suivant les périodes de la journée.

La méthodologie introduite consiste à identifier, parmi un ensemble de motifs tirés de configurations locales de trafic, celui captant au mieux les caractères de la structure spatio-temporelle du réseau pour répondre aux besoins de prévision. Un unique critère générique triant les sections du réseau suivant leur degré d'appartenance au motif est alors défini, puis seuls les $\hat{n}$ liens y contribuant le plus sont conservés en entrée du modèle de prévision. Les $\hat{n}$ sections sélectionnées sont alors qualifiées, pour la période évaluée, de sections critiques du réseau.

Les critères de sélection de variables tirés de la méthodologie proposée (type LDA) répondent aux attentes en termes d'amélioration de la robustesse et de réduction (par trois, en moyenne) du temps d'éxécution de la prévision.
Comparés aux critères alternatifs testés, les procédés de type LDA se révèlent souvent bénéfiques suivant l'indicateur de performance choisi. En conditions usuelles de trafic, dès que l'horizon excède 30 minutes, les critères LDA proposent l'**erreur de prévision la plus faible**. C'est

également le cas en conditions non récurrentes, pour tout horizon de prévision, ce qui **confirme l'hypothèse** formulée quant au **lien entre comportement critique des sections, structure spatio-temporelle et résilience**. Résultant de l'extension d'une méthodologie plus simple (M3), les critères LDA nécessitent de tester moins de critères potentiels et présentent une plus forte probabilité de **garantir l'unicité du critère** de sélection de variables.
Enfin, il résulte de l'analyse sur les cas étudiés que la **définition de la criticité** depuis des motifs latents est **pratiquement générique**, *i.e.* inchangée d'un réseau à un autre. Seule la répartition spatiale des sections critiques diverge suivant le réseau considéré, mais les rôles joués par les sections restent les mêmes. En revanche, l'hypothèse de l'**évolution dynamique des sections** n'est que **partiellement vérifiée**, puisqu'une grande part des sections restent critiques à tout instant.

▶ Limites

Le procédé de sélection de variables n'est évalué que sur la seule méthode MKNN. Une confirmation sur d'autres approches multivariées (*MIMO*) de type MSVR ou MKernel est à envisager.
La transition vers une identification dynamique des sections critiques n'est pas encore atteignable. La stratégie à fenêtre glissante proposée annihile le sens physique des motifs latents et donc l'intérêt de la méthode. Une meilleure définition de la période temporelle reste à déterminer pour une application en ligne et dynamique du procédé.
En outre, une hypothèse de stationnarité des liens entre la section d'étude et ses fils/pères est formulée lors de la construction des configurations d'état. Les décalages temporels et pondérations permettant la modélisation de toute configuration sont évalués depuis une méthode de corrélation croisée. Dans une perspective d'évaluation en temps-réel des sections critiques, les configurations pourraient être affinées par l'usage d'un procédé d'évaluation en temps-réel de ces paramètres (usage de l'information mutuelle ou *graphical LASSO* par exemple).

▶ Perspectives

L'identification des sections critiques constitue un moyen d'éviter ou de limiter le problème des données manquantes ou aberrantes dans certaines situations. En effet, la prévision ne s'appuyant plus que sur une quantité restreinte de sections, les informations manquantes ou erronnées sur les sections non sélectionnées n'influencent aucunement la prévision. Un procédé de complétion hors-ligne peut ensuite être appliqué pour que le cliché du réseau intégrant la base de données soit exploitable. En revanche, en cas d'informations manquantes sur une section critique, le problème reste entier.

5

Identification et intégration de phénomènes non récurrents perturbant le comportement nominal du réseau

> *Les embouteillages, c'est pratique : ça fait une bonne excuse pour arriver en retard.*
>
> **— Michèle Bernier**
> Le Petit Livre de Michèle Bernier

Abstract :
This chapter deals with the non-recurrent events impacting the nominal behavior of the network.
A first step contributes to the detection of unusual phenomena. A brand new detection method replacing the usual inputs (traffic states) by a clustering indicator (an entropy measurement) is introduced. Such input modification aims to better distinguish the non-recurrent events from the recurrent ones. In comparison to basic methods, its detection accuracy is good.
The second step aims to integrate permanently an incident database to the classic historical database used for learning. This incident database is regularly updated by including the detected incidents strongly impacting the prediction accuracy. The process enables to increase the variability of cases within the learning database without futilely growing the historical size. A long-term improvement of the prediction accuracy is targeted.

Il est usuellement observé que certaines composantes du trafic se répètent avec une régularité prédéterminée (pic de demande en accord avec les déplacements pendulaires, signalisation ...). Ils définissent les motifs récurrents du trafic. Néanmoins, il est reconnu que de tels motifs sont modifiés d'une période de temps à une autre, suivant alors des saisonnalités. Ces phénomènes sont dûs à des variations de comportement des usagers que ce soit lié à des périodes de vacances ou aux jours de la semaine. Il est généralement observé une hausse du volume de trafic les lundis ou vendredis et des chutes de trafic sur les déplacements domicile-travail durant le mois d'août. Néanmoins, ces motifs réguliers ou saisonniers 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 [204]. Certains événements non-récurrents sont planifiés de longue date (travaux, manifestations), auquel cas des stratégies de régulation sont applicables en amont, mais, pour la plupart, il s'agit d'événements impromptus. De tels phénomènes inopinés trouvent diverses causes : incidents, accidents, intervention d'urgence ou encore météorologie inclémente. 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.

Après l'introduction du cadre prédictif propre à la démarche globale (**chapitre** 2) et son évaluation sur diverses méthodes de prévision (**chapitre** 3), le **chapitre** 4 s'est intéressé à l'identification des sections critiques pour délimitation des informations essentielles permettant, entre autres, de mieux faire face à des conditions non récurentes. Poursuivant dans la veine de la robustesse de prévision, ce chapitre vise à identifier et mieux cibler, au sein du procédé de prévision, les périodes de congestion non-récurrentes affectant le fonctionnement nominal du réseau routier. Après une phase de détection de tels phénomènes à partir des seules caractéristiques du trafic disponibles, leur intégration au sein du procédé de prévision est étudiée dans une perspective d'implémentation en ligne.

De même que pour le chapitre 4, le procédé de prévision est considéré comme une boîte indifférenciée (*Blackbox*), si bien que chacune des méthodes issues de l'approche globale introduite en **chapitre** 2 pourrait être appliquée. Néanmoins, par souci de concision et de simplicité, seuls les résultats produits à partir de la méthode MKNN sont exposés par la suite.

5.1 Les événements non récurrents dans la littérature du trafic

5.1.1 Identification des événements non récurrents

À défaut de pouvoir anticiper ces congestions non-récurrentes, des méthodes de détection en temps-réel ont été développées à partir des données trafic au cours de ces dernières décennies. La méthode pionnère (*California*), étendue, depuis les travaux de Kahn, par Payne et Tignor

[205], se base sur la détermination de seuils à partir de la comparaison des densités de trafic entre amont et aval. Outre de telles méthodes s'appuyant sur la physique du système [206, 207], des approches statistiques visant à détecter les différences significatives entre les valeurs observées et des composantes de trafic prévues ont été largement employées par la suite. Ces méthodes s'appuient, notamment, sur des approches par régression paramétriques (ARIMA, [208]), des réseaux bayésiens [209, 210] ou encore des lissages exponentiels [211]. Elles présentent le principal désavantage de détecter tout changement d'état dû aux variations journalières et non les seules congestions non récurrentes, ce qui induit de multiples fausses alarmes et réduit l'efficacité du procédé.

L'accroissement des performances informatiques, des bases de données et le développement de l'apprentissage mènent à assimiler l'identification d'incidents à une problématique de classification d'états de trafic. Cette approche par apprentissage supervisé est encore aujourd'hui la plus efficace et la plus répandue dans la littérature. Elle consiste en une première phase d'apprentissage des classifieurs à partir d'un historique de données trafic et de recueils d'incidents, puis en une phase de test sur une nouvelle base de données trafic pour validation.
Depuis les années 1990, une grande variété de méthodes issues de l'apprentissage ont été développées aussi bien à travers les réseaux de neurones et la logique floue [212, 213, 214], les régressions partielles des moindres carrés [215], les arbres de décision [216] ou plus récemment les forêts aléatoires [217] et les machines à vecteurs supports [218].
Néanmoins, les méthodes les plus récurrentes restent les méthodes à réseaux de neurones prenant de multiples formes. Srinivasan *et al.* [219] et Cheu *et al.*[213] ont évalué trois méthodes basées sur de tels réseaux et ont abouti à des performances prometteuses, notamment à partir d'un réseau de neurones probabiliste constructiviste. Bien qu'extrêmement performantes, ces méthodes s'avèrent difficilement généralisables. Chaque réseau est calibré et paramétré pour un cas d'étude spécifique. De plus, le fonctionnement en boîte noire de la méthode ne facilite pas la compréhension du procédé de paramétrage. Scientifiquement très intéressantes, ces méthodes sont difficilement exploitables pour les gestionnaires d'infrastructures.
La principale approche concurrente est celle des méthodes à noyaux et plus spécifiquement les classifieurs à vecteurs supports (SVM, [108]). Depuis la méthode SVM usuelle appliquée à la détection d'incidents par Yuan et Cheu [218] en 2003, différentes variantes ont vu le jour avec pour objectif de pallier le problème de sensibilité du choix des paramètres. Les performances en généralisation s'avèrent dépendantes de la sélection des noyaux et de leurs paramètres. Chen *et al.* [220] ont proposé une méthode ensembliste à SVM limitant le problème du choix des noyaux, puisque chacun des noyaux est potentiellement utilisé, et améliorant les performances. Dans cette même veine, Xiao et Liu [221] ont proposé une machine à vecteur support, dont le noyau est une combinaison des noyaux usuels, et obtiennent des performances équivalentes aux approches ensemblistes, tout en limitant le temps d'apprentissage.
Face à la sensibilité du choix des paramètres communs à la plupart des méthodes d'apprentissage, Lu *et al.* [222] proposent une approche liant apprentissage et programmation logique. Les performances de reconnaissance des incidents, certes stables, restent moindres que celles observées à partir des méthodes habituelles. Enfin, certains prétraitements usuels tels que la

décomposition en ondelettes sont également constatés et efficaces ([223]). Dans le domaine, l'usage direct de la décomposition en ondelettes en tant que méthode de détection d'incidents comme c'est le cas en informatique [224, 225] reste anecdotique [226].

Il est à noter que les méthodes introduites précédemment sont celles habituellement rencontrées et produites à partir de données usuelles de trafic telles que les boucles électromagnétiques, les lectures automatiques de plaques... Néanmoins, certaines méthodes de détection se basent sur d'autres sources d'informations tantôt bien réelles à travers la détection automatiques d'incidents par traitement d'images et de vidéos ([227, 228]), tantôt en perspectives. Depuis quelques années, certaines technologies de communications inter-véhiculaires ou véhicule-infrastructure sont à l'étude et préparent l'avènement des véhicules connectés. Des méthodologies de détection d'incidents sont d'ores et déjà développées, notamment autour de la technologie VANET (*Vehicular Ad-Hoc Networks*) [229]. Bien que la plupart se basent sur des simulateurs de trafic [230, 231], leurs performances et leur rapidité de détection sont à souligner. Néanmoins, outre le fait que ces méthodes reposent sur des dispositifs disponibles à plus ou moins long terme, les méthodes développées requièrent un grand nombre d'informations trafic fournies localement par une grande quantité de capteurs. De telles informations ne sont pas disponibles en permanence, ni sur toute infrastructure. L'usage de données issues de véhicules traceurs permet de pallier cette problématique spatiale. La plupart des méthodes basées sur de telles données visent à identifier les véhicules ayant un comportement anormal et permettent non seulement de détecter les ralentissements inhabituels, mais également les vitesses excessives correspondant à des périodes inhabituelles (jours fériés) [232, 233].

5.1.2 De la mesure à l'anticipation d'impact des incidents

Au-delà de la détection en temps-réel des incidents en vue d'interventions rapides des gestionnaires d'infrastructures, la prévision des effets et de la propagation de telles congestions non-récurrentes constitue une seconde étape clé. Pendant un temps, l'estimation d'impact s'est basée sur des modèles de théorie du trafic ou de simulations. On distingue deux approches basées sur les théories déterministes de files d'attentes et à ondes de congestion [234, 235], les modèles de simulation microscopique ou à heuristiques [32, 236, 237]. Bien que pertinentes, ces méthodes se fondent sur des théories physiques de propagation et de suivi de véhicules pas toujours vérifiées sur le terrain.

Avec la mise à disposition de grandes quantités de données réelles, les méthodes statistiques de prévision d'impact se sont répandues dans la littérature. Les approches consistent souvent en une première phase de qualification et classification des incidents (de léger à fort), puis en une phase de prévision des catégories des nouveaux événements. À cette fin, les méthodes classiques de classification non supervisée sont employées, à savoir les arbres de décision [238, 239], les classifieurs bayésiens [240] ou encore les k plus proches voisins [241]. Ces approches ne fournissent, cependant, que des informations qualitatives sur l'évolution des perturbations. Au sein de chaque catégorie, une grande variabilité de comportements est alors

observée. Certaines études plus récentes [242, 243, 244] visent à caractériser non seulement la zone affectée, mais également la chute de vitesse à prévoir. Une méthode très récente et particulièrement fournie en variables caractérisant les incidents est même parvenue à prévoir les potentiels profils de vitesses sur les sections supposément impactées à partir d'un historique [245].

Ces méthodes, apportant des informations complémentaires quant au fonctionnement du réseau, restent considérées comme des modules indépendants des méthodes de prévision. L'une des perspectives des travaux introduits vise justement à intégrer directement au sein de la méthode de prévision les informations relatives aux incidents.

5.2 Objectifs méthodologiques : un procédé en deux phases

La méthodologie proposée est appliquée aux données FCD recueillies sur la zone du Pont de Cheviré à Nantes. Elle vise à répondre à deux objectifs :

- **proposer un procédé original d'identification d'événements non récurrents basé sur un indicateur de qualité de classification**, indicateur propre aux procédés de classifications issus de modèles probabilistes génératifs : la perplexité. Cette méthode est confrontée à d'autres approches classiques de détection d'incidents tels que le SVM, ses variantes ou une méthode basique à seuils.
- **intégrer les informations fournies par les procédés de détection d'incidents en vue de prévision d'impacts**. Le procédé consiste alors en l'inclusion dans l'historique d'apprentissage des incidents perturbant de façon conséquente la prévison en vue d'améliorer les performances à venir. À cette fin, l'historique glissant, permettant l'évaluation des futurs états de trafic (voir chapitre 2), est complété par un historique d'événements non récurrents. Ce dernier est régulièrement mis à jour et complété à partir des phénomènes inhabituels les plus impactants (*i.e.* les plus utiles à intégrer). Afin de limiter l'expansion de l'historique d'incidents et d'éviter toute répétition d'événements aux comportements proches, seuls les événements à fort impact prédictif sont intégrés à l'historique.

Le procédé suivi est illustré par le schéma 5.1.

Outre l'identification des incidents, la méthodologie proposée est en mesure d'améliorer à long terme la prévision en mettant à jour en temps-réel la base d'incidents utiles.

Il est à préciser qu'aucune vérité terrain n'est disponible pour le cas d'étude de la zone du Pont de Cheviré. Seule une analyse experte et manuelle des diagrammes espace-temps a permis de définir les événements considérés comme incidents. Outre l'incertitude de mesure, une incertitude quant à la réalité terrain est donc à prendre en considération. Ne pouvant être complètement entériné, le procédé exposé par la suite ne peut donc prétendre qu'au titre de preuve de concept.

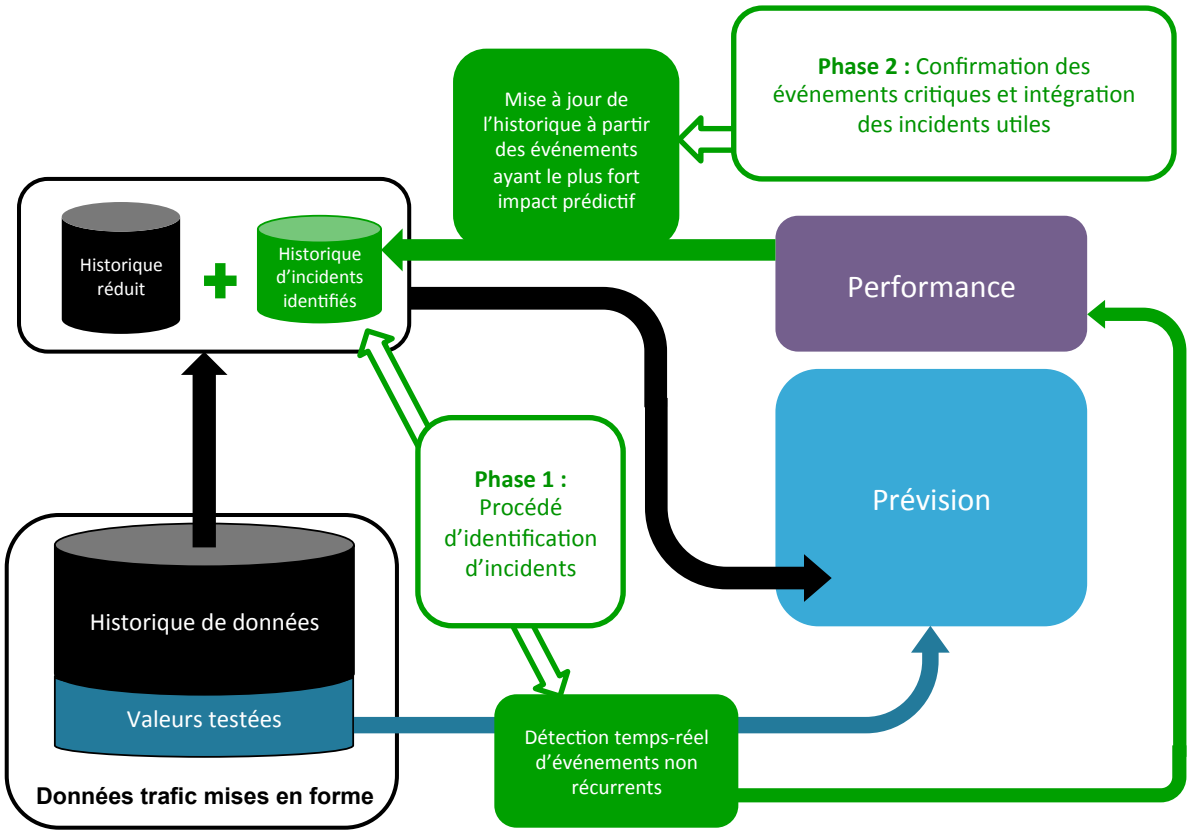


Fig. 5.1.: Illustration de la démarche de détection et intégration des événements non récurrents au sein d'une approche globale de prévision.

Par souci de facilité de lecture, bien qu'il s'agisse en réalité d'événements non récurrents, le terme "incident" sera néanmoins employé pour désigner de tels phénomènes inhabituels.

5.3 Phase d'identification des incidents

L'état de l'art met en lumière deux principales approches pour la détection d'incidents : des approches à seuil basées sur des composantes physiques du trafic et des approches de type classification à partir de méthodes d'apprentissage supervisé. La nouvelle approche proposée combine ces deux aspects. Elle s'appuie tout à la fois sur une méthode de classification issue de l'apprentissage, mais conclut à partir de seuils.

A titre de comparaison, une méthode basée sur la MAD (*Median Absolute Deviation*) pour les approches physiques et deux méthodes basées sur le SVM (Séparateurs à Vastes Marges ou *Support Vector Machine*) pour le compte des approches par classification sont également introduites.

Les quatre méthodes sont paramétrées ou apprises sur l'ensemble du mois de septembre 2013 (1440 pas de temps) ou, suivant les exigences du procédé, seulement sur les pas de temps sans incident. Elles sont ensuite testées sur le mois de novembre 2013 (1440 pas de temps). Le

mois d'octobre expérimente peu d'incidents et est réservé à un usage de base de données de référence.

5.3.1 Une méthode originale combinant classification et seuils

La méthode introduite repose sur deux phases. Une première phase vise à construire un indicateur se substituant à la variable de trafic usuellement employée pour l'identification d'incidents. La seconde phase consiste à appliquer un procédé de détection par seuils à ce nouvel indicateur.

5.3.1.1 Hypothèse et méthodologie

Le procédé repose sur l'observation que les méthodes de classification non supervisée fournissent, certes, une catégorisation des individus en entrée, mais sont également accompagnées d'un évaluateur de bien-être des individus classés. Un tel indicateur prend la forme des silhouettes pour la plupart des méthodes non supervisées telles que les K-means ou encore de la perplexité dans les cas de modèles génératifs probabilistes. Cet évaluateur peut également être perçu comme une mesure de la difficulté à classer un individu par rapport aux autres.
L'hypothèse fondamentale alors formulée est la suivante :

Hypothèse 1 *Si, dans des conditions usuelles semblables, un individu est plus ardu à classer que les autres, alors cet individu possède un comportement divergent de ceux habituellement observés. Un tel individu est qualifié de non récurrent.*

Depuis cette hypothèse, le procédé d'identification d'incidents s'effectue en deux phases :

- **Estimation via un indicateur associé à une méthode de classification non supervisée** : mesure $p_{l,t}$ en temps-réel de la difficulté de classement de l'état de trafic expérimenté par la section l à l'instant t ;
- **Filtrage par seuil des valeurs indicatrices** : comparaison de la mesure $p_{l,t}$ à un seuil $g_{l,t}$ prédéfini et qualification de l'aspect inhabituel du phénomène.

5.3.1.2 Usage de la perplexité

La méthodologie est testée à partir de la méthode générative probabiliste introduite en chapitre 4 : le LDA. La perplexité est employée afin de caractériser de façon continue la difficulté de classement de chacun des individus. Cette dernière se distingue des indicateurs de bien-être tels que la silhouette, de par son potentiel à estimer la capacité du modèle à reconstruire les individus classés. Au-delà du bien-être, la perplexité caractérise véritablement la difficulté éprouvée par le modèle issu de la classification à catégoriser les individus testés.

Mesurant une forme d'incertitude, sa définition pour un ensemble de n individus intègre une formulation proche de celle de l'entropie du système :

$$perplexité(\{i_1, ..., i_n\}) = e^{-\frac{\sum_{k=1}^{n} log(p(i_k))}{n}},$$

où n est le nombre d'individus testés, i_k un individu classé et $p(i_k)$ sa probabilité d'après le modèle défini.

5.3.1.3 Construction d'individus suivant les contraintes

L'usage de méthodes génératives probabilistes en vue d'identifier des événements non récurrents n'est pas nouveau. Kinoshita *et al.* [232] ont utilisé une telle approche de recherche de thèmes latents en vue de classer les comportements des usagers et ont associé une des tendances aux phénomènes inhabituels. Le procédé introduit diffère à l'usage, puisque la perplexité est exploitée et non le résultat direct de classification. Néanmoins, la construction d'un modèle de classification reste requise.
L'objectif étant d'estimer en temps-réel la perplexité à un instant donné t pour une section l, le modèle LDA aboutit naturellement à une classification de tels individus suivant des thèmes latents.
Par la suite, les individus en entrée du modèle sont qualifiés d'individus de trafic et sont notés : $\iota_{l,t}$, où $l \in [1, N]$ et $t \in [1, T]$ avec N le nombre de sections du réseau et T le nombre de pas de temps de la période d'estimation. Leur définition est alors contrainte par le contexte temps-réel comme par la méthode :

- **Contrainte imposée par le LDA :** l'individu à tester est vu comme une distribution sur des éléments unitaires.
- **Contrainte imposée par l'objectif temps-réel :** l'individu doit caractériser l'état de trafic à l'instant t d'évaluation en évitant tout décalage temporel lors de la détection.

Afin de répondre à de telles exigences, chaque individu $\iota_{l,t}$ est construit, pour moitié, à partir de la distribution de vitesses expérimentées sur la période glissante $[t - sw, t]$ (où sw est la longueur de fenêtre) et, pour moitié, depuis une distribution de vitesses typiques du moment de la journée associé à l'instant t. L'introduction de la fenêtre glissante de longueur sw répond aux besoins du LDA, tandis que l'usage de valeurs typiques réduit tout retard temporel en limitant sw. Les valeurs typiques de vitesses sont évaluées à partir de $\mathrm{T}_{typical} = 14$ journées sans incident du mois d'octobre 2013. De plus, afin de construire un même modèle pour toute section l ($l \in [1, N]$), les vitesses sont centrées et réduites à partir de moyennes et écart-types de référence.
Chacune des vitesses centrées-réduites caractérisant l'individu $\iota_{l,t}$ est alors associée à une classe de vitesses centrées-réduites u_j ($j \in [U_{Inf}, U_{Sup}]$). Ces classes de vitesses définissent le cadre commun permettant la comparaison sur toute section du réseau.

Transposée dans le contexte génératif probabiliste, la définition des individus en entrée est vue via l'analogie suivante liant la version originale du LDA et son application :

- **Corpus de documents** : un ensemble d'individus de trafic $\boldsymbol{\iota_{t,l}}$, $\forall t \in [1,T], \forall l \in [1,N]$
- **Document** : une distribution sur l'ensemble des classes de vitesses centrées-réduites caractérisant l'individu de trafic sur la section l à l'instant t : $\boldsymbol{\iota_{l,t}}$. Le jeu de variables aléatoires $\Upsilon_{l,t} = \{v_{l,t-sw}^{real-time}, ..., v_{l,t}^{real-time}, v_{l,t}^{typical,1}, ..., v_{l,t}^{typical,\mathrm{T}_{typical}}\} \in \{u_j, j \in [U_{Inf}, U_{Sup}]\}^{sw+\mathrm{T}_{typical}}$ se compose pour moitié des classes de vitesses centrées-réduites récemment expérimentées par la section l et pour moitié de valeurs typiques de l'instant t. L'individu $\boldsymbol{\iota_{l,t}}$ résulte des proportions $(Card(V_{l,t} = u_{U_{Inf}}), ..., Card(\Upsilon_{l,t} = u_j), ..., Card(\Upsilon_{l,t} = u_{U_{Sup}}))$.
- **Mot** : une classe de vitesse centrée-réduite v. Le vocabulaire associé est l'ensemble de catégories possibles entre deux bornes de vitesses centrées réduites définies avec un pas fixé. Une analyse rapide des valeurs observées conduit à une borne inférieure de $U_{Inf} = -7$ et une borne supérieure de $U_{Sup} = +3$ avec un pas de 0,25.

5.3.1.4 Application du LDA

Avant toute application du LDA, une optimisation de ses paramètres (K_{lda} le nombre de motifs latents et sw la longueur de fenêtre) est requise. Cette dernière est opérée à partir des données du mois de septembre 2013. Aucune hypothèse n'est formulée quant à la présence ou non d'incidents dans l'historique de test.

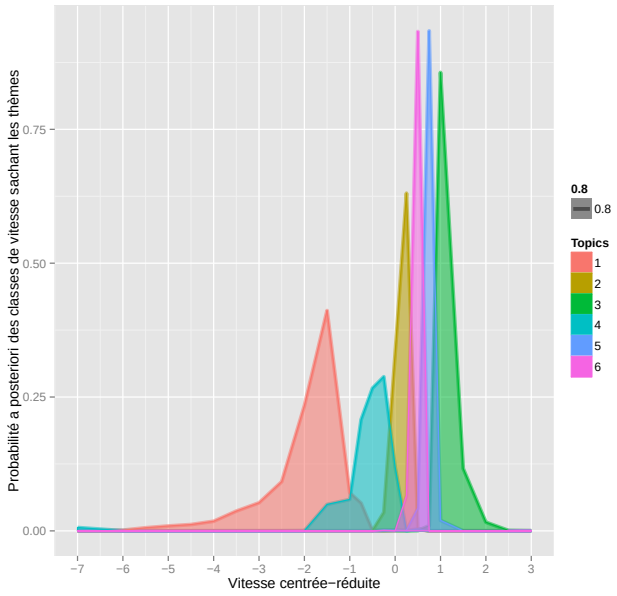


Fig. 5.2.: Les $K_{lda} = 6$ thèmes latents produits lors de la phase d'apprentissage du LDA.

Suite à la construction des individus de trafic, deux paramètres sont à optimiser : la longueur de fenêtre glissante sw et le nombre K_{lda} de thèmes latents identifiés par le LDA. Le procédé d'optimisation usuel fournit une longueur de fenêtre $sw = 5$ pas de temps, soit une latence

de 2h30, et $K_{lda} = 6$. Les 6 thèmes latents sont représentés en Figure 5.2. Il est à noter que chaque motif latent correspond à une catégorie de vitesses : le motif 3 est associé au trafic très fluide, tandis que le motif 1 est identifié aux états très congestionnés.

Ainsi paramétrée, la méthode de classification du LDA est alors appliquée à chacun des individus $\iota_{l,t}$, puis la perplexité de chacun d'eux, notée $p_{l,t}$ est estimée pour chaque section l et pas de temps t. Les résultats de classification issus du procédé LDA ne sont pas exploités en soi.

5.3.1.5 Vérification de la pertinence de l'hypothèse fondamentale

L'hypothèse 1 fondamentale suppose qu'un lien existe et est identifiable entre les phénomènes non récurrents et la perplexité associée aux états de trafic expérimentant de tels événements. Afin de vérifier la pertinence d'une telle supposition, les distributions des perplexités recueillies durant le mois de septembre 2013 pour chaque pas de temps d'une journée ont été estimées pour toute section l du réseau de Cheviré et distinguées au regard de l'identification manuelle des incidents.

