l'université nantes angers le mans
COMMUNAUTÉ D'UNIVERSITÉS ET ÉTABLISSEMENTS

UNIVERSITÉ DE NANTES

Thèse de Doctorat

David GUILBERT

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le label de l'Université de Nantes Angers Le Mans

École doctorale : Sciences et technologies de l'information, et mathématiques

Discipline : Électronique, section CNU 61
Spécialité : Traitement du signal
Unité de recherche : IETR UMR 6164

Soutenue le 20 janvier 2015

Analyse et classification des signatures des véhicules provenant de capteurs magnétiques pour le développement des algorithmes « Intelligents » de gestion du trafic

JURY

Présidente :	**M^me^ Christine BUISSON**, Directrice de Recherche/HDR, IFSTTAR/ENTPE LICIT, Vaulx-en-Velin
Rapporteurs :	**M. Salah BOURENNANE**, Professeur, Institut Fresnel, École Centrale de Marseille
	M. Ali KHENCHAF, Professeur, ENSTA Bretagne, Brest
Examinateurs :	**M. Patrick AKNIN**, Directeur Scientifique/HDR, SNCF, Paris
	M^me^ Sylvie CHARBONNIER, Maître de Conférences/HDR, Université Joseph Fourier, Polytech Grenoble
	M. Sio-Song IENG, Chercheur, IFSTTAR, Marne-la-Vallée
	M. Cédric LE BASTARD, Chargé de Recherche, Cerema, Les Ponts-de-Cé
Directeur de thèse :	**M. Yide WANG**, Professeur, Université de Nantes

Remerciements

Un immense merci à Sio-Song Ieng pour m'avoir suggéré et encouragé à me lancer dans l'aventure d'une thèse. Il aura fallu environ deux ans et la complicité de Cédric Le Bastard pour monter ce projet et le financer. Merci, Cédric d'avoir su passer outre les complexités administratives pour que je puisse faire cette thèse. Je ne peux que remercier Yide Wang qui est venu encadrer cette thèse sur la demande de Cédric. Déjà avant même de commencer cette thèse, je ne pouvais que vous être reconnaissant de vous investir dans ce projet. Je tiens aussi à vous remercier tous les trois de m'avoir fait confiance ainsi que de m'avoir laissé une grande autonomie et liberté durant cette thèse. La perspicacité de chacun m'a amené à approfondir, expliquer et réexpliquer mes raisonnements jusqu'à ce qu'ils atteignent une structure et une cohérence suffisantes. Cette thèse à vos côtés fût très enrichissante tant sur le plan personnel que professionnel.

Mes remerciements s'adressent à Salah Bourennane, Professeur à l'École Centrale de Marseille et à Ali Khenchaf, Professeur à l'École Nationale Supérieure de Techniques Avancées (ENSTA) de Bretagne pour l'honneur qu'ils me font en acceptant le rôle de rapporteur, et ainsi consacrer une partie de leur temps précieux à l'analyse de mon mémoire et la rédaction de leur rapport.

Je suis honoré de la présence de M. Patrick Aknin, Directeur Scientifique de la direction Innovation & Recherche à la Société nationale des chemins de fer français (SNCF) et de Sylvie Charbonnier, Maître de Conférences – HDR à l'Université Joseph Fourier de Grenoble pour l'intérêt qu'ils portent à mon travail en acceptant d'être examinateur. Je souhaite remercier Christine Buisson, Directrice de Recherche à l'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR) qui a su m'intégrer au projet MOCoPo. Outre ses qualités relationnelles indéniables, je lui exprime ma profonde reconnaissance pour ses conseils et le partage de ses connaissances.

Je tiens aussi à remercier les étudiants qui ont contribué lors de leur stage aux résultats de cette thèse : Antoine Delepoulle et Clélia Lopez. Tous mes encouragements à Clélia pour sa thèse.

Je n'oublie évidemment pas l'ensemble des personnels du Cerema, de l'IFSTTAR, de l'IETR et de l'école doctorale qui a contribué de façon plus ou moins directe à la réalisation de cette thèse.

Enfin, je n'omettrai pas « le truc en plus » : je remercie toute ma famille et mes amis pour leur soutien durant ces années.

Table des matières

1 Introduction **15**
- 1.1 Contexte . . . 15
- 1.2 Problématique . . . 15
- 1.3 Objectif de la thèse . . . 19
- 1.4 Plan du mémoire . . . 19

2 Différentes technologies de capteurs **21**
- 2.1 Véhicules traceurs (capteurs embarqués) . . . 21
  - 2.1.1 Système de positionnement et de datation par satellite . . . 22
  - 2.1.2 Données des téléphones mobiles . . . 22
  - 2.1.3 Avantages et inconvénients . . . 23
- 2.2 Suivi de véhicules point à point . . . 24
  - 2.2.1 Capteurs non intrusifs . . . 24
  - 2.2.2 Capteurs intrusifs . . . 28
- 2.3 Choix des capteurs . . . 30

3 Capteurs magnétiques **33**
- 3.1 Boucle inductive . . . 33
- 3.2 Magnétomètre . . . 38
  - 3.2.1 Champ magnétique terrestre . . . 38
  - 3.2.2 Choix du magnétomètre . . . 38
- 3.3 Conclusion . . . 43

4 Principe du suivi anonyme de véhicules **45**
- 4.1 Problématique du suivi de véhicules . . . 45
- 4.2 Analyse par les pelotons de véhicules . . . 47
- 4.3 Suivi individuel des véhicules . . . 49
  - 4.3.1 Détection . . . 49
  - 4.3.2 Prétraitements . . . 50
  - 4.3.3 Réidentification . . . 50
- 4.4 Évaluation et indicateurs . . . 51
  - 4.4.1 Introduction . . . 51
  - 4.4.2 Méthode d'évaluation . . . 52
- 4.5 Conclusion . . . 53

5 Détection **55**
- 5.1 Réalisation de la détection . . . 55
- 5.2 Problèmes rencontrés . . . 61
- 5.3 Proposition d'un algorithme de détection . . . 63
- 5.4 Évaluation des méthodes de détection . . . 66
  - 5.4.1 Évaluation . . . 66

5.4.2 Conclusion . . . 70
5.5 Conclusion . . . 70

6 Prétraitements **73**
6.1 Différents prétraitements . . . 74
6.1.1 Normalisation . . . 74
6.1.2 Réduction du nombre de données . . . 74
6.1.3 Déconvolution dans le cadre de la boucle inductive . . . 76
6.2 Propositions de sélection de données . . . 81
6.2.1 Informations recueillies & signatures . . . 81
6.2.2 Étude réalisée . . . 83
6.2.3 Population de VL . . . 85
6.2.4 Population de PL . . . 85
6.2.5 Population de VL–PL . . . 85
6.2.6 Conclusion . . . 85
6.3 Propositions d'algorithmes de déconvolution . . . 86
6.3.1 Contexte . . . 86
6.3.2 Définition du problème . . . 86
6.3.3 Algorithmes utilisés . . . 89
6.3.4 Estimation de la signature « réelle » . . . 94
6.3.5 Conclusion . . . 96
6.4 Conclusion . . . 96

7 Réidentification **99**
7.1 Présentation . . . 99
7.1.1 Méthodes sans apprentissage . . . 100
7.1.2 Méthodes avec apprentissage . . . 102
7.2 Propositions de classifieurs . . . 105
7.2.1 Distance de Canberra . . . 105
7.2.2 Séparateur à vaste marge . . . 105
7.3 Détermination d'un seuil et utilisation de plusieurs méthodes . . . 111
7.3.1 Introduction . . . 111
7.3.2 Étude sur une base idéale . . . 111
7.3.3 Étude sur une base perturbée . . . 117
7.3.4 Conclusion . . . 125
7.4 Nouvelle méthode d'évaluation . . . 125
7.4.1 Contexte . . . 125
7.4.2 Classification et comparaisons . . . 125
7.4.3 Réidentification et comparaisons . . . 131
7.5 Conclusion . . . 137

8 Expérimentations **139**
8.1 Temps de parcours . . . 140
8.1.1 Contexte . . . 140
8.1.2 Présentation des sites . . . 140
8.1.3 Processus . . . 142
8.1.4 Temps de parcours individuels . . . 143
8.1.5 Conclusion . . . 143
8.2 Matrice Origine – Destination . . . 145
8.2.1 Contexte . . . 145
8.2.2 SAROT1 . . . 145

8.2.3 SAROT2 . . . 147
8.2.4 Conclusion . . . 150
8.3 Projet MOCoPo . . . 150
8.3.1 Contexte . . . 150
8.3.2 Présentation du site et des capteurs . . . 151
8.3.3 Acquisition . . . 153
8.3.4 Base de données . . . 154
8.3.5 Matrice origine – destination . . . 155
8.3.6 Conclusion . . . 156
8.4 Conclusion . . . 157

9 Conclusion **159**
9.1 Bilan . . . 159
9.2 Perspectives . . . 161

Liste des tableaux

2.1 Types de mesures réalisées par différentes technologies. 31
2.2 Coût du cycle de vie pour différentes technologies. 31

3.1 Comparaison entre les générations de boucle inductive. 38

5.1 Détection en entrée de l'échangeur du Rondeau. 67
5.2 Détection en sortie de l'échangeur du Rondeau. 67
5.3 Fausses détections en entrée de l'échangeur du Rondeau. 68
5.4 Fausses détections en sortie de l'échangeur du Rondeau. 68

6.1 Estimation de la longueur des véhicules. 78

7.1 TBR par validation croisée pour le maximum de vraisemblance et l'approche floue. . 112
7.2 Paramètres établis pour la méthode SVM. 113
7.3 TBR en validation croisée sur la base d'Angers. 113
7.4 TBR en validation croisée sur la base de Rennes. 113
7.5 Angers - TBR sans fenêtrage dans le cadre d'une mesure sur la base totale. 114
7.6 Rennes - TBR sans fenêtrage dans le cadre d'une mesure sur la base totale. 114
7.7 TBR maximaux potentiels avec le fenêtrage utilisé. 114
7.8 Angers - TBR avec fenêtrage dans le cadre d'une mesure sur la base totale. 115
7.9 Rennes - TBR avec fenêtrage dans le cadre d'une mesure sur la base totale. 115
7.10 Performances des quatre approches de réidentification. 119
7.11 Réidentification sur la base de test avec fenêtre temporelle. 120
7.12 Valeur de seuil pour un pourcentage donné de véhicules légers (Site d'Angers). . . . 121
7.13 Valeur de seuil pour un pourcentage donné de poids lourds (Site d'Angers). 122
7.14 Angers - Performances témoins. 123
7.15 Angers - filtrage « Origine à multiples destinations ». 123
7.16 Angers - Post-traitement seuil à 90 %. 123
7.17 Angers - Post-traitement seuil à 80 %. 124
7.18 Angers - Post-traitement seuil à 70 %. 124
7.19 Angers - Post-traitement seuil à 70 % et origines à multiples destinations. 124
7.20 AUC des courbes en fonction de la fréquence d'échantillonnage et de la distance utilisée. 128
7.21 Aire en-dessous de la courbe pour la classification par DTW (distance euclidienne). . 128
7.22 Aire en-dessous de la courbe pour la classification par DTW (distance de Manhattan). 129
7.23 AUC pour la classification par l'addition des distances. 130
7.24 AUC pour la classification par la multiplication des distances. 131

8.1 Tableau récapitulatif des caractéristiques des sites de récupération de données 143
8.2 Matrice origine – destination de référence R de la base de test. 146
8.3 La matrice origine – destination T de la base de test. 146
8.4 La matrice d'erreur E pour la base de test. 146
8.5 Matrice origine – destination de référence R. 148

8.6 Matrice origine – destination 148
8.7 Matrice origine – destination 148
8.8 Matrice origine – destination 149
8.9 Matrice origine – destination 149
8.10 Matrice origine – destination 149
8.11 Erreur relative globale en fonction du nombre de candidats supplémentaires 150
8.12 Matrice origine – destination de référence R pour la première session. 155
8.13 Matrice origine – destination estimée T pour la première session. 156
8.14 Matrice d'erreur E pour la première session. 156

Table des figures

2.1 Utilisation des satellites de positionnement. . . . 22
2.2 Principe de localisation par triangulation. . . . 23
2.3 Capteur vidéo. . . . 25
2.4 Capteur micro-ondes. . . . 26
2.5 Capteur infrarouge passif. . . . 26
2.6 Capteur infrarouge à balayage. . . . 27
2.7 Capteur acoustique SAS. . . . 27
2.8 Tuyau pneumatique. . . . 29
2.9 Câble piézoélectrique. . . . 30
2.10 Fibre optique. . . . 30

3.1 Boucles inductives. . . . 33
3.2 Schéma simplifié des composants d'une boucle inductive. . . . 34
3.3 Schéma implantation d'une boucle. . . . 34
3.4 Schéma d'un circuit RLC. . . . 35
3.5 Exemple de signature inductive. . . . 35
3.6 Schéma d'interaction entre le véhicule et la boucle. . . . 36
3.7 Signature d'une voiture à partir d'une boucle inductive de 10 cm de largeur. . . . 37
3.8 Modèle simplifié du champ magnétique terrestre. . . . 38
3.9 Perturbation du champ magnétique terrestre. . . . 39
3.10 Résolution des capteurs magnétiques. . . . 39
3.11 Principe du capteurs à magnétorésistance. . . . 40
3.12 Principe du capteurs à magnétorésistance géante. . . . 41
3.13 Capteur magnétomètre. . . . 41
3.14 Définition des axes du repère. . . . 42
3.15 Effet des changements de température sur une journée pour la mesure magnétique. . . . 42
3.16 Mesure magnétique en fonction de la température. . . . 43

4.1 Problématique de la réidentification. . . . 46
4.2 Trois signatures d'un même véhicule issues de la boucle inductive. . . . 47
4.3 Différentes signatures de VL et PL issues de la boucle inductive. . . . 48
4.4 Matrice des véhicules associés. . . . 48
4.5 Processus de suivi de véhicules. . . . 50
4.6 Enregistrement vidéo. . . . 52
4.7 Principe de la validation croisée pour k segments. . . . 52

5.1 ATDA. . . . 57
5.2 ATDA modifié. . . . 59
5.3 Automate d'états de Chinrungrueng et coauteurs. . . . 60
5.4 Déformation d'une signature par la vitesse. . . . 61
5.5 Détection d'une signature au lieu de deux. . . . 62

5.6 Signature tronquée. . . . 63
5.7 Signal brut et lissé d'un véhicule. . . . 64
5.8 Automate d'états proposé pour la détection des véhicules. . . . 65
5.9 Signal et signature extraite. . . . 66
5.10 Véhicule réalisant un changement de voie entre la VR et la VL. . . . 69
5.11 Situation de freinage brusque sur les capteurs. . . . 69
5.12 Trois véhicules de front sur deux voies. . . . 70

6.1 Les signaux des paires de boucles inductives. . . . 75
6.2 Compression de la signature en 20 sections moyennées. . . . 75
6.3 Compression de la variation magnétique selon l'axe z. . . . 76
6.4 Principe de la convolution et de la déconvolution. . . . 77
6.5 Déconvolution de Godard. . . . 79
6.6 Déconvolution par moindres carrés de Kwon. . . . 81
6.7 Déconvolution par la méthode de Godard. . . . 82
6.8 Exemple de signature d'un véhicule de classe 10. . . . 82
6.9 Exemple de signature d'un véhicule de classe 1. . . . 82
6.10 Représentation des données pour les PL. . . . 84
6.11 Diagramme pour estimer $\tilde{\mathbf{h}}$ et $\tilde{\mathbf{e}}$. . . . 86
6.12 Changement de repère. . . . 87
6.13 Estimation de $\mathbf{h}$. . . . 90
6.14 La moyenne géométrique mobile pour l'estimation. . . . 91
6.15 Les solutions numériques de $\tilde{\mathbf{h}}$. . . . 92
6.16 Les courbes de Picard. . . . 93
6.17 Courbe en L : $\|\tilde{\mathbf{h}}\|_{\mathbf{2}}$ versus $\|\mathbf{s} - \tilde{\mathbf{E}}_{\mathbf{n-1}}\tilde{\mathbf{h}}_{\mathbf{n}}\|_2$. . . . 95
6.18 Fonctions transferts estimées pour 2 boucles différentes et 2 véhicules différents. . . . 95
6.19 Voiture de tourisme avec remorque « 1 » sur le capteur UD1J (vitesse estimée 76 km h^{-1}). 96
6.20 Voiture de tourisme « 2 » sur le capteur UD1k (vitesse estimée 116 km h^{-1}). . . . 97
6.21 Signatures sans et avec déconvolution pour les voitures de tourisme « 1 » et « 2ˆ. . . 97