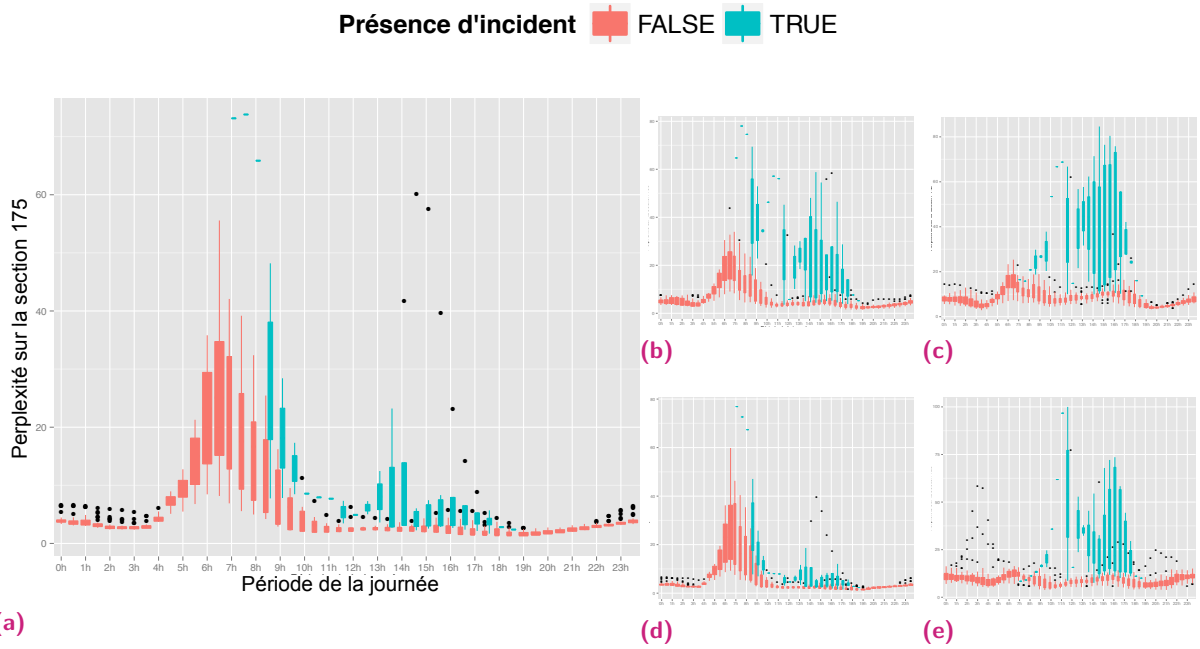


Fig. 5.3.: Représentation des distributions de perplexité en fonction du pas de temps de la journée et de la présence ou non d'incidents : a) sur la section 175 b) sur la section 160 c) sur la section 142 d) sur la section 166 e) sur la section 128 du réseau du Pont de Cheviré.

La Figure 5.3 montre les résultats de quelques sections particulièrement représentatives. Force est de constater que la plupart des événements inhabituels expérimentent effectivement des perplexités plus élevées que celles usuellement observées. Certaines valeurs particulièrement inhabituelles de perplexité sont néanmoins identifées comme valeurs usuelles de trafic. Ces

dernières peuvent être dues à des événements non identifiés lors du traitement manuel d'identification des incidents. Ces représentations sont également l'occasion de commenter le comportement usuel des sections. Les périodes de pic avérées (pic du matin sur la zone de Cheviré) sont effectivement celles expérimentant le plus de transitions d'une catégorie de vitesse à une autre et donc les plus fortes perplexités de la journée.

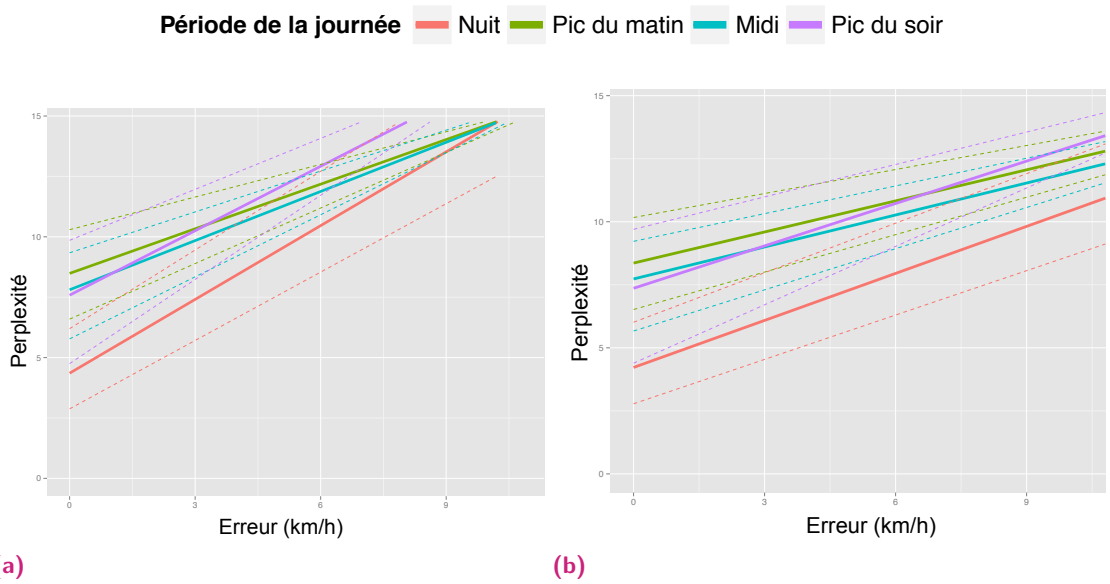


Fig. 5.4.: Courbe de tendances linéaires liant la perplexité à l'erreur de prévision par MKNN : a) Horizon de prévision à 30 min b) Horizon de prévision à 1h30. Les tendances médianes sont présentées en traits pleins, tandis que les premiers et derniers quartiles sont illustrés en pointillés.

Les courbes de tendance représentées sur la Figure 5.4 montrent bien que la perplexité augmente quand l'erreur de prévision croît. L'analyse par période de la journée met en évidence que le phénomène est particulièrement marqué durant la nuit (période usuellement non congestionnée et plus aisée à prédire). La période de forte congestion de la journée (pic du matin) expérimente également une hausse de la perplexité face à des erreurs croissantes de prévision. Ceci confirme que la perplexité permet également de mieux distinguer les périodes de congestion plus forte qu'à l'accoutumée au milieu d'une période de congestion usuelle.

5.3.1.6 Application du filtrage par seuil

La vérification de l'hypothèse étant encourageante, un seuillage est envisagé en vue de séparer automatiquement les phénomènes récurrents des non récurrents via la perplexité. Les seuils sont définis à partir d'une période de test $[1, T_{test}]$ (septembre 2013), puis leur efficacité est vérifiée sur une période de validation (novembre 2013).

5.3.1.6.1 Définition des seuils

La définition des seuils vise à reprendre la comparaison des distributions de perplexités pour chacun des T_{day} pas de temps d'une journée (voir Figure 5.3). Un seuil de perplexité est donc

fixé pour chacun des T_{day} ($T_{day} = 48$ pour un pas de temps de 30 mn) moments d'une journée. Dictée par l'allure des distributions et la nécessité de s'adapter à chacun des pas de temps, l'expression des $N \times T_{day}$ seuils $g_{l,j}$ prend la forme suivante :

$$\forall l \in [1, N], \forall j \in [1, T_{day}], \;\; g_{l,j} = mediane(\{p_{l,k}\}_{k \in T_j}) + a^{lda} \times MAD(\{p_{l,k}\}_{k \in T_j})$$

où l est une section, j est un des T_{day} moments d'une journée, $T_j = \{t \in [1, T_{test}] \, | t \equiv j \pmod{T_{day}}\}$, a^{lda} un paramètre à estimer, MAD une fonction d'expression :
$\forall \{x_1 ... x_n\}, \; MAD(\{x_1 ... x_n\}) = mediane(\{\|x_i - mediane(\{x_1 ... x_n\})\|\}_{i \in [1,n]})$.
Par la suite, la notation $g_{l,t} = g_{l,j}$ où $t \equiv j \pmod{T_{day}}$ sera adoptée pour désigner le seuil associé à l'instant t sur la section l. Dans notre cas d'étude, $T_{day} = 48$ est une période de 24h composée de 48 pas de temps. Le paramètre a^{lda} déterminant la largeur de fenêtre du filtre est paramétré en recherchant l'inflexion de la courbe (théorie du coude) représentant le nombre de détections en fonction du a^{lda} choisi. Par lecture de la figure 5.5, le meilleur a^{lda} est estimé à 2,5 en considérant l'inflexion du nombre d'incidents détectés et celle du nombre d'incidents effectifs.

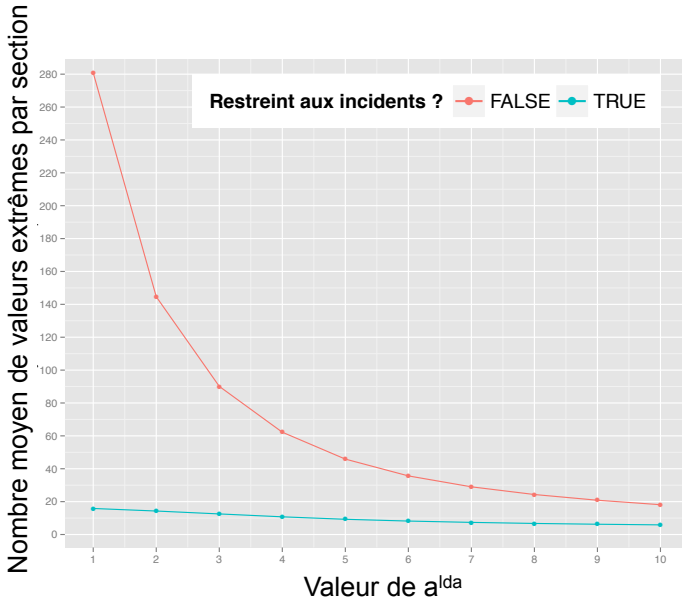


Fig. 5.5.: Illustration du procédé de sélection du paramètre a^{lda} optimal. Le nombre d'incidents détectés en fonction du a^{lda} et le nombre d'incidents réels et détectés sont conjointement représentés.

5.3.1.6.2 Comparaison aux seuils

Les seuils étant définis, à tout instant t et pour chacune des sections l du réseau, la confrontation de la valeur $p_{l,t}$ à celle du seuil $g_{l,t}$ aboutit à la classification de l'état de trafic. Les perplexités supérieures au seuil sont supposées être associées à des incidents. Le résultat du procédé est conservé dans une variable binaire $y^{lda}_{l,t}$ défini par : $y^{lda}_{l,t} = 1$ si et seulement si (ssi) $p_{l,t} > g_{l,t}$, sinon, $y^{lda}_{l,t} = 0$. Les résultats d'une telle classification sont exposés sur quelques sections représentatives en Figure 5.6.

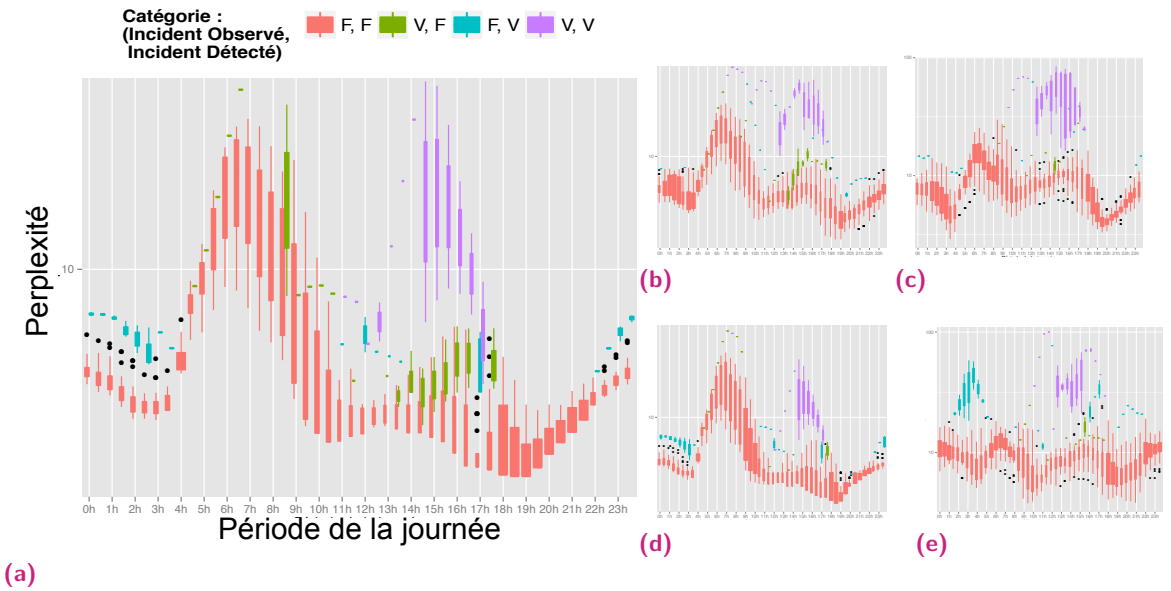


Fig. 5.6.: Représentation des distributions de perplexités en fonction du pas de temps de la journée : a) sur la section 175 b) sur la section 160 c) sur la section 142 d) sur la section 166 e) sur la section 128 du réseau du Pont de Cheviré (novembre 2013). L'identification manuelle d'incidents est confrontée à la détection automatique par LDA par croisement des informations (Table Vrai (V) - Faux (F).

Ces premiers résultats montrent qu'une grande part des incidents significatifs au sens de la perplexité sont bien identifiés. Néanmoins, il apparaît que certains incidents se trouvent être à la limite des seuils fixés et ne sont pas identifiés, bien que visuellement la perplexité soit plus élevée. Une tel phénomène est lié à la quantité d'information disponible pour caractériser les seuils. La substitution d'une méthode non paramétrique par apprentissage à la définition paramétrique actuelle des seuils pourrait améliorer l'identification de tels cas. Il est également à remarquer que certaines périodes nocturnes semblent expérimenter des perplexités plus élevées que la normale, détectées par l'approche automatique du LDA, mais non perçues lors de l'identification manuelle. Du fait, de la faible perplexité usuelle durant les périodes nocturnes, la moindre déviation des états de trafic est particulièrement sensible au procédé de détection automatique, tandis que l'identification manuelle est incapable de les distinguer. Ceci souligne donc la haute sensibilité du filtrage aux périodes usuellement très fluides, mais sa moindre sensibilité durant les périodes congestionnées. Cette remarque est confirmée par l'analyse de la distribution des vitesses en fonction des incidents expérimentés ou détectés proposée en Figure 5.7. Il est plus aisé d'identifier les phases de ralentissement en période fluide qu'en période congestionnée, malgré la transformation initiale des variables d'entrée par le LDA. La variation de vitesse durant ces périodes de congestion récurrente est si importante que l'incident s'y confond avec un état de trafic usuel.

Afin de pleinement évaluer les performances du procédé par LDA, ses performances sur la période de test et la période de validation doivent être comparées à d'autres approches usuelles. En matière de détection d'incidents, les performances sont très dépendantes de la qualité de la

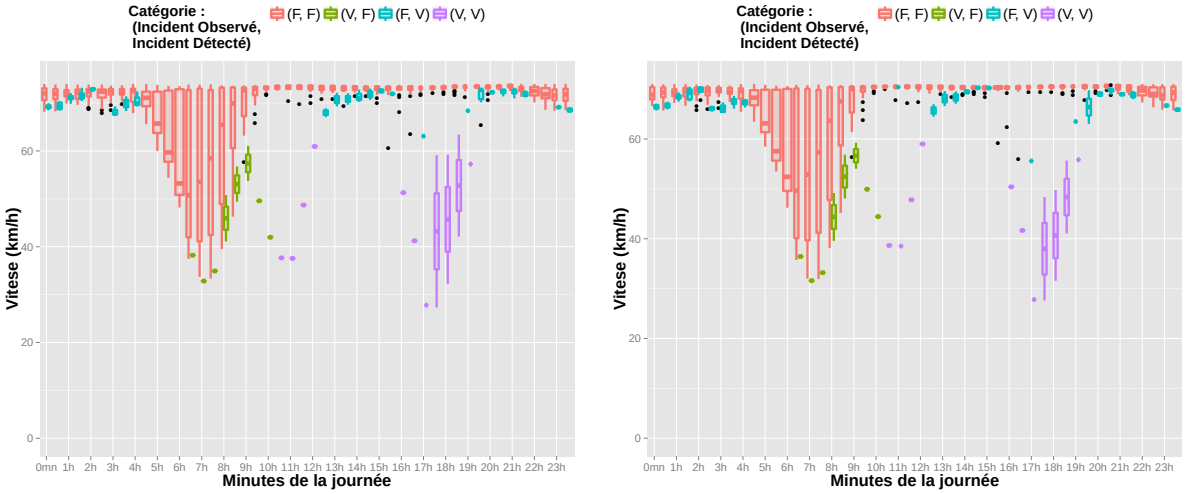


Fig. 5.7.: Représentation des distributions de vitesses en fonction du pas de temps de la journée : a) sur la section 175 b) sur la section 166 du réseau du Pont de Cheviré (novembre 2013). L'identification manuelle d'incidents est confrontée à la détection automatique par LDA par croisement des informations (Table Vrai (V) - Faux (F).

réalité terrain. Or, dans notre cas, cette dernière est estimée manuellement. Il est donc essentiel de replacer ce procédé dans son contexte.

5.3.2 Introduction d'approches alternatives

Trois approches alternatives sont proposées dans cette partie : une évaluation par seuils directement appliquée aux vitesses expérimentées, deux approches de classification supervisée issues du SVM.

5.3.2.1 Méthode de référence pour les approches physiques par seuils.

Introduite pour le compte des approches physiques à seuils, cette méthode intuitive se base sur un filtrage par seuil appliqué aux vitesses expérimentées par les sections. Proche du procédé de filtrage appliqué à la perplexité, seule la valeur testée est distincte. Les seuils sont définis à partir d'une période test $[1, T_{test}]$ (septembre 2013), puis testés sur une période de validation (novembre 2013). Les seuils définis pour chacun des T_{day} pas de temps d'une période d'estimation (correspondant à 24h) ont pour expression :

$$\forall l \in [1, N], \forall j \in [1, T_{day}], \quad g_{l,j}^{mad} = mediane(\{Z_k^l\}_{k \in T_j}) - a^{mad} \times MAD(\{Z_k^l\}_{k \in T_j})$$

où l est une section, j un instant de la période T_{day}, $T_j = \{t \in [1, T] \ | t \equiv j \pmod{T_{day}}\}$, a^{mad} un paramètre à estimer et Z_t^l la vitesse observée sur la section l à l'instant t. La valeur optimale du a^{mad}, obtenue par le même procédé qu'au paragraphe 5.3.1.6, est fixée à 2,75.

La confrontation de toute valeur de vitesse Z_t^l observée à l'instant t sur la section l au seuil associé $g_{l,t}^{mad}$ permet de conclure quant à la récurrence ou non de l'état de trafic. Le résultat du procédé est conservé dans une variable binaire $y_{l,t}^{mad}$ défini par : $y_{l,t}^{mad} = 1$ ssi $Z_t^l > g_{l,t}$, sinon, $y_{l,t}^{mad} = 0$.

5.3.2.2 Méthode de référence pour les approches par classification : une machine à vecteurs supports (SVM).

Les machines à vecteurs supports ont d'ores et déjà été introduites en section 3.1.2. Un retour à la méthode originelle ([108, 107]), classant en 2 catégories les individus introduits en entrée, est proposé pour application aux jeux de données. Le concept de la classification SVM est appliqué à la lettre, puisqu'il s'agit de rechercher un espace de grande dimension permettant de trouver un hyperplan séparant en 2 catégories les données.

Dans notre cas d'étude, les deux classes alternatives sont définies par la récurrence ou non de l'événement considéré. Comme tout procédé d'apprentissage, les paramètres sont, dans un premier temps, appris à partir d'un jeu test (septembre 2013), ce qui aboutit à définir un modèle noté SVM. La phase d'apprentissage est alors opérée sur des individus dont la catégorie est connue. Dans un second temps, le modèle est validé sur un autre jeu d'évaluation (novembre 2013).

Pour chacune des sections $l \in [1, N]$ du réseau à évaluer, un modèle SVM_l lui est associé avec ses propres hyperparamètres $(\epsilon_l, C_l, \gamma_l)$ dans le cas de noyaux gaussiens. Le paramètre ϵ_l contrôle la largeur de marge au sein de laquelle les vecteurs supports sont recherchés. Le coût C_l tend à pénaliser les mauvaises classifications en permettant un léger élargissement de la marge, tandis que γ_l caractérise la largeur de bande du noyau gaussien associé. Chacun des N modèles adopte la forme suivante : $y_{l,t}^{svm} = SVM_l(\boldsymbol{x_{l,t}^{svm}})$, où $\boldsymbol{x_{l,t}^{svm}}$ est l'individu en entrée caractérisant l'état de trafic à l'instant t et $y_{l,t}^{svm}$ est la catégorie binaire associée (1 pour incident, 0 pour événement récurrent).

Pour une section l donnée, un individu $\boldsymbol{x_{l,t}^{svm}}$ à catégoriser est caractérisé par un couple (Z_t^l, j), où Z_t^l est la vitesse à l'instant t d'évaluation et $j|t \equiv j \pmod{T_{day}}$ est l'instant d'évaluation rapporté à la périodicité T_{day}. Il est à remarquer qu'une telle définition met à égalité les différentes approches testées, puisqu'aucune ne dispose de plus d'information qu'une autre.
À tout instant t, l'application du modèle SVM associé à la section l testée aboutit donc à une variable catégorielle binaire $y_{l,t}^{svm}$ caractérisant la présence d'incidents.

5.3.2.3 Méthode alternative : le SVM novelty.

5.3.2.3.1 Cadre théorique

À notre connaissance, l'approche alternative exposée n'a pas encore été appliquée dans le domaine du trafic. Cette méthode proposée par Schölkopf *et al.* [246], un des initiateurs du

SVM, étend les possibilités de ce dernier aux cas où aucun élément de la catégorie alternative n'est disponible. L'algorithme de classification est donc appris à partir d'un historique, dont les individus appartiennent tous à la même classe. Un tel procédé est en mesure d'identifier toute valeur sortant de l'ordinaire ou nouveauté. Nommé *SVM novelty*, le procédé vise à définir un hyperplan séparant en haute dimension l'espace en 2 sous-espaces, dont l'un est associé à l'ensemble des valeurs de l'historique. L'objectif est d'isoler au mieux les données de l'origine et de toute autre valeur. L'expression du modèle SVM reste inchangée à l'exception près que le coût C pénalisant les erreurs au-delà du seuil ϵ est remplacé par un paramètre $\psi \in [0,1]$ caractérisant le taux maximum de valeurs aberrantes (i.e. n'appartenant pas à la classe originelle) et le taux minimum de vecteurs supports. La formulation quadratique à optimiser prend dès lors la forme suivante :

$$\min \quad \frac{1}{2}\|W\|^2 + \frac{1}{\psi}\sum_{i=1}^{T_{test}} \xi_i - b$$

sous les contraintes : $\forall i \in [1, T_{test}], \quad (W^t.\phi(\boldsymbol{x_i})) \geq b - \xi_i$ où $\xi_i \geq 0$ est la variable d'assouplissement des contraintes.

5.3.2.3.2 Application au cas étudié

De même que pour l'approche par SVM, un modèle $SVM^{novelty}$ est associé à chacune des N sections du réseau à évaluer. Pour toute section $l \in [1, N]$, chaque modèle $SVM_l^{novelty}$ est caractérisé par un triplet $(\epsilon_l, \frac{1}{\psi_l}, \gamma_l)$ optimisé à partir d'un jeu de test (septembre 2013, moins les principaux incidents identifiés manuellement).
Pour une section l donnée, un individu $\boldsymbol{x_{l,t}}^{svm}$ à catégoriser est caractérisé par un couple (Z_t^l, j), où Z_t^l est la vitesse à l'instant t d'évaluation et $j \ |t \equiv j \pmod{T_{day}}$ est l'instant d'évaluation rapporté à la périodicité T_{day}. À tout instant t, l'application du modèle SVM associé à la section l testée aboutit donc à une variable catégorielle binaire $y_{l,t}$ caractérisant la présence d'incidents.

5.3.3 Comparatif des méthodes de détection d'incidents

5.3.3.1 Comparaison des performances via des outils usuels

Usuellement, trois indicateurs sont rencontrés dans la littérature de la détection d'incidents : le taux de Détection (DR), le taux de fausses alarmes (FAR) et le taux de bonne classification (GCR). Ces mesures aboutissent à un panel complet des performances de chacun des procédés de détection en fournissant les éléments pour équilibrer entre fausses alertes et détection. Elles sont dépeintes par les expressions suivantes :

$$\begin{cases} DR &= \frac{Nombre\ \ d'individus\ \ classés\ \ comme\ \ incident}{Nombre\ \ total\ \ d'incidents\ \ observés} \times 100\% \\ FAR &= \frac{Nombre\ \ de\ \ fausses\ \ alarmes}{Nombre\ \ total\ \ de\ \ cas\ \ sans\ \ incident} \times 100\% \\ GCR &= \frac{Nombre\ \ d'individus\ \ bien\ \ classés}{Nombre\ \ total\ \ d'individus} \times 100\% \end{cases}$$

Ces indicateurs ont été évalués pour chacune des quatre méthodes introduites précédemment sur la période de test (septembre 2013) et la période de validation (novembre 2013) (voir tableau 5.1). L'estimation sur la période de test fournit une mesure de l'efficacité de la phase d'appentissage ou biais permanent, tandis que la période de validation apporte des informations sur la robustesse de la méthode face à de nouvelles valeurs.

Période	**septembre 2013**				**novembre 2013**			
Méthode	LDA	MAD	SVM	SVMn	LDA	MAD	SVM	SVMn
DR	72,8	68	31,4	35	65	53	50,1	62
FAR	5,5	5,3	0,04	3,6	6,4	12	0,5	5,8
GCR	93,5	93,7	97	93,7	93,2	86,8	98,9	93,8

Tab. 5.1.: Performance des 4 méthodes suivant les indicateurs usuels (en pourcentages). SVMn est noté pour SVM novelty.

L'analyse de la phase de test (septembre 2013) montre que l'apprentissage est difficile pour les méthodes à vecteurs supports. Le SVM classique sous-estime nettement le nombre d'incidents se produisant, ce qui peut-être expliqué par le déséquilibre entre nombre d'incidents observés et nombre total d'individus sur la période considérée. En règle générale, le ratio incident - événement usuel des bases d'apprentissage avoisine les 15% alors qu'il ne dépasse pas 5% dans notre cas [221]. Ce dernier point et l'incertitude quant à la réalité terrain expliquent également que, de façon générale, les performances de détections (DR) soient inférieures aux valeurs observées dans la littérature.
Les résultats de la phase de validation illustrent la robustesse du procédé tiré du LDA et des méthodes à vecteurs supports. Les résultats surprenants du SVM et SVM novelty en matière de taux de détection sont sans nul doute liés au fait que la quantité d'incidents identifiés manuellement durant le mois de novembre 2013 est plus réduite qu'en septembre 2013.
Ces outils usuels fournissent une première comparaison des méthodes, mais restent extrêmement dépendant de la réalité terrain, elle-même approchée. D'autres représentations sont donc présentées par la suite.

5.3.3.2 Comparatif sur l'axe principal de la zone du Pont de Cheviré

Afin de faciliter l'interprétation des résultats et leur représentation, la zone du Pont de Cheviré considérée est réduite à son axe principal et ses branches d'accès. Sur ces 55 sections, les vitesses expérimentées en régime fluide varient entre 70 et 100 km/h suivant les limitations, ce qui garantit la lisibilité des incidents considérés.

La figure 5.8 vise à caractériser la capacité de détection d'incidents distincts. À cette fin, le nombre d'incidents détectés est réprésenté section par section, puis par journée à incidents sur le mois de novembre 2013. Dans ce cadre, la méthode basée sur la MAD s'avère être la plus performante. Elle est, en effet, en mesure de détecter, à un instant au moins, l'ensemble des incidents expérimentés par le réseau. La méthode SVM est, de manière générale, moins performante que LDA et SVM novelty. Globalement, les méthodes sont aptes à détecter la

plupart des sections impactées par un incident, mais des disparités existent d'une section à une autre.

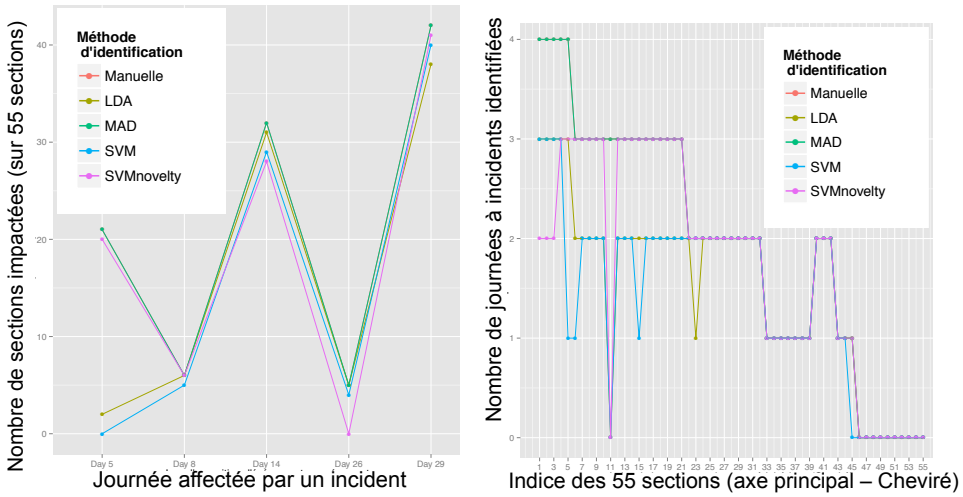


Fig. 5.8.: Comparaison des performances d'identification d'incidents distincts : a) en nombre de sections effectivement détectées suivant les jours b) en nombre de jours détectés par section durant le mois de validation (novembre 2013).

Les performances de chacune des méthodes et la réalité terrain sont représentées via les diagrammes espace-temps [1] en Figure 5.9 pour deux journées. Les jours considérés sont des journées actives caractérisées par un pic de circulation en début de matinée. Pour chacun des cas exposés, un second pic apparaît et est naturellement associé à un événement non récurrent.

Force est de constater que les méthodes à vecteurs supports et à LDA paraissent plus performantes que l'approche basée sur la MAD. En effet, les valeurs détectées sont moins disséminées (effet de grésillement) et le nombre de fausses détections est clairement plus limité. La comparaison entre approche par MAD et LDA est extrêmement intéressante, dans la mesure où l'approche par LDA est une extension de la MAD avec changement des variables d'entrée. Au lieu de juger directement depuis les caractéristiques de trafic, la détection est opérée après modification de ces dernières via la perplexité du LDA. La vérification de l'hypothèse fondamentale montrait une très bonne caractérisation des phénomènes non récurrents par la perplexité. Cette tendance est confirmée par la Figure 5.9, puisque l'effet de grésillement est amoindri. Les comportements sortant de l'ordinaire occasionnent donc une hausse de l'entropie de classification du LDA, plus facilement repérée qu'à partir des vitesses brutes.
En outre, le LDA et le SVM novelty sont en mesure d'identifier des périodes non récurrentes omises lors de l'analyse manuelle comme l'illustre la congestion matinale du 29 novembre (entre 3 et 5km, vers minuit-3h du matin).

1. Les diagrammes espace-temps constituent une représentation des trajectoires de véhicules sur une période donnée. La couleur correspond à la vitesse expérimentée à un instant donné en un point du réseau.

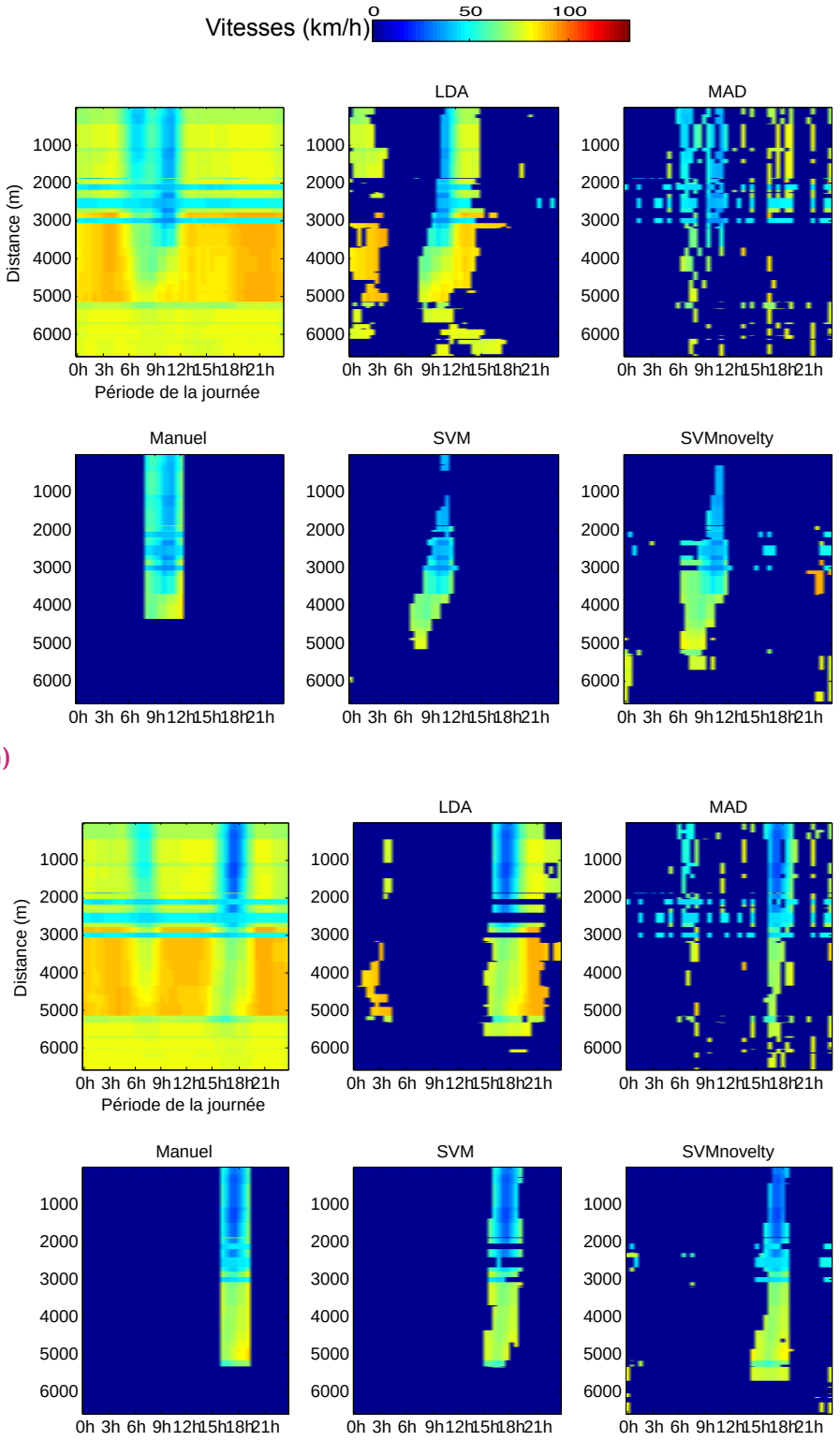


Fig. 5.9.: Comparaison des méthodes à la réalité terrain à partir des diagrammes espace-temps sur la zone réduite de Cheviré : a) durant le jeudi 14 novembre 2013 b) durant le vendredi 29 novembre 2013.

En revanche, la fenêtre glissante de l'approche LDA affecte la largeur de la fenêtre de détection. Cette dernière est plus importante que celle des méthodes alternatives. Néanmoins, ceci reste peu pénalisant en réalité, puisque cela maintient la période d'alerte sur la queue de congestion et ne semble pas retarder outre mesure la date de détection.

Face aux performances satisfaisantes de l'approche LDA, cette dernière est exploitée, par la suite, et intégrée au cadre prédictif pour amélioration des performances de prévision.

5.4 Phase de validation et d'intégration temps-réel des incidents

Cette seconde phase envisage l'intégration du procédé de détection au sein du cadre prédictif appliqué via la méthode MKNN. L'objectif avéré est de parvenir à un cadre prédictif plus robuste, d'une part, et apte à s'améliorer, d'autre part.

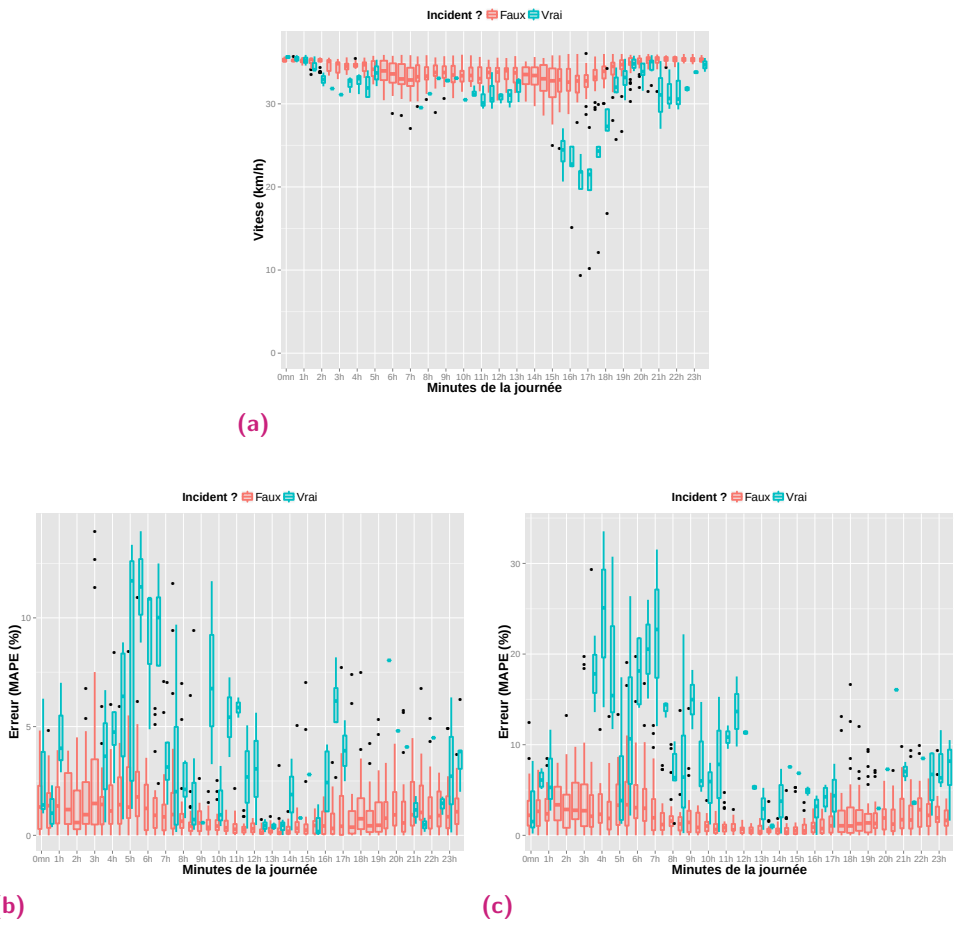


Fig. 5.10.: Représentations en fonction du pas de temps de la journée sur la section 105 du Pont de Cheviré : a) Distribution de vitesses b) Distribution d'erreurs de prévision à 30 minutes c) Distribution d'erreurs de prévision à 90 minutes.

Ainsi qu'il est mis en évidence par la Figure 5.10 sur une section du réseau du Pont de Cheviré, les congestions non récurrentes occasionnent des pertes de performances sur toute la durée du phénomène. Le pourcentage d'erreur peut ainsi se voir doubler ou tripler lors de tels événements. L'apparition de phénomènes non récurrents est souvent due à des facteurs externes et difficiles à anticiper. En revanche, mesurer l'impact de toute congestion non récurrente et son évolution est envisageable et recherché dans cette partie.

5.4.1 Procédé d'adaptation du cadre prédictif pour intégration des incidents

Illustrée en Figure 5.1 et par l'algorithme 3, la solution adoptée pour mieux prendre en considération les incidents consiste en l'ajout d'un historique permanent d'événements non récurrents (caractérisé par $\boldsymbol{Bool}_{IncidentHist}$) se greffant à l'historique glissant originel (ligne 28). La méthode de prévision est supposée parfaitement en mesure de sélectionner, parmi les deux historiques, les individus pertinents pour reconstruire l'instant évalué, puis prédire l'individu futur. L'intégration ou non d'incidents pour prédire l'instant $t + h$ repose alors pleinement sur la méthode de prévision. Ceci constitue la première phase de l'algorithme 3. Ainsi, l'historique considéré pour apprentissage du modèle et prévision intègre désormais deux composantes : une composante propre à définir les tendances régulières et saisonnières du trafic grâce à son historique glissant et une composante assurant une bonne variabilité des dynamiques du réseau par l'intégration des phénomènes les moins communs.

La seconde phase consiste en une mise à jour de l'historique d'événements non récurrents. À cette fin, le processus intervient quasiment en temps-réel avec une latence due à l'horizon de prévision h.
Tout instant expérimentant un événement non récurrent préalablement identifié (fonction $IncidentDetection$) constitue un candidat potentiel à l'intégration au sein de l'historique permanent d'incidents (ligne 15). Les détections isolées (phénomène de grésillement observé en Figure 5.9) sont évitées en ne déclenchant le processus d'évaluation que si plus de $\alpha^{incident}$ pourcent des sections sont concernées (ligne 16). Enfin, dans le but de limiter la taille de l'historique permanent et donc la complexité temporelle du modèle prédictif, les itérations d'incidents proches de par leurs impacts ou à faible impact sur le réseau doivent être évitées (ligne 18). Seuls les incidents véritablement distincts de ceux de l'historique au sens de l'erreur de prévision sont intégrés. Cette stratégie vise à garantir la résilience à venir.

Les principales adaptations du cadre prédictif sont mises en évidence en vert dans l'algorithme 3.

Algorithme 3 Intégration des incidents au cadre prédictif originel.

▷ INITIALISATION (VALEURS CONNUES) :
Soit $\boldsymbol{Hyperparameters}$ Vecteur de paramètres propres à la méthode de prévision
Soit h Valeur de l'horizon de prévision
Soit $\underline{\boldsymbol{XHist_{t-h}}}$ Historique d'individus observés (jusqu'au pas de temps $t-h-1$ (temps-réel)) - Format d'entrée : $t-h-1 \times Card(\boldsymbol{X})$
Soit $\underline{\boldsymbol{YHist_{t}}}$ Historique d'individus observés (jusqu'au pas de temps $t-1$ (temps-réel)) - Format de sortie : $t-h-1 \times Card(\boldsymbol{Y})$
Soit $\boldsymbol{Y_{t-h}}$ État de trafic mesuré à l'instant $t-h$
Soit $\widehat{\boldsymbol{Y_t}}$ Résultat de prévision des états de trafic depuis l'instant $t-h$
Soit $\boldsymbol{X_{t-h}}$ Individu à l'instant t-h
Soit $\boldsymbol{Bool_{IncidentHist}}$ Booléen de l'historique d'incidents résultant du procédé à $t-1$
▷ NOUVELLE MESURE À L'INSTANT t
$\boldsymbol{Y_t} \leftarrow$ Mesure(t)
▷ ESTIMATION DE L'ERREUR DE PRÉVISION DEPUIS L'INSTANT $t-h$:
$Error^{t}_{t-h} \leftarrow$ Error$(\boldsymbol{Y_t}, \widehat{\boldsymbol{Y_t}})$ Structure recueillant les diverses formes d'erreurs
▷ DÉTECTION DES INCIDENTS À L'INSTANT $t-h$ ET BASCULEMENT À L'ÉCHELLE GLOBALE
$\boldsymbol{BoolIsIncident_{t-h}} \leftarrow$ IncidentDetection $(\boldsymbol{Y_{t-h}}, t-h, \underline{\boldsymbol{YHist_t}})$ Vecteur booléen de taille N
$IsIncident^{Global}_{t-h} \leftarrow (Sum(\boldsymbol{BoolIsIncident_{t-h}}) > \alpha^{incident} \times N)$
▷ IDENTIFICATION DES INCIDENTS À FORT IMPACT PRÉDICTIF DE L'INSTANT T-H
$\boldsymbol{Bool_{IncidentHist}}(t-h) \leftarrow (IsIncident^{Global}_{t-h}\ \&\ Error^{t}_{t-h} > Bound_{\beta^{incident}}(t))$
▷ MISE À JOUR DES HISTORIQUES :
$\underline{\boldsymbol{YHist_t}} \leftarrow \underline{\boldsymbol{YHist_t}} \cup \boldsymbol{Y_t}$
$\underline{\boldsymbol{XHist_{t-h}}} \leftarrow \underline{\boldsymbol{XHist_{t-h}}} \cup \boldsymbol{X_{t-h}}$
▷ CRÉATION DE L'INDIVIDU D'ENTRÉE
$\boldsymbol{X_t} \leftarrow$ BuildX$(\boldsymbol{Y_t}, \underline{\boldsymbol{YHist_t}})$
▷ GESTION DES DONNÉES MANQUANTES ET ADAPTATION DE L'HISTORIQUE :
$\boldsymbol{Bool_{NotMissing}} \leftarrow$ IsNotMissing$(\underline{\boldsymbol{XHist_{t-h}}})$
$\boldsymbol{Bool_{InHist}} \leftarrow$ RestrictHist$(\underline{\boldsymbol{XHist_{t-h}}})$
$\underline{\boldsymbol{CutHistX}} \leftarrow \underline{\boldsymbol{XHist_{t-h}}}(\boldsymbol{Bool_{NotMissing}}\ \&\ (\boldsymbol{Bool_{InHist}}|\boldsymbol{Bool_{IncidentHist}}), :)$
$\underline{\boldsymbol{CutHistY}} \leftarrow \underline{\boldsymbol{YHist_t}}(\boldsymbol{Bool_{NotMissing}}\ \&\ (\boldsymbol{Bool_{InHist}}|\boldsymbol{Bool_{IncidentHist}}), :)$
▷ PRÉVISION :
$\widehat{\boldsymbol{Y_{t+h}}} \leftarrow$ Prediction $(\boldsymbol{X_t}, \underline{\boldsymbol{CutHistX}}, \underline{\boldsymbol{CutHistY}}, \boldsymbol{Hyperparameters})$
▷ SORTIES :
$\boldsymbol{Error^{t}_{t-h}}$; $\widehat{\boldsymbol{Y_{t+h}}}$; $\boldsymbol{Bool_{IncidentHist}}$

La première phase hors-ligne d'ajout d'un historique permanent d'incidents ne pose pas de réelle diffculté et se résume à l'ajout d'une condition ($\boldsymbol{Bool_{IncidentHist}}$) sur la création de l'historique courant. Les principales difficultés apparaissent lors de la mise à jour restrictive de

l'historique d'incidents et de la condition associée.
Les enjeux soulevés et traités par l'algorithme 3 relèvent de divers ordres :

- Adapter la notion locale d'incidents fournie par le procédé d'identification aux objectifs globaux de prévision à l'échelle du réseau (ligne 16) ;
- Éviter les itérations d'information au sein de l'historique permanent d'incidents. Cet aspect relève, en réalité, d'un enjeu de qualification de la significativité d'impact de l'incident sur la prévision (ligne 18).

5.4.2 Passage à l'échelle du réseau de la détection locale d'incidents

Les procédés de détection introduits en section 5.3 (appliqués ligne 15) sont en mesure de détecter l'apparition de phénomènes inhabituels pour chacune des sections indépendamment les unes des autres. Un tel procédé ne prend pas en considération la présence ou non d'incidents au voisinage spatio-temporel. L'apparition d'événements disséminés aussi bien spatialement que temporellement, comme observés pour l'approche par MAD (Figure 5.9), en est une conséquence directe conduisant à de fausses alertes.
Afin de pallier cette contrainte, Rakha *et al.* ont étendu leur procédé d'identification de congestions à un fonctionnement par zones spatio-temporelles [247]. Outre l'aspect fausse alerte, l'impact sur le réseau d'un événement isolé est généralement très limité. Jin *et al.* [223] ont développé un procédé de détection d'incidents n'activant l'alerte qu'à partir d'un certain seuil de sections concernées. Le degré d'alerte est alors accentué par paliers suivant le nombre de sections impactées.
Dans cette même veine, une approche simple est mise en oeuvre dans le but de n'activer le module d'intégration des incidents que si le réseau subit une alerte suffisante et susceptible de le perturber. À cette fin, un pourcentage $\alpha^{incident}$ de sections identifiées comme victimes d'incidents à l'instant t d'évaluation est fixé. L'intégration d'incidents à l'historique n'est activée qu'à la condition que le pourcentage de sections impactées soit supérieur à $\alpha^{incident}$. Cette condition est traduite par la ligne 16 de l'algorithme 3. Le paramètre $\alpha^{incident}$ est déterminé expérimentalement par une analyse de l'évolution du nombre d'incidents détectés manuellement en fonction du pourcentage $\alpha^{incident}$ de réseau impacté en septembre 2013. Le meilleur compromis entre limitation du nombre d'incidents éliminés et pourcentage du réseau élevé est trouvé pour $\alpha^{incident} = 7\%$, taux auquel aucun incident de la réalité terrain n'est supprimé.

5.4.3 Limitation des itérations d'informations au sein de l'historique d'incidents

Comme il est illusoire d'envisager de prévoir un incident, seul l'impact des incidents est pertinent à évaluer. De plus, seuls les incidents à fort impact prédictif sont utiles aux modèles de prévision, car garants d'une bonne variabilité des dynamiques de trafic. Cette identification des événements à fort impact prédictif constitue la clé de voûte de la mise à jour restrictive et

en temps-réel de l'historique permanent d'incidents. L'idée principale consiste à s'appuyer sur l'erreur de prévision afin de qualifier l'impact prédictif des incidents, préalablement identifiés par application du procédé de détection. Seuls les événements à fort impact prédictif sont alors intégrés à l'historique permanent.

5.4.3.1 L'erreur de prévision comme qualificatif d'impact

La phase de mise à jour se perçoit comme un nouveau procédé de détection des incidents basé sur l'erreur de prévision ($Error^t_{t-h}$) et venant se superposer au procédé de détection originel. Une telle approche appuyant sa détection sur une erreur de prévision est observable chez Wang *et al.* [248]. L'erreur de prévision, produite depuis une méthode paramétrique, y est assimilée à une mesure de l'écart entre l'état usuel (fourni par la méthode paramétrique) et l'état réellement observé.
Contrairement, à la méthode paramétrique, qui fournit un état usuel, notre approche non paramétrique produit le meilleur état à partir de son historique d'apprentissage. Pour toute erreur de prévision élevée, ni l'historique glissant usuel, ni l'historique permanent d'incidents ne sont en mesure d'expliquer convenablement la variabilité de l'état prédit. Un manque est dès lors identifié parmi les historiques et doit être comblé en intégrant l'événement à l'historique d'incidents.

5.4.3.2 Caractérisation formelle des incidents à fort impact prédictif

L'objectif est de parvenir à identifier les incidents détectés ayant un plus fort impact prédictif que les états de trafic usuellement observés en conditions similaires. La caractérisation des conditions habituelles d'erreur est opérée à partir d'un facteur de périodicité T_{day} de 24h correspondant à 48 pas de temps. La catégorisation est obtenue à partir de seuils définis pour chacun des pas de temps j à partir d'une période de paramétrisation $[1, T_{test}]$ (septembre 2013) :

$$\forall j \in [1, T_{day}],\ Bound_t(j) = median(\{Error^t_{t-h}(k)\}_{k \in T_j}) + \beta^{incident} \times MAD(\{Error^t_{t-h}(k)\}_{k \in T_j})$$

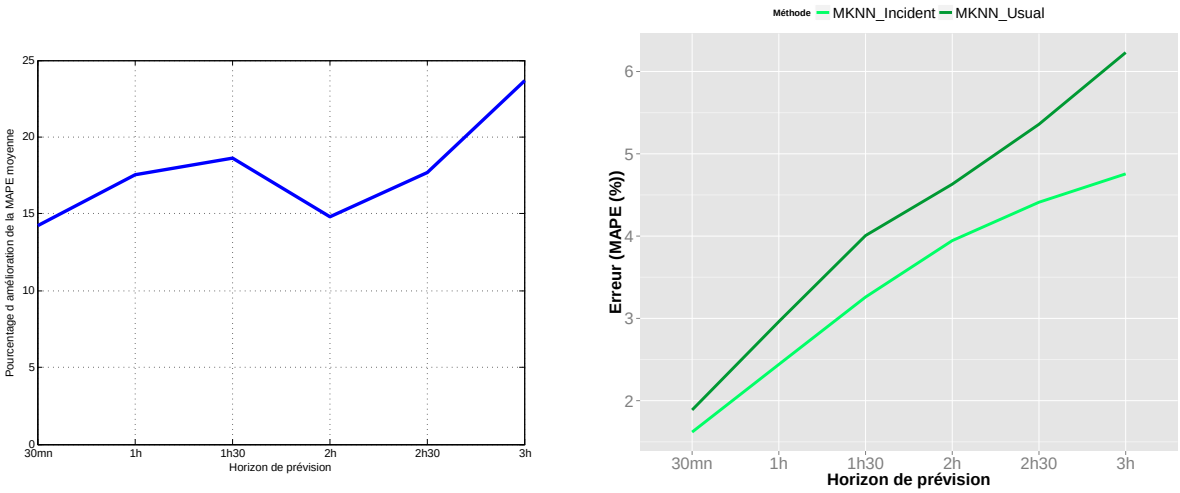
où $T_j = \{t \in [1, T], t \equiv j \pmod{T_{day}}\}$, $\beta^{incident}$ un paramètre à estimer et E^t_{t-h} l'erreur de prévision par pas de temps. Par extension, la fonction est définie à tout instant t par $Bound_t(t) = Bound_t(j)$ où $t \equiv j \pmod{T_{day}}$. Le choix du paramètre $\beta^{incident}$ est obtenu expérimentalement et fixé à 150%. Cette valeur correspond approximativement aux valeurs usuelles de filtrage par MAD et est confirmée par analyse de la variation du nombre d'incidents de la réalité terrain conservés en fonction du paramètre du filtre.

Le procédé de catégorisation se résume ainsi : tout incident se produisant à un instant t, dont l'erreur de prévision à l'horizon $t + h$ est supérieure à $Bound_t(t + h)$, est considéré à fort impact prédictif et ajouté à l'historique permanent d'incidents (voir ligne 18). Les effets d'un tel incident ne sont pas prévisibles de façon satisfaisante à partir des seules informations tirées des

deux historiques. L'ajout de l'individu associé à l'historique permanent est donc indispensable en vue d'améliorer les performances ultérieures de prévision.

5.4.4 Evaluation des performances liées à l'intégration d'un historique permanent d'incidents

L'algorithme 3 est mis en application à partir d'un historique glissant de 30 jours et d'un historique permanent d'incidents initialisé à partir des événements non-récurrents de la vérité terrain sur septembre et octobre 2013. La mise en oeuvre est évaluée sur le mois de novembre 2013 à partir des incidents identifiés en temps-réel par l'approche LDA. La prévision, elle-même, est assurée par l'approche globale MKNN.



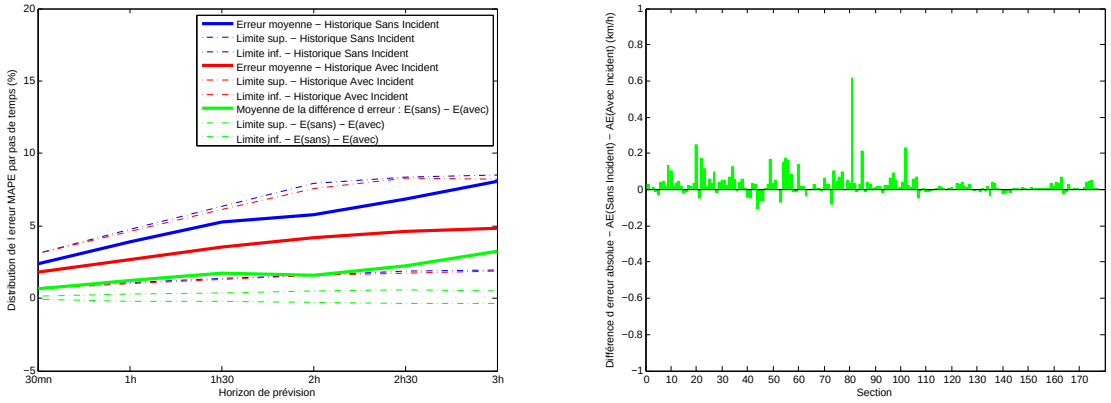
(a) (b)

Fig. 5.11.: Représentations des performances globales de prévision du MKNN suivant l'horizon de prévision et l'intégration ($MKNN_{Incident}$) ou non ($MKNN_{Usual}$) de l'historique d'incidents : a) Pourcentage d'amélioration de l'erreur globale ($MAPE_{global}$) dû à l'insertion d'un historique d'incidents b) Erreur globale ($MAPE_{global}$ en fonction de l'horizon de prévision.

L'effet de l'intégration d'un historique d'incidents se ressent spécifiquement sur les indicateurs d'erreur les plus globaux ($MAPE_{global}$). La Figure 5.11 montre que, quel que soit l'horizon de prévision, l'erreur globale est *a minima* améliorée de 15%. Il est à noter que la réduction de l'erreur observée n'est que peu liée au simple fait d'étendre l'historique. L'analyse portant sur la longueur de fenêtre optimale de l'historique (Figure 2.8, chapitre 2) a effectivement montré qu'au-delà de 30 jours, l'amélioration n'excède pas 2,5% tous les 20 jours supplémentaires.

L'analyse comparant les erreurs sur les périodes non récurrentes à partir de prévision avec & sans historique d'incidents confirme la tendance globale de réduction de l'erreur. La Figure 5.12 a) montre non seulement que la réduction d'erreur due à l'intégration de l'historique d'incident croît avec l'horizon de prévison, mais que seul un nombre restreint et ciblé d'incidents est

amélioré. Au regard du profil moyen, l'allure des premiers et derniers quartiles montre clairement que seules les erreurs conséquentes sont mieux prévues. Un tel ciblage des événements traités concorde avec l'effet attendu de l'intégration d'un historique d'incidents. Seules les situations similaires à celles de l'historique sont mieux prévues, d'où l'intérêt de bien cibler les phénomènes à intégrer. Les résultats actuels témoignent de l'insuffisance d'informations dans l'historique d'incidents pour traiter la grande variabilité d'événements, mais montrent d'ores et déjà une forte réduction d'erreur sur certains événements localisés.



(a) **(b)**

Fig. 5.12.: Ciblage de l'analyse sur les performances de prévision durant les périodes recueillant des phénomènes non récurrents : a) Distribution de l'erreur par pas de temps ($MAPE_{TimeStep}$) sur les périodes à incidents en fonction de l'horizon et du type de prévision (avec ou sans insertion des incidents) & Distribution de la différence d'erreur entre les 2 situations b) Différence d'erreur absolue entre situation sans historique d'incidents et avec historique permanent pour chacune des sections du réseau (Horizon 30 min)

L'analyse spatiale de la différence d'erreur par section ($MAPE_{section}$) produite en Figure 5.12 b) met en évidence que l'ajout de l'historique permanent d'incidents est bénéfique à la grande majorité des sections du réseau lors de phénomènes non récurrents. Seules 10 sections font exception à l'horizon 30 minutes. Ces 10 sections semblent être victimes de l'extension de l'historique, qui, dans leur cas, apporte trop de variabilité. Le rapport amélioration / perte reste tout de même largement en faveur de l'introduction d'un historique permanent d'incidents.

Enfin, la Figure 5.13 s'intéresse à mesurer l'effet de l'ajout d'un historique d'incidents sur la quantité d'incidents à fort impact prédictif identifiés. Cette figure résume la confrontation du taux d'identification des incidents à fort impact prédictif depuis un modèle avec et sans historique d'incidents. Cette représentation est particulièrement intéressante, dans la mesure où elle permet d'évaluer le véritable apport de l'intégration d'un historique d'incident sur la variabilité des comportements retenues par l'historique. À court-terme, l'intégration de l'historique d'incident permet de mieux prévoir 10% des événements non récurrents. La détérioration ou instabilité des performances avec l'horizon de prévision est probablement due aux divergences d'impact sur le long terme de phénomènes originellement similaires.

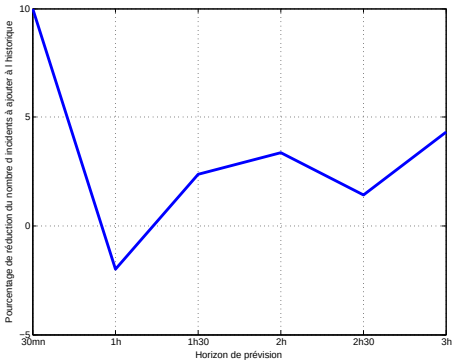


Fig. 5.13.: Pourcentage de réduction du nombre d'incidents à intégrer à l'historique permanent d'incidents (à fort impact prédictif) entre prévision avec & sans historique d'incidents.

5.5 Perspectives et discussion

▶ Synthèse

Cette section fait l'objet d'un double objectif : introduire et tester de nouvelles stratégies de détection d'incidents, puis intégrer ces stratégies au sein du coeur même du procédé de prévision, en vue d'aboutir à une meilleure robustesse de la prévision face à des phénomènes peu communs.

La nouvelle stratégie de détection introduite consiste en une transformation de l'espace d'entrée de l'algorithme de détection usuellement centré sur les états de trafic. La modification de l'espace, basée sur un procédé de classification non supervisée, aboutit à un nouvel indicateur permettant de bien discriminer événements non récurrents et tendances usuelles. Les performances du nouveau procédé sont confrontées à des méthodes classiques et montrent un bon comportement de ce dernier.

Suite aux travaux d'identification d'incidents, une stratégie d'extension de l'historique usuel est proposée. L'objectif est de créer un historique permanent d'incidents garantissant une plus grande variabilité des comportements saisis, puis de mettre à jour cet historique à partir des phénomènes non récurrents, préalablement identifiés, occasionnant encore des erreurs conséquentes. L'intérêt de l'intégration d'un tel historique est confirmé sur le cas d'étude du Pont de Cheviré.

▶ Limites

Le procédé d'identification d'incidents introduit et basé sur le LDA déplore quelques limites : la détection tend à s'étendre dans le temps, du fait de l'usage de fenêtres temporelles, et la saisonnalité des données n'est pas prise en considération. Un recalibrage régulier (tous les mois, par exemple) s'avère donc nécessaire afin de conserver des résultats pertinents.