7.1 Exemple d'un problème de discrimination binaire linéairement séparable. . . . 106
7.2 Représentation des deux notions de marge . . . 107
7.3 Représentation des vecteurs supports . . . 108
7.4 Représentation de la notion d'écart . . . 108
7.5 Projection des données . . . 110
7.6 TBR en fonction de p_2 pour l'approche floue. . . . 112
7.7 Matrice des distances réalisée à partir des données d'Angers. . . . 116
7.8 Matrice des distances réalisée à partir des données de Rennes. . . . 116
7.9 TBR en fonction de la répartition poids lourds - véhicules légers. . . . 117
7.10 TBR et TRi en fonction du seuil de décision . . . 118
7.11 Évolution du pourcentage de véhicules légers en fonction du seuil (floue). . . . 121
7.12 Évolution du pourcentage de véhicules légers en fonction du seuil (MV). . . . 121
7.13 Évolution du pourcentage de véhicules légers en fonction du seuil (SVM-d). . . . 122
7.14 Signatures sans et avec déconvolution pour les voitures de tourismes « 1 » et « 2 ». . 126
7.15 Courbes ROC : Déconvolution. . . . 127
7.16 Classification avec les magnétomètres par DTW (distance euclidienne). . . . 129
7.17 Classification avec les magnétomètres par DTW (distance de Manhattan). . . . 130
7.18 Classification avec les magnétomètres par l'addition des distances de Manhattan. . . 130
7.19 Classification avec les magnétomètres par la multiplication des distances de Manhattan 131
7.20 Taux de bonne réidentification : déconvolution. . . . 134
7.21 TBR et TR déconvolution avec vote à l'unanimité. . . . 135

7.22 TBR et TR déconvolution avec vote à l'unanimité. 136
7.23 TBR en fonction du nombre de candidats supplémentaires. 137
7.24 TBR et TR en fonction du nombre de candidats supplémentaires. 138

8.1 Angers - Site Quai–Berge. 140
8.2 Angers - Répartition des véhicules recensés en fonction de la vitesse. 141
8.3 Rennes - Périphérique. 141
8.4 Rennes - Répartition des véhicules recensés en fonction de la vitesse. 142
8.5 Estimation des temps de parcours pour des VL. 144
8.6 Estimation des temps de parcours pour des PL. 144
8.7 Expérimentation sur le site SAROT pour la base de donnée SAROT1. 145
8.8 Schéma simplifié pour l'expérimentation avec la déconvolution. 147
8.9 Organisation du projet MOCoPo. 151
8.10 Zones d'implantation des réseaux de capteurs sur la rocade sud de Grenoble. 151
8.11 Schéma de la rocade sud de Grenoble. 152
8.12 Entrées – sorties de l'échangeur du Rondeau. 152
8.13 Réseaux de capteurs de Alpexpo à l'entrée de l'échangeur du Rondeau. 152
8.14 Réseaux de capteurs magnétomètres. 153
8.15 Réseaux de capteurs boucles inductives - magnétomètre. 154
8.16 Temps de parcours des véhicules entre l'origine et la destination. 155

1

Introduction

Sommaire

1.1 Contexte 15
1.2 Problématique 15
1.3 Objectif de la thèse 19
1.4 Plan du mémoire 19

1.1 Contexte

Cette thèse a été menée durant la période 2011–2014 au sein de l'équipe de recherche Techniques Physiques avancées du Département Laboratoire et Centre d'Études et de Conception de Prototypes d'Angers, structure dépendant du Centre d'études et d'expertise sur les risques, la mobilité et l'aménagement (Cerema) et de l'Institut d'Électronique et de Télécommunications de Rennes – UMR CNRS 6164. Elle s'est développée en collaboration avec le Laboratoire Exploitation, Perception, Simulateurs et Simulations (LEPSIS) et le Laboratoire d'Ingénierie Circulation Transport (LICIT) de l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR) au travers du projet Mesure et mOdélisation de la COngestion et de la POllution (MOCoPo).

Le rédacteur de la thèse présentée est titulaire d'un Diplôme d'Études Approfondies (ancien master II recherche) en automatique et informatique industrielle. Après cinq années passées au sein de l'équipe de recherche Techniques Physiques avancées en tant que chargé d'études, l'auteur a saisi l'opportunité de réaliser cette thèse dans le cadre d'une formation qualifiante. Le financement de cette thèse a été pris en charge par la Direction de la Recherche et de l'Innovation du Ministère de l'Écologie, du Développement durable et de l'Énergie (MEDE) pendant quatre ans à hauteur de 75 %. L'auteur a donc consacré 75 % de son temps à la réalisation de cette thèse et 25 % de son temps aux problématiques liées à son travail de chargé d'études durant ces quatre années.

1.2 Problématique

Les réseaux de transports se sont progressivement développés pour permettre à l'Homme de se déplacer et de réaliser ainsi ses différentes activités. Les demandes en matière de mobilité ne

cessent d'augmenter. Elles intègrent de plus en plus de contraintes. Les préoccupations de la société se traduisent, en effet, au travers des problématiques d'aménagement du territoire, de mobilité des individus et des biens, de lutte contre l'insécurité routière, d'une prise de conscience écologique. Le réseau de transport est complexe de par cette diversité des usages à laquelle il doit répondre mais aussi de par la pluralité des modes de transport.

D'après le rapport de l'office parlementaire d'évaluation des choix scientifiques et technologiques de janvier 2014 sur la mobilité [1], le développement actuel de notre société provoque des problèmes quant à la pollution, la santé et les inégalités sociales et internationales.

Dans ce contexte, la France a pris l'engagement de diminuer par quatre ses émissions de gaz à effet de serre d'ici 2050 par rapport à ceux émis en 1990 [2]. Cette décision politique a été prise en 2003. Elle est désignée sous le nom de facteur 4. Le secteur du transport est l'un des principaux contributeurs de gaz à effet de serre avec une estimation provisoire pour l'année 2012 de 27,2 % des émissions nationales d'après le rapport sur les chiffres clés du transport 2014 [3]. Toujours d'après ce rapport, le secteur du transport représente 35,2 % des émissions de CO_2, l'un des principaux gaz à effet de serre. De plus, les transports émettent des particules fines et de NO_x qui ont un impact avéré sur la santé selon l'étude de l'Organisation Mondiale de la Santé [4]. Les risques de décès dus à des maladies cardiovasculaires et pulmonaires sont accrus avec la pollution de l'air. Elle favorise aussi le développement des allergies. Les chiffres clés du transport 2014 [3] évoquent une part d'environ 60 % du secteur du transport dans les émissions françaises de NO_x et environ 20 % pour les particules fines. En regardant ces chiffres de manière plus précise, la route représente la majorité de la répartition du transport intérieur de voyageurs en France avec 87,9 % du trafic, suivi par le ferroviaire avec 10,6 % et l'aérien avec 1,4 %. En ce qui concerne le transport de marchandises, avec 83,6 % du trafic, la route reste prépondérante vis-à-vis du ferroviaire, du fluvial et de l'aérien. En ce qui concerne la pollution, la route est le principal contributeur à l'émission du CO_2 avec environ 95 %. L'ensemble de ces statistiques montre l'importance de l'impact écologique et sanitaire de la route.

La route a aussi le triste record d'être le transport le plus accidentogène pour l'année 2012 avec 62 250 accidents répertoriés, loin devant les 11 080 du maritime et les 136 du ferroviaire. Le nombre de tués sur les routes, même s'il ne cesse de diminuer, passant de 8412 en 1995 à 3842 en 2012, reste conséquent. Baisser ce nombre pour tendre vers zéro reste un objectif des pouvoirs publics.

Les sociologues, d'après le rapport sur les nouvelles mobilités des sénateurs Baupin et Keller [1], font part d'un rétrécissement de l'espace grâce aux nouvelles technologies de l'information et de la communication. En effet, les échanges d'informations sont instantanés avec l'ensemble de la planète. La communication avec une personne éloignée géographiquement est possible instantanément ce qui donne l'impression de réduire l'espace. Par contre, la mobilité des personnes et des biens nécessite plus de temps et se retrouve parfois entravée, particulièrement dans les grandes agglomérations. Le temps semble alors disproportionné au regard de la distance à effectuer. Le déséquilibre entre le temps et la distance à parcourir entraîne l'impression de « perdre du temps », de l'inquiétude, de la souffrance, un sentiment d'exclusion suivant [1]. Pour la route, la commission européenne a estimé à 80 milliers de milliards d'euros par an le coût de la congestion au niveau de l'Europe.

Pour retrouver plus de sérénité dans la mobilité et répondre à des contraintes croissantes en matière de pollution, d'exigences sanitaires et de sécurité, le rapport [1] de l'office parlementaire d'évaluation des choix scientifiques et technologiques préconise plusieurs orientations, dont les suivantes :

- Privilégier les infrastructures existantes à la construction de nouvelles ;
- Apporter à l'usager des informations en temps réel ;
- Élaborer au niveau européen le développement d'une « route plus intelligente ».

En résumé, il s'agit surtout de donner à l'usager ayant besoin de se déplacer le choix du mode le plus pratique, le plus pertinent, le plus rapide et le plus serein.

Pour répondre à l'ensemble des besoins en mobilité et faciliter le choix des usagers, les « Systèmes de Transport Intelligent » (STI) se développent grâce aux nouvelles technologies. Leur objectif est de fournir un service innovant multimodal (c'est à dire qui tient compte de plusieurs modes de transports). Le déplacement peut être vu comme une demande de mobilité : transport de marchandises ou de personnes d'un lieu à un autre. Pour répondre à cette demande, une offre est disponible : véhicules, infrastructures et services. Lorsque la demande est supérieure à l'offre, une inéquation apparaît et se traduit par une situation de congestion voire d'immobilité. L'objectif des STI est d'optimiser l'utilisation de l'offre pour répondre à la demande sous l'effet de contraintes intrinsèques (pollution, congestion, incident...). Grâce aux STI, un suivi des événements est réalisé, ce qui amène des optimisations structurelles. Ils permettent d'adapter par exemple les horaires des transports et les lieux d'arrêts aux besoins des usagers, ou de coordonner les différents modes de transport entre eux. Les STI sont aussi des moyens d'actions sur la demande des usagers et le transport de marchandises. La possibilité de mettre des péages à l'entrée des villes pour limiter la circulation, la réduction des tarifs en dehors des heures de pointe pour les transports en commun sont des exemples concrets pour contraindre les comportements des usagers. Les STI encouragent aussi l'utilisation d'alternatives pour répondre à une demande de déplacement en diffusant de l'information. Lors de congestions ou d'incidents sur un mode de transport, l'information diffusée offre la possibilité de changer d'itinéraire ou de mode de transport. L'information sur les émissions des gaz à effet de serre suivant les modes de transport pour faire un même déplacement va influencer le choix de certaines personnes. Les services délivrés par les STI sont vastes. La norme ISO 14813-1:2007 (« Systèmes intelligents de transport (ITS) – Architecture(s) de modèle de référence pour le secteur ITS – partie 1 : Domaine de service, groupes de service et services ITS ») les divise en 11 domaines : l'information aux usagers, la gestion du trafic et l'opérationnel, les véhicules, le transport de marchandises, les transports publics, les urgences, les transports en relation avec le paiement électronique, la sécurité routière, les conditions météorologiques et environnementales, les désastres, la sécurité nationale. Le STI, quel que soit le service à rendre, fonctionne schématiquement en trois phases : l'acquisition des données, le traitement des données, l'utilisation et la diffusion des données. Il faut impérativement collecter des informations pertinentes en temps réel au niveau des différents réseaux : routier, ferroviaire, maritime, fluvial, aérien...

L'un des critères important que l'usager va exploiter pour faire son choix de transport va être la notion de temps. Elle tend à remplacer la notion de distance d'après [1]. La notion de temps est, à la fois une notion facilement compréhensible par l'usager et est également souvent considérée comme un indicateur de performance pour le gestionnaire. Comme vu précédemment, la route représente le mode de transport le plus utilisé, notre intérêt va se porter sur ce mode de transport, ainsi ce manuscrit se focalisera sur les temps de parcours sur le réseau routier. Une situation de congestion entraîne une augmentation du temps de parcours et a un impact sur l'environnement. En effet, les conditions de circulation en saturation provoquent des arrêts et redémarrages, ce qui augmente la consommation d'énergie et la pollution de l'air. Pour éviter cette situation, une politique de régulation de la vitesse aura pour conséquence de limiter les arrêts et redémarrages. En plus, d'après [5] l'usager acceptera d'autant mieux une situation de congestion s'il connaît son temps de parcours. Sa conduite sera plus apaisée et il ressentira moins de stress. En recevant l'information, il aura l'opportunité d'adapter son itinéraire, de différer son déplacement ou de changer de mode de transport.

Il convient aussi de connaître l'origine et la destination des usagers, de manière à proposer des alternatives. La mise en place de réseaux de transports en commun sera ainsi optimisée, et il sera possible de déterminer des aires de covoiturage par exemple. Le gestionnaire pourra également contraindre une partie du trafic à utiliser un itinéraire secondaire pour fluidifier un itinéraire surchargé. De nombreuses solutions existent, mais pour mettre en place les plus adéquates, les gestionnaires de réseaux et les usagers doivent disposer d'estimations fiables et précises.

Wang et coauteurs [6] discernent deux types de données pour le domaine routier : la surveillance des flux et la surveillance des routes. D'un côté, la surveillance des flux utilise les données macrosco-

piques collectées en un point telles que le taux d'occupation (c'est le temps durant lequel un point de la chaussée est occupé par un véhicule), la vitesse, le comptage... De l'autre, la surveillance des routes exploite les trajets individuels de chacun des véhicules détectés et suivis.

Pour déterminer les temps de parcours, trois familles de méthodes existent : celles qui utilisent les données macroscopiques, celles qui utilisent les données de suivi de véhicules, et celles qui utilisent la fusion de ces deux types de données.

Les différentes méthodes pour le calcul des temps de parcours à partir des données macroscopiques sont décrites dans [5] et [7]. Parmi elles, la méthode de la vitesse locale individuelle et la méthode des vitesses locales moyennes sont plus particulièrement adaptées à des configurations autoroutières, mais ne conviennent pas à un réseau urbain. Les méthodes sur les lois du trafic sont fondées sur la théorie des files d'attente ou sur les lois d'écoulement du trafic d'après [5, 7]. Elles sont plus fiables que les méthodes précédentes en ce qui concerne les phases de transition (formation et résorption de la congestion). Les méthodes statistiques (approches stochastiques décrites dans [8]) sont plus pertinentes pour les réseaux urbains. Elles utilisent principalement les données de taux d'occupation et de débit. Le temps de parcours est estimé par régression. Les paramètres du modèle sont évalués pour un réseau donné et ne peuvent pas être généralisés pour d'autres. La méthode de Bonvallet et Robin – Prévallée (BRP) ([5] et [7]) par exemple, fondée sur une relation empirique entre temps de parcours et taux d'occupation, nécessite quant à elle de nombreux paramétrages pour être opérationnelle.

Un panel d'approches différentes ([5] et [7]) permet d'estimer les temps de parcours mais ces méthodes diffèrent selon le type de réseau étudié et sont parfois difficiles à paramétrer. Une mesure directe des temps de parcours individuel par le suivi de véhicules éviterait les erreurs d'estimations des modèles indirects. Actuellement plusieurs types de capteurs permettent le suivi de véhicules avec plus ou moins de succès et délivrent ainsi les temps de parcours individuels.

Des méthodes de fusion des données de suivi de véhicules et de données macroscopiques existent. L'un des principaux avantages de la fusion de données est de pallier les limites d'un capteur par un autre. Le but de la fusion est d'optimiser la complémentarité des différentes sources ainsi que la redondance d'informations pour obtenir la meilleure information possible. Le problème de fusion de données est délicat d'après [5], au vu de la différence de nature des données. En effet, celles-ci doivent être dans un même espace de représentation (par exemple cohérence spatiale, temporelle).

Cette thèse va plus particulièrement s'intéresser aux méthodes de suivi de véhicules. En effet, pour l'ensemble des réseaux que ce soit urbain ou interurbain, les méthodes de suivi de véhicules développées dans ce manuscrit restent identiques contrairement à d'autres méthodes. De plus, le suivi de véhicules est un moyen de répondre dynamiquement à l'analyse de réseaux et de fournir des informations pertinentes pour les matrices origine – destination, à savoir d'où vient un véhicule et où il va. La connaissance des matrices origine – destination se fait souvent grâce aux différents types d'enquêtes. Ces enquêtes ont un coût non négligeable et par conséquent ne sont pas réalisées fréquemment. Les données sont donc souvent obsolètes. Elles sont partielles car elles ne couvrent que des périodes données sur un jour précis et retranscrivent principalement le trajet domicile travail [7]. Il est donc très difficile d'obtenir des matrices origine – destination précises avec uniquement des enquêtes ou des données de trafic macroscopiques mais différentes techniques ont été développées en ce sens. La plupart des techniques d'estimation de matrice origine – destination utilisent des informations *a priori* qui sont désignées comme matrice origine – destination cible [9]. Il est donc très important d'avoir une matrice origine – destination cible qui soit la plus précise possible pour utiliser ces différentes méthodes [10–12]. Le suivi de véhicules permet d'établir directement l'origine et la destination d'un véhicule en dynamique. À partir de cette mesure, la reconstruction des matrices origine – destination précises serait plus simple.