Concernant le processus d'extension de l'historique, les principaux verrous résident dans l'affinement des paramètres de passage à l'échelle globale de la notion d'incident et ceux de caractérisation de l'impact prédictif. De plus, une étude plus approfondie du potentiel apporté

par la mise à jour de l'historique d'incident permettrait d'entériner le concept introduit. La vérification de la réduction du nombre d'incidents introduits dans l'historique reste à confirmer à plus grande échelle temporelle. Les principaux effets à long-terme devraient être observés à partir de bases de données s'étalant sur six mois ou un an.

▶ Perspectives

En ce qui concerne les procédés d'identification d'incidents, aucune réalité terrain n'est disponible pour confrontation des performances de détection des divers procédés exposés. Néanmoins une identification manuelle a permis de définir une référence terrain. Les résultats exposés mériteraient donc d'être ré-évalués au regard d'une véritable réalité terrain et, éventuellement, à partir d'une base d'incidents plus conséquente pour la méthode SVM.
La stratégie de détection introduite se décompose en deux phases indépenantes : une transformation de l'individu à classer, puis l'application d'un procédé de détection, filtrage à seuil dans notre cas d'étude. Tout autre procédé de détection pourrait prendre la place du procédé actuel basé sur des seuils. En outre, il serait également intéressant d'évaluer d'autres transformations des individus à classer, notamment via la décomposition en ondelettes.

Enfin, l'intégration de la détection d'incidents est présentée depuis la méthode MKNN, mais pourrait se voir adaptée sans modification à une régression non paramétrique à noyaux. L'adaptation à l'approche MSVR requiert une ré-évaluation de ses vecteurs supports uniquement lors de l'ajout d'éléments à l'historique d'incidents. Une telle proposition d'adaptation s'inspire de la version en ligne des régressions à vecteurs support proposée par [115], montrant avec succès le potentiel des méthodes SVR à s'adapter en ligne face à des événements peu récurrents.
Un ultime point à approfondir réside dans l'analyse de l'historique d'incidents et, plus spécifiquement, des incidents intégrés à l'historique. Une étude de propagation d'impact de ces incidents permettrait d'identifier des typologies de scenarii et de modéliser les mécanismes en jeu.

6

Vers une intégration de l'effet des conditions météorologiques dans le cadre prédictif

> *Il y a des gens qui pourraient parler des heures pour ne rien dire, ce sont ceux qui dissertent sur la météo, et qui seraient capables de faire une thèse sur un nuage, tout ça pour ne pas avoir à rentrer chez eux.*
>
> **— David Fœnkinos**
> Les souvenirs de David Fœnkinos

Abstract :
This last chapter is an attempt to integrate the weather impact to the global forecasting framework. The collection of road specific weather information leads to the first analysis with reduced bias due to spatial or temporal lags between traffic and weather measurements. Some statistical and expert local analyses are introduced to assess the relationship between adverse weather and flow speed reduction or error increase. Even if a wide range of weather factors is available, the main effects are statistically highlighted according to the rain intensity, which is the most recurrent adverse weather phenomenon. Some local assessments point out an effect of the rainy periods on the flow speed as a reduction of 5 km/h is observed by comparing similar days under various weather conditions. The statistical quantification of the weather impact suffers from a lack of adverse weather situations, but some integration process to the forecasting framework are expected to be promising with a wider range of data.

Après avoir analysé la possibilité de mieux de répondre aux perturbations induites par l'apparition d'événements non récurrents, cet ultime chapitre envisage ces phénomènes sous un tout autre point de vue en considérant leurs causes. Les phénomènes non récurrents identifiés et intégrés au modèle de prévision au cours du **chapitre** 5 peuvent en effet connaître diverses origines que ce soit un accident survenu sur le bord de route, une manifestation, des travaux ou encore des phénomènes naturels tels que des conditions météorologiques dégradées. Ce changement de perspective est illustré par le schéma présenté en Figure 6.1. Tandis que l'impact sur le fonctionnement nominal du réseau est pris en considération dans le **chapitre** 5, l'influence de facteurs externes sur l'apparition de telles congestions non récurrentes est évaluée dans ce chapitre. Une cause spécifique de perturbations récurrente dans la littérature du trafic est considérée : les conditions météorologiques dégradées.

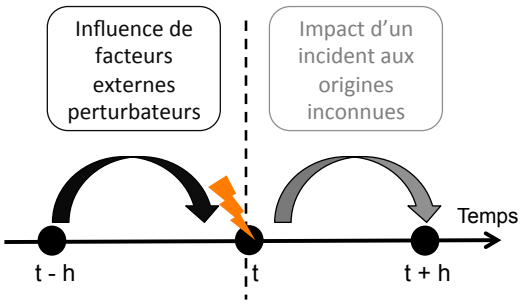
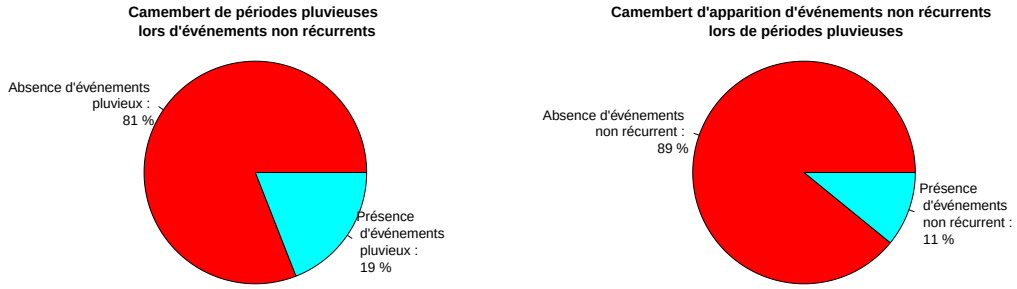


Fig. 6.1.: Transition de l'analyse d'impact à l'intégration de facteurs exogènes.

Comme l'illustre la Figure 6.2, pratiquement un quart des incidents se produisent par temps pluvieux et un sixième des jours pluvieux expérimentent des événements non récurrents. Bien que de telles proportions n'indiquent rien quant aux éventuelles corrélations existant entre apparition d'incidents et conditions météorologiques dégradées, elles montrent néanmoins que l'une et l'autre coexistent.



(a) **(b)**

Fig. 6.2.: Camemberts illustrant l'interaction entre météorologie dégradée et phénomènes non récurrents sur le mois de septembre 2013. a) Occurence de pluies durant des événements non récurrents b) Occurence d'événements non récurrents durant les périodes pluvieuses.

Ces résultats ne manquent pas de rappeler les expériences d'embouteillages de tout un chacun face à des chutes de neige ou de gros orages sur les réseaux routiers. Des conditions météorologiques dégradées sont susceptibles d'entraîner une réduction de la visibilité, de l'adhérence et une détérioration de la chaussée, puis, par extension, une diminution des vitesses expérimentées. L'effet de la météorologie sur les variables macroscopiques de trafic a été étudié dès les années 50 [249] et montre généralement un impact non négligeable sur les comportements au volant.
Les informations météorologiques en certaines sections de route ont été relevées sur le réseau du Pont de Cheviré à Nantes. Contrairement aux informations usuellement prélevées dans les aéroports à proximité, les données recueillies sont extrêmement précises car évaluées sur le terrain par le service route de MétéoFrance. L'analyse des effets de conditions dégradées sur les caractéristiques de trafic (vitesses tirées de FCD) constituent la première phase de cette étude. La seconde étape consiste à analyser le comportement de l'erreur de prévision face aux composantes météorologiques pour conclure quant à leur potentielle intégration dans la structure du cadre prédictif.

6.1 La météorologie, un facteur exogène récurrent de la littérature du trafic

Outre une légère modification des modes et demandes en déplacement, les intempéries imposent des contraintes conséquentes aux usagers des infrastructures de transport en réduisant la visibilité et l'adhérence sur les routes [250]. Il n'est pas rare que des accidents soient la conséquence directe de conditions dégradées. Il n'est donc pas étonnant que les premières études de l'impact de la météorologie sur le trafic remontent à l'époque de la popularisation des automobiles dans les années 50 [249]. Depuis, la littérature sur l'estimation des effets météorologiques sur le trafic a été bien alimentée, mais la faisabilité de l'intégration de tels phénomènes aux méthodes de prévision reste encore peu étudiée.

6.1.1 Estimation des effets météorologiques sur le trafic

Suite aux premières analyses de Tanner au début des années 50 [249], l'impact de la météorologie sur le trafic a connu un regain d'intérêt dans les années 70 avec Jones *et al.* [251] conduisant par la suite à une littérature conséquente [252, 253]. Bien que la plupart des études menées concluent à un impact de conditions météorologiques dégradées sur le trafic routier, il n'existe pas de consensus quant à la caractérisation des effets constatés. L'impact, évalué expérimentalement sur des cas d'étude bien distincts, est très dépendant du contexte et donc des nombreuses autres variables et contraintes de l'infrastructure routière (topographie, comportements régionaux des usagers par exemple). Les études menées se distinguent également par les échelles d'évaluation du comportement des usagers reprenant celles propres aux caractéristiques de trafic :

– l'échelle **microscopique** s'intéresse au comportement individuel des usagers de l'infrastructure traduit par la vitesse individuelle, le temps et la distance inter-véhiculaire entre autres. Les typologies de véhicules peuvent être prises en compte [252] et des analyses de répartition et vitesses par voie sont également menées [253, 254]. Il est alors mis en évidence que, par croissance de l'intensité des précipitations, la voie lente expérimente de plus fortes modifications des comportements des usagers (réduction de vitesse et modification de la répartition poids-lourd / véhicule léger) que les autres voies. De même, les distances inter-véhiculaires tendent à croître pour plus de sécurité, mais ce comportement est restreint aux distances entre véhicules légers. Le temps intervéhiculaire entre poids-lourds et véhicules légers est inchangé, comportement dû au respect naturel de la distance de sécurité dans ce genre de situations.

– l'échelle **mésoscopique** s'intéresse au comportement de groupes ou pelotons de véhicules. Il a été constaté une accentuation de la tendance à former des pelotons de véhicules lors de phases de météorologie dégradée. Seuls les travaux de Billot [252] ont étendu à ce jour l'analyse à cette échelle. Il s'avère que la configuration des pelotons de véhicules est dépendante des variables météorologiques. Le nombre moyen de véhicules formant un peloton, ainsi que leur vitesse médiane accusent des diminutions lors de l'intensification des précipitations.

– l'échelle **macroscopique** s'intéresse à l'impact sur les états de trafic agrégés par périodes de temps ou sur une section de réseau. Il s'agit de déterminer si un lien existe entre les conditions météorologiques et les variables macroscopiques de trafic : le débit (la capacité), la densité et la vitesse. Les comparaisons sont moins aisées qu'aux autres échelles, car les variables macroscopiques agrègent les informations et tendent à lisser les disparités de comportement temporellement et/ou spatialement. Les études menées à cette échelle sont également celles produisant les résultats les plus variés, voire conflictuels.
  Aucun effet n'est constaté sur la densité maximale, puisque cette caractéristique est propre au nombre maximum de véhicules pouvant occuper l'infrastructure. En revanche, une chute de capacité de 8,5% est dénotée par Hall & Barrow [255] pour de faibles précipitations, tandis que Ibrahim & Hall [256] relèvent une réduction de 10 à 20% pour de fortes pluies et jusqu'à 48% lors de chutes de neige. Les ordres de grandeur de telles réductions de capacité sont confirmés par les mesures relevées par Jones *et al.* [251] et Smith *et al.* [257]. L'augmentation des distances de sécurité entre les véhicules induit une réduction du débit maximal, mais également un ralentissement global des véhicules. Concernant les vitesses moyennes, le manuel américain sur l'exploitation des autoroutes (*Highway Capacity Manual*) [258] annonce des réductions de vitesses de 9,5km/h lors de pluies légères contre 19,5km/h pour de fortes précipitations. En revanche, des études similaires menées par Ibrahim & Hall [256] sur une autoroute canadienne font état d'une chute de vitesses cantonnée à 2km/h pour des pluies légères, atteignant 5-10km/h par croissance de l'intensité pluviométrique. Les habitudes de conduite sous conditions dégradées et l'équipement sont sans doute à l'origine de telles différences. Cependant, les

tendances restent inchangées d'une étude à l'autre : la réduction de vitesse décroît de façon exponentielle en fonction de la dégradation des conditions météorologiques.
Enfin, les relations liant l'ensemble des caractéristiques de trafic sont également modélisées par analyse des allures de diagrammes fondamentaux [252, 259].

Les diverses études menées mettent en évidence que les précipitations affectent à la fois le comportement des conducteurs (microscopique) et les conditions de trafic (macroscopique) lors de comparaison à des temps secs [260]. Les travaux de El-Faouzi *et al.* [253], ainsi que la thèse de Romain Billot [252] (chapitres 1 & 2) fournissent un état de l'art complet et détaillé de l'estimation de l'impact météorologique sur les conditions de trafic à toute échelle.

Ne disposant que de vitesses moyennes spatiales issues de données FCD, seule l'échelle macroscopique est accessible à cette étude et, plus spécifiquement, la quantification des effets météorologiques sur la vitesse moyenne. En tant qu'indicateur du niveau de service, la vitesse du flux de trafic est également la caractéristique de trafic macroscopique reportant les impacts les plus conséquents lors de précipitations [261]. Son étude sur le réseau du Pont de Cheviré semble dès lors pertinente.
De plus, contrairement aux analyses menées à ce jour, les informations météorologiques recueillies correspondent aux véritables conditions expérimentées par l'infrastructure routière. Les informations disponibles sont variées et s'étendent du degré de précipitation comme la plupart des études [253] à la visibilité, en passant par la force du vent et la température au sol.

Après avoir quantifié les effets de conditions dégradées sur le trafic, l'étape suivante conduit naturellement à tenter d'intégrer de tels effets au sein des modèles de prévision. L'existence d'une grande variété de modèles de prévision assure des stratégies d'intégration distinctes.

6.1.2 Intégration des effets météorologiques aux méthodes de prévision

L'intégration de l'impact de conditions météorologiques dégradées aux modèles de prévision est inégalement étudiée suivant les approches adoptées. La nécessité de disposer à la fois de données de trafic et météorologiques de qualité avantage indéniablement les approches s'appuyant le moins sur les données, à savoir les méthodes de type *model-driven* [259, 252, 262], puis, dans une moindre mesure, les méthodes statistiques paramétriques [60, 263, 264, 265]. Quelques approches apprenant la structure de l'historique de type Réseaux de Neurones Artificiels sont également rencontrées, mais font exception [266]. De façon générale, bien que les méthodes de prévision par apprentissage (*data-driven*) aient fait l'objet de multiples études, l'intégration de paramètres externes reste peu évaluée.

Le procédé d'intégration des phénomènes météorologiques diffère en fonction des modèles employés ou des points de vue considérés. Les principales approches sont les suivantes :

– **Modification des paramètres du modèle suivant les conditions météorologiques** : cette approche se rencontre particulièrement pour les méthodes de type *model-driven* et vise à faire évoluer les hypothèses formulées en fonction de l'intensité de précipitation. L'information météorologique peut se voir transformée en facteurs d'impact sur les paramètres du modèle [259, 267] ou nouveau paramètre prenant part aux équations du modèle [252] pour les modèles macroscopiques, tandis que les modèles microcopiques tendent à modifier les distributions des caractéristiques de trafic ou à introduire des pénalités de temps de parcours [262]. L'information météorologique n'est pas directement intégrée, mais ses effets sont quantifiés pour adapter les paramètres ou équations régulant le modèle. Certains modèles à Réseaux de Neurones [266] proposent d'apprendre des modèles sur des historiques distincts suivant le degré de précipitation et d'activer les modèles suivant les prévisions météorologiques.

– **Ajout de l'information météorologique brute comme variable d'entrée exogène** : cette approche se rencontre particulièrement pour les méthodes de type *data-driven*. Le procédé consiste simplement à ajouter des variables d'entrées décrivant les conditions météorologiques pour être prises en considération par le modèle comme facteurs explicatifs de la variance de la variable expliquée. Les modèles issus de la famille ARIMA (ARMAX, VARMAX ou STM) [60] ou les Réseaux de Neurones Artificiels [266, 268] sont les principaux représentants de telles approches. L'ajout de variables exogènes se manifeste sur les méthodes paramétriques par l'intégration de nouvelles composantes expliquant l'état prédit au sein de l'expression usuelle [60]. L'impact des composantes exogènes est préalablement calibré en fonction du degré d'explication de la variance de l'erreur. Le facteur météorologique est donc connu et maîtrisé. Pour les méthodes à Réseaux de Neurones, l'ajout de variables exogènes est géré par le processus de pondération et de création de neurones intermédiaires [266]. La quantification de la prise en considération des facteurs météorologiques reste cachée et inaccessible.

– **Évaluation de coefficients correctifs** : cette approche n'est pas spécifique à une typologie de modèles et se retrouve sous diverses formes. Ainsi, Rakha *et al.* [259], puis Mahmassani *et al.* [267] développent des facteurs d'ajustements multipliant les paramètres de modèles macroscopiques/mésoscopiques évalués par temps sec. Lu [263] propose un facteur correctif, adapté à une régression paramétrique, prenant la forme d'un produit de convolution entre la variation de vitesse usuelle et le taux de pluviométrie. Enfin, Billot *et al.* [264] proposent des facteurs pénalisant directement les temps de parcours prédits par un modèle paramétrique sec.

Quel que soit le procédé employé, une amélioration des performances de prévision est constatée [269] sur les modèles intégrant l'information météorologique. Plus fine est l'information, meilleure est la performance de prévision. Il est recommandé d'employer des informations météorologiques disposant d'une résolution inférieure à 1 heure. Peu de comparaisons d'approches véritablement distinctes sont proposées, si bien qu'il est difficile de préférer un procédé d'intégration plutôt qu'un autre. Seul le contexte d'utilisation et la qualité des données recueillies jouent un rôle.

Néanmoins, l'intégration de facteurs météorologiques pour les approches par apprentissage reste peu représentée et évaluée dans la littérature [265]. L'insertion de tels facteurs dans le cadre prédictif global constitue également un défi à relever. Défi d'autant plus délicat que, d'une part, les données météorologiqes doivent être connues au moins localement pour la plupart des sections du réseau et, d'autre part, le modèle doit être adapté de façon hétérogène en fonction des conditions météorologiques locales.

L'impact de la météorologie sur les états de trafic observés sur le Pont de Cheviré entre septembre et novembre 2013 fait l'objet d'une analyse approfondie avant de s'intéresser aux potentiels effets sur l'erreur de prévision.

6.2 Qualification des informations météorologiques

Les informations météorologiques sont analysées et confrontées aux vitesses expérimentées par le réseau du Pont de Cheviré entre septembre et octobre 2013 pour caractérisation des effets.

6.2.1 Descriptif des données météorologiques et cas d'étude

Les données météorologiques à disposition présentent une forme particulière, puisqu'il s'agit de relevés directement fournis par le service route de MétéoFrance. Contrairement aux informations usuellement tirées des stations météorologiques des aéroports à proximité, les données produites sont locales et spécifiques à l'infrastructure considérée. Une grande variété de mesures sont effectuées avec un pas de temps de 5 minutes depuis six capteurs répartis sur le réseau du Pont de Cheviré (Figure 6.3).



Fig. 6.3.: Positionnement des 6 capteurs MétéoFrance sur le réseau du Pont de Cheviré à Nantes.

Les mesures météorologiques disponibles se composent de l'occurence de précipitation, du type de précipitation, de la catégorie d'intensité de précipitation, de la température de l'air et du sol,

de la force du vent et des rafales, ainsi que d'un indicateur binaire de visibilité.
Néanmoins, l'intégralité de ces informations n'est pas exploitable pour diverses raisons : catégories jamais atteintes, manque d'intérêt de la variable dans le contexte par exemple. Seules certaines variables météorologiques permettent une véritable analyse. Une pré-sélection manuelle des variables météorologiques considérées est effectuée et restreint le nombre de variables d'intérêt à quatre : la catégorie de précipitation, la température au sol, la force des rafales et la visibilité.
La température du sol est préférée à la température de l'air, car elle constitue *a priori* une meilleure représentation du gel éventuel de la chaussée et n'est séparée de la température de l'air que par un biais systématique +3 °C. De même, la force moyenne du vent n'est pas retenue, car peu discriminante sur la période considérée avec des vents uniquement inférieurs à 30 km/h. La force des rafales est donc préférée, car irrégulière et *a priori* plus perturbante. Enfin, les variables indicatrices de la typologie de précipitation ne sont pas retenues, car peu de catégories sont distinctes et l'interprétation est redondante avec les informations apportées par le degré de précipitation, de la température du sol et de la visibilité.

En outre, afin de comparer des données homogènes, les informations météorologiques, originellement disponibles avec une résolution de 5 minutes, sont agrégées pour atteindre un pas de temps de 30 minutes, malgré tout conforme aux recommandations de la littérature [269]. Dans cette même veine, seuls six capteurs météorologiques fournissent des informations pour l'intégralité du réseau. Un procédé d'appariement topographique pour lier données trafic et météorologique (*map-matching*) est mis en oeuvre pour association à tout instant de données météorologiques à chacune des sections du réseau. Chaque section routière reçoit les données issues de la station météorologique la plus proche à vol d'oiseau. L'hypothèse que les données météorologiques locales soient invariantes sur de petites zones du réseau routier est donc formulée. Cette supposition est justifiée par le choix d'un pas de temps de 30 minutes effaçant les éventuels décalages temporels dus à la distance séparant la section d'étude et le capteur météorologique. Des décalages temporels entre impact sur la vitesse et mesure de pluviométrie sont effectivement régulièrement constatés dans la littérature [263, 268] et atteignent entre 15 minutes et 1 heure suivant la proximité de la station météorologique et les mouvements atmosphériques. De plus, à l'échelle du réseau du Pont de Cheviré, peu de disparités sont effectivement constatées d'une station météorologique à une autre.

6.2.2 De l'information disponible à l'information exploitable

Seules quatre variables météorologiques sont retenues pour analyse d'impact sur le trafic, à savoir deux variables catégorielles : le degré de précipitation et la visibilité, puis deux variables continues : la température au sol et la force des rafales de vent. Afin de faciliter l'usage d'outils de comparaison et mieux cibler les zones d'intérêt, les deux variables continues sont discrétisées pour aboutir à des catégories. L'ensemble des mesures météorologiques disponibles se présente sous la forme suivante :

– l'**intensité de précipitation** (I) partagée en quatre catégories correspondant aux classes usuellement employées en France pour distinguer les typologies de précipitations :
  1. Temps sec : $I = 0mm/h$.
  2. Pluie légère : $I \leq 2,4mm/h$
  3. Pluie intermédiaire : $2,4 < I \leq 7,2mm/h$
  4. Pluie forte : $I > 7,2mm/h$
– la **température au sol** (T_{sol}) scindée en trois catégories correspondant à des cas susceptibles d'affecter les conditions routières :
  1. Températures favorisant l'apparition de brouillard ou de gel : $T_{sol} \leq 7°C$
  2. Températures usuelles : $7°C < T_{sol} \leq 27°C$
  3. Fortes températures : $T_{sol>27°C}$
– la **force des rafales de vent** (B_R) partagée en deux catégories suivant un seuil fixé à 28 km/h. Des vents supérieurs à 28 km/h sont en effet susceptibles d'avoir un impact non négligeable sur la visibilité lors de précipitations [261].
– la **visibilité** (Vis) séparée en deux catégories par un seuil fixé à 200 m.

6.2.2.1 Analyse de la répartition des observations sur les catégories pour évaluation de la significativité

L'ensemble de ces catégories météorologiques n'est exploitable qu'à condition de disposer d'échantillons significatifs au sein de chaque catégorie. Une analyse préliminaire de la répartition sur chacune des catégories associées aux mesures météorologiques est proposée en Figure 6.4.

Force est de constater que les conditions météorologiques sont majoritairement clémentes sur la période d'étude. La mise en évidence d'effets sur le trafic s'avère dès lors délicate dans la mesure où, statistiquement, peu d'observations sont disponibles pour caractériser les périodes inclémentes. Le climat tempéré de Nantes, dû à la proximité côtière, limite les températures fraîches avoisinant les zéros et donc les gels de chaussée. De plus, les précipitations semblent plus s'apparenter à des crachins passagers qu'à de véritables pluies abondantes. Enfin, la visibilité inférieure à 200 m n'est atteinte que pour 2 % des observations, ce qui implique peu de brouillards. De plus, la mauvaise visibilité n'est jamais liée à des précipitations, ce qui limite son impact au seul phénomène de la présence de brume.

Le croisement de facteurs météorologiques possède, d'après la littérature [253], un impact sur le trafic plus conséquent. Associé aux précipitations, des vents supérieurs à 24 km/h sont susceptibles d'amplifier la baisse des vitesses [261]. Un tel croisement des variables étant envisagé, une analyse succinte des proportions de données pour chacune des catégories (voir Table 6.1) met en évidence l'impossibilité d'exploiter un tel procédé dans notre contexte d'étude. En effet, le croisement des catégories montre que les cas pertinents, i.e. confrontant

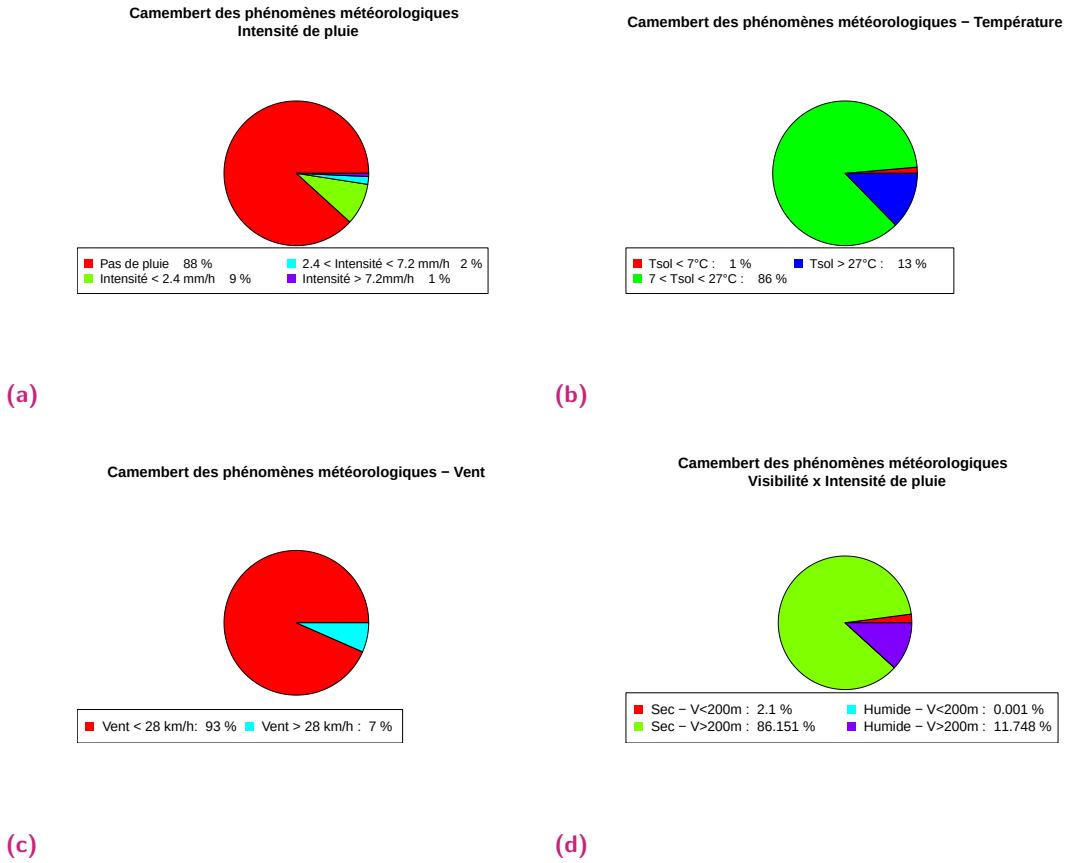


Fig. 6.4.: Camemberts illustrant la proportion de données recueillies au sein de chacune des catégories : a) pour l'intensité pluviométrique b) pour la témpérature au sol c) pour la force des rafales de vent d) pour la visibilité croisée avec la pluie. (Période de septembre à novembre 2013)

des précipitations importantes à d'autres facteurs dégradés, font exception durant la période analysée. Seuls 3 % des cas font état de précipitations accompagnées de fortes rafales de vent, dont 2 % pour des précipitations faibles. Les cas de températures inférieures à $7°C$ sont logiquement atteints par temps clair, lorsque le ciel est dégagé et n'excède par 1% des situations rencontrées. Finalement, cette analyse préliminaire montre que la variable précipitation est probablement la seule en mesure de mettre en évidence des différences significatives de vitesse de par son échantillonnage sur les catégories disponibles.

6.2.2.2 Mesure de significativité de l'impact par analyse des variances

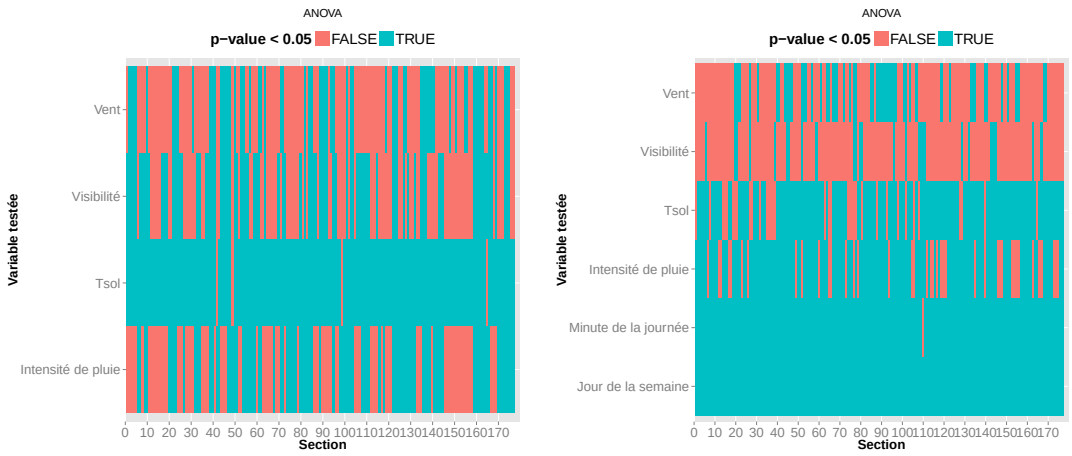
À titre de vérification de la pertinence de l'usage des quatre variables météorologiques pour explication des variations de vitesse, une analyse de variance (ANOVA) est menée sur les 177 portions du réseau du Pont de Cheviré sur la période septembre - octobre 2013.
L'analyse de variance consiste en un test statistique visant à vérifier que plusieurs échantillons sont issus d'une même population (Hypothèse H_0). Ce test est appliqué pour évaluer une ou plusieurs variables explicatives catégorielles ayant supposément de l'influence sur la distribution

T_{sol}		$T_{sol} \leq 7°C$		$\leq 7°C < T_{sol} \leq 28°C$		$T_{sol} > 28°C$	
Rafales (B_R, km/h)		$B_R \leq 28$	$B_R > 28$	$B_R \leq 28$	$B_R > 28$	$B_R \leq 28$	$B_R > 28$
$I(mm/h)$	$I = 0$	1 %	0 %	70 %	4 %	12 %	0 %
	$0 < I \leq 2,4$	0 %	0 %	7 %	2 %	0,1 %	0 %
	$2,4 < I \leq 7,2$	0 %	0 %	1 %	0,6 %	0 %	0 %
	$I > 7,2$	0 %	0 %	0,9 %	0,4 %	0 %	0 %

Tab. 6.1.: Tableau croisé de proportions pour chacune des catégories météorologiques sur la période septembre - novembre 2013.

d'une variable continue à expliquer. Un risque de 5 % est considéré pour les tests statistiques pratiqués. Une variable explicative contribue à scinder en deux populations et possède donc un impact sur la variable expliquée, si, sous l'hypothèse H_0, la probabilité de trouver une valeur plus extrême que celle observée (*p-value*) est inférieure à 5 %. En pratique, l'hypothèse nulle consiste à supposer que la variable explicative (météorologique par exemple) testée n'a pas d'influence sur la variable expliquée (la vitesse du flux).

L'analyse de variance, illustrée en Figure 6.5, vise à évaluer le degré d'explication de la vitesse produit par chacune des quatre variables météorologiques, ainsi que par des variables temporelles en guise de comparaison.



(a) (b)

Fig. 6.5.: Résultats d'analyse de variances (ANOVA) menée sur les 177 sections du Pont de Cheviré entre septembre et octobre 2013 : a) Variables explicatives uniquement météorologiques b) Variables explicatives météorologiques et temporelles.

Le premier constat porte sur l'effet indéniable, observé en Figure 6.5 b), de la typologie de jour de la semaine et de la période de la journée sur la distribution de vitesses du flux. L'impact statistiquement significatif confirme l'existence de périodes de pointe et de profils de demande

journaliers variant suivant les jours de la semaine. En outre, la comparaison entre Figure 6.5 a) & b) met en évidence la nécessité de comparer les effets météorologiques sur des périodes temporelles similaires. En effet, la comparaison de distributions de vitesses sans distinction temporelle (a) montre que peu de variables météorologiques ont un impact statistiquement significatif. En revanche, dès que la temporalité est introduite, l'intensité de précipitation devient de plus en plus pertinente. Il est à remarquer le fort degré de significativité de la température au sol résultant du biais introduit par le cycle de températures sur toute journée. La température du sol tire sa significativité de sa capacité à représenter l'heure de la journée. Enfin, la visibilité et le vent semblent peu significatifs, ce qui est sans doute lié à la faible quantité d'observations disponibles et leur disparité temporelle. Finalement, **seul le degré de précipitation paraît disposer d'un véritable effet significatif sur la vitesse du flux de trafic à condition de mener une analyse par période temporelle et suivant les jours de semaine**. Les conditions inclémentes étant limitées sur la période d'évaluation, il est naturel de ne pouvoir observer des effets météorologiques qu'en menant une analyse fine et locale.

6.3 Quantification des effets météorologiques sur la vitesse

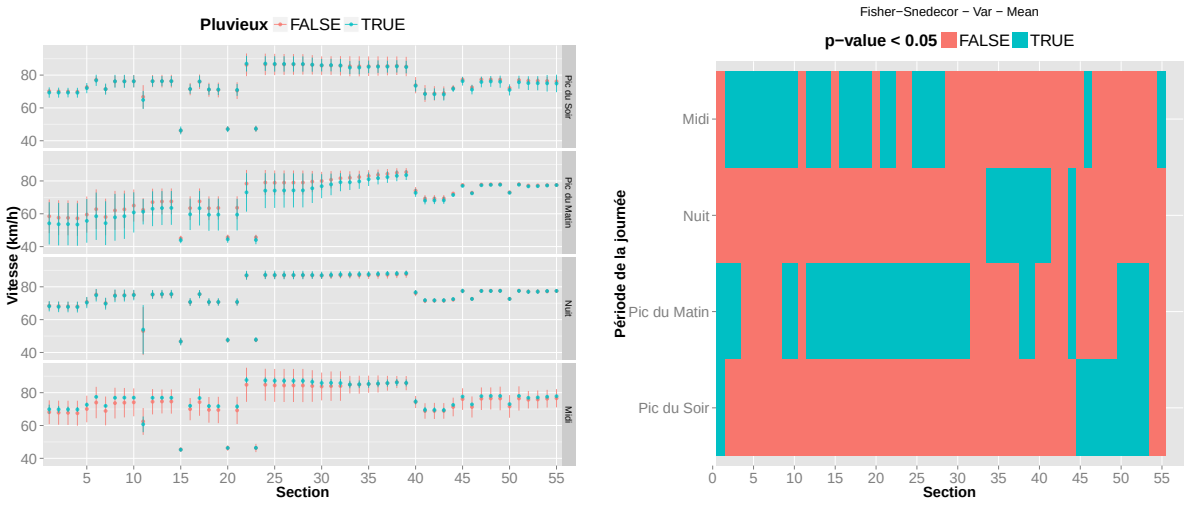
La quantité d'observations en conditions inclémentes est limitée sur la période d'évaluation. La quantification de l'impact des précipitations sur la vitesse du flux s'avère alors difficile, mais pas impossible. Si peu de fortes précipitations avec fort vent ou faibles températures sont observées, des impacts statistiquement significatifs ont été montrés pour le facteur précipitation sur une grande part des sections. Néanmoins, les effets ne sont observables qu'en comparant des conditions de trafic semblables. Or, la donnée débit n'est pas disponible, seules les informations temporelles nous renseignent. Ainsi, une première analyse d'impact est proposée par période de la journée, puis une seconde confronte deux journées similaires au regard des typologies de jour.

6.3.1 Analyse d'impact par période de la journée

L'analyse par période de la journée est menée par l'intermédiaire d'un test de Fisher-Snedecor évaluant la distinction entre les distributions de vitesses par temps sec ou humide, puis via une analyse graphique des distributions de vitesses sur une période de 24 heures.

Le test de Fisher-Snedecor consiste en deux tests statistiques visant à comparer deux distributions et formule l'hypothèse H_0 que ces deux distributions suivent une même distribution normale (même variance, puis même moyenne). L'hypothèse H_0 est vérifiée si la p-value excède le risque, traditionnellement fixé à 5 %.
Les deux distributions évaluées correspondent aux distributions de vitesses par temps sec et humide, testées sur chacune des quatre périodes typiques d'une journée (nuit, pic du matin, midi, pic du soir). La zone d'étude est restreinte aux 55 principales portions de l'axe Nord-Sud périphérique du réseau du Pont de Cheviré, car ces sections, expérimentant de plus grandes

vitesses en conditions normales, sont susceptibles de subir des dégradations plus marquées sous conditions inclémentes.



(a) (b)

Fig. 6.6.: Identification d'impacts du temps sec ou humide sur la vitesse du flux par période de la journée à partir de 55 sections du Pont de Cheviré entre septembre et octobre 2013 : a) Distributions de vitesses par période de la journée en fonction du temps sec ou humide b) Résultat du test de significativité de Fisher-Snedecor par période temporelle.

Les distributions testées sont représentées pour chacune des 55 sections et toute période temporelle en Figure 6.6 a). Seule la période du pic du matin expérimente sur certaines sections des vitesses plus faibles par temps de pluie, si bien que la congestion récurrente du début de matinée est renforcée. La période du milieu de journée, au contraire, tire bénéfice des conditions inclémentes. Cette dernière observation est à confirmer par une analyse du profil de demande, mais correspond sans doute à une réduction de l'usage de l'infrastructure les midis pluvieux.

Ces observations sont renforcées par l'analyse des résultats du test de Fisher-Snedecor (Figure 6.6 b)) montrant que la plupart des différences significatives sont localisées sur la période du pic du matin, puis celle de midi. Toutes les sections de la zone du périphérique ne permettent pas de rejeter l'hypothèse nulle, si bien que des disparités de comportements sont observables. Ces disparités sont sans doute liées à l'ordre de grandeur des vitesses expérimentées et à la rigidité du risque fixé à 5 %. Les sections expérimentant les plus fortes vitesses sont marquées par de plus fortes réductions par temps pluvieux.

Finalement, **l'analyse statistique n'aboutit pas à une conclusion définitive quant à l'impact sur les vitesses du flux. Suivant les périodes considérées, des effets sont notables sur certaines sections. Un accroissement de la congestion durant la période de pic du matin est notamment constaté**, tandis que toute autre période de la journée ne paraît pas affectée.

Une analyse plus fine et horaire sur le réseau du Pont de Cheviré, illustrée en Figure 6.7, met essentiellement en évidence que **l'intensité de précipitation possède un effet observable**

sur la distribution de vitesses. Cette observation confirme que la force des rafales et même la visibilité n'influencent pas significativement la distribution de vitesses, ce qui est probablement dû à la quantité d'informations disponible. Les faibles et hautes températures au sol tendent à expérimenter des vitesses plus élevées et non des ralentissements. La significativité des différences de distributions constatées par analyse de variance ne traduit donc pas un impact dû à des conditions dégradées pour la température du sol (T_{sol}).

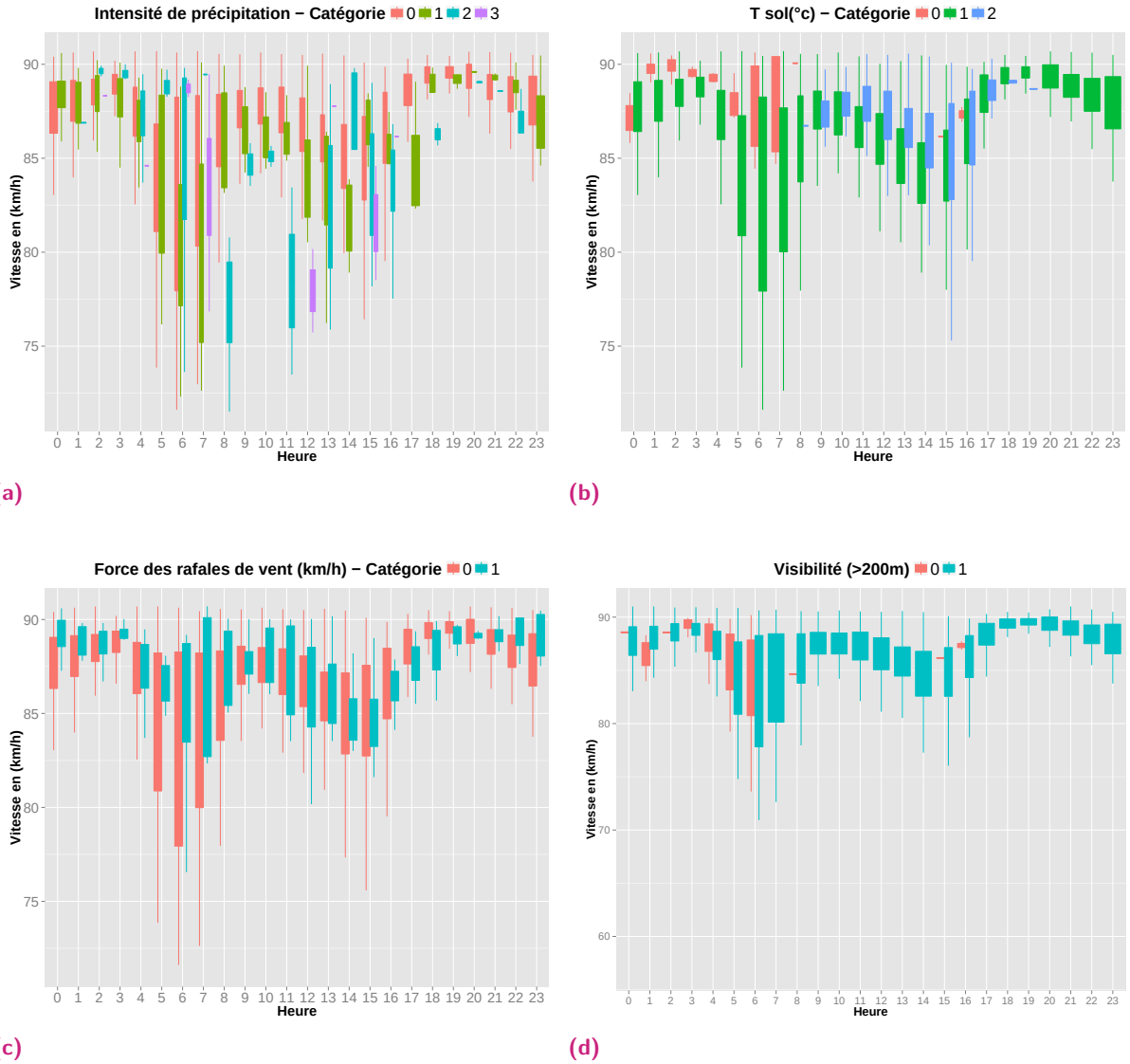


Fig. 6.7.: Représentation des distributions de vitesses en fonction des catégories associées aux variables météorologiques sur la section 176 (septembre-octobre 2013) : a) Catégorie d'intensité de précipitation b) Catégorie de température au sol c) Catégorie de force des rafales de vent d) Catégorie de visibilité.

Outre un comparatif des impacts de chacune des composantes météorologiques sur les distributions de vitesses horaires, la Figure 6.7 met en évidence que les pluies supérieures à 2,8

mm/h (moyennes à fortes) induisent des réductions de vitesses du flux plus conséquentes. Cette observation est cohérente avec les analyses produites au cours de la thèse de Romain Billot [252] en chapitre 2 sur le cas Suisse. Cependant, une observation attentive révèle que cet effet n'est pas systématique : certaines périodes de très fortes pluies (fin de congestion matinale, fin d'après-midi sur la Figure 6.7 a)) expérimentent des vitesses de flux habituelles. Ces valeurs particulières résultent probablement de la faible quantité d'observations en conditions dégradées, si bien que la quantification des effets météorologiques reste instable : des réductions de la vitesse du flux sont effectivement observées, mais nécessitent une confirmation sur une base de données plus conséquente ou par une analyse plus locale par confrontation de journées similaires.

6.3.2 Confrontation de journées similaires

Après une analyse globale sur diverses périodes temporelles plus ou moins finement découpées, des profils de vitesses comparables [1] sont comparés sous des conditions de précipitations distinctes. Seule la composante intensité de précipitation varie.
Les analyses de variances (ANOVA) ont montré que les périodes de la journée et les jours de la semaine constituent des critères significatifs discriminant les vitesses moyennes expérimentées par les sections. Une telle tendance montre que la mise en évidence de l'impact des conditions météorologiques sur le trafic est conditionné par la comparaison de situations similaires, i.e. profils identiques de demande. Ne disposant pas des renseignements quant aux profils de demande ou de débit, seules les typologies de jours sont susceptibles de produire des profils de demande similaires. Partant de cette observation, des couples journées pluvieuses / ensoleillées sont comparés et analysés en supposant les demandes identiques sur un même jour de la semaine, hors vacances scolaires. Les résultats sont exposés en Figure 6.8 sur des journées travaillées et d'autres de repos.

Quelles que soient la typologie de journée considérée et la durée des précipitations subies, une réduction de vitesse sur les profils humides est constatée. Elle se traduit simplement par une congestion plus conséquente (perte de 5 km/h) lors de précipitations éparses durant la période de pointe du matin. En revanche, lors de jours chômés (dimanches) subissant des périodes continues de précipitations, l'effet est notoire et se traduit par une réduction de 5 km/h sur les périodes supposées de plus forte activité. **Cette analyse très locale confirme l'existence d'impacts du degré de précipitations sur la vitesse du flux, effets extrêmement localisés les rendant sensibles à l'agrégation temporelle**.

Finalement, cette étude météo-sensible sur le Pont de Cheviré met en évidence l'existence d'un effet de la météorologie sur le trafic. Cependant, depuis les données recueillies, seule une analyse très locale est apte à fournir une quantification de l'impact. De plus, la présence de phénomènes météorologiques à un instant donné semble ne pas être le seul facteur d'impact. **Des**

1. profils partageant une même typologie de jour en dehors des congés

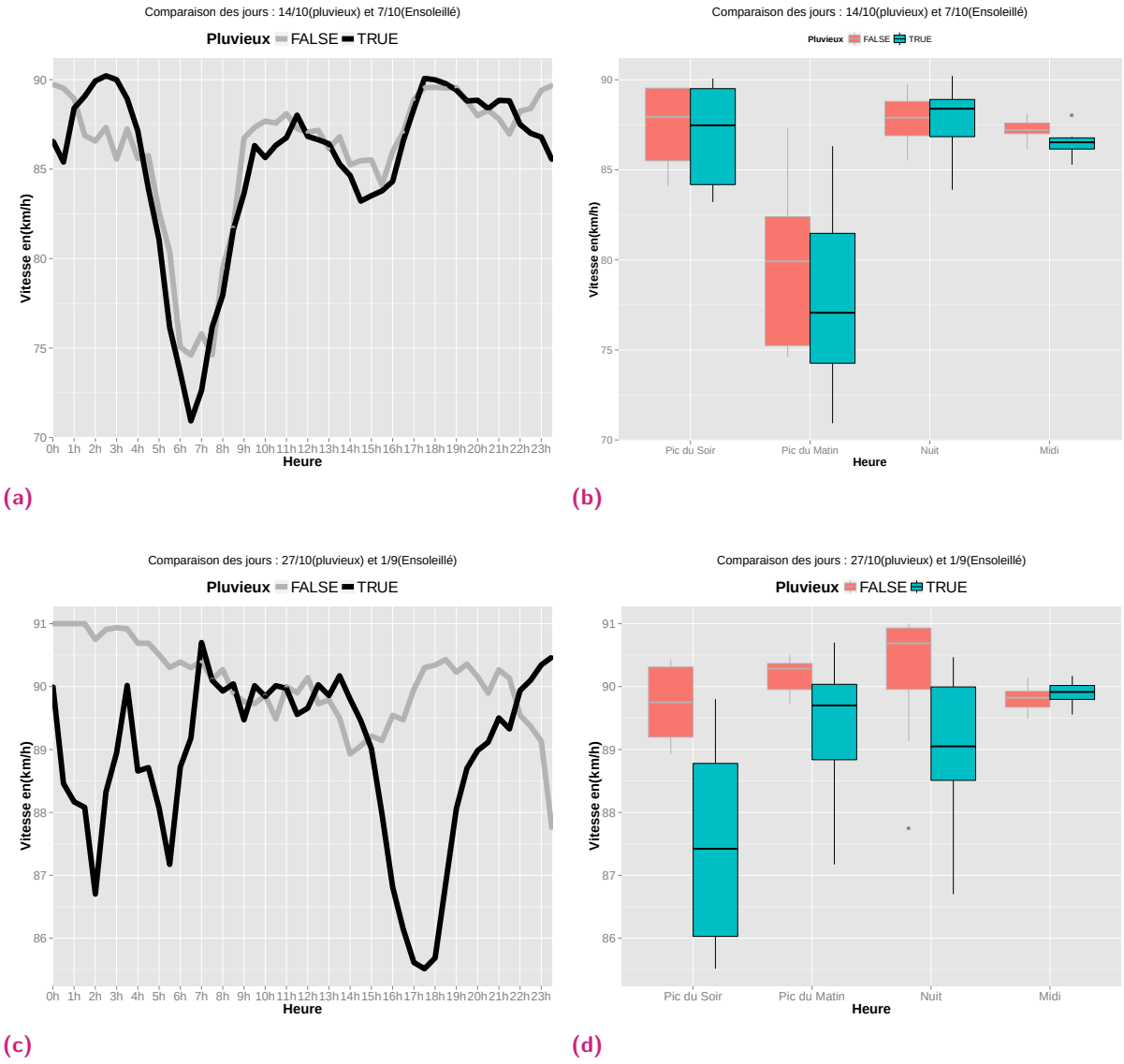


Fig. 6.8.: Confrontation de journées recevant des demandes de trafic similaires, mais expérimentant des conditions météorologiques distinctes sur la section : a) Comparaison des profils de vitesses des lundis 7 et 14 octobre 2013 (Précipitations éparses de 3h à 8h) b) Comparaison des distributions de vitesses des lundis 7 et 14 octobre 2013 c) Comparaison des profils de vitesses des dimanches 1er septembre et 27 octobre 2013 (Précipitations continues de 1h à 7h, puis 14h-21h) d) Comparaison des distributions de vitesses des dimanches 1er septembre et 27 octobre.

informations complémentaires sont sans doute à prendre en considération. Les longues phases pluvieuses, même de faible intensité, occasionnent des ralentissements conséquents (Figure 6.8 b)), au même titre que des précipitations de courte durée, mais intenses (Figure 6.8 a)). La seule occurence de précipitations influence certes la visibilité, mais les antécédents expérimentés par l'infrastructure (sol mouillé, saturation du sol) sont probablement des paramètres à prendre en considération. **Des effets de cumuls ou de retardement constituent ainsi de probables facteurs aggravant la détérioration des conditions de trafic.** À terme, ces fac-

teurs seraient à intégrer comme partie prenante de l'analyse, mais seraient nettement plus aisés à considérer depuis une variable continue d'intensités de précipitation. Des corrélations croisées ou des évaluations de l'information mutuelle entre variables de trafic et météorologiques aboutiraient à une quantification de la durée d'impact d'épisodes de précipitations.
Les observations météorologiques à disposition pèchent par une quantité limitée de conditions inclémentes rendant l'interprétation statistique compliquée. Seule l'intensité de précipitation permet une distinction statistique pertinente des distributions de vitesses pour la plupart des sections du réseau. Néanmoins, la confrontation locale de journées similaires est encourageante et met en évidence une réduction effective des vitesses suivant les conditions météorologiques observées.

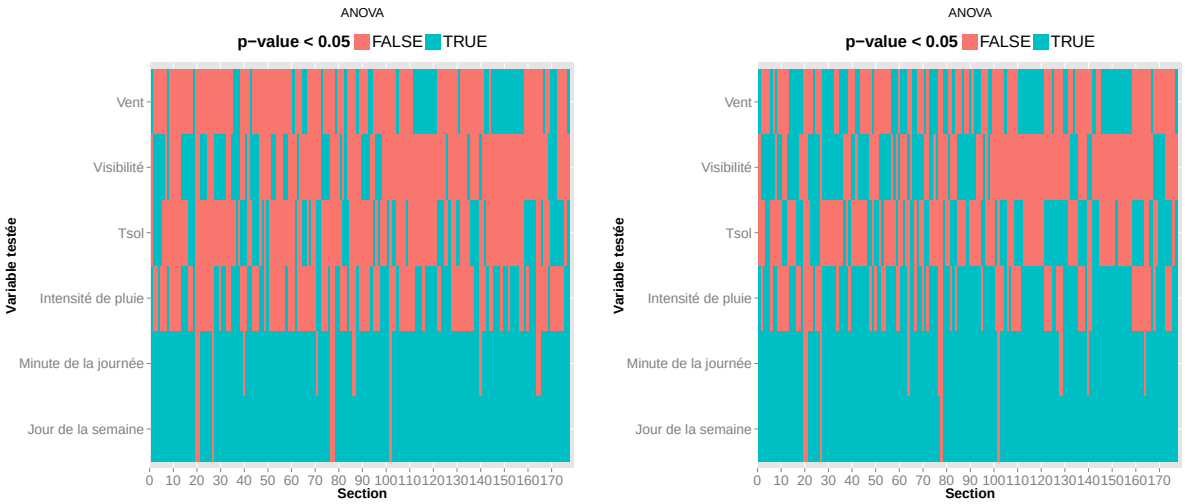
6.4 Évaluation de l'impact du degré de précipitation sur l'erreur de prévision

Après analyse de l'effet de la météorologie sur le trafic, l'estimation de l'impact sur l'erreur de prévision est envisagée. À cette fin, l'erreur de prévision à six horizons variant de 30 minutes à 3 heures est évaluée depuis l'approche globale appliquée via MKNN. Mises à part les périodes pluvieuses contribuant à l'historique d'apprentissage selon le procédé usuel, aucune information spécifique à la météorologie n'est intégrée au cadre prédictif. Cette analyse vise véritablement à évaluer l'éventuel effet de conditions inclémentes sur le procédé de prévision originel. L'estimation est effectuée sur le mois de novembre 2013 expérimentant des précipitations sur les deux premières semaines.

6.4.1 Analyse statistique de l'erreur de prévision en fonction des conditions météorologiques

Au même titre que la vitesse du flux, l'erreur de prévision est une variable aléatoire dont les variations sont explicables en partie par des facteurs externes. Une analyse de variance (ANOVA) est menée afin d'identifier les variables météorologiques et temporelles pour lesquelles l'hypothèse nulle peut se voir rejetée avec un risque à 5 %. L'hypothèse H_0 suppose encore l'indifférenciation des distributions en fonction des valeurs prises par les variables explicatives. Les résultats du test effectué sur l'intégralité de la zone du Pont de Cheviré sont exposés en Figure 6.9.

Une nouvelle fois, la période temporelle (typologie et heure de la journée) explique en grande partie l'erreur. Il est effectivement moins aisé de prévoir les états de trafic durant les périodes de forte activité que lors des phases nocturnes. Les variables météorologiques contribuent peu à la différenciation de l'erreur de prévision à 30 minutes, mais leur degré d'explication de la variabilité de l'erreur croît avec l'horizon de prévision. **L'intensité de précipitation reste le facteur le plus récurrent ayant des effets significatifs sur l'erreur de prévision, mais**



(a) (b)

Fig. 6.9.: Résultats d'analyses de variances menées sur l'erreur de prévision sur l'intégralité de la zone du Pont de Cheviré : a) ANOVA à l'horizon 30 minutes b) ANOVA à l'horizon 3h.

le taux de significativité de la visibilité croît avec l'horizon de prévision. L'augmentation d'observations par faible visibilité restitue à ce paramètre son impact sur le trafic et l'erreur de prévision.
Au final, pour de faibles horizons de prévision, peu de sections expérimentent des erreurs significativement différentes lorsque les conditions météorologiques se dégradent. Pour autant, l'existence de liens entre phénomène météorologique et ampleur de l'erreur de prévision n'est pas à rejeter. En effet, à long terme, l'erreur de prévision tend naturellement à croître, si bien que les liens existants entre erreur et facteurs externes sont exacerbés pour atteindre les seuils de significativité.

Les tests statistiques étant menés sur des modèles théoriques simples liant les variables explicatives à la variable expliquée, seule l'explication de la variance du modèle est estimée. Une vérification visuelle directe des effets mesurés sur l'erreur de prévision s'impose. À cette fin, une analyse des distributions du pourcentage d'erreurs en fonction de l'intensité de précipitation sur deux sections est proposée en Figure 6.10. Pour la section 170, le degré de précipitation permet de rejeter l'hypothèse nulle selon l'analyse des variances, tandis que la section 17 ne permet pas de distinguer les distributions d'erreur suivant l'intensité de précipitation. Les deux sections se différencient par des ordres de grandeur d'erreurs et de vitesses distincts.
Quel que soit l'horizon considéré, malgré quelques exceptions, les cas de croissance de l'erreur en fonction des conditions de précipitations sont effectivement plus marqués pour la section 170. La Figure 6.10 e) & f) proposant la distribution des vitesses observées sur la même période et sous les mêmes conditions inclémentes met en évidence qu'une meilleure distinction entre les vitesses est atteinte sur la section 170 que sur la section 17. L'absence de différenciation sur la section 17 est donc essentiellement due à l'absence d'impact des conditions météorologiques

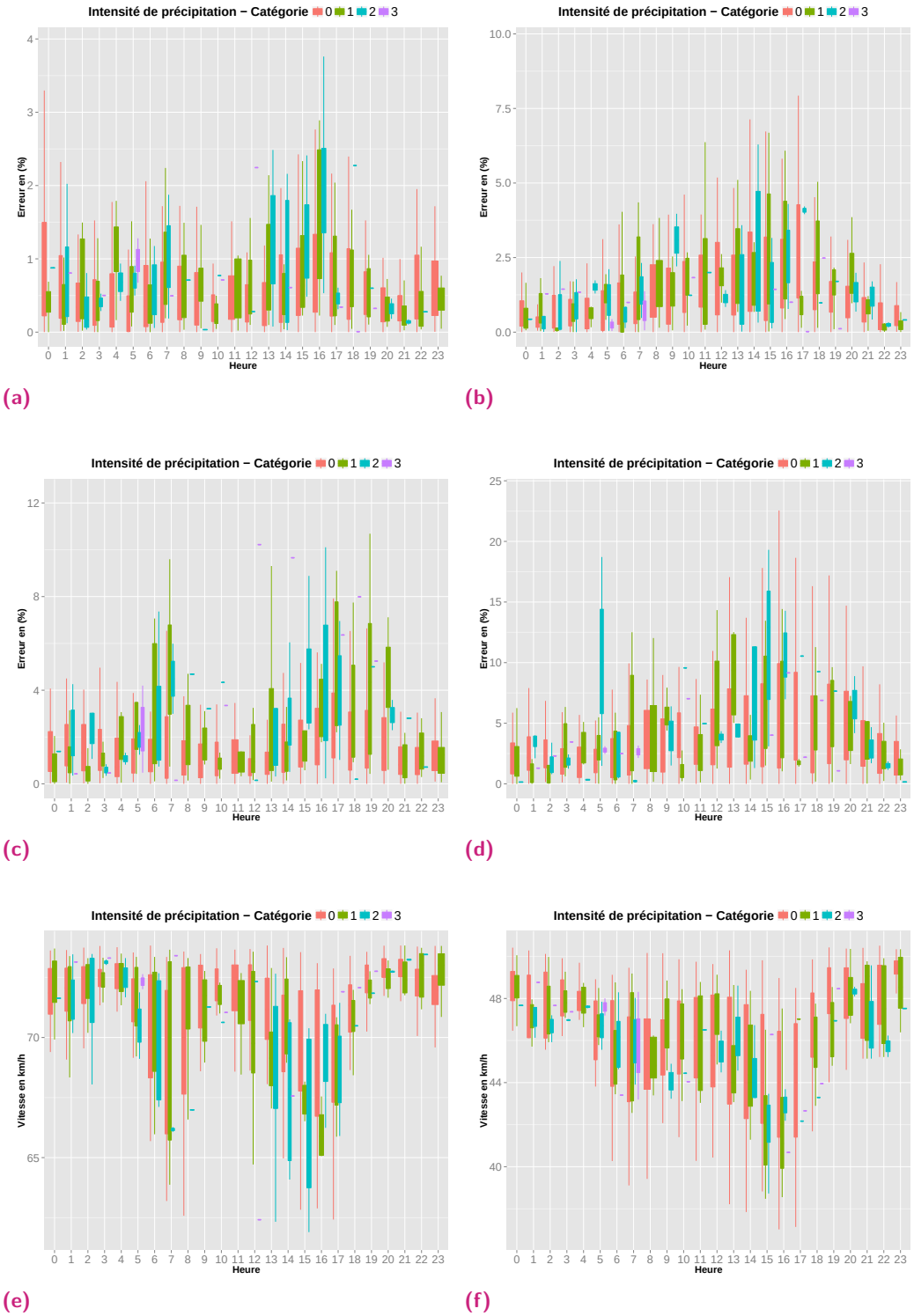


Fig. 6.10.: Distributions des erreurs de prévision en fonction de l'intensité de précipitation et de l'heure de la journée : a) Erreur sur la section 170 à l'horizon 30 minutes b) Erreur sur la section 17 à l'horizon 30 minutes c) Erreur sur la section 170 à l'horizon 3h d) Erreur sur la section 17 à l'horizon 3h e) Vitesse sur la section 170 f) Vitesse sur la section 17.

expérimentées sur les vitesses et non à une meilleure capacité de prévision de la méthode. Néanmoins, quelle que soit la section considérée, aucune tendance ne se dessine quant aux effets de la météorologie sur l'erreur de prévision. Certaines périodes expérimentent des erreurs plus conséquentes que d'autres, mais de nombreuses prévisions restent malgré tout dans la norme. Les effets de l'intensité de précipitation sur l'erreur semblent nécessiter une analyse plus locale ou l'intervention de facteurs externes tels que la durée des précipitations ou l'état antérieur des chaussées pour aboutir à une véritable quantification. Les effets sur l'erreur de prévision sont probablement cachés et lissés par l'agrégation temporelle de phénomènes se produisant sous des conditions de trafic légèrement différentes.