Les applications de temps de parcours et de matrice origine – destination sont des indicateurs importants pour planifier le transport. Elles permettent d'assurer la régulation et la fluidité du trafic. Dans ce mémoire, ces deux indicateurs sont des applications du suivi de véhicules. Ces mesures directes offrent la perspective d'obtenir des résultats précis et généralisables à l'ensemble des réseaux.

1.3 Objectif de la thèse

L'objectif principal de cette thèse est d'optimiser les méthodes de suivi anonyme de véhicules afin d'obtenir des informations pour les applications de temps de parcours et de matrices origine – destination. Le suivi de véhicules sera effectué à partir des données issues de la boucle inductive (oscillateur) ou du magnétomètre.

Ces méthodes ont fait l'objet de publications. Pour celles utilisant la boucle inductive, un des apports de la thèse va se situer dans l'optimisation des paramètres pour la réidentification. L'autre apport toujours pour ce capteur est la transformation du signal. La boucle inductive a une zone de mesure surfacique. Cette largeur a pour effet de lisser le signal et provoque ainsi une perte de détail. Ainsi, un algorithme de déconvolution aveugle est proposé pour annihiler l'effet lisseur du détecteur et ainsi retrouver les détails de la signature « réelle ».

Pour les méthodes utilisant les données du magnétomètre, l'analyse des résultats de détection de véhicules a conduit au développement d'une extension d'un algorithme existant. Les développements proposés pour le magnétomètre permettent d'améliorer la détection des véhicules. Les méthodes utilisées pour la boucle inductive pour la réidentification sont aussi adaptées aux données provenant du magnétomètre. La réidentification des véhicules à l'aide du magnétomètre est réalisée à partir d'un seul capteur en temps réel.

L'ensemble des algorithmes développés pour la boucle inductive a pu être validé sur les données d'expérimentations réalisées sur des voies proches de la ville d'Angers. Même si les algorithmes ont été testés en laboratoire, ils peuvent être facilement implémentés pour une utilisation opérationnelle.

De même, les algorithmes proposés pour le magnétomètre ont été développés et validés dans le cadre du projet PREDIT Mesure et mOdélisation de la COngestion et de la POllution (MOCoPo, http ://mocopo.ifsttar.fr/).

1.4 Plan du mémoire

Le chapitre 2 présente une liste des principales technologies utilisées pour réaliser le suivi de véhicules. Ces technologies sont décrites succinctement ainsi que leurs qualités et défauts majeurs. Ce chapitre éclaire le lecteur sur les motivations qui nous ont conduit à sélectionner les capteurs magnétiques pour le suivi de véhicules. Ensuite, le principe de fonctionnement des capteurs magnétiques et leurs caractéristiques sont développés au chapitre 3.

Le chapitre 4 explicite comment le suivi de véhicules est accompli à partir des capteurs magnétiques. Ce chapitre évoque les trois étapes successives pour suivre un véhicule :

- la première étape, « la détection », est détaillée au chapitre 5 ;
- la seconde étape, « les prétraitements », est développée dans le chapitre 6 ;
- la troisième étape, « la réidentification », est expliquée dans le chapitre 7.

Une dernière étape concerne la réalisation d'application pour la gestion du trafic à partir des informations du suivi de véhicules. Le chapitre 4 se termine sur l'évaluation et les indicateurs de performances du suivi de véhicules.

Le chapitre 5 répertorie les différentes méthodes de détection utilisées. Le signal délivré par les magnétomètres ne convenant pas, le choix d'une de ces méthodes de détection est effectué. Des propositions sont apportées à cet algorithme pour répondre aux difficultés rencontrées lors de nos expérimentations avec les magnétomètres.

Le chapitre 6 évoque les prétraitements opérés pour s'affranchir des déformations du signal, pour compresser les informations. Une partie de ce chapitre s'intéresse plus particulièrement à la sélection de données pour la réalisation du suivi de véhicules. Les véhicules de catégories distinctes ont des

signatures variées alors qu'au sein d'une même catégorie les signaux sont « semblables ». Nous suggérons dans le cadre de la sélection de variables de différencier les catégories de véhicules pour améliorer la réidentification. L'autre partie concerne la déconvolution du signal de la boucle inductive pour retrouver plus d'informations sur le véhicule. Nous proposons dans ce cadre un nouvel algorithme de déconvolution aveugle qui ne nécessite pas une nouvelle estimation à chaque passage de véhicule.

Le chapitre 7 se concentre sur les différentes mesures de similarité entre deux signaux. La réidentification est abordée comme un cas particulier de classification. Nous avons adapté les algorithmes de classifications pour répondre aux besoins de la réidentification. Nous avons aussi proposé l'association de différentes mesures de similarité ou de classifieur pour réaliser la réidentification.

L'aboutissement de ce travail est valorisé par la mise en pratique d'applications issues du suivi de véhicules. Le chapitre 8 est consacré aux expérimentations et démontre les améliorations apportées par les méthodes développées. Dans un premier temps, les expérimentations du chapitre se focalisent sur les temps de parcours individuels, issus du suivi de véhicules en différenciant les catégories de véhicules avec des boucles inductives. Dans un second temps, les expérimentations du chapitre traitent des matrices origine – destination à partir du suivi de véhicules avec des boucles inductives. Deux expérimentations sont réalisées : l'une compare le maximum de vraisemblance, la logique floue et l'association des deux méthodes ; l'autre compare la réidentification à partir de signaux déconvolués et non déconvolués. La dernière partie du chapitre, décrit le projet MOCoPo et l'expérimentation réalisée. Le suivi de véhicules est effectué à partir de magnétomètre. La mise en place de notre algorithme de détection est aussi évaluée.

Nous terminons ce mémoire, au chapitre 9, par un bilan sur le le travail réalisé et les différentes perspectives de ce travail.

2

Différentes technologies de capteurs

Sommaire

2.1 Véhicules traceurs (capteurs embarqués) **21**
2.1.1 Système de positionnement et de datation par satellite 22
2.1.2 Données des téléphones mobiles 22
2.1.3 Avantages et inconvénients 23
2.2 Suivi de véhicules point à point **24**
2.2.1 Capteurs non intrusifs 24
2.2.2 Capteurs intrusifs 28
2.3 Choix des capteurs **30**

Historiquement, pour connaître les déplacements effectués, des enquêtes terrains étaient menées auprès des usagers en leur demandant de renseigner leur origine et leur destination par questionnaire. Elles s'effectuaient soit en adressant les questionnaires à une population déterminée par sa localisation géographique, soit en plaçant des enquêteurs sur le réseau concerné entraînant une gêne à la circulation. Ces enquêtes sont parfois encore utilisées mais restent très ponctuelles. L'évolution a été de relever les plaques d'immatriculation manuellement en plaçant des personnes aux origines et destinations étudiées. Ces méthodes restent fastidieuses et ponctuelles. Elles donnent une photographie de la surveillance des routes à un instant donné mais l'évolution des temps de parcours, des matrices origine – destination n'est pas connue en temps réel. De plus, ces enquêtes ne sont pas anonymes.

Le principal objectif de ce manuscrit est de réaliser le suivi anonyme de véhicules tout en offrant les mesures essentielles de trafic pour un gestionnaire, à savoir : le comptage, la vitesse, la classification et le taux d'occupation. Différentes technologies de capteurs permettent actuellement de réaliser le suivi de véhicules en temps réel. Deux situations de suivi de véhicules se distinguent : le suivi de véhicules à partir de véhicules traceurs c'est-à-dire de véhicules équipés d'un dispositif transmettant leur position ; le suivi de véhicules à partir de l'heure de passage du véhicule en différents points.

2.1 Véhicules traceurs (capteurs embarqués)

Les véhicules traceurs sont les véhicules transmettant au moins une information sur leur position grâce à un équipement embarqué. Les envois d'information peuvent se faire en continu et ne

nécessitent pas de passer en un point précis du réseau. Suivant la technologie utilisée, l'information transmise peut être plus ou moins complète. Ainsi dans certains cas, la vitesse et la classe du véhicule (bus, poids lourds...), voire la météorologie (capteurs de température, de détection de pluie...) sont également transmises [13].

2.1.1 Système de positionnement et de datation par satellite

Le véhicule est équipé d'un système de positionnement par satellite pour calculer sa géolocalisation (GPS : « Global Positionning System », GLONASS : « Système global de navigation satellitaire », Galiléo, Beidou). Le système permet de relever la position du véhicule à intervalles réguliers (généralement toutes les secondes [13]). La précision du système étant insuffisante pour positionner précisément le véhicule, d'autres capteurs (odomètre, gyromètre...) peuvent être utilisés pour améliorer la précision. Les véhicules embarquent un système de communication pour transmettre l'ensemble des informations à un centre de traitement des données comme le montre la figure 2.1. Les données vont être filtrées et projetées sur le réseau routier (Map-Matching) pour l'estimation de temps de parcours, l'analyse de la vitesse moyenne, la détection de congestion. Cette analyse est limitée par le nombre de véhicules traceurs et le manque d'information sur les sections sans véhicule traceur. L'analyse ne permet pas non plus de distinguer les différentes voies d'un axe.

Dans ce cadre, soit le gestionnaire du réseau routier possède ses propres véhicules qu'il envoie dans le flux de la circulation sur les itinéraires l'intéressant, soit il achète les données à des entreprises disposant d'une flotte de véhicules équipés. Les véhicules ainsi suivis peuvent transmettre des données précises sur leur itinéraire mais il existe une dépendance à un tiers et seulement une partie des véhicules est suivie. De plus, la question de la fréquence d'envoi des données se pose.

Cette technologie n'est pas retenue car le but est de réaliser un suivi de véhicules représentatif du trafic et d'offrir les principales mesures de trafic.

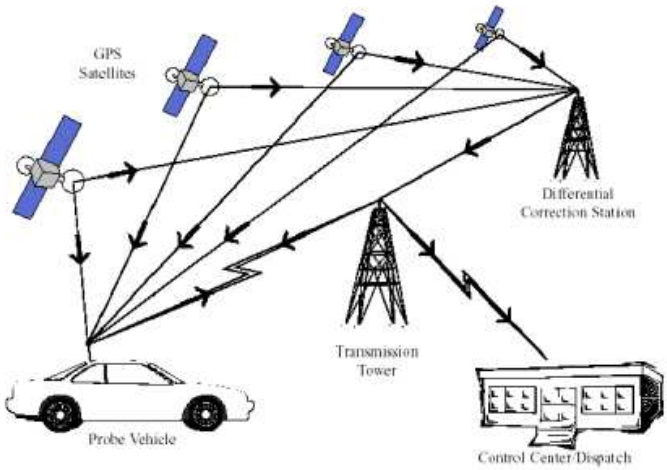


FIGURE 2.1 – Utilisation des satellites de positionnement (extrait de [14]).

2.1.2 Données des téléphones mobiles

Actuellement une grande partie de la population française est équipée de téléphones mobiles. Chaque téléphone portable est identifié par un numéro unique appelé IMEI (« International Mobile Equipment Identity ») composé de 15 chiffres. Lorsque le téléphone portable est allumé celui-ci cherche à se connecter à une cellule ou relais, pour cela il envoie son IMEI au relais. Une fois cette connexion réalisée, les informations personnelles de l'utilisateur sont envoyées par la carte SIM (« Subscriber Identity Module »). Ainsi l'itinéraire d'un téléphone portable peut être reconstitué. Une

cellule couvre une zone de transmission donnée, les cellules adjacentes recouvrent partiellement cette zone afin d'assurer une continuité des communications lorsque les utilisateurs se déplacent. De fait, le téléphone communique avec plusieurs cellules à la fois afin de faciliter le transfert intercellulaire sans interruption. L'opérateur téléphonique a ainsi la possibilité de localiser les téléphones lors de ces événements. Après utilisation d'algorithmes de filtrage et de repositionnement sur le réseau routier, les informations de temps de parcours, de détections de congestions sont établies. D'après [13], la précision est faible et cette technique nécessite que les utilisateurs soient en communication sur la route.

Lorsque les utilisateurs ne sont pas en communication, d'autres techniques existent pour géolocaliser les téléphones. Celui-ci se connecte à une antenne relais lorsqu'il se situe dans une zone couverte par le réseau. La précision de la localisation va alors dépendre de la densité d'antennes relais dans le secteur, d'après [13], en zone urbaine la précision de la géolocalisation varie entre 100 m et 700 m et en zone rurale jusqu'à 10 km. Une estimation un peu plus précise est réalisable si un programme a été installé au préalable sur la carte SIM du téléphone. Comme le montre la figure 2.2, une triangulation est effectuée entre trois antennes relais lorsque le téléphone se déplace. D'après [13], la précision en zone urbaine est de 150 m et de 5 km en zone rurale.

Le gestionnaire routier dépend des opérateurs de téléphonie pour l'achat des données. Les données vont permettre la détection de congestion et donner les temps de parcours mais n'offrent pas l'ensemble des principales mesures de trafic.

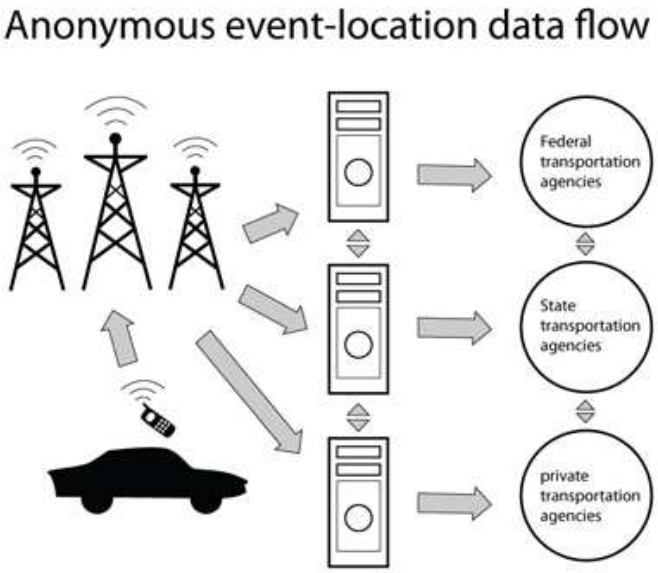


FIGURE 2.2 – Principe de localisation par triangulation.

2.1.3 Avantages et inconvénients

Les données issues de véhicule traceurs ont l'avantage de ne pas générer d'investissement au niveau de l'infrastructure puisque cette technologie utilise des infrastructures existantes. Un autre intérêt des véhicules traceurs est de bénéficier de l'itinéraire complet d'un véhicule. Suivant la méthode utilisée, cet itinéraire sera plus ou moins précis. En fonction du taux de pénétration de l'équipement, la répartition spatiale des véhicules équipés est variable et certaines classes de véhicules peuvent être sous-représentées. Pour les données telles que les temps de parcours et la détection de congestion, ces technologies sont bien adaptées. Mais ces technologies ne délivrent pas d'informations sur des données importantes pour le gestionnaire de réseaux telles que le taux d'occupation, la classification... Ces technologies sont complémentaires à d'autres capteurs afin que le gestionnaire de réseaux puisse

disposer de l'ensemble des principales données de trafic. De plus, certaines de ces méthodes ne garantissent pas l'anonymat de l'usager même si les données peuvent être cryptées. Ces méthodes ne seront par conséquent pas retenues pour le suivi de véhicules dans ce manuscrit.

2.2 Suivi de véhicules point à point

Dans le cas du suivi point à point, les méthodes consistent à identifier un véhicule en un point origine et à pouvoir le réidentifier à un point destination. Afin de réaliser la réidentification, il faut être en mesure d'extraire les caractéristiques uniques d'un même véhicule. Suivant le capteur utilisé, les caractéristiques servant à l'identification du véhicule vont être différentes.

2.2.1 Capteurs non intrusifs

Les capteurs non intrusifs ne nécessitent pas d'être placés sur ou dans la chaussée. Ils sont positionnés : en bordure de la chaussée, sur des mâts, sur des portiques ou en-dessous des ponts.

Capteur vidéo

Deux processus sont développés à partir de la vidéo pour les mesures de trafic. Le premier processus consiste à détecter les plaques d'immatriculation des véhicules et ensuite à utiliser des algorithmes de reconnaissance de caractères. Afin d'améliorer et d'accélérer la détection des plaques d'immatriculation, une région spécifique de l'image est souvent déterminée. Pour réaliser la lecture automatique de plaque, deux configurations de matériel sont utilisées :

- soit une caméra à champ proche est utilisée, dans ce cas, la caméra est dédiée à une seule voie. Le coût de la caméra est abordable et les traitements d'images sont simplifiés mais cela nécessite l'achat d'autant de caméras que de voies à surveiller ;
- soit une caméra haute définition est utilisée, engendrant un coût d'achat supérieur mais des coûts de maintenance et d'installation moindres. Les performances d'après [13] sont variables suivant la densité du trafic à surveiller et les algorithmes utilisés.
- La plaque d'immatriculation d'un véhicule étant unique, le suivi de véhicules utilise cette caractéristique. Ceci pose des problèmes d'acceptabilité de la part de l'usager. Même si la plaque d'immatriculation est cryptée, l'usager peut ressentir une atteinte à la vie privée.

Le second processus consiste à détecter les véhicules en analysant les images vidéos comme le montre la figure 2.3. Par exemple, la présence de groupes de pixels différents de l'image de fond (la route sans véhicule) représente les véhicules détectés. D'autres traitements sont appliqués pour améliorer la détection : le suivi des mouvements de groupes de pixels sur plusieurs images successives, analyse colorimétrique, suppression du bruit (l'ombre des véhicules...). Différentes caractéristiques sont extraites pour un véhicule, par exemple la forme, la couleur, la longueur...

Pour les deux processus, dans le cas de faible luminosité, l'acquisition de la vidéo peut se faire dans le domaine de l'infrarouge mais les informations colorimétriques ne sont plus disponibles et l'image est acquise en niveau de gris.

Les limites de fonctionnement pour la vidéo sont : les conditions météorologiques (pluie, neige, brouillard...), l'interdistance entre les véhicules (masquage par le véhicule précédent...), les flous de bougés (occasionnés par le support, par exemple les mâts subissent parfois le vent ou des vibrations dues au passage de certains véhicules), les conditions de luminosité (ensoleillement, nuit...). Pour ce dernier facteur limitant, des caméras dites jour/nuit sont aujourd'hui capables de basculer automatiquement suivant un seuil de luminosité du domaine du visible au domaine de l'infrarouge.

Les informations délivrées par les deux processus sont le débit, le taux d'occupation, le temps intervéhiculaire, la vitesse. Le suivi de véhicules est possible en utilisant la plaque d'immatriculation dans un cas et les caractéristiques des véhicules dans l'autre cas. L'avantage du deuxième processus est de donner en plus la distance intervéhiculaire et de différencier au moins les véhicules légers des poids lourds. Suivant les algorithmes implémentés, le deuxième processus peut distinguer les véhicules en quatre classes et les deux roues motorisées. D'après [13], il n'est pas possible de distinguer un autocar d'un camion. Les résultats obtenus varient suivant l'implantation (en bord de voie, terre-plein central ou en surplomb), les conditions météorologiques et les algorithmes implémentés. Une maintenance est à prévoir pour le nettoyage des vitres des caissons des caméras. La mise en place de l'équipement est parfois chère, néanmoins, l'ensemble des mesures de trafic est disponible. Cependant une partie des usagers a un sentiment d'atteinte à sa liberté individuelle, surtout si la lecture de plaques minéralogiques est pratiquée.

FIGURE 2.3 – Capteur vidéo (extrait de Macq Electronique – iCAR-CLASS).

Capteurs micro-ondes

Les capteurs micro-ondes sont plus communément appelés RADAR pour « RAdio Detection And Ranging ». Ce capteur utilise une antenne directive pour émettre des ondes électromagnétiques. Comme le montre la figure 2.4, dans le domaine routier, cette même antenne va recevoir le signal réfléchi par tout véhicule traversant le faisceau d'émission. Il existe principalement deux types de capteurs micro-ondes pour les équipements routiers ([15]) : le radar à ondes continues (radar Doppler) et le radar à ondes continues modulées en fréquence.

Le radar Doppler calcule la différence de fréquences appelée « fréquence Doppler » entre l'onde émise et l'onde réfléchie. Cette différence est proportionnelle à la vitesse radiale du véhicule. À partir de la vitesse radiale, la vitesse du véhicule est estimée. Il est à noter que ce type de capteur ne peut pas détecter les véhicules arrêtés. L'amplitude et la forme du signal reçu sont liées à la silhouette du véhicule et à sa surface de réflexion, ce qui permet de distinguer les poids lourds des véhicules légers. Les facteurs limitant pour ce type de capteurs sont les véhicules à faible vitesse (inférieure à $20\,\mathrm{km\,h^{-1}}$), une très forte densité de véhicules, les petits gabarits de véhicules, les fortes accélérations et décélérations. Le pourcentage d'erreur au niveau de la vitesse d'après [14] est estimé à 7,9 %.

Le radar à ondes continues modulées en fréquence, comme son nom l'indique, utilise la modulation en fréquence. La différence de temps entre l'émission et la réception du signal détermine la distance entre le radar et le véhicule, en mouvement ou à l'arrêt. La mesure de la vitesse nécessite deux zones de détection. Ces capteurs estiment très bien la vitesse et ils ne sont pas affectés par les conditions météorologiques d'après [14]. En revanche, ils nécessitent une maintenance particulière afin de vérifier la bonne orientation du capteur.

Capteurs infrarouges

Deux types de capteurs infrarouges se distinguent : le capteur infrarouge passif et le capteur infrarouge actif.

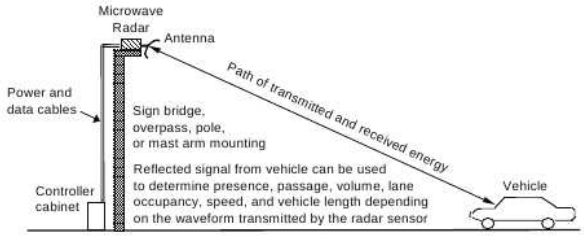


FIGURE 2.4 – Capteur micro-ondes (extrait de [16]).

Le capteur infrarouge passif est sensible à la chaleur émise par les véhicules et les piétons. Lorsqu'un véhicule entre dans le champ de mesure d'un détecteur passif, des changements d'énergie sont détectés en présence du véhicule. La différence d'énergie détectée créée par le véhicule est décrite par la théorie du transfert radiatif. Lorsque le capteur ne comporte qu'une seule zone de détection, le passage du véhicule permet de mesurer le débit et le taux d'occupation. Les capteurs avec de multiples zones de détections, dont l'intervalle des zones est connu, mesurent en plus la vitesse et la longueur du véhicule. Les capteurs infrarouges passifs offrent une faible précision au niveau de la vitesse et de la longueur, d'après [15], l'erreur sur la mesure de la vitesse est supérieure à 10 %. Un autre inconvénient de ce type de capteur est leur faisceau étroit qui ne permet pas de détecter tous les véhicules notamment les deux roues motorisées. Dans le cadre de l'utilisation d'un capteur vidéo sensible à l'infrarouge, les méthodes de traitement de l'image sont alors utilisées.

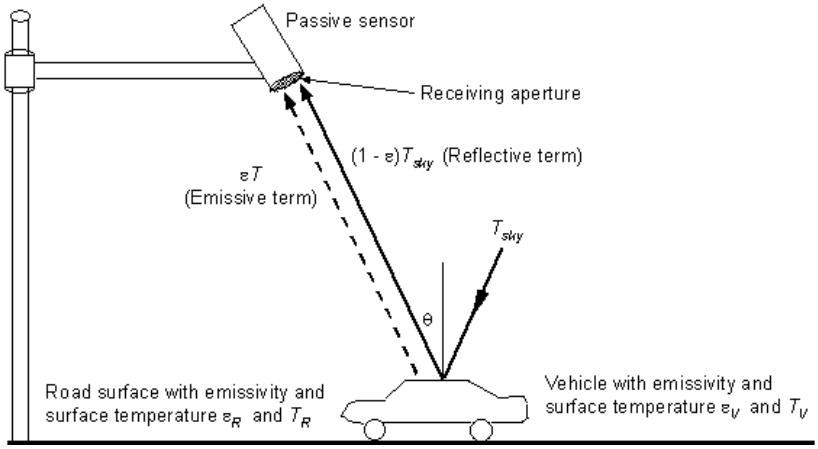


FIGURE 2.5 – Dispositif d'un capteur infrarouge passif (extrait de [16]).

Le capteur infrarouge actif fonctionne sur le même principe que les radars impulsionnels. Un émetteur envoie une impulsion qui se réfléchit comme le montre la figure 2.5. La détection d'un véhicule est effectuée grâce à la mesure du temps de parcours entre l'émission et la réception. D'après [14], ce type de capteur reste sensible à la météorologie, à la présence de particules atmosphériques et la pose peut s'avérer délicate.

De manière à couvrir la voie, il est possible d'utiliser de multiples sources de diodes laser pour émettre un certain nombre de faisceaux fixes ou d'utiliser des capteurs à balayage. Ces capteurs sont susceptibles de couvrir plusieurs voies. Un exemple d'une configuration de faisceau à balayage est représenté sur la figure 2.6. Les capteurs infrarouges les plus modernes peuvent reconstituer une image des véhicules en deux voire trois dimensions. La silhouette d'un véhicule est ainsi reconstituée et la catégorie du véhicule est obtenue. Comme pour la vidéo, des problèmes de masquage entre les différents véhicules peuvent arriver surtout lors de l'observation de multiples voies.

Capteurs acoustiques

Les capteurs acoustiques se divisent également en deux catégories : actif et passif.

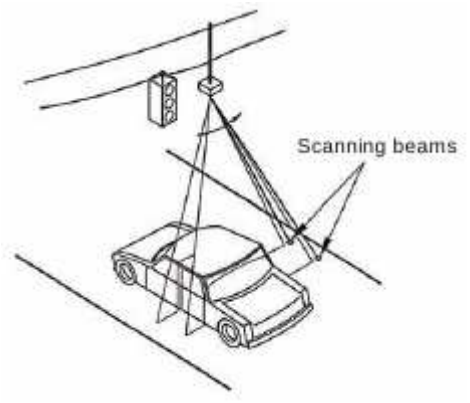


FIGURE 2.6 – Capteur infrarouge à balayage (extrait de [14]).

FIGURE 2.7 – Capteur acoustique SAS (extrait de www.smarteksys.com).

Le capteur acoustique actif est composé d'émetteurs et de récepteurs à ultrasons. L'émetteur envoie périodiquement un signal impulsionnel en direction de la chaussée. Lors du passage d'un véhicule le signal est réfléchi puis traité par le récepteur. A partir de la mesure du temps de parcours entre l'émission et la réception, la détection, le comptage et le taux d'occupation sont effectués et estimés. La présence d'un deuxième émetteur récepteur à proximité donne l'estimation de la vitesse. Ce capteur est très utilisé au Japon. Il offre de bons résultats en congestion mais il présente l'inconvénient d'une perte de précision selon les conditions météorologiques (brouillard et pluie) [13].

Le capteur acoustique passif mesure le bruit environnant. Puis, lors du passage d'un véhicule dans la zone de détection, une augmentation du niveau sonore est détecté. Il est composé d'une grille de microphones afin de quadriller une zone. Il est possible de détecter les véhicules, d'estimer leur vitesse, et de déterminer leur classification en appliquant quelques traitements supplémentaires. Il a l'avantage de pouvoir couvrir plusieurs voies, mais ses performances sont affectées par la température et le vent, la détection devient difficile pour les véhicules à faible allure d'après [14].

Radio-identification

La radio-identification est composée de lecteurs et de marqueurs ou transpondeurs. Le transpondeur est équipé d'une puce et d'une antenne. Lorsque le transpondeur est à proximité d'un lecteur, celui-ci interroge le transpondeur. Le transpondeur émet en réponse *a minima* son identifiant unique, il peut aussi émettre d'autres informations. Pour le suivi de véhicules, le véhicule est équipé d'un transpondeur qui l'identifie au passage à proximité des lecteurs. La radio-identification permet d'avoir des données fiables et précises pour les temps de parcours, les origines et les destinations mais nécessite d'équiper les véhicules. Il est à noter qu'une partie des usagers est déjà équipée de transpondeur puisque cette technologie est utilisée pour le péage automatique des autoroutes en France. La radio-

identification s'effectue généralement à des vitesses faibles et l'émission de certains badges à des puissances plus élevées endommage les lecteurs [13].

Collecte des adresses MAC (Media Access Control)

De plus en plus de véhicules et de personnes sont équipés d'appareils électroniques disposant du Bluetooth et du wi-fi. En installant des détecteurs le long d'un réseau, il est possible de collecter les adresses MAC des appareils. L'adresse MAC d'un appareil électronique (téléphone, GPS...) étant unique, le suivi de l'appareil est réalisé et le temps de parcours, l'origine et la destination de l'appareil sont calculés. Les limites de fonctionnement sont liées au taux de représentation des équipements à bord des véhicules. Plusieurs appareils peuvent être présents dans un même véhicule, et à l'inverse certains véhicules ne possèdent aucun appareil, ce qui rend difficile le comptage des véhicules par exemple. Dans des zones urbaines, la discrimination entre les piétons et véhicules est parfois difficile. Une ligne de transport en commun à proximité de l'antenne peut aussi avoir un impact sur la précision des mesures [13] faites à partir du Bluetooth. Les bornes wi-fi sont surtout présentes en milieu urbain et la précision de la localisation est inférieure à 50 m d'après [13].

Avantages et inconvénients

Les technologies telles que la collecte des adresses MAC et la radio-identification ne fournissent pas l'ensemble des données et seule une partie du trafic est analysée avec ces capteurs (véhicules équipés seulement). Mais ces technologies sont appelées à se développer. Ainsi, SCOOP@F est un projet de déploiement de systèmes de transport intelligents coopératifs. Ce projet vise à définir les normes de communication entre les véhicules et l'infrastructure pour l'échange d'information. Mais la mise en place et le taux de pénétration des équipements prendront du temps. Le lancement de l'expérimentation en grandeur nature commencera en 2016 avec environ 3000 véhicules et 2000 km de réseau instrumenté (d'après la fiche projet éditée en mai 2014 par la Direction générale des Infrastructures, des Transports et de la Mer).

La vidéo peut obtenir de bons résultats pour le suivi de véhicules et réaliser l'ensemble des mesures souhaitées, mais comme évoqué précédemment dans cette thèse, nous souhaitons que le suivi soit anonyme.

Le capteur acoustique passif a une précision moindre dans des situations de congestion. [14] relève 8 à 16 % d'erreur au niveau de la détection des véhicules. Le capteur acoustique actif est très sensible aux conditions météorologiques qui altèrent le fonctionnement et la précision des mesures.

Cheung et coauteurs [14] ont constaté, avec le capteur infrarouge passif, une faible précision des mesures avec 10 % d'erreur au niveau de la détection et de la vitesse des véhicules. Le capteur infrarouge actif a des difficultés à détecter des véhicules très sombres, et la précision des mesures se dégrade par temps pluvieux, en cas de brouillard et par des projections d'eau lors d'une chaussée humide.

La détection des véhicules à faible vitesse (inférieure à $20\,\mathrm{km\,h^{-1}}$) devient difficile pour la technologie radar. Les difficultés de détection sont aussi dues à une interdistance faible des véhicules. Le pourcentage d'erreur au niveau de la vitesse d'après [14] est estimé à 7,9 % pour la technologie radar.

D'après [13], une précision équivalente à l'ensemble des mesures fournies par la boucle inductive n'est atteinte par aucun des capteurs de trafic non intrusifs. Par contre, ils sont capables de donner de meilleurs résultats pour certaines mesures spécifiques.

2.2.2 Capteurs intrusifs

Les capteurs intrusifs sont des capteurs placés sur ou dans la chaussée. Leur mise en place nécessite la fermeture de la voie concernée.

Tuyau pneumatique

Le tuyau est un tube en caoutchouc qui est placé perpendiculairement au sens de la circulation sur la chaussée comme le montre la figure 2.8. Il est maintenu la plupart du temps par des brides fixées au sol aux deux extrémités grâce à des clous spéciaux. Une des extrémités est raccordée au capteur, l'autre étant bouchée. Le passage d'une roue sur le tuyau produit une surpression d'air se propageant dans le tuyau. Ce signal est transformé en impulsion électrique par l'intermédiaire d'un détecteur pneumatique. Ce capteur compte le nombre d'essieux lorsqu'un seul tuyau est en place. L'utilisation d'un deuxième tuyau espacé d'un mètre permet de déterminer le nombre de véhicules et les vitesses, ainsi que de réaliser la classification de ces véhicules, entre véhicule léger et poids lourd. Cette distinction est possible en considérant qu'un véhicule ayant une distance inter-essieux de plus de 3,45 m est un poids lourd. Ce capteur est facile à poser et peu onéreux, et il conserve de bonnes performances. Cependant, subissant de fortes sollicitations mécaniques au passage des essieux notamment des poids lourds (un trafic de poids lourds supérieur à 30 % accélère le vieillissement de l'installation), il a une durée de vie courte. Il est en général utilisé dans le cas de campagnes ponctuelles de mesures. D'après [13], il n'est pas recommandé de l'installer pour des campagnes de mesures supérieures à deux mois, et pour des axes présentant un trafic journalier supérieur à 10 000 véhicules par jour.

FIGURE 2.8 – Tuyau pneumatique (extrait de http ://www.iprocia.fr).