Le lien entre erreur de prévision et facteur météorologique n'est pas clairement établi depuis les analyses menées à l'heure actuelle. Néanmoins, la perspective de l'intégration des phénomènes météorologiques a été envisagée.

6.4.2 Vers une intégration des phénomènes météorologiques au sein du cadre prédictif

Deux méthodes de quantification et d'intégration de l'information météorologique au cadre prédictif ont été proposées. Toutes deux se sont révélées peu efficaces, du fait d'un manque de caractérisation des liens entre trafic et météorologie, mais l'une d'elles pourrait être prometteuse à l'avenir.

Les deux approches de quantification et intégration des phénomènes météorologiques ont consisté en :

- **une approche à coefficients correctifs** : pour toute typologie de jour et pas de temps de la journée, un facteur multiplicatif est appliqué à la prévision par temps sec de l'approche globale MKNN. Ce facteur est évalué depuis la confrontation de modèles de régression linéaire construits sous diverses conditions météorologiques (intensité de précipitation). Le facteur multiplicatif est défini, pour toute catégorie météorologique, par le ratio de la prévision linéaire en conditions dégradées sur la prévision par temps sec.
- **une approche par assimilation gaussienne [270] de deux modèles prédictifs** : un procédé d'assimilation des prévisions fournies par MKNN et un modèle météo-sensible est appliqué en cas de précipitations prévues à l'horizon de prévision. L'assimilation gaussienne suppose que chacun des modèles prédictifs apporte une information plus ou moins fiable quant à la réalité de l'état de trafic expérimenté, puis modélise l'ensemble des états maximisant la vraisemblance par une distribution gaussienne.
  Dans le cas étudié, le modèle météo-sensible est construit par estimation des valeurs typiques (moyenne) de vitesses par typologie de jour, pas de temps de la journée et conditions météorologiques (sec/humide). La prévision à l'horizon h de ce modèle est produite par lecture directe de la table de vitesses moyennes en tenant compte des conditions à l'instant $t + h$. L'erreur globale est fixe quel que soit l'horizon de prévision et

vaut 4,5 %.

L'assimilation consiste alors à procéder à une moyenne pondérée entre les deux valeurs prédites par chacun des modèles. La pondération est estimée à partir du degré de fiabilité de chacune des valeurs prédites. La fiabilité de la prévision est caractérisée par la variance des K proches voisins pour MKNN et la variance typique pour le modèle météo-sensible.

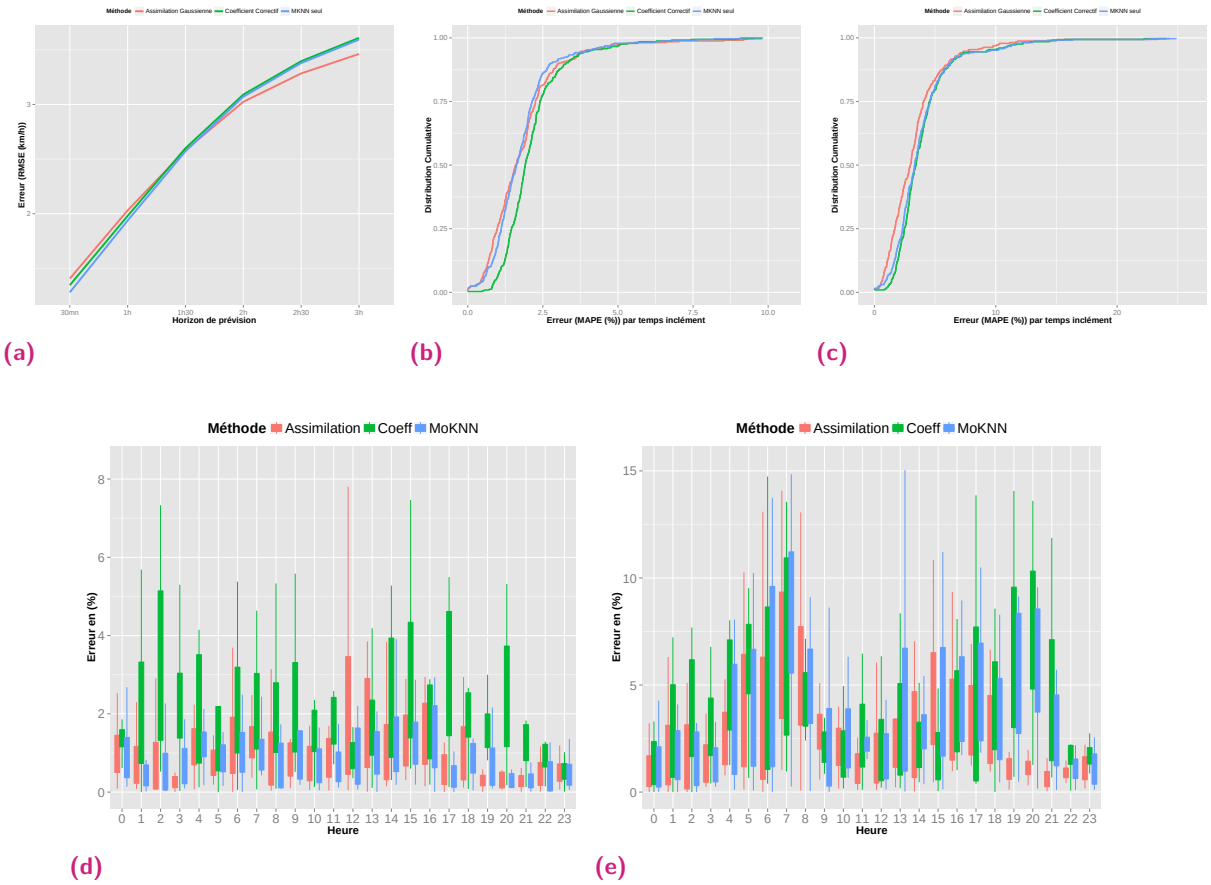


Fig. 6.11.: Illustration des performances des tentatives d'intégration de facteurs météorologiques au sein du cadre prédictif : a) Erreur globale ($RMSE_{global}$) en fonction de l'horizon de prévision b) Distribution cumulée de $MAPE_{TimeStep}$ à l'horizon 30 minutes c) Distribution cumulée de $MAPE_{TimeStep}$ à l'horizon 1h30 d) Distribution du pourcentage d'erreur en conditions humides à l'horizon 30 minutes e) Distribution du pourcentage d'erreur en conditions humides à l'horizon 3h.

Aucune des deux approches ne permet une amélioration globale de l'erreur de prévision sur des horizons de prévisions inférieurs à 1h30 comme le met en évidence la Figure 6.11 a). Au-delà de 2h, l'assimilation permet une légère réduction de l'erreur due, en grande partie, aux performances de la méthode météo-sensible, qui prend l'avantage sur MKNN en proposant une erreur fixe d'un peu plus de 3 km/h. Néanmoins, le seuil de 3 km/h n'étant pas atteint avant l'horizon 2 heures par MKNN, un léger bénéfice de l'approche par assimilation est tout de même à noter dès 1h30. Ce léger avantage est manifeste sur la distribution cumulée des erreurs par pas de temps proposée en Figure 6.11 c).

Finalement, une analyse plus localisée révèle que le procédé à coefficients correctifs améliore rarement les performances de prévision en conditions humides durant les périodes de forts trafic, mais détériore grandement l'erreur durant les phases nocturnes (Figure 6.11 d) & e)). L'explication trouve sa source dans la différence de comportement entre méthodes prédictives linéaires et non-linéaires. L'erreur des approches linéaires possède des biais systématiques expliqués par un même facteur qu'il est impossible de retrouver sur les approches non-linéaires. L'approche par assimilation se révèle efficace durant les périodes nocturnes, mais irrégulière lors des pics de trafic. L'assimilation tire donc bénéfice de l'usage de deux modèles prédictifs distincts et non de l'introduction d'informations météorologiques.

Ces résultats confirment le défaut de quantification des effets de la météorologie sur la vitesse du flux et l'erreur de prévision à partir des seules données disponibles. À l'heure actuelle, la quantification des effets reste inaccessible, prévenant alors toute intégration de facteurs météorologiques dans la structure prédictive, mais la méthode par assimilation promet de belles perspectives en cas d'affinement des informations météorologiques. Une analyse plus fine, menée sur une base de données mieux fournie en conditions dégradées et intégrant l'état des chaussées ou les événements météorologiques passés, constitue un projet à entreprendre en vue d'affiner les conclusions. En outre, une réduction du pas temporel d'agrégation, fixé à 30 minutes, contribuerait à limiter l'effet de lissage effaçant les différences dues à la météorologie.

6.5 Discussion et perspectives

► Synthèse

Cet ultime chapitre d'ouverture à la prise en considération de facteurs exogènes météorologiques propose une analyse statistique, puis une étude ciblée des liens existant entre phénomènes météorologiques et conditions de trafic.

Pour la première fois, les données météorologiques exploitées sont directement mesurées sur les infrastructures routières et apportent une information locale, *a priori* dénuée de biais temporels ou spatiaux. Une grande variété d'indicateurs météorologiques sont disponibles et analysés au regard de leur contribution.

Les phénomènes pluvieux étant les plus récurrents et impactants, seule l'intensité de pluie possède un effet statistiquement mesurable. À partir des 3 mois observés, les autres facteurs mesurés n'atteignent que trop rarement des valeurs associées à des conditions dégradées, ce qui empêche de conclure du point de vue statistique. Néanmoins, le cas du facteur visibilité montre qu'une légère hausse de la quantité de conditions dégradées (en novembre 2013) suffit pour l'observation d'effets statistiquement confirmés.

En ce qui concerne la quantification des effets sur le trafic, une grande variabilité de la sensibilité aux précipitations est observée sur les profils de vitesses en fonction des sections étudiées et suivant les typologies de jours considérés. Néanmoins, pour une majorité de sections, les précipitations accentuent les congestions du pic du matin. Si la forte variabilité suivant les sections et périodes temporelles limite les conclusions statistiques, les analyses locales dévoilent,

lors de pluies intenses ou légères mais continues, des chutes de vitesse de près de 5 km/h pour des sections expérimentant des vitesses de 90 km/h. De façon similaire, l'impact sur l'erreur de prévision est observé lors d'études locales, mais effacé lors d'agrégations temporelles.
L'intégration de facteurs météorologiques au sein du procédé de prévision reste très dépendante de la quantification d'impact, mais des stratégies intéressantes sont proposées et testées.

▶ Limites

Ce chapitre pointe la difficulté de quantifier les effets de la météorologie sur le trafic et contribue à expliquer la grande variété de conclusions observées dans la littérature suivant les cas d'étude. Ainsi, la quantification des effets météorologiques sur la vitesse du flux et, par extension, l'erreur de prévision s'avère être, pour les faibles précipitations observées, extrêmement délicate. Les tendances sont difficiles à déceler, du fait de l'agrégation temporelle, de la discrétisation en catégorie des informations et de la quantité limitée de conditions dégradées observées. De plus, l'analyse locale, menée sur des journées similaires, laisse supposer que la confrontation directe des informations météorologiques et trafic ne suffit pas à saisir toute la finesse de l'impact météorologique. La mesure de l'effet d'accumulation des précipitations est sans doute à prendre en considération.

▶ Perspectives

Outre la sélection de conditions météorologiques moins clémentes pour l'obtention de phénomènes plus marqués et l'usage de facteurs météorologiques continus, les perspectives d'étude doivent s'orienter vers :

- la prise en considération des événements météorologiques antérieurement expérimentés par l'infrastructure pour caractérisation de l'état de la chaussée ;
- la classification par durée et intensité des événements météorologiques pour mieux cibler les effets et quantifier pour chaque classe les impacts sur le trafic ;
- l'intégration, sous diverses approches, au sein du cadre prédictif d'informations météorologiques plus riches tirées d'une meilleure quantification des effets.

Les perspectives d'amélioration de l'étude actuelle sont variées et promettent une caractérisation fine des effets météorologiques. Cependant, la question de l'intégration de tels phénomènes au sein du cadre prédictif reste entière, mais certaines propositions introduites sont encourageantes. L'approfondissment et l'affinement du procédé d'intégration par assimilation pourrait aboutir à des performances intéressantes. La principale difficulté consiste à affiner la pondération entre modèle sec et humide pour favoriser, à tout instant, le meilleur modèle en fonction des conditions de trafic et météorologiques expérimentées.

Conclusion générale et perspectives

Synthèse des contributions

Le cheminement suivi tout au long de cette thèse vise à fournir des éléments contribuant à répondre à trois grands enjeux de la prévision du trafic : les nouvelles sources de données issues des ITS (Systèmes de Transports Intelligents) sont-elles exploitables pour une application à la prévision court-terme du trafic ? À quel prix peut-on parvenir à des modèles de prévision plus parcimonieux ou efficaces ? Comment parvenir à une meilleure résilience des modèles face à des événements inhabituels ?

Ces trois problématiques ont été abordées à divers égards en faisant preuve d'un réel souci de reproductibilité des résultats. Les analyses sont menées depuis diverses sources de données, à de multiples horizons de prévision, grâce à une grande variété de méthodes et sur différents réseaux aux dynamiques urbaines à interurbaines. La chance d'avoir disposé d'une telle diversité de cas d'étude aboutit à des résultats confirmés ou modérés suivant les échelles temporelles, spatiales et la précision des données. Cette modulation des résultats en fonction des circonstances constitue une grande richesse de cette étude et contribue à ancrer pleinement les conclusions dans la réalité du terrain connue des exploitants de réseaux routiers.

L'analyse multi-sources accroît non seulement la diversité des cas d'étude, mais évalue également le potentiel des technologies à rendre la prévision court-terme possible. En tant que nouvelles technologies de recueil sur le marché, l'usage des données Bluetooth ou issues de GPS proposent des perspectives de déploiements importantes. Tandis que les travaux de Master, rappelés dans le chapitre 1, ont montré la capacité des données Bluetooth à restituer les principales caractéristiques spatiales et temporelles du trafic, l'opportunité de prévoir vitesses et débits est estimée au cours de cette thèse. Les exigences de la prévision en temps-réel quant à la représentativité des données sont extrêmement élevées, puisqu'il s'agit de pouvoir saisir la finesse des variations pour anticiper celles à venir. Dans le contexte des données fournies par la ville de Brisbane, l'usage des données Bluetooth en entrée des algorithmes de prévision s'avère limité et conduit à des erreurs de prévision conséquentes. Les difficultés sont liées aux variations du taux de pénétration et à la latence de détection. En revanche, les données traceurs

sur la zone nantaise fournissent des vitesses moyennes spatiales de qualité suffisante pour la pratique de prévison sur de courts horizons.

Une analyse fine de l'état de l'art sur la prévision du trafic nous a conduit à observer que, malgré les évolutions des méthodes face aux enjeux des grandes bases de données, aucun effort n'est mené quant à la délimitation de l'objectif de prévision. Malgré quelques contributions sur la recherche de parcimonie ou d'unification des modèles, la fonction objectif se concentre sur la prévision d'une unique section du réseau. Pourtant une grande parcimonie des modèles est accessible en unifiant spatialement les modèles de prévision via des considérations à l'échelle du réseau. Le chapitre 2 pose les bases d'un cadre prédictif adapté à une approche globale garantissant parcimonie, rapidité d'exécution et des performances de prévision au moins équivalentes aux approches usuelles, dites locales. De plus, l'usage d'approches globales n'introduit pas de disparité spatiale visible de l'erreur de prévision, ce qui garantit une bonne qualité de prévision pour toute section. Les performances comparées des approches locales et globales ont été, dans un premier temps, évaluées depuis une méthode de régression des K plus proches voisins nécessitant peu d'adaptations. Afin de garantir la significativité des performances estimées, diverses méthodes de prévision ont été adaptées, puis appliquées en maintenant un contexte d'application commun aux approches globales comme locales. L'analyse comparative du chapitre 3 est éloquente à plusieurs titres. Outre la confirmation de l'équivalence entre approche locale et globale en terme d'erreur de prévision, chacune des méthodes dévoile son domaine spatial ou temporel de prédilection. L'analyse multi-échelle et multi-horizon fait ressortir des tendances et typologies d'usage des modèles exploitables dans une perspective de prévision ensembliste.

L'introduction d'un nouveau cadre prédictif propice au passage à l'échelle du réseau nécessite des adaptations méthodologiques, mais offre également de nouvelles opportunités à évaluer quant aux performances de prévision et de résilience. Les modifications méthodologiques et l'analyse des performances se révèlent riches en contributions, illustrées en Figure 6.12.

En ce qui concerne l'adaptation méthodologique, différents procédés de prévision s'appuyant sur des méthodes à noyaux ont été mis en œuvre et adaptés à l'approche globale. Les régressions non linéaires (KNN ou Kernel) nécessitent peu d'adaptations. En revanche, le chapitre 3 met en évidence les modifications de la fonction objectif pour application de l'approche globale depuis l'une des meilleures méthodes de prévision actuelles : la Régression à Vecteurs Supports (SVR). La méthode multi-variée résultante (MSVR) est directement tirée de la littérature informatique et a été modifiée pour une première application au trafic. L'analyse des erreurs de prévision montre la bonne performance de cette méthode en particulier quand l'horizon de prévision croît.
Au cours du chapitre 4, dans une optique d'optimisation de l'approche globale, un effort a été porté sur l'intégration fine des relations spatio-temporelles entre les sections du réseau. Si l'approche globale, introduite au chapitre 2, assure d'ores et déjà une bonne résilience en conservant l'intégralité des informations portées par les sections, une caractérisation plus fine

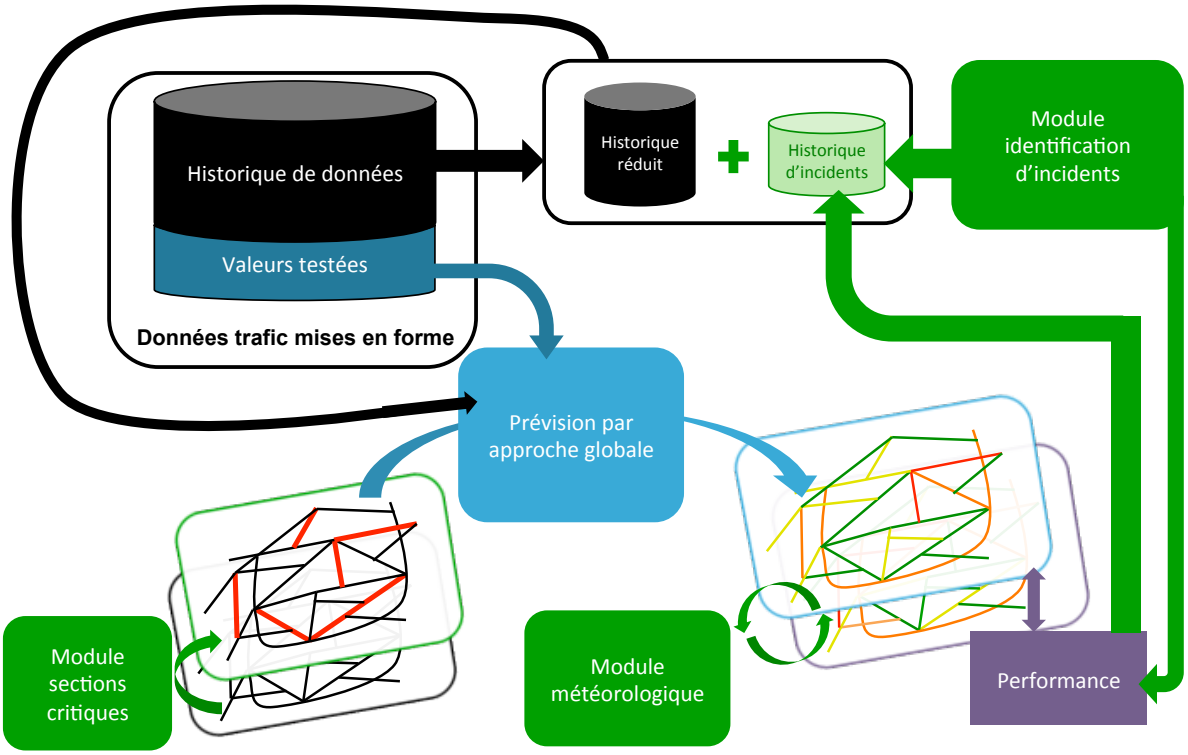


Fig. 6.12.: Illustration des principales contributions et objets d'étude rencontrés au cours de cette thèse

de la structure spatio-temporelle aboutit à une limitation du fléau de la dimension et un gain de temps d'exécution conséquent pour des performances légèrement améliorées. Le procédé employé mène de surcroît à une méthodologie d'identification des sections critiques en fonction des périodes temporelles pouvant être étendue à un processus dynamique. L'hypothèse du lien existant entre comportement local des sections et configuration de la structure spatio-temporelle est alors confirmée. À terme, une meilleure appréhension du fonctionnement du réseau est possible par interprétation de l'évolution des sections les plus critiques. La confrontation de différents critères de criticité des sections montre que la caractérisation introduite garantit une meilleure résilience.

En vue de parfaire la caractérisation de l'approche globale, l'intérêt de l'étude se porte finalement sur un aspect mettant souvent à mal les procédés usuels de prévision, à savoir les événements non récurrents. La problématique est traitée sous deux aspects : la détection et l'estimation d'impacts de comportements inhabituels perturbant le fonctionnement nominal du réseau, puis la potentielle prise en considération, en amont, de facteurs réputés pour leur impact sur le trafic.
Un procédé original de détection des incidents est développé au cours du chapitre 5. Il consiste à changer d'indicateur pour la détection en passant des états de trafic usuellement étudiés à un indicateur tiré d'une mesure d'entropie issue d'un procédé de classification. Les phases récurrentes et non récurrentes sont mieux distinguées et permettent une meilleure identificaton. Les événements identifiés par un tel procédé sont alors évalués au regard de leur impact sur

l'erreur de prévision et intégrés ou non à un historique permanent d'incidents. La présence de ce module d'identification d'incidents conduit à un filtrage de l'historique garantissant à long terme le maintien et l'amélioration des performances de prévision de l'approche globale. Le gain constaté est variable suivant les horizons de prévision et ciblé sur certains phénomènes, mais présent.
Enfin, la mesure des effets météorologiques sur le trafic et la possible intégration d'un tel facteur au sein du cadre prédictif ont été envisagées en chapitre 6. La confrontation directe des mesures de trafic et météorologiques (intensité de précipitation, force des rafales par exemple) a permis l'identification d'effets localisés par comparaison dans des contextes similaires. Néanmoins, à partir des données disponibles, la granularité temporelle des informations de trafic et météorologiques, ainsi que la quantité de données disponibles contraignent la significativité statistique des analyses. En effet, l'agrégation spatiale ou temporelle des phénomènes en vue de généralisation et quantification des impacts efface les différences observées localement. Le regroupement selon les seules mesures météorologiques usuelles ne paraît dès lors pas constituer le facteur principal d'impact, lequel résulte sans doute des conditions antérieures, de la durée et intensité des phénomènes expérimentés. L'intégration des phénomènes météorologiques nécessite donc une analyse plus fine et moins naïve. Néanmoins, les opportunités d'intégration au cadre prédictif restent prometteuses, puisque diverses approches sont envisageables.

Limites et Perspectives

Les travaux présentés ouvrent le champ à des perspectives de divers ordres.

L'approche globale a mis en avant l'opportunité d'aboutir à des prévisions rapides et efficaces à l'échelle de tout un réseau. Cependant, l'analyse du réseau urbain, dont une portion est tronquée, confirme que l'erreur de prévision reste dépendante de la taille, de la découpe et de la structure du réseau. La question du dimensionnement des réseaux à considérer constitue une perspective particulièrement intéressante. En effet, s'il a été possible de montrer, sur des réseaux bien délimités, l'opportunité de généraliser les paramètres du modèle prédictif pour toutes les sections, qu'en est-il lorsque le réseau est plus étendu ou mal découpé ? De même, la recherche du découpage optimal constitue en soi une question de recherche abordée, entre autres, dans les travaux de Geroliminis *et al.* [39] sur le Diagramme Fondamental Macroscopique. Un approfondissement de ces aspects constitue une thématique de recherche pertinente.

Concernant l'analyse du potentiel de l'approche globale, divers aspects ont été négligés ou pourraient être affinés et améliorés.
Le traitement des données manquantes à l'issue du recueil d'informations n'est que très peu abordé. Comme pour tout processus de fouille, un procédé de complétion des informations est développé, mais le traitement en temps-réel des ressources manquantes [103] n'est pas envisagé. La question des données manquantes est propre au contexte actuel de prévision, mais devrait peu à peu disparaître avec le développement du phénomène *Big Data* dans le trafic et

la massification des données. Le pari d'un déploiement de grande envergure des technologies est pris, conduisant à une forte croissance du taux de pénétration. Le réel problème porterait alors moins sur les données manquantes que sur la qualification de ces masses de données. Néanmoins, l'approche globale se prêterait facilement à des extensions traitant les données manquantes ou les capteurs défaillants. La sélection de variables par identification des sections critiques a montré que l'intégration de l'ensemble des informations n'est pas nécessaire pour prévoir à l'échelle du réseau, donc que l'absence d'informations sur certaines sections n'est pas systématiquement problématique. Le potentiel des sections critiques, complété par les démarches de prévision à partir de données tronquées [103], peut conduire à un procédé original assurant l'efficacité de toute approche globale en dépit de données manquantes.
En termes de prétraitement et transformation des données d'entrée, aucun procédé n'est testé au cours de cette étude. La décomposition du signal en ondelettes constitue une perspective intéressante à évaluer, d'autant que certaines composantes du signal sont sans doute communes à toutes les sections et d'autres propres à une influence locale. Outre l'application d'une décomposition en ondelettes au cadre prédictif introduit, la classification des composantes contribuerait pleinement à une meilleure appréhension de la dynamique du réseau. L'étude de ces ondelettes est susceptible d'influencer aussi bien l'identification de sections critiques que la détection d'incidents.
Enfin, la mise en œuvre de l'approche globale est évaluée pour les seules méthodes à noyaux, mais est envisageable pour des approches paramétriques ou à Réseaux de Neurones moyennant des adaptations à étudier et clarifier. De tels travaux ont d'ores et déjà été initiés avec la tenue d'un stage encadré par nos soins en collaboration avec VirginiaTech.

L'analyse des données FCD et Bluetooth s'est certes avérée riche, mais l'intégralité de leur potentiel n'est pas encore exploitée. La granularité du réseau, donc la dimension des sections pourrait être réduite à condition de disposer de suffisamment d'informations trafic, ce qui requiert un taux de pénétration plus important. En affinant les sections, l'approche globale pourrait concurrencer un modèle microscopique et les discrétisations numériques associées [32]. La discrétisation spatiale et temporelle pourrait en effet être réduite au point de saisir les dynamiques de type remontées de congestion. L'affinement de la granularité spatiale et temporelle grâce à la croissance du taux d'échantilonnage des FCD promet une meilleure représentativité et une amélioration des performances. Un tel affinement serait également profitable à l'intégration de facteurs exogènes au trafic tels que la météorologie. Cette dernière mérite d'ailleurs une analyse plus approfondie mettant mieux en relief les paramètres affectant le trafic.

Une nouvelle perspective réside dans l'opportunité fournie par l'approche globale de développer un procédé de prévision de type ensembliste depuis des méthodes de prévision plus complexes que celles employées par les approches ensemblistes usuelles. En effet, le gain de temps d'exécution des méthodes de prévision par approche globale est conséquent et suffisant pour permettre leur exécution simultanée pour une prévision en temps-réel. Un tel processus entérinerait un peu plus le gain et l'intérêt apporté par l'adoption d'une approche globale.

Applications envisageables

Pour achever les perspectives de ces travaux, les applications concrètes sur des outils appartenant désormais au quotidien sont introduites. Waze propose une mise à jour des itinéraires en fonction des conditions de trafic, GoogleMaps propose un itinéraire dépendant des conditions de trafic, mais aucun n'est actuellement en mesure de proposer un itinéraire en fonction de conditions de trafic prévues réaffectant dynamiquement les véhicules sur l'itinéraire optimal. Les avantages seraient multiples : d'une part, l'itinéraire de départ serait optimisé par rapport à un temps de trajet prévu, d'autre part, en fonction des évolutions du trafic (pas de quinze minutes par exemple), les prévisions pourraient être mises à jour conduisant à une éventuelle réaffectation dynamique sur d'autres itinéraires. Contrairement à Waze, la prévision apporterait également un garde-fou au procédé de réaffectation, puisque la saturation de l'itinéraire alternatif ou de délestage serait analysée, puis anticipée pour proposer un autre itinéraire.
D'un point de vue pratique, la prévision des états de trafic pour chacune des sections serait extrêmement gourmande en puissance comme en temps de calcul. L'adoption de l'approche globale apporte une première réponse en vue de réduire la complexité temporelle et permettre l'évaluation sur de petits processeurs. En outre son fonctionnement par réseaux permet d'envisager les chemins alternatifs en limitant le besoin de nouvelles prévisions sur les sections environnantes. En revanche, un travail conséquent sur le choix d'itinéraire dynamique à partir de données prédites reste à fournir.