Câble piézoélectrique

L'effet piézoélectrique est la propriété de certains corps à se polariser sous l'effet d'une contrainte mécanique. Plusieurs matériaux ayant cette propriété existent, les plus couramment utilisés d'après [13] pour la réalisation de câbles piézoélectriques sont : la céramique, le quartz et les polymères. Ce capteur est enrobé de résine puis introduit dans la chaussée comme le montre la figure 2.9. Lorsqu'une pression est exercée sur le câble piézoélectrique par les pneus du véhicule, une tension électrique proportionnelle à la pression exercée est générée. Il mesure la charge d'un essieu et compte le nombre d'essieux. Il est alors possible de déterminer le débit de véhicules ainsi que la classification de ces véhicules. Il est alors possible de vérifier pour les poids lourds que leur charge à l'essieu est inférieure à la limite légale en vigueur. Mais ce type de capteur est sensible à l'hétérogénéité de l'uni de la chaussée, à sa déformation mécanique sous la charge, variant sous l'effet de la température et des vibrations. L'installation des câbles piézoélectriques génère une interruption du trafic. D'après [13], les véhicules présentant de faible charge à l'essieu sont difficiles à détecter suivant le type de chaussée. Les véhicules dont l'interdistance est faible ainsi que ceux possédant des remorques peuvent perturber les mesures. D'après Isaksson [17], ces capteurs engendrent des coûts d'installation et de maintenance élevés.

Fibre optique

Le mode de propagation de la lumière dans une fibre optique est modifié par sa déformation. Une fibre optique est installée dans la chaussée avec à une extrémité une diode émettrice et à l'autre

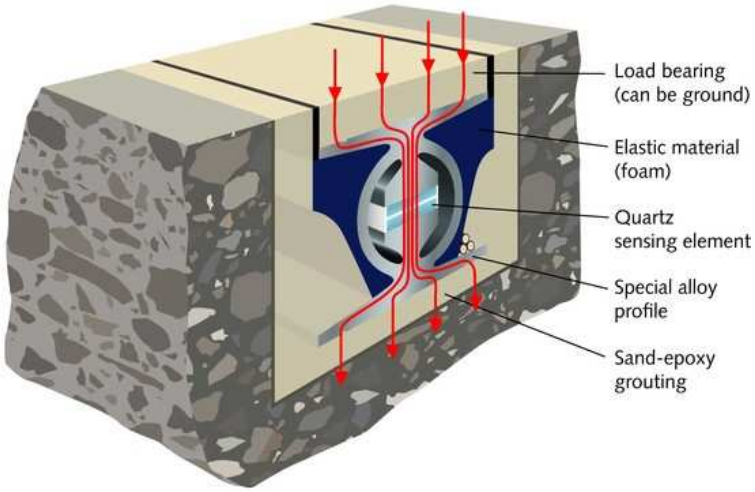


FIGURE 2.9 – Câble piézoélectrique (extrait de http ://www.invicom.com).

extrémité une diode réceptrice. Une modification du signal reçu est provoquée par le passage d'un essieu sur la fibre. Les différents algorithmes utilisés pour faire les mesures dépendent du type de capteur : modulation d'intensité, de phase, de polarisation, de fréquence. Ce capteur est polyvalent car un grand nombre de données peut être obtenu : le pesage, la détection, la classification selon le nombre d'essieux, la vitesse. La fibre optique est aussi précise et délivre un grand nombre de mesures de trafic ; le temps de réponse est rapide. Mais le coût d'installation reste élevé, des pertes sont engendrées par la courbure de la fibre et les connecteurs sont fragiles [15].

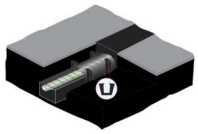


FIGURE 2.10 – Fibre optique (extrait de http ://www.sensorline.de).

Avantages et inconvénients

Le principal inconvénient de ces capteurs est l'interruption plus ou moins longue de la circulation pour leur mise en place. La solution utilisant le tuyau pneumatique n'est pas durable dans le temps. Ainsi ce capteur est dédié aux campagnes ponctuelles de mesures. Les autres capteurs intrusifs mentionnés délivrent en permanence et en temps réel les principales informations de trafic aux gestionnaires de réseaux. Ils respectent l'ensemble des conditions émises dans ce manuscrit.

2.3 Choix des capteurs

Les véhicules traceurs ne sont pas retenus suite aux inconvénients constatés précédemment, les principales mesures de trafic ne sont pas mesurées. Il en est de même pour la radio-identification et la collecte des adresses MAC.

Le tableau 2.1 permet de récapituler les différentes mesures réalisées par type de capteurs. Deux capteurs intrusifs n'ont pas été présentés précédemment et sont évoqués dans ce tableau : la boucle inductive et le magnétomètre. Ces deux capteurs seront présentés plus en détail dans la suite de ce manuscrit. D'après ce tableau seul deux capteurs permettent de réaliser l'ensemble des mesures : le câble piézoélectrique et la fibre optique.

Technologie	Comptage	Vitesse	Classification	Taux d'occupation	Poids
Boucle inductive	✔	✔	✔	✔	✖
Tuyaux pneumatiques	✔	✔	✔	✔	✖
Câbles piézoélectriques	✔	✔	✔	✔	✔
Radar à effet Doppler	✔	✔	✔	✔	✖
Radar à ondes continues modulées en fréquence	✔	✔	✔	✔	✖
Capteur infrarouge passif	✔	✔	✔	✔	✖
Capteur infrarouge actif	✔	✔	✔	✖	✖
Capteur acoustique passif	✔	✔	✔	✔	✖
Capteur acoustique actif	✔	✔	✔	✔	✖
Vidéo	✔	✔	✔	✔	✖
Fibre optique	✔	✔	✔	✔	✔
Magnétomètre	✔	✔	✔	✔	✖

Tableau 2.1 – Types de mesures réalisées par différentes technologies.

Un autre critère entre aussi en jeu pour le choix de la technologie, notamment la prise en compte du coût sur leur cycle de vie. Il est défini en reprenant les coûts d'acquisition, d'installation et de maintenance. Le tableau 2.2 de [14] représente le coût estimé pour une application de comptage et d'estimation de vitesse sur une autoroute deux fois trois voies. Lors de l'installation de certains capteurs, une interruption de la circulation est parfois nécessaire. L'étude de [14] n'inclut pas les coûts dus à cette interruption. Les boucles inductives ainsi que les magnétomètres nécessitent obligatoirement un arrêt du trafic. Les autres capteurs cités dans le tableau 2.2 peuvent dans certains cas nécessiter une interruption du trafic dépendant des conditions d'installation des mâts, portiques, et aussi suivant les sites. Le coût du cycle de vie pour les boucles inductives est celui de l'installation de paires de boucles. Actuellement il est possible de mettre une seule boucle par voie pour obtenir autant d'informations qu'avec une paire de boucles inductives, ce qui diminuerait le coût. Le magnétomètre, comme le capteur acoustique actif, semble présenter un coût avantageux par rapport à la majorité des autres technologies.

Technologie	Coût du cycle de vie en $	Durée de vie (en année)
Boucle inductive	1810	10
Radar à effet Doppler	2130	7
Radar à ondes continues modulées en fréquence	1370	7
Capteur infrarouge passif	1500	7
Capteur infrarouge actif	7130	7
Capteur acoustique passif	1700	7
Capteur acoustique actif	510	7
Vidéo	1730	10
Magnétomètre	890	10

Tableau 2.2 – Coût du cycle de vie pour différentes technologies (extrait de [14]).

Le dernier facteur pris en compte pour le choix du capteur est la précision des mesures. La précision de chaque technologie dépend du type de mesures et aussi des conditions dans lesquelles les mesures sont effectuées. La fibre optique offre une bonne précision, mais ce capteur a un coût élevé et est très peu déployé en France.

Le capteur le plus répandu en France, mais aussi aux États-Unis, est la boucle inductive. Elle présente l'avantage d'être performante, robuste et de récolter les principales données pour l'analyse du trafic en temps réel. Son inconvénient est principalement l'interruption de la circulation lors des opérations d'installation et de maintenance. Il est donc envisageable d'améliorer les performances de la boucle inductive, c'est-à-dire de pouvoir réaliser le suivi de véhicules au vu du patrimoine disponible.

Le magnétomètre pourrait remplacer efficacement la boucle inductive puisqu'il fournit les mêmes informations sur le trafic et qu'il nécessite une interruption de trafic moindre pour son installation et sa maintenance. De plus, le coût de déploiement massif des magnétomètres sur l'échelle d'un réseau serait moindre. L'inconvénient majeur de ce capteur est la correction à apporter due à l'influence de la température de la chaussée sur les mesures. La similitude entre les signaux de la boucle et du magnétomètre laisse présager que des techniques de traitement similaires peuvent être utilisées pour l'identification et la réidentification anonyme de véhicules (c'est-à-dire seulement à partir des caractéristiques magnétiques).

Les capteurs retenus dans ce manuscrit pour développer des méthodes de suivi de véhicules sont la boucle inductive et le magnétomètre. Le chapitre suivant expose plus précisément ces deux capteurs.

3

Capteurs magnétiques

Sommaire

3.1 Boucle inductive **33**
3.2 Magnétomètre **38**
3.2.1 Champ magnétique terrestre 38
3.2.2 Choix du magnétomètre 38
3.3 Conclusion **43**

3.1 Boucle inductive

FIGURE 3.1 – Boucles inductives.

La boucle inductive (figure 3.1) a fait son apparition dans les années 1960 en tant que capteur de trafic. Depuis, ce capteur est toujours utilisé et est l'un des plus répandus en Europe [13] et aux États-Unis [16] pour la gestion du trafic. Comme le montre la figure 3.2, les principaux composants de ce capteur sont :

- comme son nom l'indique, une boucle d'une ou plusieurs spires de fil conducteur implantée dans le revêtement de la chaussée ;

- un fil d'entrée de la boucle qui relie soit directement la boucle au détecteur soit la boucle à une boite de jonction ;
- un câble de liaison quand celui-ci est nécessaire pour relier la boite de jonction au détecteur ;
- un détecteur électronique stocké dans une armoire électrique en bordure de la chaussée.

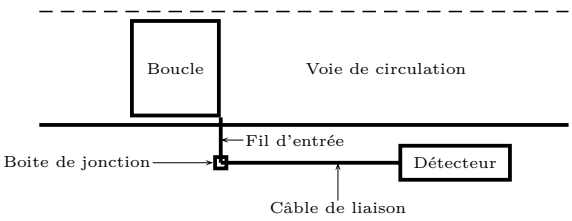


FIGURE 3.2 – Schéma simplifié des composants d'une boucle inductive.

Suivant les applications et les gestionnaires, une grande variété de formes d'enroulement pour les boucles est possible [16] : circulaire, orthogonale, trapézoïdale, losange, rectangulaire... Au niveau de la France, la norme PR NF P99-301 (« Données routières : élaboration, stockage, diffusion – Capteurs à boucles inductives – Définitions, caractéristiques et mise en œuvre ») définit les caractéristiques et la mise en œuvre des capteurs à boucles inductives. Ainsi, comme le montre la figure 3.3a, la boucle selon la norme PR NF P99-301 est un rectangle de longueur B et de largeur A avec un nombre de spires variant de trois à quatre. La largeur A est comprise entre 0,8 m et 3 m et la longueur B est supérieure ou égale à 0,8 m. Dans ce manuscrit les boucles utilisées ont une largeur A de 1,5 m et une longueur B de 2 m avec trois spires, c'est l'un des formats les plus répandus en France. Une saignée dans laquelle une boucle est insérée à une profondeur variant de 50 à 70 mm pour une largeur de 12 mm $\pm$ 5 mm est illustrée sur la figure 3.3b. En général la boucle est posée sur un lit de sable sec d'environ 10 mm d'épaisseur.

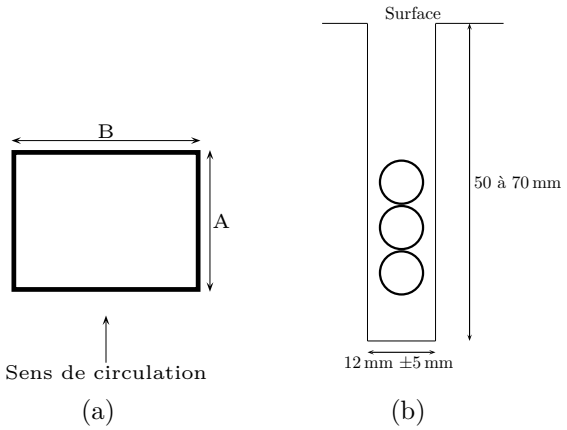


FIGURE 3.3 – Schéma d'implantation d'une boucle : (a) format rectangulaire ; (b) Coupe de la saignée.

La boucle inductive et le détecteur forment un circuit RLC. C'est un circuit électronique oscillant dont la fréquence est fonction de son inductance et de sa capacitance. Le capteur a le schéma électrique équivalent de la figure 3.4.

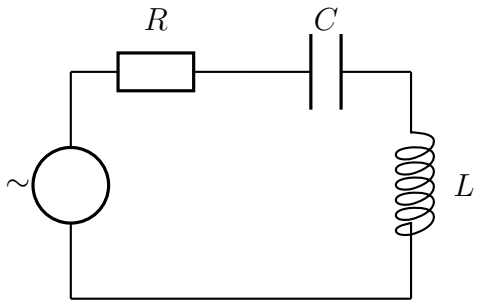


FIGURE 3.4 – Schéma d'un circuit RLC.

Suivant les modèles de détecteur, le signal électrique sinusoïdal appliqué est de l'ordre d'une dizaine de millivolts et la fréquence du signal est comprise entre 50 Hz et 150 kHz. La fréquence de résonance du circuit est donnée par l'équation suivante :

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{3.1}$$

avec f la fréquence du circuit en Hertz, L l'inductance totale (c'est-à-dire de la boucle et des câbles) et C la capacité totale du circuit (boucle et circuit électronique du détecteur). L'ensemble des calculs pour déterminer l'inductance totale et la capacité totale ont été décrits dans l'article [18] par Mills et repris ensuite dans le rapport technique [16]. Quand aucun objet métallique n'est au-dessus de la boucle, la fréquence d'oscillation est constante et est considérée comme la référence. Elle est notée f_{sv}. Le courant qui parcourt la boucle engendre un champ magnétique au voisinage de sa surface. Le champ magnétique généré par la boucle est exprimé par l'équation $H_b = \frac{N_b I_b}{l_b}$ avec H_b le champ magnétique en Henry, N_b le nombre de spires de la boucle, I_b l'intensité du courant en ampère et l_b la longueur en mètre de la boucle. Le flux magnétique est proportionnel au champ magnétique : $B_b = \mu H_b = \frac{\mu N_b I_b}{l_b}$ avec $\mu = \mu_0\mu_r$, μ_0 étant la perméabilité de l'espace vide et μ_r la perméabilité relative du matériau. Le flux ϕ_b à travers la surface S_b de la boucle s'exprime par l'équation $\phi_b = B_b S_b = \frac{\mu N_b I_b S_b}{l_b}$. L'inductance de la boucle est obtenue par : $L_b = \frac{N_b \phi_b}{I_b} = \frac{\mu N_b^2 S_b}{l_b}$. Dans le cas du capteur de trafic, le milieu n'est pas uniforme [16] et un facteur correcteur F_b est appliqué tel que : $L_b = \frac{\mu N_b^2 S_b F_b}{l_b}$.

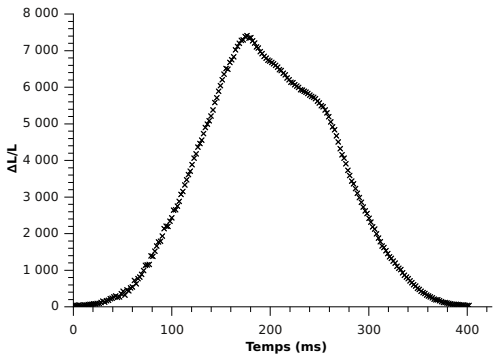


FIGURE 3.5 – Exemple de signature inductive.

Un détecteur à boucle inductive détecte la présence d'un objet métallique par la perturbation

de son champ électromagnétique rayonné. Selon les auteurs de [16], deux phénomènes opposés se produisent :

- la masse ferromagnétique de l'objet se comporte comme un noyau vis à vis d'une bobine et l'inductance de la bobine augmente en entraînant une diminution de la fréquence ;
- la masse métallique du véhicule est le siège de courants de Foucault qui par opposition induisent une diminution du champ électromagnétique. L'inductance de la bobine diminue et la fréquence de l'oscillateur augmente. C'est donc un phénomène équivalent à celui des pertes par couplage magnétique entre deux circuits.