Articles de l'auteur et valorisation

Revues Internationales

Pierre-Antoine Laharotte, Romain Billot, Etienne Come, Latifa Oukhellou, Alfredo Nantes, and Nour-Eddin El Faouzi. Spatiotemporal analysis of bluetooth data : application to a large urban network.
IEEE Transactions on Intelligent Transportation Systems 16(3): 1439 – 1448, 2015.

Conférences Internationales

Pierre-Antoine Laharotte, Romain Billot, Nour-Eddin El Faouzi, and Hesham Rakha. Network-Wide Traffic State Prediction Using Bluetooth Data.
In Transportation Research Board 94th Annual Meeting (No. 15-3022), 2015.

Distinctions

Pierre-Antoine Laharotte. Analyse de l'apport des données Bluetooth pour la caractérisation du trafic : Traitement du réseau semi-urbain de Brisbane. ENTPE,
Prix international ABERTIS 2014 catégorie Master/PFE, Puerto Rico, 2014.

Pierre-Antoine Laharotte. Analyse de l'apport des données Bluetooth pour la caractérisation du trafic : Traitement du réseau semi-urbain de Brisbane. ENTPE,
Prix national ABERTIS 2014 catégorie Master/PFE, Paris, 2014.

Présentations

Présentations en conférence

Pierre-Antoine Laharotte, Romain Billot, Nour-Eddin El Faouzi, Latifa Oukhellou and Etienne Come. How to filter transportation modes with Bluetooth data?
4th International Symposium of Transport Simulation (ISTS), 2014.

Séminaires

Pierre-Antoine Laharotte. Use of Bluetooth Data for traffic characterization : Case study of the Brisbane's semi-urban network.
Abertis Semninar, Civil Engineering of BarcelonaTech, Universitat Politecnica de Catalunya, Barcelone, Espagne, 2014.

Rapports

Pierre-Antoine Laharotte. Introduction to a 2-steps filtering method.
Processing of ASFINAG Data: Filtering & Prediction, Projet VORAB, Vienne, Autriche 2014.

Articles soumis

Pierre-Antoine Laharotte, Romain Billot, and Nour-Eddin El Faouzi. Mining critical links for large scale transport network state prediction.

Pierre-Antoine Laharotte, Romain Billot, and Nour-Eddin El Faouzi. Detection of non-recurrent road traffic events based on clustering indicators,
ESANN17.

Bibliographie

[1] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2) :167–256, 2003.

[2] Jason Wohlgemuth and Mihaela Teodora Matache. *Small World Properties of Facebook Group Networks*. PhD thesis, University of Nebraska at Omaha.

[3] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. Complex systems analysis of series of blackouts : Cascading failure, critical points, and self-organization. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 17(2) :026103, 2007.

[4] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1) :47, 2002.

[5] Zengwang Xu and Daniel Z Sui. Small-world characteristics on transportation networks : a perspective from network autocorrelation. *Journal of Geographical Systems*, 9(2) :189–205, 2007.

[6] William L Garrison. Connectivity of the interstate highway system. *Papers in Regional Science*, 6(1) :121–137, 1960.

[7] Bruce Douglas Greenshields. The photographic method of studying traffic behavior. In *Highway Research Board Proceedings*, volume 13, 1934.

[8] Laharotte P.A. Analyse de l'apport des données bluetooth pour la caractérisation du trafic : Traitement du réseau semi-urbain de brisbane. Master's thesis, ENTPE, 2013.

[9] Gabriel Michau. *Link Dependent Origin-Destination Matrix Estimation : Nonsmooth Convex Optimisation with Bluetooth-Inferred Trajectories*. PhD thesis, Lyon, 2016.

[10] Shaun Quayle, Peter Koonce, Darryl DePencier, and Darcy Bullock. Arterial performance measures with media access control readers : Portland, oregon, pilot study. *Transportation Research Record : Journal of the Transportation Research Board*, (2192) :185–193, 2010.

[11] Francois Dion and Hesham Rakha. Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B : Methodological*, 40(9) :745–766, 2006.

[12] J Barceló, L Montero, L Marqués, and C Carmona. Travel time forecasting and dynamic od estimation in freeways based on bluetooth traffic monitoring. *Transportation Research Board 89th Annual Meeting*, 570, 2010.

[13] Darryl D Puckett and Michael J Vickich. Bluetooth®-based travel time/speed measuring systems development. Technical report, 2010.

[14] Yibing Wang and Markos Papageorgiou. Real-time freeway traffic state estimation based on extended kalman filter : a general approach. *Transportation Research Part B : Methodological*, 39(2) :141–167, 2005.

[15] Michael W Szeto and Denos C Gazis. Application of kalman filtering to the surveillance and control of traffic systems. *Transportation Science*, 6(4) :419–439, 1972.

[16] Denos Gazis and Chiu Liu. Kalman filtering estimation of traffic counts for two network links in tandem. *Transportation Research Part B : Methodological*, 37(8) :737–745, 2003.

[17] Yufei Yuan, Aurélien Duret, and Hans Van Lint. Mesoscopic traffic state estimation based on a variational formulation of the lwr model in lagrangian-space coordinates and kalman filter. *Transportation Research Procedia*, 10 :82–92, 2015.

[18] Kaveh Farokhi Sadabadi, Masoud Hamedi, and Ali Haghani. Evaluating moving average techniques in short-term travel time prediction using an avi data set. In *Transportation Research Board 89th Annual Meeting*, number 10-3144, 2010.

[19] Bahar Namaki Araghi, Kristian Skoven Pedersen, Lars Tørholm Christensen, Rajesh Krishnan, and Harry Lahrmann. Accuracy of travel time estimation using bluetooth technology : Case study limfjord tunnel aalborg. *International Journal of Intelligent Transportation Systems Research*, 13(3) :166–191, 2015.

[20] Bahar Namaki Araghi, Lars Tørholm Christensen, Rajesh Krishnan, and Harry Lahrmann. Application of bluetooth technology for mode-specific travel time estimation on arterial roads : Potentials and challenges. In *Proceedings from the Annual Transport Conference at Aalborg University*, 2012.

[21] Pierre-Antoine Laharotte, Romain Billot, Nour-Eddin El-Faouzi, Latifa Oukhellou, and Etienne Come. How to filter transportation modes with bluetooth data ? In *4th International Symposium of Transport Simulation (ISTS)*, 2014.

[22] Pierre-Antoine Laharotte, Romain Billot, Etienne Come, Latifa Oukhellou, Alfredo Nantes, and Nour-Eddin El Faouzi. Spatiotemporal analysis of bluetooth data : application to a large urban network. *IEEE Transactions on Intelligent Transportation Systems*, 16(3) :1439–1448, 2015.

[23] Chenyi Chen, Yin Wang, Li Li, Jianming Hu, and Zuo Zhang. The retrieval of intra-day trend and its influence on traffic prediction. *Transportation research part C : emerging technologies*, 22 :103–118, 2012.

[24] Jin Wang and Qixin Shi. Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory. *Transportation Research Part C : Emerging Technologies*, 27 :219–232, 2013.

[25] Aude Hofleitner, Ryan Herring, and Alexandre Bayen. Arterial travel time forecast with streaming data : A hybrid approach of flow modeling and machine learning. *Transportation Research Part B : Methodological*, 46(9) :1097–1122, 2012.

[26] Mohammed Elhenawy, Hao Chen, and Hesham A Rakha. Dynamic travel time prediction using data clustering and genetic programming. *Transportation Research Part C : Emerging Technologies*, 42 :82–98, 2014.

[27] Michael J Lighthill and Gerald Beresford Whitham. 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*, volume 229, pages 317–345. The Royal Society, 1955.

[28] GCK Wong and SC Wong. A multi-class traffic flow model–an extension of lwr model with heterogeneous drivers. *Transportation Research Part A : Policy and Practice*, 36(9) :827–841, 2002.

[29] A Kotsialos and M Papageorgiou. The importance of traffic flow modeling for motorway traffic control. *Networks and Spatial Economics*, 1(1-2) :179–203, 2001.

[30] Harold J Payne. Models of freeway traffic and control. *Mathematical models of public systems*, 1971.

[31] Dirk Helbing. Gas-kinetic derivation of navier-stokes-like traffic equations. *Physical Review E*, 53(3) :2366, 1996.

[32] Carlos F Daganzo. The cell transmission model : A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B : Methodological*, 28(4) :269–287, 1994.

[33] Yibing Wang, Markos Papageorgiou, and Albert Messmer. Renaissance–a unified macroscopic model-based approach to real-time freeway network traffic surveillance. *Transportation Research Part C : Emerging Technologies*, 14(3) :190–212, 2006.

[34] Yufei Yuan, JWC Van Lint, R Eddie Wilson, Femke van Wageningen-Kessels, and Serge P Hoogendoorn. Real-time lagrangian traffic state estimator for freeways. *IEEE Transactions on Intelligent Transportation Systems*, 13(1) :59–70, 2012.

[35] Aude Hofleitner, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. Learning the dynamics of arterial traffic from probe data using a dynamic bayesian network. *IEEE Transactions on Intelligent Transportation Systems*, 13(4) :1679–1693, 2012.

[36] Martin Treiber and Arne Kesting. Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns. *Transportation research part C : emerging technologies*, 21(1) :31–41, 2012.

[37] James Haworth. *Spatio-temporal forecasting of network data*. PhD thesis, UCL (University College London), 2014.

[38] Hao Liu. *Travel time prediction for urban networks*. TU Delft, Delft University of Technology, 2008.

[39] Nikolas Geroliminis and Carlos F Daganzo. Existence of urban-scale macroscopic fundamental diagrams : Some experimental findings. *Transportation Research Part B : Methodological*, 42(9) :759–770, 2008.

[40] Serge P Hoogendoorn and Piet HL Bovy. 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) :283–303, 2001.

[41] Masako Bando, Katsuya Hasebe, Akihiro Nakayama, Akihiro Shibata, and Yuki Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, 51(2) :1035, 1995.

[42] Rui Jiang, Qingsong Wu, and Zuojin Zhu. Full velocity difference model for a car-following theory. *Physical Review E*, 64(1) :017101, 2001.

[43] Hesham Rakha and Brent Crowther. Comparison of greenshields, pipes, and van aerde car-following and traffic stream models. *Transportation Research Record : Journal of the Transportation Research Board*, (1802) :248–262, 2002.

[44] Boris S Kerner, Sergey L Klenov, Gerhard Hermanns, and Michael Schreckenberg. Effect of driver over-acceleration on traffic breakdown in three-phase cellular automaton traffic flow models. *Physica A : Statistical Mechanics and its Applications*, 392(18) :4083–4105, 2013.

[45] M Cremer and J Ludwig. A fast simulation model for traffic flow on the basis of boolean operations. *Mathematics and Computers in Simulation*, 28(4) :297–303, 1986.

[46] Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. *Journal de physique I*, 2(12) :2221–2229, 1992.

[47] Jaime Barceló and Jordi Casas. Dynamic network simulation with aimsun. In *Simulation approaches in transportation analysis*, pages 57–98. Springer, 2005.

[48] Fabien BONNEFOI, Julien SOULA, Ludovic Leclercq, and Cécile Bécarie. Simulation multi-echelle du trafic routier.

[49] Steven L Jones, Andrew J Sullivan, Naveen Cheekoti, Michael D Anderson, and D Malave. Traffic simulation software comparison study. *UTCA Report*, 2217, 2004.

[50] I. Prigogine, R. Herman, and R.S. Schechter. Kinetic theory of vehicular traffic. *Systems, Man and Cybernetics, IEEE Transactions on*, 2(2), 1972.

[51] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis : forecasting and control*. John Wiley & Sons, 2015.

[52] Mohammed S Ahmed and Allen R Cook. *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*. Number 722. 1979.

[53] Mohammad M Hamed, Hashem R Al-Masaeid, and Zahi M Bani Said. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, 121(3) :249–254, 1995.

[54] Jianhua Guo and Brian L Smith. *Short term speed variance forecasting using linear stochastic modeling of univariate traffic speed series*. University of Virginia Center for Transportation Studies, 2007.

[55] Brian L Smith, Billy M Williams, and R Keith Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C : Emerging Technologies*, 10(4) :303–321, 2002.

[56] Ning Zhang, Yunlong Zhang, and Haiting Lu. Seasonal autoregressive integrated moving average and support vector machine models : prediction of short-term traffic flow on freeways. *Transportation Research Record : Journal of the Transportation Research Board*, (2215) :85–92, 2011.

[57] Xinyu Min, Jianming Hu, Qi Chen, Tongshuai Zhang, and Yi Zhang. Short-term traffic flow forecasting of urban network based on dynamic starima model. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6. IEEE, 2009.

[58] Ryan Herring, Aude Hofleitner, Saurabh Amin, T Nasr, A Khalek, Pieter Abbeel, and Alexandre Bayen. Using mobile phones to forecast arterial traffic through statistical learning. In *89th Transportation Research Board Annual Meeting, Washington DC*, 2010.

[59] Xinyu Min, Jianming Hu, and Zuo Zhang. Urban traffic network modeling and short-term traffic flow forecasting based on gstarima model. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1535–1540. IEEE, 2010.

[60] Lykourgos Tsirigotis, Eleni I Vlahogianni, and Matthew G Karlaftis. Does information on weather affect the performance of short-term traffic forecasting models ? *International Journal of Intelligent Transportation Systems Research*, 10(1) :1–10, 2012.

[61] Tiep Mai, Bidisha Ghosh, and Simon Wilson. Multivariate short-term traffic flow forecasting using bayesian vector autoregressive moving average model. In *Transportation Research Board 91st Annual Meeting*, number 12-3728, 2012.

[62] Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE Transactions on Intelligent Transportation Systems*, 10(2) :246–254, 2009.

[63] Srinivasa Ravi Chandra and Haitham Al-Deek. Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems*, 13(2) :53–72, 2009.

[64] WY Szeto, Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Multivariate traffic forecasting technique using cell transmission model and sarima model. *Journal of Transportation Engineering*, 135(9) :658–667, 2009.

[65] Joe Whittaker, Simon Garside, and Karel Lindveld. Tracking and predicting a network traffic process. *International Journal of Forecasting*, 13(1) :51–61, 1997.

[66] Anthony Stathopoulos and Matthew G Karlaftis. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C : Emerging Technologies*, 11(2) :121–135, 2003.

[67] Hao Liu, Henk van Zuylen, Hans van Lint, and Maria Salomons. Predicting urban arterial travel time with state-space neural networks and kalman filters. *Transportation Research Record : Journal of the Transportation Research Board*, (1968) :99–108, 2006.

[68] Byoungjo Yoon and Hyunho Chang. Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting. *Journal of Transportation Engineering*, 140(7) :04014027, 2014.

[69] Marco Lippi, Matteo Bertini, and Paolo Frasconi. Short-term traffic flow forecasting : An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 14(2) :871–882, 2013.

[70] Howard R Kirby, Susan M Watson, and Mark S Dougherty. Should we use neural networks or statistical models for short-term motorway traffic forecasting ? *International Journal of Forecasting*, 13(1) :43–50, 1997.

[71] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[72] Hao Chen, Hesham A Rakha, and Catherine C McGhee. Dynamic travel time prediction using pattern recognition. In *20th World Congress on Intelligent Transportation Systems, Tokyo, Japan*, 2013.

[73] Mark Dougherty. A review of neural networks applied to transport. *Transportation Research Part C : Emerging Technologies*, 3(4) :247–260, 1995.

[74] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Short-term traffic forecasting : Where we are and where we're going. *Transportation Research Part C : Emerging Technologies*, 43 :3–19, 2014.

[75] Eric Davalo and Patrick Naïm. *Des réseaux de neurones*. Eyrolles, 1990.

[76] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. Traffic flow prediction with big data : a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2) :865–873, 2015.

[77] Jwe Van Lint and H Van Zuylen. Monitoring and predicting freeway travel time reliability : using width and skew of day-to-day travel time distribution. *Transportation Research Record : Journal of the Transportation Research Board*, (1917) :54–62, 2005.

[78] Kit Yan Chan, Tharam S Dillon, Jaipal Singh, and Elizabeth Chang. Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg–marquardt algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 13(2) :644–654, 2012.

[79] Kit Yan Chan, Tharam Dillon, Elizabeth Chang, and Jaipal Singh. Prediction of short-term traffic variables using intelligent swarm-based neural networks. *IEEE Transactions on Control Systems Technology*, 21(1) :263–274, 2013.

[80] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Optimized and meta-optimized neural networks for short-term traffic flow prediction : a genetic approach. *Transportation Research Part C : Emerging Technologies*, 13(3) :211–234, 2005.

[81] Eleni I Vlahogianni. Prediction of non-recurrent short-term traffic patterns using genetically optimized probabilistic neural networks. *Operational Research*, 7(2) :171–184, 2007.

[82] Shiliang Sun, Rongqing Huang, and Ya Gao. Network-scale traffic modeling and forecasting with graphical lasso and neural networks. *Journal of Transportation Engineering*, 138(11) :1358–1367, 2012.

[83] Tao Ji, Qingle Pang, and Xinyun Liu. Study of traffic flow forecasting based on genetic neural network. In *Sixth International Conference on Intelligent Systems Design and Applications*, volume 1, pages 960–965. IEEE, 2006.

[84] Shiliang Sun and Changshui Zhang. The selective random subspace predictor for traffic flow forecasting. *IEEE Transactions on intelligent transportation systems*, 8(2) :367–373, 2007.

[85] Guy Leshem and Yaacov Ritov. Traffic flow prediction using adaboost algorithm with random forests as a weak learner. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 19, pages 193–198. Citeseer, 2007.

[86] Benjamin Hamner. Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. In *2010 IEEE International Conference on Data Mining Workshops*, pages 1357–1359. IEEE, 2010.

[87] Leo Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.

[88] Leo Brieman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. Classification and regression trees. *Monterey, CA : Wadsworth & Brooks*, 1984.

[89] Wenhao Huang, Guojie Song, Haikun Hong, and Kunqing Xie. Deep architecture for traffic flow prediction : deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems*, 15(5) :2191–2201, 2014.

[90] Shiliang Sun, Changshui Zhang, and Yi Zhang. Traffic flow forecasting using a spatio-temporal bayesian network predictor. In *International Conference on Artificial Neural Networks*, pages 273–278. Springer, 2005.

[91] Catriona M Queen and Casper J Albers. Forecasting traffic flows in road networks : A graphical dynamic model approach. In *Proceedings of the 28th International Symposium of Forecasting, International Institute of Forecasters*, 2008.

[92] Changshui Zhang, Shiliang Sun, and Guoqiang Yu. A bayesian network approach to time series forecasting of short-term traffic flows. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pages 216–221. IEEE, 2004.

[93] Xing-Yu Chen, Hsing-Kuo Pao, and Yuh-Jye Lee. Efficient traffic speed forecasting based on massive heterogenous historical data. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 10–17. IEEE, 2014.

[94] Zuduo Zheng and Dongcai Su. Short-term traffic volume forecasting : A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transportation Research Part C : Emerging Technologies*, 43 :143–157, 2014.

[95] Michael May, Dirk Hecker, Christine Körner, Simon Scheider, and Daniel Schulz. A vector-geometry based spatial knn-algorithm for traffic frequency predictions. In *2008 IEEE International Conference on Data Mining Workshops*, pages 442–447. IEEE, 2008.

[96] Taehyung Kim, Hyoungsoo Kim, and David J Lovell. Traffic flow forecasting : overcoming memoryless property in nearest neighbor non-parametric regression. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 965–969. IEEE, 2005.

[97] R Keith Oswald, William T Scherer, and Brian L Smith. Traffic flow forecasting using approximate nearest neighbor nonparametric regression. *Final project of ITS Center project : Traffic forecasting : non-parametric regressions*, 2000.

[98] Hongyu Sun, Henry Liu, Heng Xiao, Rachel He, and Bin Ran. Use of local linear regression model for short-term traffic forecasting. *Transportation Research Record : Journal of the Transportation Research Board*, (1836) :143–150, 2003.

[99] Xiaoyan Gong and Feiyue Wang. Three improvements on knn-npr for traffic flow forecasting. In *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, pages 736–740. IEEE, 2002.

[100] Mohammad Arif Rasyidi, Jeongmin Kim, and Kwang Ryel Ryu. Short-term prediction of vehicle speed on main city roads using the k-nearest neighbor algorithm. *Journal of Intelligence and Information Systems*, 20(1) :121–131, 2014.

[101] Brian L Smith and Michael J Demetsky. Traffic flow forecasting : comparison of modeling approaches. *Journal of transportation engineering*, 123(4) :261–266, 1997.

[102] Nour-Eddin El Faouzi and Pascal Sarda. Régression non paramétrique : méthode du noyau et régression locale pondérée. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 318(1) :69–72, 1994.

[103] James Haworth and Tao Cheng. Non-parametric regression for space-time forecasting under missing data. *Computers, Environment and Urban Systems*, 36(6) :538–550, 2012.

[104] Sébastien Blandin, Laurent El Ghaoui, and Alexandre Bayen. 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*, pages 4360–4365. IEEE, 2009.

[105] Shiliang Sun and Qiaona Chen. Kernel regression with a mahalanobis metric for short-term traffic flow forecasting. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 9–16. Springer, 2008.

[106] Ziwen Leng, Junwei Gao, Yong Qin, Xin Liu, and Jing Yin. Short-term forecasting model of traffic flow based on grnn. In *2013 25th Chinese Control and Decision Conference (CCDC)*, pages 3816–3820. IEEE, 2013.

[107] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3) :273–297, 1995.

[108] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[109] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[110] Wei-Chiang Hong. Traffic flow forecasting by seasonal svr with chaotic simulated annealing algorithm. *Neurocomputing*, 74(12) :2096–2107, 2011.

[111] Wei-Chiang Hong, Yucheng Dong, Feifeng Zheng, and Chien-Yuan Lai. Forecasting urban traffic flow by svr with continuous aco. *Applied Mathematical Modelling*, 35(3) :1282–1291, 2011.

[112] Chun-Hsin Wu, Jan-Ming Ho, and Der-Tsai Lee. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, 5(4) :276–281, 2004.

[113] Lelitha Vanajakshi and Laurence R Rilett. 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*, pages 194–199. IEEE, 2004.

[114] Yunlong Zhang and Yuanchang Xie. Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record : Journal of the Transportation Research Board*, (2024) :92–99, 2008.

[115] Manoel Castro-Neto, Young-Seon Jeong, Myong-Kee Jeong, and Lee D Han. Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert systems with applications*, 36(3) :6164–6173, 2009.

[116] DG Krige. *A statistical approach to some mine valuation and allied problems on the Witwatersrand : By DG Krige*. 1951.

[117] Kit Yan Chan, Tharam S Dillon, and Elizabeth Chang. An intelligent particle swarm optimization for short-term traffic flow forecasting using on-road sensor systems. *IEEE Transactions on Industrial Electronics*, 60(10) :4714–4725, 2013.

[118] Wei-Chiang Hong, Yucheng Dong, Feifeng Zheng, and Shih Yung Wei. Hybrid evolutionary algorithms in a svr traffic flow forecasting model. *Applied Mathematics and Computation*, 217(15) :6733–6747, 2011.

[119] Daniel Boto-Giralda, Francisco J Díaz-Pernas, David González-Ortega, José F Díez-Higuera, Míriam Antón-Rodríguez, Mario Martínez-Zarzuela, and Isabel Torre-Díez. Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 25(7) :530–545, 2010.

[120] Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Random process model for urban traffic flow using a wavelet-bayesian hierarchical technique. *Computer-Aided Civil and Infrastructure Engineering*, 25(8) :613–624, 2010.

[121] Xiaomo Jiang and Hojjat Adeli. Dynamic wavelet neural network model for traffic flow forecasting. *Journal of transportation engineering*, 131(10) :771–779, 2005.

[122] Li Qu, Li Li, Yi Zhang, and Jianming Hu. Ppca-based missing data imputation for traffic flow volume : a systematical approach. *IEEE Transactions on intelligent transportation systems*, 10(3) :512–522, 2009.

[123] Ciprian Alecsandru and Sherif Ishak. Hybrid model-based and memory-based traffic prediction system. *Transportation Research Record : Journal of the Transportation Research Board*, (1879) :59–70, 2004.

[124] Shiliang Sun. Traffic flow forecasting based on multitask ensemble learning. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, pages 961–964. ACM, 2009.

[125] Tao Cheng, James Haworth, and Jiaqiu Wang. Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, 14(4) :389–413, 2012.

[126] Wanli Min and Laura Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C : Emerging Technologies*, 19(4) :606–616, 2011.

[127] Shanhua Wu, Zhongzhen Yang, Xiaocong Zhu, and Bin Yu. Improved k-nn for short-term traffic forecasting using temporal and spatial information. *Journal of Transportation Engineering*, 140(7) :04014026, 2014.

[128] Stephen Clark. Traffic prediction using multivariate nonparametric regression. *Journal of transportation engineering*, 129(2) :161–168, 2003.

[129] James Haworth and Tao Cheng. Graphical lasso for local spatio-temporal neighbourhood selection. In *Proceedings the GIS Research UK 22nd Annual Conference. Presented at the GISRUK*, pages 425–433, 2014.

[130] Yiannis Kamarianakis and Poulicos Prastacos. Space–time modeling of traffic flow. *Computers & Geosciences*, 31(2) :119–133, 2005.

[131] Yiannis Kamarianakis, Wei Shen, and Laura Wynter. Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso. *Applied stochastic models in business and industry*, 28(4) :297–315, 2012.

[132] Tao Cheng, Jiaqiu Wang, James Haworth, Benjamin Heydecker, and Andy Chow. A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *Geographical Analysis*, 46(1) :75–97, 2014.

[133] Justin Dauwels, Aamer Aslam, Muhammad Tayyab Asif, Xinyue Zhao, Nikola Mitro Vie, Andrzej Cichocki, and Patrick Jaillet. Predicting traffic speed in urban transportation subnetworks for multiple horizons. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pages 547–552. IEEE, 2014.

[134] Srinivasa Chandra and Haitham Al-Deek. Cross-correlation analysis and multivariate prediction of spatial time series of freeway traffic speeds. 2061 :64–76.

[135] Corrado De Fabritiis, Roberto Ragona, and Gaetano Valenti. Traffic estimation and prediction based on real time floating car data. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 197–203. IEEE.

[136] V. Fabrizi and R. Ragona. A pattern matching approach to speed forecasting of traffic networks. 6(3) :333–342.

[137] Marco Lippi, Matteo Bertini, and Paolo Frasconi. Collective traffic forecasting. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 259–273. Springer, 2010.

[138] Pierre-Antoine Laharotte, Romain Billot, El Faouzi, Hesham A Rakha, et al. Network-wide traffic state prediction using bluetooth data. In *Transportation Research Board 94th Annual Meeting*, number 15-3022, 2015.

[139] Mohammadreza Saeedmanesh and Nikolas Geroliminis. Clustering of heterogeneous networks with directional flows based on "snake" similarities. *Transportation Research Part B : Methodological*, 91 :250–269, 2016.

[140] Aude Hofleitner, Ryan Herring, Alexandre Bayen, Yufei Han, Fabien Moutarde, and Arnaud De La Fortelle. 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.

[141] Alessandra Pascale, Dimitrios Mavroeidis, and Hoang Thanh Lam. Spatiotemporal clustering of urban networks : Real case scenario in london. *Transportation Research Record : Journal of the Transportation Research Board*, (2491) :81–89, 2015.

[142] Shamas Ul Islam Bajwa, Edward Chung, and Masao Kuwahara. Performance evaluation of an adaptive travel time prediction model. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 1000–1005. IEEE.

[143] J. Y. Wang, K. I. Wong, and Y. Y. Chen. 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*, pages 582–587. IEEE.

[144] JWC Van Lint, SP Hoogendoorn, and Henk J van Zuylen. Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C : Emerging Technologies*, 13(5) :347–369, 2005.

[145] Yongsong Qin, Shichao Zhang, Xiaofeng Zhu, Jilian Zhang, and Chengqi Zhang. Semi-parametric optimization for missing data imputation. *Applied Intelligence*, 27(1) :79–88, 2007.

[146] Ming Zhong, Pawan Lingras, and Satish Sharma. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transportation Research Part C : Emerging Technologies*, 12(2) :139–166, 2004.

[147] Heribert Kirschfink. Knowledge-based system for the completion of traffic data. *European journal of operational research*, 71(2) :247–256, 1993.

[148] D Belomestny, V Jentsch, and M Schreckenberg. Completion and continuation of nonlinear traffic time series : a probabilistic approach. *Journal of Physics A : Mathematical and General*, 36(45) :11369, 2003.

[149] Yang Zhang and Yuncai Liu. Data imputation using least squares support vector machines in urban arterial streets. *IEEE Signal Processing Letters*, 16(5) :414–417, 2009.

[150] Daiheng Ni, John D Leonard, Angshuman Guin, and Chunxia Feng. Multiple imputation scheme for overcoming the missing values and variability issues in its data. *Journal of transportation engineering*, 131(12) :931–938, 2005.

[151] Jianwei Wang, Nan Zou, and Gang-Len Chang. Travel time prediction : Empirical analysis of missing data issues for advanced traveler information system applications. *Transportation Research Record : Journal of the Transportation Research Board*, (2049) :81–91, 2008.

[152] Huachun Tan, Guangdong Feng, Jianshuai Feng, Wuhong Wang, Yu-Jin Zhang, and Feng Li. A tensor-based method for missing traffic data completion. *Transportation Research Part C : Emerging Technologies*, 28 :15–27, 2013.

[153] Mark E Whitlock and Catriona M Queen. Modelling a traffic network with missing data. *Journal of Forecasting*, 19(7) :561–574, 2000.

[154] Haibo Chen, Susan Grant-Muller, Lorenzo Mussone, and Frank Montgomery. A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Computing & Applications*, 10(3) :277–286, 2001.

[155] Aliaga Frederic. Analyse des effets de la météorologie sur le trafic à partir de données bluetooth.

[156] Baradel Fabien. Méthode d'apprentissage statistique appliquée au trafic routier.

[157] Rob J Hyndman. Measuring forecast accuracy, 2014.

[158] Marco Dorigo, Vittorio Maniezzo, and Alberto Colorni. Ant system : optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1) :29–41, 1996.

[159] Vladimir Vapnik. Pattern recognition using generalized portrait method. *Automation and remote control*, 24 :774–780, 1963.

[160] Vapnik N Vladimir and V Vapnik. The nature of statistical learning theory, 1995.

[161] Matilde Sánchez-Fernández, Mario de Prado-Cumplido, Jerónimo Arenas-García, and Fernando Pérez-Cruz. Svm multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *IEEE transactions on signal processing*, 52(8) :2298–2307, 2004.

[162] Fernando Pérez-Cruz, Gustavo Camps-Valls, Emilio Soria-Olivas, Juan José Pérez-Ruixo, Aníbal R Figueiras-Vidal, and Antonio Artés-Rodríguez. Multi-dimensional function approximation and regression estimation. In *International Conference on Artificial Neural Networks*, pages 757–762. Springer, 2002.

[163] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[164] Devis Tuia, Jochem Verrelst, Luis Alonso, Fernando Pérez-Cruz, and Gustavo Camps-Valls. Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geoscience and Remote Sensing Letters*, 8(4) :804–808, 2011.

[165] Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International Conference on Computational Learning Theory*, pages 416–426. Springer, 2001.

[166] Parisel Camille. Panorama de méthodes d'apprentissage automatique appliquées à la prédiction de la circulation en milieu urbain.

[167] Yanru Zhang and Ali Haghani. A gradient boosting method to improve travel time prediction. *Transportation Research Part C : Emerging Technologies*, 58 :308–324, 2015.

[168] Michael GH Bell. A game theory approach to measuring the performance reliability of transport networks. *Transportation Research Part B : Methodological*, 34(6) :533–545, 2000.

[169] J.L. Sullivan, D.C. Novak, L. Aultman-Hall, and D.M. Scott. Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links : A link-based capacity-reduction approach. 44(5) :323–336.

[170] Darren M. Scott, David C. Novak, Lisa Aultman-Hall, and Feng Guo. Network robustness index : A new method for identifying critical links and evaluating the performance of transportation networks. 14(3) :215–227.

[171] Pamela M. Murray-Tuite and Hani S. Mahmassani. Methodology for determining vulnerable links in a transportation network. 1882(1) :88–96.

[172] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500) :2323–2326, 2000.

[173] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3 :1157–1182, 2003.

[174] Ron Kohavi. Wrappers for feature subset selection. 1996.

[175] Lei Yu and Huan Liu. Feature selection for high-dimensional data : A fast correlation-based filter solution. In *ICML*, volume 3, pages 856–863, 2003.

[176] Yiming Yang and Jan O Pedersen. A comparative study on feature selection in text categorization. In *ICML*, volume 97, pages 412–420, 1997.

[177] Fadi Dornaika, Elena Lazkano, and Basilio Sierra. Improving dynamic facial expression recognition with feature subset selection. *Pattern Recognition Letters*, 32(5) :740–748, 2011.

[178] Abhimanyu Das and David Kempe. Submodular meets spectral : Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1057–1064, 2011.

[179] Pratik Jawanpuria, Manik Varma, and Saketha Nath. On p-norm path following in multiple kernel learning for non-linear feature selection. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 118–126, 2014.

[180] Feng Nan, Joseph Wang, and Venkatesh Saligrama. Feature-budgeted random forest. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 1983–1991, 2015.

[181] Huang Xiao, Battista Biggio, Gavin Brown, Giorgio Fumera, Claudia Eckert, and Fabio Roli. Is feature selection secure against training data poisoning ? In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1689–1698, 2015.

[182] Matteo Cassotti and Francesca Grisoni. Variable selection methods : an introduction. 2012.

[183] Alan Miller. *Subset selection in regression*. CRC Press, 2002.

[184] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[185] James Kennedy. Particle swarm optimization. In *Encyclopedia of machine learning*, pages 760–766. Springer, 2011.

[186] Mohammad Goodarzi, Matheus P Freitas, and Richard Jensen. Ant colony optimization as a feature selection method in the qsar modeling of anti-hiv-1 activities of 3-(3, 5-dimethylbenzyl) uracil derivatives using mlr, pls and svm regressions. *Chemometrics and intelligent laboratory systems*, 98(2) :123–129, 2009.

[187] Riccardo Leardi. *Nature-inspired methods in chemometrics : genetic algorithms and artificial neural networks*, volume 23. Elsevier, 2003.

[188] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[189] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3) :432–441, 2008.

[190] Grzegorz Swirszcz and Aurelie C Lozano. Multi-level lasso for sparse multi-task regression. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 361–368, 2012.

[191] Xiaohui Chen, Xinghua Shi, Xing Xu, Zhiyong Wang, Ryan Mills, Charles Lee, and Jinbo Xu. A two-graph guided multi-task lasso approach for eqtl mapping. In *AISTATS*, volume 22, pages 208–217, 2012.

[192] Rochd El Mehdi, Quafafou Mohamed, and Aznag Mustapha. Encoding local correspondence in topic models. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, pages 602–609. IEEE, 2013.

[193] Li Li, Ding Wen, and Danya Yao. A survey of traffic control with vehicular communications. *IEEE Transactions on Intelligent Transportation Systems*, 15(1) :425–432, 2014.

[194] Katayoun Farrahi and Daniel Gatica-Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1) :3, 2011.

[195] Tâm Huynh, Mario Fritz, and Bernt Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 10–19. ACM, 2008.

[196] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision*, 79(3) :299–318, 2008.

[197] Raul Montoliu. Discovering mobility patterns on bicycle-based public transportation system by using probabilistic topic models. In *Ambient Intelligence-Software and Applications*, pages 145–153. Springer, 2012.

[198] Côme Etienne and Oukhellou Latifa. Model-based count series clustering for bike sharing system usage mining : A case study with the vélib'system of paris. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3) :39, 2014.

[199] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan) :993–1022, 2003.

[200] Francis Bach and Guillaume Obozinski. Modeles graphiques probabilistes. 2006.

[201] Thijs Westerveld and Arjen P de Vries. Generative probabilistic models for multimedia retrieval : query generation against document generation. *IEE Proceedings-Vision, Image and Signal Processing*, 152(6) :852–858, 2005.

[202] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[203] Todd K Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6) :47–60, 1996.

[204] Alexander Skabardonis, Pravin Varaiya, and Karl Petty. Measuring recurrent and nonrecurrent traffic congestion. (1856) :118–124.

[205] Howard J Payne and Samuel C Tignor. Freeway incident-detection algorithms based on decision trees with states. *Transportation Research Record*, (682), 1978.

[206] Wei-Hua Lin and Carlos F Daganzo. A simple detection scheme for delay-inducing freeway incidents. *Transportation Research Part A : Policy and Practice*, 31(2) :141–155, 1997.

[207] Jing Jin and Bin Ran. Automatic freeway incident detection based on fundamental diagrams of traffic flow. *Transportation Research Record : Journal of the Transportation Research Board*, (2099) :65–75, 2009.

[208] Samir A Ahmed and Allen R Cook. *Application of time-series analysis techniques to freeway incident detection*. Number 841. 1982.

[209] Moshe Levin and Gerianne M Krause. Incident detection : A bayesian approach. *Transportation Research Record*, 682 :52–58, 1978.

[210] Conrad L Dudek, Carroll J Messer, and Nelson B Nuckles. Incident detection on urban freeways. *Transportation Research Record*, 495 :12–24, 1974.

[211] Athanasios P Chassiakos and Yorgos J Stephanedes. Smoothing algorithms for incident detection. *Transportation Research Record*, 1394 :8–16, 1993.

[212] H. Xu, C. M. Kwan, L. Haynes, and J. D. Pryor. Real-time adaptive on-line traffic incident detection. In *Intelligent Control, 1996., Proceedings of the 1996 IEEE International Symposium on*, pages 200–205. IEEE.

[213] Ruey Long Cheu, Dipti Srinivasan, and Wee Hoon Loo. Training neural networks to detect freeway incidents by using particle swarm optimization. *Transportation Research Record : Journal of the Transportation Research Board*, (1867) :11–18, 2004.

[214] Shu-Yan Chen, Wei Wang, and Gao-Feng Qu. Traffic incident detection based on rough sets approach. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 7, pages 3734–3739. IEEE, 2007.

[215] Wei Wang, Shuyan Chen, and Gaofeng Qu. Comparison between partial least squares regression and support vector machine for freeway incident detection. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 190–195. IEEE, 2007.

[216] Shuyan Chen and Wei Wang. Decision tree learning for freeway automatic incident detection. *Expert systems with applications*, 36(2) :4101–4105, 2009.

[217] Qingchao Liu, Jian Lu, and Shuyan Chen. Traffic incident detection using random forest. In *Transportation Research Board 92nd Annual Meeting*, number 13-1610, 2013.

[218] Fang Yuan and Ruey Long Cheu. Incident detection using support vector machines. *Transportation Research Part C : Emerging Technologies*, 11(3) :309–328, 2003.

[219] Dipti Srinivasan, Xin Jin, and Ruey Long Cheu. Adaptive neural network models for automatic incident detection on freeways. *Neurocomputing*, 64 :473–496, 2005.

[220] Shuyan Chen, Wei Wang, and Henk Van Zuylen. Construct support vector machine ensemble to detect traffic incident. *Expert systems with applications*, 36(8) :10976–10986, 2009.

[221] Jianli Xiao and Yuncai Liu. Traffic incident detection using multiple-kernel support vector machine. *Transportation Research Record : Journal of the Transportation Research Board*, (2324) :44–52, 2012.

[222] Jian Lu, Shuyan Chen, Wei Wang, and Bin Ran. Automatic traffic incident detection based on nfoil. *Expert Systems with Applications*, 39(7) :6547–6556, 2012.

[223] Ying Jin, Jing Dai, and Chang-Tien Lu. Spatial-temporal data mining in traffic incident detection. In *Proc. SIAM DM 2006 Workshop on Spatial Data Mining*, volume 5, 2006.

[224] DW Kwon, K Ko, Marina Vannucci, ALN Reddy, and S Kim. Wavelet methods for the detection of anomalies and their application to network traffic analysis. *Quality and Reliability Engineering International*, 22(8) :953–969, 2006.

[225] Chin-Tser Huang, Sachin Thareja, and Yong-June Shin. Wavelet-based real time detection of network traffic anomalies. In *Securecomm and Workshops, 2006*, pages 1–7. IEEE, 2006.

[226] Asim Karim and Hojjat Adeli. Incident detection algorithm using wavelet energy representation of traffic patterns. *Journal of Transportation Engineering*, 128(3) :232–242, 2002.

[227] Jutaek Oh, Joonyoung Min, Myungseob Kim, and Hanseon Cho. Development of an automatic traffic conflict detection system based on image tracking technology. *Transportation Research Record : Journal of the Transportation Research Board*, (2129) :45–54, 2009.

[228] Panos G Michalopoulos, Richard A Fundakowski, Meletios Geokezas, and Robert C Fitch. Vehicle detection through image processing for traffic surveillance and control, July 11 1989. US Patent 4,847,772.

[229] Andrea Baiocchi, Francesca Cuomo, Mario De Felice, and Gaetano Fusco. Vehicular ad-hoc networks sampling protocols for traffic monitoring and incident detection in intelligent transportation systems. *Transportation Research Part C : Emerging Technologies*, 56 :177–194, 2015.

[230] Christoph Sommer, David Eckhoff, Reinhard German, and Falko Dressler. A computationally inexpensive empirical model of ieee 802.11 p radio shadowing in urban environments. In *Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on*, pages 84–90. IEEE, 2011.

[231] Kai Yin, Xiubin Bruce Wang, and Yunlong Zhang. Vehicle-to-vehicle connectivity on two parallel roadways with a general headway distribution. *Transportation Research Part C : Emerging Technologies*, 29 :84–96, 2013.

[232] Akira Kinoshita, Atsuhiro Takasu, and Jun Adachi. Traffic incident detection using probabilistic topic model. In *EDBT/ICDT Workshops*, pages 323–330. Citeseer.

[233] Henry Y.T. Ngan, Anthony G.O. Yeh, and Nelson H.C. Yung. Outlier detection in traffic data based on the dirichlet process mixture model. 9(7) :773–781.

[234] Tim Lawson, David Lovell, and Carlos Daganzo. Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck. *Transportation Research Record : Journal of the Transportation Research Board*, (1572) :140–147, 1997.

[235] S Chandana Wirasinghe. Determination of traffic delays from shock-wave analysis. *Transportation Research*, 12(5) :343–348, 1978.

[236] Raktim Pal and Kumares C Sinha. Simulation model for evaluating and improving effectiveness of freeway service patrol programs. *Journal of Transportation Engineering*, 128(4) :355–365, 2002.

[237] Zhuojin Wang, Pamela M Murray-Tuite, et al. A cellular automata approach to estimate incident-related travel time on interstate 66 in near real time. 2010.

[238] Woon Kim, Suhasini Natarajan, and Gang-Len Chang. Empirical analysis and modeling of freeway incident duration. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 453–457. IEEE, 2008.

[239] Kevin Smith and B Smith. Forecasting the clearance time of freeway accidents. *Center for Transportation Studies, University of Virginia*, 2001.

[240] Stephen Boyles, David Fajardo, and S Travis Waller. A naive bayesian classifier for incident duration prediction. In *86th Annual Meeting of the Transportation Research Board, Washington, DC*, 2007.

[241] Jaimyoung Kwon, Michael Mauch, and Pravin Varaiya. Components of congestion : Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Transportation Research Record : Journal of the Transportation Research Board*, (1959) :84–91, 2006.

[242] Bei Pan, Ugur Demiryurek, and Cyrus Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 595–604. IEEE, 2012.

[243] Younshik Chung and Wilfred W Recker. A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents. *Intelligent Transportation Systems, IEEE Transactions on*, 13(3) :1454–1461, 2012.

[244] Mahalia Miller and Chetan Gupta. Mining traffic incidents to forecast impact. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 33–40. ACM, 2012.

[245] Bei Pan, Ugur Demiryurek, Cyrus Shahabi, and Chaitali Gupta. Forecasting spatiotemporal impact of traffic incidents on road networks. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 587–596. IEEE, 2013.

[246] Bernhard Schölkopf, Robert C. Williamson, Alexander J. Smola, John Shawe-Taylor, John C. Platt, and others. Support vector method for novelty detection. In *NIPS*, volume 12, pages 582–588. Citeseer.

[247] Mohammed Elhenawy and Hesham A Rakha. Automatic congestion identification with two-component mixture models. *Transportation Research Record : Journal of the Transportation Research Board*, (2489) :11–19, 2015.

[248] Jiawei Wang, Xin Li, Stephen Shaoyi Liao, and Zhongsheng Hua. A hybrid approach for automatic incident detection. *Intelligent Transportation Systems, IEEE Transactions on*, 14(3) :1176–1185, 2013.

[249] JC Tanner. Effect of weather on traffic flow. 1952.

[250] Paul Pisano and Lynette Goodwin. Research needs for weather-responsive traffic management. *Transportation Research Record : Journal of the Transportation Research Board*, (1867) :127–131, 2004.

[251] E Roy Jones and Merrell E Goolsby. The environmental influence of rain on freeway capacity. *Highway Research Record*, (321), 1970.

[252] Romain Billot. *Analyse et modélisation de l'impact de la météorologie sur le trafic routier*. PhD thesis, Ecole Centrale Paris, 2010.

[253] Nour-Eddin El Faouzi, Olivier de Mouzon, Romain Billot, and Jacques Sau. Assessing the changes in operating traffic stream conditions due to weather conditions. *Advances in Transp. Studies J. A*, 21 :33–46, 2010.

[254] Eleni I Vlahogianni and Matthew G Karlaftis. Comparing traffic flow time-series under fine and adverse weather conditions using recurrence-based complexity measures. *Nonlinear Dynamics*, 69(4) :1949–1963, 2012.

[255] Fred L Hall and Deanna Barrow. *Effect of weather on the relationship between flow and occupancy on freeways*. Number 1194. 1988.

[256] Amal T Ibrahim and Fred L Hall. *Effect of adverse weather conditions on speed-flow-occupancy relationships*. Number 1457. 1994.

[257] Brian L Smith, Kristi G Byrne, Rachel B Copperman, Susan M Hennessy, and Noah J Goodall. An investigation into the impact of rainfall on freeway traffic flow. In *83rd annual meeting of the Transportation Research Board, Washington DC*. Citeseer, 2004.

[258] Highway Capacity Manual. Highway capacity manual. *Washington, DC*, 2000.

[259] Hesham Rakha, Mohamadreza Farzaneh, Mazen Arafeh, and Emily Sterzin. Inclement weather impacts on freeway traffic stream behavior. *Transportation Research Record : Journal of the Transportation Research Board*, (2071) :8–18, 2008.

[260] Cambridge Systematics. Inc. weather-responsive traffic management : Concept of operations, 2003.

[261] Michael Kyte, Zaher Khatib, Patrick Shannon, and Fred Kitchener. Effect of weather on free-flow speed. *Transportation research record : Journal of the transportation research board*, (1776) :60–68, 2001.

[262] William HK Lam, Hu Shao, and Agachai Sumalee. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation research part B : methodological*, 42(10) :890–910, 2008.

[263] Haiyun Lu. Short-term traffic prediction using rainfall. *Int. J. Signal Process. Syst*, 2(1) :70–73, 2014.

[264] N-E El Faouzi, Romain Billot, and Salim Bouzebda. Motorway travel time prediction based on toll data and weather effect integration. *IET intelligent transport systems*, 4(4) :338–345, 2010.

[265] Declan McHugh. Traffic prediction and analysis using a big data and visualisation approach, 2015.

[266] Stephen Dunne and Bidisha Ghosh. Weather adaptive traffic prediction using neurowavelet models. *IEEE Transactions on Intelligent Transportation Systems*, 14(1) :370–379, 2013.

[267] Tian Hou, Hani Mahmassani, Roemer Alfelor, Jiwon Kim, and Meead Saberi. Calibration of traffic flow models under adverse weather and application in mesoscopic network simulation. *Transportation Research Record : Journal of the Transportation Research Board*, (2391) :92–104, 2013.

[268] Shan-Huen Huang and Bin Ran. *An application of neural network on traffic speed prediction under adverse weather condition*. PhD thesis, University of Wisconsin–Madison, 2003.

[269] Shane Butler, John Ringwood, and Damien Fay. Use of weather inputs in traffic volume forecasting. In *Submitted to ISSC 2007 Conference*, 2007.

[270] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.

A

Mise en oeuvre algorithmique du procédé d'optimisation des hyperparamètres

Cette annexe présente l'algorithme général de mise en oeuvre du cadre prédictif temps-réel lors de la phase d'optimisation des hyperparamètres. L'algorithme a été rédigé de manière à pouvoir traiter la plupart des méthodes d'optimisation rencontrées. Il se partage en deux phases correspondant à la validation-croisée s'appuyant sur un échantillonnage de l'historique d'apprentissage originel, puis la mise en oeuvre du procédé de recherche du jeu de paramètres optimum.

Algorithme 4 Mise en oeuvre du cadre prédictif durant la phase d'optimisation des hyperparamètres.

▷ APPLICATION D'UNE STRATÉGIE D'ÉCHANTILLONNAGE OU VALIDATION CROISÉE :
Définition d'un jeu d'optimisation et d'un jeu de test Set_{test}
Soit Set_{Hist} l'ensemble des indices du jeu d'apprentissage $[1, T_{Hist}]$
Soit $\underline{XHist_{Hist}}$ Ensemble des Individus du jeu d'apprentissage
Soit $\underline{YHist_{Hist}}$ Ensemble des Labels du jeu d'apprentissage
$Set_{test} \leftarrow$ SplitSet(Set_{Hist}) Partage du jeu d'apprentissage en un historique d'optimisation et un jeu de test
Soit $\underline{Set_{Par}}$ la sélection des jeux de paramètres à tester suivant la stratégie adoptée
for all $t \in Set_{test}$ **do**
▷ USAGE D'UNE APPROCHE D'OPTIMISATION DES HYPERPARAMÈTRES :
Procédé de sélection des sets de paramètres à estimer pour cartographier l'erreur associée au jeu de test
for Idx = 1 $\leftarrow$ Card($\underline{Set_{Par}}$) **do**
$Hyperparameters \leftarrow \underline{Set_{Par}}(Idx, :)$
$\widehat{Y_t} \leftarrow \underline{\widehat{Y_{Pred}}}(t-h, Idx, :)$
$[\underline{E_{T_{Hist}}}(t, Idx), Y_{t+h}] \leftarrow Algo\ 1\ (Hyperparameters, h, \underline{XHist_{Hist}}, \underline{YHist_{Hist}}, \widehat{Y_t})$
$\underline{\widehat{Y_{Pred}}}(t, Idx, :) \leftarrow \widehat{Y_{t+h}}$
Remove $\widehat{Y_{t+h}}, \widehat{Y_t}$
end for
end for
$OptimParameters \leftarrow$ Mapping($\underline{E_{T_{Hist}}}$)

B

Implémentation sous Matlab du modèle MSVR

Cette annexe présente l'implémentation sous Matlab du modèle SVR généralisé aux cas de sorties multidimensionnelles ou MSVR. Le code résulte d'une modficiation de l'alglorithme IRWLS proposé par Perez-Cruz et al. [162] en vue de pallier la difficulté du passage aux dimensions.

```
% Multioutput SVR
%
%   We have m labeled examples, d dimensions and k outputs to predict.
%
% inputs:   - x : training patterns (m x d),
%           - y : training targets (m x k),
%           - ker : kernel type ('lin', 'poly', 'rbf'),
%           - C : cost parameter,
%           - par : kernel parameter (see function 'kernelmatrix') ,
%           - tol : tolerance.
%

function [rho, NSV, K, SV, bias] = MoSvr(x,y,ker,C,epsi,par,tol)


% * Define dimensions of the problem :
n_m=size(x,1); % n [Number of Individuals / Observations]
n_d=size(x,2); % d [Number of input variables]
n_k=size(y,2); % Q [Number of output variables]

if size(y, 1)~=n_m
    error('x & y have not the same number of observations')
end


% * Build the kernel matrix on the labeled samples
K = Kernelmatrix_MoSVR(ker, x, x, par);



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% INITIALISATION
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%
```

```
%%% STEP 1 : Optimization of Beta & b
%
% * Initialize the Beta (i.e. W) and b parameters
rho = zeros(n_m, n_k);
b = zeros(n_m, n_k);
bias = zeros(1, n_k);
% * Assess the prediction error :
% ** ei = prediction error per output (n_m * n_k)
E = y - K*rho - b;
% ** ui = RSE (n_m x 1)
u = sqrt(sum(E.^2, 2));  % Root squared error

%
%%% STEP 2 : Modification of ai
%
% * Assess the Support Vectors (i.e. update the ai)
% ** Points for which prediction error is larger than epsilon (i.e Future Support
%Vectors )
NewSV = find(u >= epsi);
SV = NewSV;
% ** Set initial values of ai
a = 2*C*(u-epsi)./u;  % Taylor Development at the First Order for the Loss Function Lp(W, b)
% *** Notice : Normally, only the SV values of ai must be non-null values !!
% * Assess the Initial Loss (normally it is produced before the new SV assessment)
L = zeros(size(u));
% Modify only entries for which  u > epsi. with the sq slack
%L(i) = (RSE(i) - epsi).^2
L(SV) = u(SV).^2 - 2*epsi*u(SV) + epsi^2;
%Lp is the quantity to minimize (sq norm of parameters + slacks)
% Lp = .5*sum(||wj||^2) + C*sum(L(ui))
Lp(1,1) = sum(diag(rho'*K*rho))/2 + C*sum(L);

%
%%% Further Initialization
%
k=2;
ToDo=1;




%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% ITERATIONS ON OPTIMIZATION OF Beta & b, THEN ALTER ai
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


while(ToDo)
    %
    %%% SAVE OLD VALUES FOR 2nd LOOP (if required)
    %
    rho_a = rho;
```

```
bias_a = bias;
b_a = b;
E_a = E;
u_a = u;
SV_a = SV;

%
%%% UPDATE THE SUPPORT VECTORS
%
SV = NewSV;

%
%%% STEP 1 : BUILD A NEW beta & b FROM PREVIOUS SUPPORT VECTORS : (OptimValues)
%

% * Assess the linear matrix to inverse
% M1 = K + D_a + 1 = y (only for obs SV. see above)
%  The term 10^(-10) prevent any prematurate end of the algorithm.
M1 = [K(SV,SV) + diag(1./a(SV))] + 1e-10*eye(length(a(SV)));
V1 = ones(length(SV), 1);
V2 = a(SV)'*K(SV, SV);
M = [M1 V1; V2 sum(a(SV))];

% * Compute new betas(k) and b(k) from SV(k-1) et ai(k-1) :
OptimValues = inv(M) * [y(SV,:); a(SV)'*y(SV, :)];

NewBeta = OptimValues(1:end-1, :);
Newb = ones(n_m, 1) * OptimValues(end, :);

rho = zeros(size(rho)); % Get Beta(k) associate to SV(k-1)
rho(SV,:) = NewBeta; % Get b(k) associate to SV(k-1)
b = Newb;
bias = b(1, :);

% * Assess the prediction error
% Error ei :
E = y - K*rho - b;
% RSE ui
u = sqrt(sum(E.^2, 2));

%
%%% UP-DATE SUPPORT VECTORS: NewSV for step (k+1)
%
NewSV = find(u>=epsi);


%
%%% ASSESS THE LOSS FUNCTION AT STEP k
%
% Compute L(ii) = Eta(i)
L=zeros(size(u));
```

```
L(NewSV) = u(NewSV).^2 - 2*epsi*u(NewSV) + epsi^2;
% Recompute the loss function Lp
Lp(k,1) = sum(diag(rho'*K*rho))/2 + C*sum(L) ;  % Lp = .5*sum(||wj||^2) + C*sum(L(ui))

%
%%% BACKTRACKING STEP
%
if k > 2
    %%%%
    % Backtracking Step : IF REQUIRED (i.e.If the actual loss is higher than the previous one)
    %%%%
    nu = 1;

    % Seek for intermediate optimal values
    while(Lp(k,1) > Lp(k-1,1))

        Beta_s = rho;
        b_s = b;

        nu = nu/10;
        SV = SV_a;

        % * Assess intermediate optimum (w & b):
        rho = zeros(size(rho));
        rho(SV,:) = nu*Beta_s(SV, :) + (1-nu)*rho_a(SV,:);
        b = nu*b_s + (1-nu)*b_a;
        bias = b(1, :);

        % * Assess intermediate prediction error :
        E = y - K*rho - b;
        u = sqrt(sum(E.^2, 2));

        % * Define associated future SV :
        NewSV = find(u>=epsi);

        % * Assess Loss
        L = zeros(size(u));
        L(NewSV) = u(NewSV).^2 - 2*epsi*u(NewSV) + epsi^2;
        Lp(k,1) = sum(diag(rho'*K*rho))/2 + C*sum(L);

        %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        %%% Stopping criterion #1 -> NO CONVERGENCE
        %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        if(nu < 10^-16)
            % 'Stop 1 : Lp cannot be improved !
            Lp(k,1) = Lp(k-1,1)-10^-15;
            rho = rho_a;
            bias = bias_a;
            u = u_a;
            SV = SV_a;
            ToDo = 0;
```

```
            end
            %%%%
```

FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'ÉCOLE NATIONALE DES TRAVAUX PUBLICS DE L'ÉTAT

NOM : LAHAROTTE **DATE DE SOUTENANCE :** 16/12/2016
(avec précision du nom de jeune fille le cas échéant)

Prénoms : Pierre-Antoine, Eugène, Charles, Claude

TITRE : Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier

École doctorale : Mécanique, Énergétique, Génie civil, Acoustique (MEGA)

NATURE : Doctorat **Numéro d'ordre :** 2016LYSET013

Spécialité : Génie civil

Code B.I.U. - / et bis **CLASSE :**

RÉSUMÉ : La maturité de la télématique et des technologies de l'information et la communication (TIC), ainsi que l'avènement du big data dans le transport ont conduit à des développements foisonnants dans le domaine des systèmes de transports intelligents (ITS), aussi bien sur le plan des technologies de recueil que du traitement innovant de l'information. Il est désormais possible de connaître les conditions de circulation et les états de trafic sur la plupart des sections d'un réseau routier sans avoir recours à des infrastructures intrusives de collecte de données, de transmettre l'information résultante via des réseaux sans fil et de traiter rapidement toutes ces données multi-sources disponibles.

La constitution de grandes bases de données a naturellement fait évoluer la pratique de gestion du trafic et plus particulièrement les méthodes de prévision. Ces méthodes ont connu un renouveau en s'inspirant des travaux produits en apprentissage statistique. Néanmoins, la façon d'appréhender le problème de la prévision est restée à une échelle locale. Pour chaque section de route, un modèle de prévision est adapté et optimisé. Notre travail de thèse présente un cadre de prévision du trafic routier qui aborde la question à l'échelle du réseau. L'étude menée au sein de ces travaux de thèse vise à exposer et évaluer cette nouvelle approche, dite globale, au regard d'approches usuelles, puis à analyser sa sensibilité vis-à-vis de divers facteurs.

Après un positionnement par rapport à l'état de l'art en théorie du trafic, le cadre prédictif fondé sur des méthodes de prévision multi-variées par apprentissage est détaillé. Une version multidimensionnelle des k plus proches voisins, modèle parcimonieux et simple, est évaluée sur divers cas d'études. L'originalité réside dans l'exploitation de données issues de méthodes innovantes de collecte (*e.g.* Bluetooth, véhicules traceurs, véhicules connectés).

Par la suite, les performances de l'approche initiale sont comparées à d'autres méthodes d'apprentissage. Un effort particulier est porté sur l'adaptation de méthodes à noyaux au cadre prédictif global. Les performances obtenues laissent entrevoir une typologie des méthodes en fonction des caractéristiques spatio-temporelles du réseau. Afin d'améliorer les performances en prévision et de réduire les temps de calcul, une méthode d'identification et de sélection des sections critiques du réseau est proposée. Les résultats prouvent qu'un sous-ensemble restreint de sections est en effet suffisant pour garantir des performances satisfaisantes en généralisation.

Enfin, la résilience du cadre prédictif est évaluée au regard des événements non récurrents affectant le fonctionnement nominal du réseau, comme des incidents ou des conditions météorologiques dégradées. Les résultats soulignent l'impact de ces conditions non récurrentes sur la prévision temps-réel de la dynamique court-terme d'un réseau et permettent de dresser une feuille de route pour l'élaboration d'un cadre prédictif résilient et opérationnel. Cette nouvelle vision de la prévision s'inscrit dans les perspectives actuelles en terme d'applications sur les modules embarqués et les objectifs des gestionnaires d'infrastructures.

MOTS-CLES : *big data*, Bluetooth, *Floating Car Data* (FCD), prévision, apprentissage, réseau, résilience, événement non récurrent, sections critiques.

Laboratoire(s) de recherche : Laboratoire d'Ingénierie Circulation Transport (LICIT)

Directeur de thèse : Nour-Eddin El Faouzi

Président de jury :

Composition du jury : Vincent Aguilera (Mines Pont ParisTech), Rapporteur
Liming Chen (Ecole Centrale de Lyon), Rapporteur
Latifa Oukhellou (Université de Marne-la-Vallée), Examinatrice
Philippe Goudal (Mediamobile), Examinateur
Nour-Eddin El Faouzi (Université de Lyon), Directeur de thèse
Romain Billot (Télécom Bretagne), Co-directeur de thèse