L'effet des courants de Foucault est prépondérant par rapport à l'effet ferromagnétique d'après [16], dans le cas des boucles étudiées dans ce manuscrit. Par conséquent la présence d'un véhicule sur la boucle va produire *in fine* une diminution de son inductance. La variation de l'inductance ΔL est proportionnelle à la variation de la fréquence Δf selon l'équation suivante :

$$\frac{\Delta f}{f} = -\frac{1}{2}\frac{\Delta L}{L} = -\frac{1}{2}S \tag{3.2}$$

avec S la sensibilité du capteur. La majorité des détecteurs mesure la variation de la fréquence et les différentes technologies utilisées sont décrites dans [16]. Les fluctuations d'inductance ou de fréquence sont exprimées en valeur relative $\Delta L/L$ ou $\Delta f/f$ et sont fonction du temps. L'ensemble des points de mesures d'une détection représente la signature inductive du véhicule (figure 3.5) qui est affectée par la composition du châssis du véhicule.

L'interaction entre la boucle et le véhicule est modélisée par la figure 3.6. D'après [16], l'inductance mutuelle M entre la boucle et le véhicule est donnée par :

$$M = \frac{\mu N_b S_v F_b}{d} \tag{3.3}$$

avec S_v la surface du véhicule parallèle à la surface de la boucle en mètre carré et d la distance entre la boucle et le véhicule en mètre.

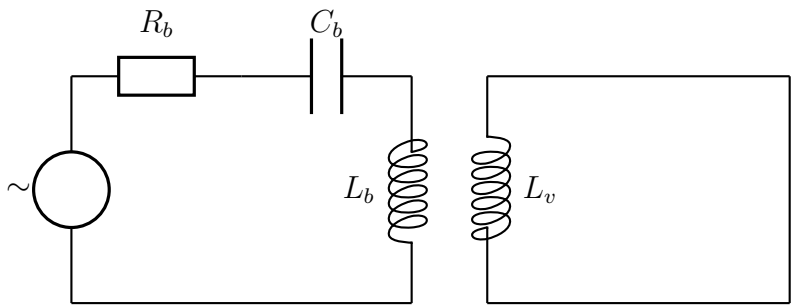


FIGURE 3.6 – Schéma d'interaction entre le véhicule et la boucle.

La sensibilité du détecteur est définie par la plus petite variation relative de l'inductance lors du passage du véhicule sur la boucle et qui provoque sa détection.

$$S = \frac{\Delta L}{L} = \frac{L_{sv} - L_{av}}{L_{sv}} = \frac{M}{L_b L_v} \tag{3.4}$$

avec L représentant la valeur de l'inductance ; L_{sv} et L_{av} l'inductance sans la présence de véhicule et avec la présence d'un véhicule, respectivement ; L_b l'inductance propre de la boucle ; L_v l'inductance

propre d'une spire court-circuitée modélisant le véhicule. D'après [16], la sensibilité peut s'exprimer comme suit :

$$S = \frac{S_v l_b l_v F_b}{S_b d^2 F_v} \tag{3.5}$$

La sensibilité du capteur se trouve être inversement proportionnelle au carré de la distance. La réponse de la boucle va diminuer fortement avec la distance. Les véhicules dont la garde au sol est élevée, comme par exemple les poids lourds, vont provoquer une diminution de l'inductance plus faible que les véhicules possédant une garde au sol faible. La réponse de la boucle est aussi fonction du ratio entre la surface de la boucle et la surface du véhicule. Quand la surface de la boucle est grande par rapport au véhicule, la sensibilité du capteur diminue. Ceci explique la difficulté à détecter des véhicules dont la surface est faible comme les deux roues avec la géométrie de capteur adoptée dans ce manuscrit. La sensibilité du capteur est aussi diminuée si le véhicule ne passe pas dans l'axe de la boucle, le ratio entre la surface du véhicule et de la boucle étant moindre. L'idéal serait d'avoir une boucle inductive de la longueur B équivalente à la largeur du véhicule et d'une largeur A suffisamment étroite. D'après l'étude de Gajda et coauteurs [19], les boucles inductives de petite largeur conservent mieux les détails du signal que les boucles inductives de taille standard. Due à un effet de moyenne glissante plus important, la signature issue de boucle inductive de taille standard ne contient donc pas les détails de la variation de l'inductance ([20]). Comme le montre la figure 3.7, une boucle inductive de plus petite largeur, 10 cm au lieu de 1,50 m par exemple, permet de mettre en évidence les essieux du véhicule.

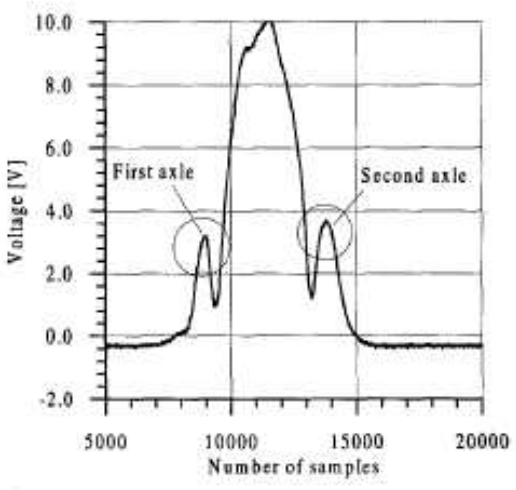


FIGURE 3.7 – Signature d'une voiture à partir d'une boucle inductive de 10 cm de largeur (extrait de [19]).

Les anciennes générations de détecteur transforment le signal électrique en un signal électrique tout ou rien, représentatif de la présence d'un véhicule. Cette variation se traduit par un créneau de tension dont la longueur est liée à celle du véhicule et à son temps de passage, limitant ainsi les informations obtenues. Les nouvelles générations intègrent une centrale d'acquisition qui analyse le signal numérique (Cf. Tableau 3.1). Avec une seule boucle inductive de la nouvelle génération au lieu de deux boucles de l'ancienne génération, plus d'informations sur le trafic sont obtenues.

L'intérêt est de travailler avec la signature de boucle inductive pour développer des méthodes de suivi de véhicules, le signal issu de la boucle inductive est noté s_x. Nous rappelons que ce signal provient d'un des formats les plus répandus en France de boucles inductives, à savoir un rectangle de 1,5 m de largeur et de 2 m de longueur. L'autre capteur utilisé au cours de nos expérimentations est le magnétomètre.

Mesure	Ancienne génération		Nouvelle génération
	1 boucle	2 boucles	1 boucle
Débit	✔	✔	✔
Taux d'occupation	✔	✔	✔
Intervalle intervéhiculaire	✔	✔	✔
Vitesse du véhicule	✖	✔	✔
Longueur du véhicule	✖	✔	✔
Distance intervéhiculaire	✖	✔	✔
Catégorie de véhicule	✖	✖	✔

Tableau 3.1 – Comparaison entre l'ancienne et la nouvelle génération de capteurs à boucle inductive

3.2 Magnétomètre

3.2.1 Champ magnétique terrestre

Le champ magnétique terrestre dépend de la localisation. La figure 3.8 représente un schéma simplifié du champ magnétique terrestre. Bien que le champ magnétique terrestre soit en perpétuelle évolution, cette variation au cours du temps est très faible, de l'ordre de 5 à 15 % sur environ 150 ans [17]. Dans le cadre de la détection de véhicule cela ne pose donc pas de problème, la variation est en effet négligeable lors du passage d'un véhicule.

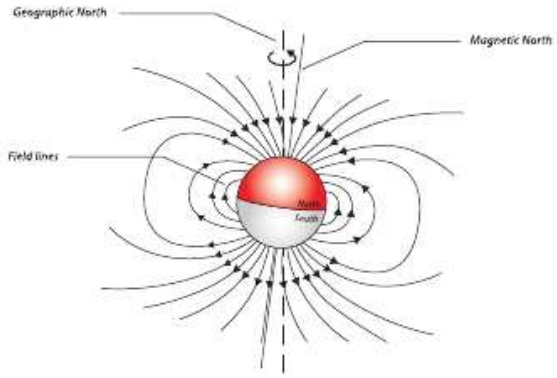
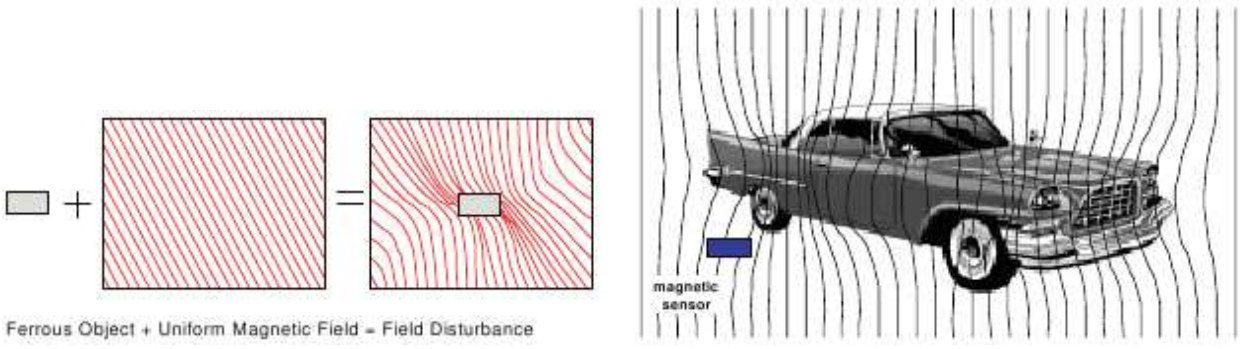


FIGURE 3.8 – Modèle simplifié du champ magnétique terrestre (extrait de [17]).

Pour la détection de véhicules, le but n'est pas de mesurer le champ magnétique terrestre mais la variation de celui-ci. Ensuite, les différents paramètres liés au véhicule sont estimés. En effet lorsqu'un véhicule traverse un champ magnétique, il perturbe ce champ et en provoque une variation. Tous les objets métalliques (figure 3.9a) perturbent le champ magnétique et peuvent être considérés comme des dipôles [14, 17]. Un véhicule est considéré comme un ensemble de dipôles qui perturbe le champ magnétique comme le montre la figure 3.9b.

3.2.2 Choix du magnétomètre

Actuellement, il existe plusieurs technologies pour mesurer l'intensité du champ magnétique terrestre. La figure 3.10 montre les plages de mesures des principaux capteurs magnétiques. La plupart des systèmes développés sont basés sur les phénomènes électromagnétiques. Il est difficile de comparer les capteurs magnétiques entre eux car chacun possède ses avantages et inconvénients. La technologie de capteur est aussi choisie en fonction de l'application visée. D'après Tumanski [22], quatre technologies sont prépondérantes actuellement : SQUID (Superconducting Quantum Interference Device), flux-gate, à effet Hall et magnétorésistance. De nouvelles technologies (comme les magnétorésistance



(a) (b)

FIGURE 3.9 – Perturbation du champ magnétique terrestre (extrait de [21]) : (a) par un objet métallique ; (b) par un véhicule.

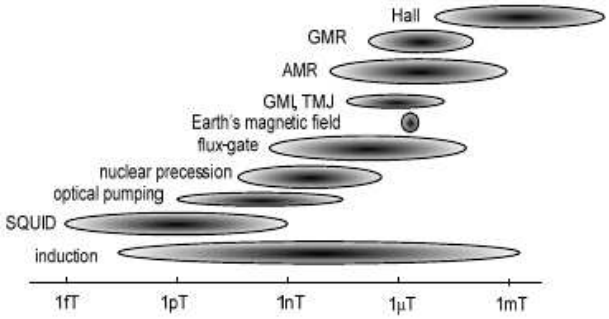


FIGURE 3.10 – Les résolutions des différentes technologies de capteurs magnétiques (extrait de [22]).

tunnel) se développent mais elles ne sont pas forcément disponibles au niveau des industriels. Sur les quatre technologies disponibles, l'amélioration des paramètres et la miniaturisation des composants sont toujours en cours de développement. Selon Tumanski [22], les différents paramètres à prendre en compte dans le choix du magnétomètre sont : la sensibilité, la linéarité, la portée, la bande de fréquence et les dimensions.

Les magnétomètres à SQUID sont parmi les plus sensibles pour la mesure des champs magnétiques mais plusieurs obstacles à leur utilisation pour la détection de véhicules existent actuellement. Leur fonctionnement est basé sur les propriétés électromagnétiques lors du refroidissement de certains matériaux en-dessous d'une température de transition supraconductrice. Selon l'article [22], ces magnétomètres nécessitent des températures très basses 9,3 K ou 120 K suivant la technique utilisée. Ceci n'est donc pas réalisable pour l'application souhaitée. De plus, les magnétomètres SQUID sont onéreux.

Les magnétomètres flux-gate sont performants et présentent de nombreux avantages. Ils ont une bonne sensibilité ($10\,\mathrm{mV\,nT^{-1}}$) pour une résolution de 10 pT. Les problèmes de dérive du zéro sont simples à résoudre d'après [22]. L'inconvénient de ce capteur est son coût par rapport à un déploiement à grande échelle, il reste par conséquent un capteur dispendieux.

Les capteurs à effet Hall sont faciles à réaliser, cela permet une miniaturisation et un faible coût du capteur. Ces capteurs sont linéaires sur une large plage de mesures. Cependant, ils ne sont pas très sensibles (au maximum $5\,\mathrm{mV\,mT^{-1}}$).

Le principe de base de la magnétorésistance est la variation de la résistivité d'un matériau ou d'une structure en fonction d'un champ magnétique externe. La trajectoire du courant est déviée par le champ magnétique, ce qui produit un allongement de la longueur du trajet du courant. Cet allongement est en fait une augmentation de la résistance effective. Cette définition inclut différents mécanismes qui produisent cet effet macroscopique. Mais actuellement d'après [22], seulement

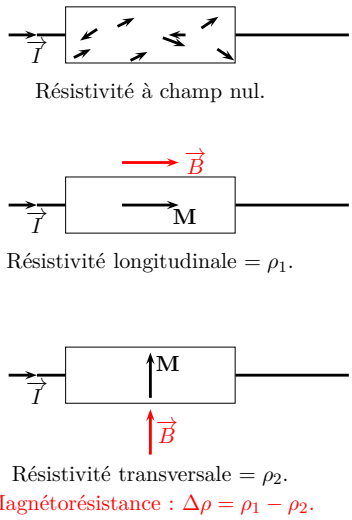


FIGURE 3.11 – Principe du capteur à magnétorésistance (AMR) : variation de la résistance électrique d'un matériau en fonction de l'orientation de l'aimantation et du champ magnétique.

trois effets principaux sont couramment utilisés : magnétorésistance anisotropes (AMR : Anisotropic MagnetoResistance), magnétorésistance géante (GMR : Giant MagnetoResistance), et jonction magnétique à effet tunnel (TMR : Tunnel MagnetoResistance). La magnétorésistance anisotrope est un effet typique des matériaux ferromagnétiques et a été découverte en 1857 par William Thomson. C'est à partir des années 1970–1980 que la technologie des capteurs à magnétorésistance a pu se développer. Ce capteur se compose d'une couche ferromagnétique très fine déposée sur une couche non ferromagnétique conductrice. Il peut détecter les champs statiques à courant continu, ainsi que la force et la direction du champ. Lors de la fabrication, la couche ferromagnétique est déposée sous un champ magnétique fort. Ce champ définit l'orientation du vecteur aimantation **M** appelé axe facile. Le vecteur **M** est parallèle à la longueur de la résistance. Lorsqu'un courant électrique traverse le matériau, les moments magnétiques de ses électrons s'alignent dans la direction de l'aimantation qui varie selon le champ magnétique appliqué. La résistance est maximale lorsque le courant circule parallèlement à **M**, elle est minimale lorsque les deux directions sont perpendiculaires. L'angle θ entre le courant et l'aimantation est modifié par le champ magnétique extérieur, ce qui entraîne une modification de la résistivité. La magnétorésistance s'exprime en fonction des résistivités longitudinale et latérale. Les propriétés physiques de la couche mince provoquent un changement de sa résistance en présence d'un champ magnétique. La variation relative de la résistance est faible ne dépassant pas 2 % d'après Tumanski [22]. Le capteur AMR est un bon moyen de mesurer à la fois les positions linéaire et angulaire et le déplacement dans le champ magnétique terrestre.

Les capteurs GMR découverts dans les années 1980 utilisent un autre effet. Les matériaux métalliques constitués d'un empilement de couches ferromagnétiques et non magnétiques ont leur résistance électrique qui diminue fortement en fonction de l'intensité du champ magnétique. La transition de l'état antiparallèle initial de l'aimantation à l'état parallèle s'accompagne par un changement de résistance. Ce modèle a été remplacé par des magnétomètres à vanne de spin. Dans les capteurs à vanne de spin, l'aimantation antiparallèle est obtenue artificiellement par dépôt d'une couche supplémentaire de matériau antiferromagnétique. De cette façon, un film ferromagnétique (spin) est magnétisé d'une autre manière que la seconde couche dite libre. Les capteurs TMR sont similaires avec une différence près, à la place du conducteur, il est utilisé une fine couche isolante. En changeant l'empilement des couches et en s'assurant de la possibilité de changement des orientations relatives de l'aimantation, une amélioration des paramètres des capteurs est possible. Ainsi la variation relative de la magnétorésistance augmente.

Le magnétomètre AMR a été retenu dans ce manuscrit. Ce magnétomètre présente un encombrement réduit. Les capteurs AMR peuvent paraître dépassés par les nouveaux capteurs GMR et TMR

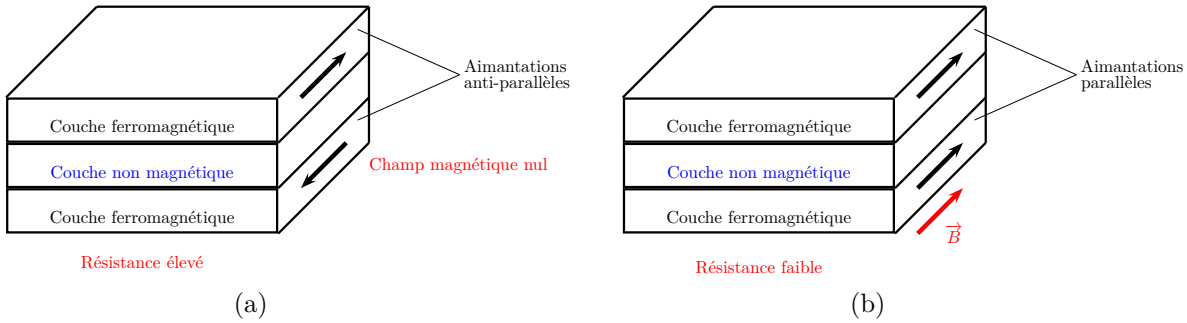


FIGURE 3.12 – Principe du capteur à magnétorésistance géante (GMR) : diminution de la résistance électrique d'un matériau en appliquant un champ magnétique.

mais ils ont, d'après [22] encore des avantages intéressants : ils sont simples à préparer et donc peu onéreux et ils ont une meilleure sensibilité que les capteurs GMR. Leurs principaux inconvénients sont les suivants : une faible variation relative de la résistance, ne dépassant pas 2 % ; la possibilité de démagnétisation par un champ magnétique élevé ; la sensibilité à la température de la chaussée qui doit être prise en compte.

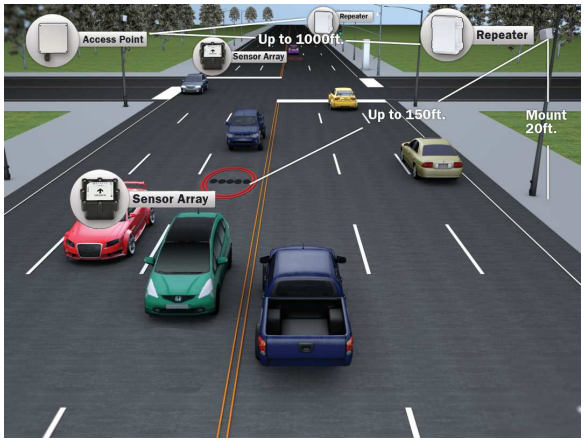


FIGURE 3.13 – Capteur magnétomètre (extrait de www.sensysnetworks.com).

La pose de ce capteur nécessite un carottage. Pour le capteur choisi, la transmission des données est réalisée par une liaison sans fil vers un émetteur récepteur (figure 3.13). L'optimisation de la consommation électrique du capteur est un point important pour sa durée de vie. Comme pour la boucle inductive, l'association de plusieurs magnétomètres permet d'obtenir des mesures telles que la vitesse, la classe du véhicule... Dans la majorité des études, les capteurs sont positionnés et intégrés sur les différentes voies. Suivant l'application visée (simple comptage de véhicules, détermination de la vitesse, réidentification des véhicules...), un ou plusieurs capteurs sont disposés sur la voie. Les capteurs magnétomètres les plus récents pour le trafic comportent trois axes. Les trois axes définissent un référentiel (x,y,z) orienté comme le montre la figure 3.14.

La circulation des véhicules sur la figure 3.14 se fait selon l'axe des x. L'axe y est perpendiculaire au sens de circulation de la voie, cet axe permet l'estimation de la position latérale du véhicule par rapport au capteur. L'axe z est dirigé vers le ciel. L'origine du repère est le centre du capteur. Les signaux issus du magnétomètres seront notés s_x, s_y et s_z pour les axes x, y et z, respectivement.

L'un des inconvénients de ce type de capteur est la sensibilité des magnétomètres anisotropes à la température. La figure 3.15 extraite de [14] montre l'évolution de la mesure magnétique pour

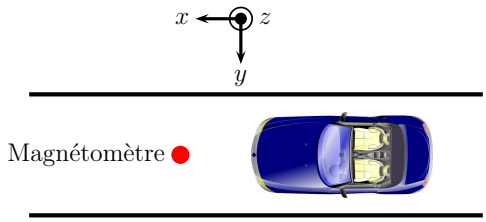


FIGURE 3.14 – Définition des axes du repère.

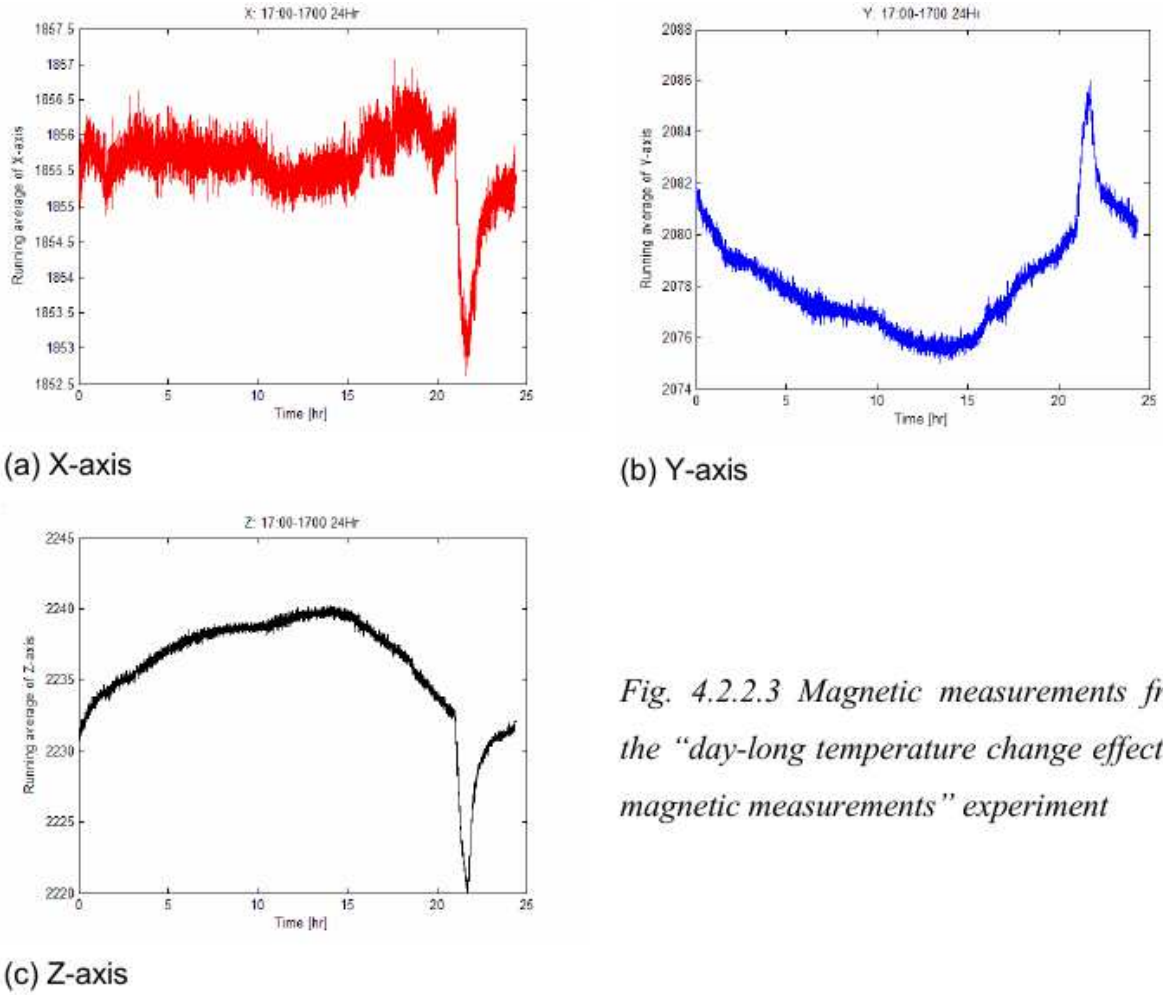


Fig. 4.2.2.3 Magnetic measurements from the "day-long temperature change effect on magnetic measurements" experiment

FIGURE 3.15 – Effet des changements de température sur une journée pour la mesure magnétique, les mesures commencent à 17 h (extrait de [14]).

un capteur placé à l'extérieur pendant 24 heures sur les trois axes. Dans cette situation, le capteur est placé de telle sorte qu'aucun véhicule ne puisse perturber la mesure. Les mesures ont commencé à 17 heures et se sont terminées le lendemain à 17 heures. Pour l'axe z, la mesure augmente dans un premier temps due à une température décroissante (passage du jour à la nuit) ensuite la mesure décroît avec l'augmentation de la température (durant la journée). L'axe y a un comportement inverse. Pour les trois axes, un pic est observé, correspondant d'après les auteurs à une soudaine exposition au soleil. Cette expérimentation montre l'effet de la température sur la mesure du champ magnétique terrestre par des magnétomètres AMR. La valeur sur l'axe x a varié entre 1852 et 1857, celle de l'axe y entre 2075 et 2086, et celle de l'axe z entre 2220 et 2240.

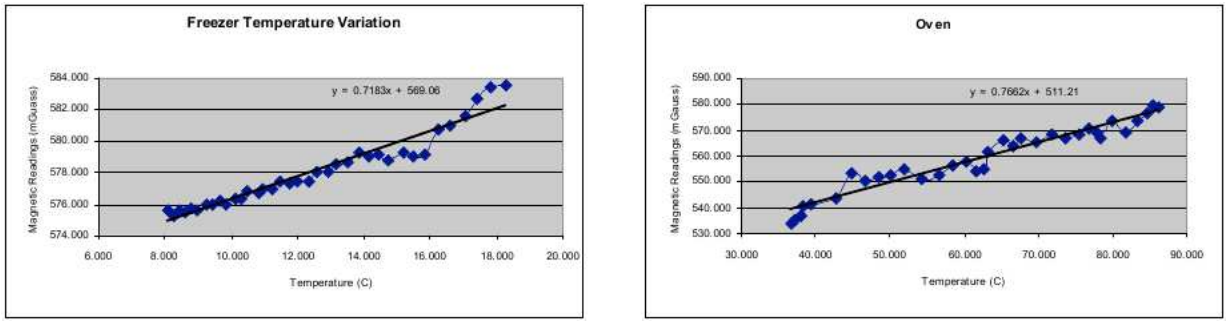


FIGURE 3.16 – Mesure magnétique en fonction de la température (extrait de [23]).

Pour éliminer l'effet de la température, Mesa et coauteurs dans [23] proposent une relation entre la variation de la température et la variation de la mesure du champ magnétique. Les courbes de la figure 3.16 représentent la variation du champ magnétique pour des températures comprises entre 8 et 18 degrés Celsius pour la première, 35 et 85 degrés Celsius pour la seconde. L'auteur a déduit une relation linéaire entre les deux paramètres à partir de plusieurs expérimentations. Ces variations de la mesure en fonction de la température devront être prises en compte lors du processus de suivi de véhicules.

3.3 Conclusion

Le développement de boucle inductive fiable et robuste est réalisable grâce à une parfaite maîtrise des propriétés majeures et la compréhension du principe de la boucle inductive. Le signal obtenu par la boucle inductive dépend principalement de la surface de celle-ci, de la surface du véhicule à détecter et de la garde au sol du véhicule à détecter. La géométrie la plus utilisée en France, un rectangle de 1,50 m de largeur sur 2 m de longueur, présente l'inconvénient de ne pas pouvoir mettre en évidence des caractéristiques du véhicule, telles que les essieux par exemple.

Le magnétomètre dont le développement est plus récent se trouve valorisé par une mise en œuvre plus rapide et moins dispendieuse. Cependant les mesures effectuées par magnétomètre sont sensibles à la température environnante. Néanmoins le magnétomètre est un capteur en pleine évolution dont la sensibilité, la linéarité, la portée, la bande fréquence et les dimensions ne cessent de progresser.

4

Principe du suivi anonyme de véhicules

Sommaire

4.1 Problématique du suivi de véhicules **45**
4.2 Analyse par les pelotons de véhicules **47**
4.3 Suivi individuel des véhicules **49**
4.3.1 Détection 49
4.3.2 Prétraitements 50
4.3.3 Réidentification 50
4.4 Évaluation et indicateurs **51**
4.4.1 Introduction 51
4.4.2 Méthode d'évaluation 52
4.5 Conclusion **53**

4.1 Problématique du suivi de véhicules

Le problème à résoudre n'est pas de suivre en continu un véhicule, mais d'identifier et réidentifier un véhicule lors de son passage sur des capteurs localisés à des endroits différents. Ainsi sur une section équipée de capteurs en entrée (Origine) et en sortie (Destination), il s'agit de comparer les caractéristiques d'un véhicule sortant aux caractéristiques d'un ensemble de véhicules entrés en amont et de trouver les deux qui sont identiques. Comme le montre la figure 4.1, une caractéristique des véhicules pourrait être la couleur de celui-ci. L'utilisation uniquement de la couleur va engendrer des incertitudes, par exemple un véhicule rouge passé à la destination peut être associé avec le bus rouge ou la voiture rouge à l'origine. La précision et la sensibilité de la mesure vont aussi avoir un rôle important, par exemple la couleur verte pour un véhicule peut être attribuée à plusieurs véhicules et nécessite de plus amples précisions pour caractériser un véhicule. Le fait que l'ensemble des entrées et des sorties ne soit pas instrumenté implique que certains véhicules n'ont pas d'appariement entre l'origine et la destination, c'est le cas pour la voiture mauve à la destination. De plus, même si l'ensemble des entrées et des sorties est instrumenté, des non détections de véhicules ou fausses détections vont engendrer des problèmes lors du suivi de véhicules. Le problème du suivi de véhicules

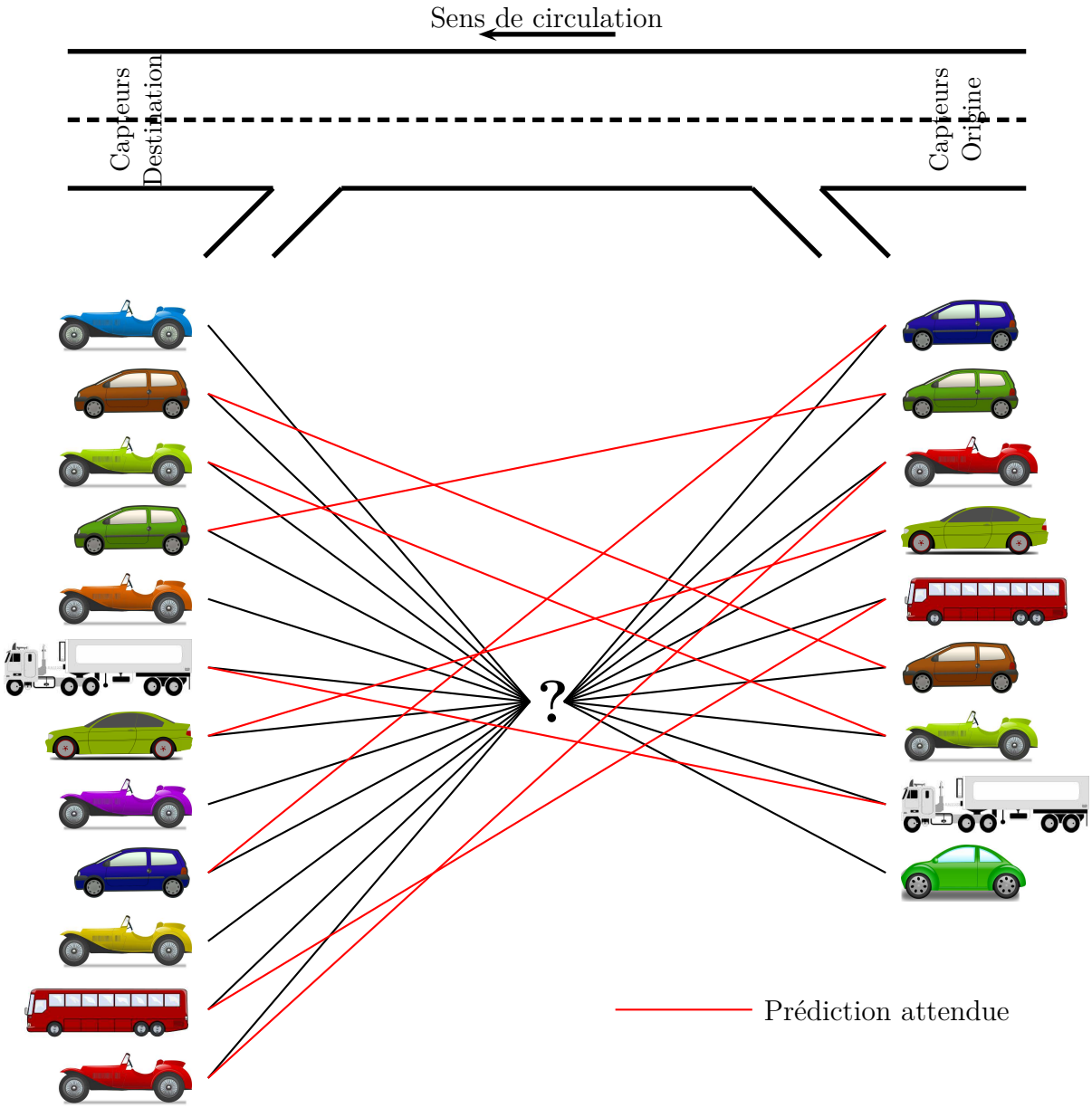


FIGURE 4.1 – Problématique de la réidentification.

n'est donc pas simple et nécessite d'avoir suffisamment d'informations pertinentes pour caractériser les véhicules de manière à les identifier puis les réidentifier ensuite.

Dans ce manuscrit, les capteurs employés sont les boucles inductives et les magnétomètres AMR. Lors de la détection d'un véhicule, l'ensemble des points de mesures est collecté et représente la signature du véhicule. L'appariement entre les signatures d'un même véhicule acquises par des capteurs de même nature à des localisations différentes est complexe. Lors de l'acquisition de la signature, des déformations apparaissent :

- en régime d'écoulement perturbé il n'est pas exclu que la signature soit déformée (par exemple, accélération – décélération sur le capteur) ;
- sensibilité du capteur ;
- passage dans l'axe ou non du véhicule ;
- ...

La figure 4.2 illustre la déformation de la signature d'un même véhicule pour un signal issu d'une boucle inductive. La signature représentée par la courbe noire a une durée inférieure aux deux autres signatures. La vitesse du véhicule lors de l'acquisition de la signature devait être plus élevée que lors de l'acquisition des deux autres signatures. L'amplitude inférieure de la signature verte peut s'expliquer par un passage du véhicule en-dehors de l'axe central de la voie.

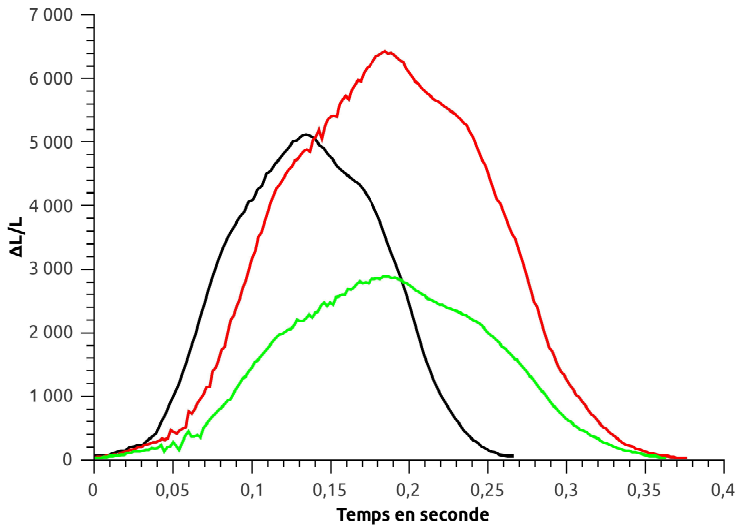


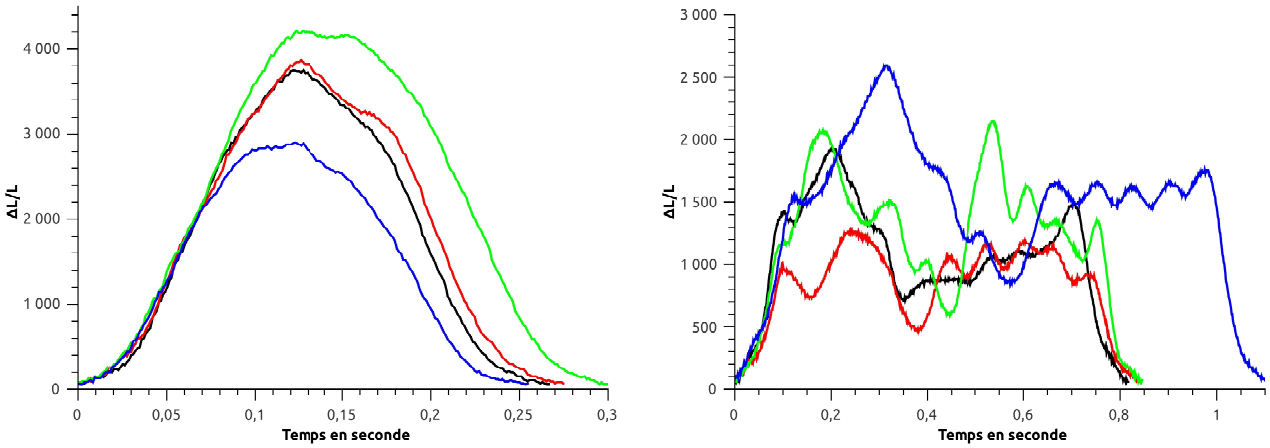
FIGURE 4.2 – Trois signatures d'un même véhicule issues de la boucle inductive.

Le risque de confusion est évident pour les véhicules légers du fait qu'ils présentent souvent des signatures très ressemblantes. La figure 4.3a illustre cette difficulté en montrant quatre signatures d'une même voiture issues de la boucle inductive. Au contraire, comme le montre la figure 4.3b, les poids lourds ont des signatures beaucoup plus diversifiées et par voie de conséquence plus faciles à apparier, d'autant plus qu'ils sont moins nombreux dans un trafic. Cependant, ils roulent plus lentement et ils ne sont pas forcément représentatifs de l'écoulement du trafic total.

Deux principes ont été développés pour le suivi de véhicules : l'analyse par les pelotons de véhicules et l'analyse par les caractéristiques individuelles des véhicules.

4.2 Analyse par les pelotons de véhicules

Les méthodes sur l'analyse des pelotons de véhicules ont surtout été développées par Coifman et coauteurs [24–27]. Une séquence de longueur de véhicules donne selon les auteurs plus d'informations



(a) (b)

FIGURE 4.3 – Différentes signatures de Véhicules Légers (VL) (a) et Poids Lourds (PL) (b) issues de la boucle inductive.

que les mesures individuelles des véhicules. Les véhicules origine et destination dont la longueur est approximativement la même sont associés ensemble. Les voitures ont souvent à peu près la même longueur, ce qui a pour conséquence d'associer plusieurs fois les véhicules. Des associations de véhicules se trouvent être des faux positifs.

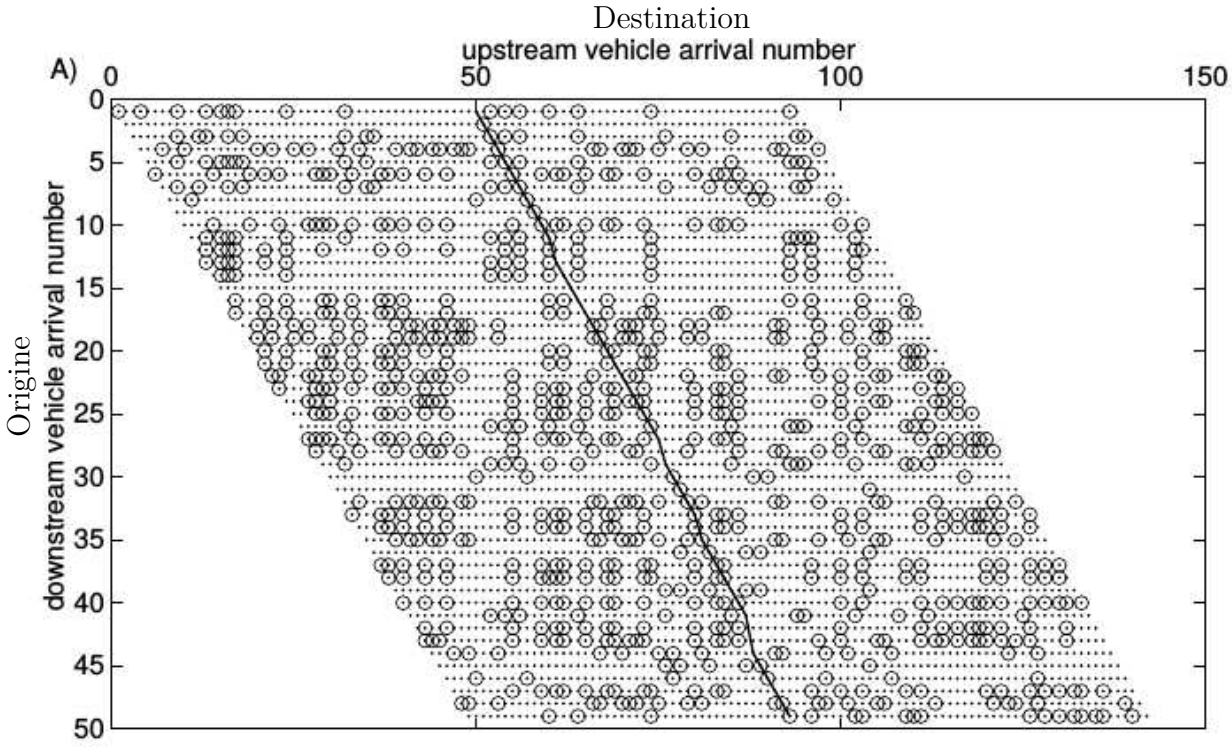


FIGURE 4.4 – Matrice des véhicules associés (extrait de [24]).

La figure 4.4 illustre les associations réalisées. L'ordre d'arrivée à l'origine des véhicules est représenté en ordonnée, et celui de la destination en abscisse. La véritable association entre l'origine et la destination est réalisée manuellement à partir de vidéos. Cette référence est représentée par le trait noir sur la figure 4.4. Les ronds représentent l'ensemble des associations possibles par rapport à la longueur du véhicule par la méthode proposée par Coifman et coauteurs [24]. En analysant les séquences

de véhicules, les pelotons de véhicules sont réidentifiés. Mais les véhicules à l'intérieur du peloton ne sont pas forcément associés correctement. Ensuite, dans [28–30] Coifman et coauteurs ont amélioré cet algorithme en filtrant les longueurs erronées. La dernière amélioration apportée par Coifman et coauteurs est basée sur la répartition des longueurs de véhicules. Dans [31], ils ont constaté que les véhicules de 7 à 25 m ne représentent que 10 % des observations du trafic, ces véhicules sont donc plus faciles à réidentifier. L'algorithme proposé dans [31] cherche d'abord à réidentifier les véhicules longs pour améliorer les performances. Li dans [32] a aussi travaillé sur la réidentification à partir de pelotons de véhicules en utilisant quatre mesures statistiques différentes sur les séquences de la longueur des véhicules. Soit une séquence de n véhicules à l'origine représentée par $\mathbf{l_o} = \{l_{o,i}\}$ et à la destination par $\mathbf{l_d} = \{l_{d,i}\}$ avec $i \in 1,2,...,n$. Les mesures statistiques proposées par Li dans [32] sont les suivantes : une mesure relative $RPS = \sum_{i=1}^{n} 2 \times \frac{l_{d,i}-l_{o,i}}{l_{d,i}+l_{o,i}}$; une mesure par la moyenne $APS = \sum_{i=1}^{n} \frac{l_{d,i}-l_{o,i}}{n}$; le coefficient de corrélation $CPS = \frac{\sum_{i=1}^{n}(l_{d,i}-\bar{l}_d)(l_{o,i}-\bar{l}_o)}{\sqrt{\sum_{i=1}^{n}(l_{d,i}-\bar{l}_d)^2}\sqrt{\sum_{i=1}^{n}(l_{o,i}-\bar{l}_o)^2}}$ avec $\bar{l}_o$ et $\bar{l}_d$ les valeurs moyennes de $\mathbf{l_o}$ et de $\mathbf{l_d}$, respectivement ; une mesure par la division $DPS = \sum_{i=1}^{n} \frac{l_{d,i}}{l_{o,i}}$. Ces différentes méthodes développées ont l'avantage d'être utilisables par l'ensemble des capteurs mesurant la longueur individuelle des véhicules. Elles ont été développées et testées à partir des mesures de boucles inductives. Elles pourraient aussi être utilisées avec les magnétomètres. À noter que ces méthodes sont particulièrement adaptées pour des situations de congestion où la séquence des véhicules à l'intérieur d'un peloton évolue peu. Pour Li [32], une limite de cette approche est le taux de réidentification insuffisant dans le cas où la distance entre deux boucles inductives successives est supérieure à 2 km. La longueur des véhicules est en fait une des caractéristiques mesurées par le capteur, il serait intéressant d'utiliser d'autres informations mesurées par le capteur pour améliorer la réidentification. Il est logique que la longueur est suffisamment discriminatoire pour les véhicules longs étant donné qu'ils sont peu représentés en temps normal au sein du trafic. Pour discerner les autres véhicules entre eux, d'autres caractéristiques des véhicules devraient être plus saillantes. L'hypothèse est de considérer que la signature de chaque véhicule est unique.

4.3 Suivi individuel des véhicules

Dans ce manuscrit, nous reprenons l'hypothèse de considérer que chaque signature de véhicule est unique qu'elle soit obtenue à partir d'une boucle inductive (Ritchie et coauteurs [33], Tabib et coauteurs [34]) ou d'un magnétomètre (Cheung et coauteurs [14]). Chaque véhicule possède des caractéristiques distinctes, c'est-à-dire que chaque véhicule possède sa propre signature. Le premier capteur d'un tronçon fournit les caractéristiques à l'origine des véhicules, et le capteur à l'autre extrémité du tronçon fournit les caractéristiques à la destination des véhicules. La notation $\mathbf{v}_o$ représente les caractéristiques d'un véhicule origine et $\mathbf{v}_d$ celles d'un véhicule destination. Les caractéristiques d'un véhicule sont considérées comme unique et les acquisitions sont noyées dans un bruit blanc gaussien centré $\mathbf{n}_o$ pour l'origine et $\mathbf{n}_d$ pour la destination :

$$\mathbf{v}_o = \mathbf{v} + \mathbf{n}_o \tag{4.1}$$

$$\mathbf{v}_d = \mathbf{v} + \mathbf{n}_d \tag{4.2}$$

À partir de cette hypothèse, le suivi de véhicules peut être décomposé en plusieurs étapes (figure 4.5).

4.3.1 Détection

Cette étape consiste non seulement à détecter le véhicule mais aussi à segmenter correctement la signature. Les non détections et les fausses détections vont perturber le suivi de véhicules. Ces fausses informations peuvent engendrer des erreurs dans le suivi de véhicules. La segmentation est

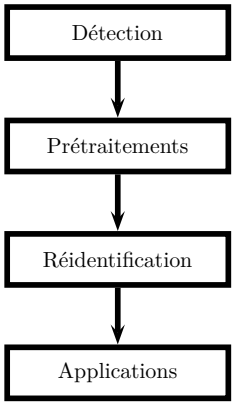


FIGURE 4.5 – Processus de suivi de véhicules.

importante car une mauvaise segmentation peut entraîner une perte d'information et par conséquent le suivi de véhicules sera plus difficile.

4.3.2 Prétraitements

Après avoir effectué la détection, les prétraitements visent à corriger les déformations du signal. Les causes de la déformation du signal sont multiples : la sensibilité différente de chaque capteur, la vitesse de passage du véhicule, un passage plus ou moins dans l'axe du capteur... Parmi les prétraitements, une sélection de paramètres peut être effectuée pour limiter les données.

D'autres prétraitements cherchent une réduction spatiale et temporelle. La réduction spatiale vise à ne considérer que les véhicules issus d'origines physiquement possibles, c'est-à-dire dans le cas d'un feu rouge les véhicules venant de cet axe ne sont pas des candidats possibles pendant les périodes ou le feu est rouge. La réduction temporelle consiste à établir une fenêtre temporelle pour sélectionner les candidats possibles. La fenêtre temporelle est établie par une période de temps raisonnable et réalisable de sorte que le véhicule correct peut être inclus dans l'ensemble des véhicules candidats. Pour délimiter les bornes inférieure et supérieure de la fenêtre temporelle, les limitations de vitesse sur l'itinéraire ainsi que les conditions de trafic peuvent être prises en compte. La fenêtre temporelle sera de préférence dynamique et tiendra compte de l'évolution des conditions de trafic. Les techniques de réduction peuvent faire appel à des algorithmes de modélisation du trafic. Cette étape permet de réduire l'ensemble de véhicules candidats.

Les prétraitements sont une étape importante pour préparer les données qui vont servir à la réidentification.

4.3.3 Réidentification

La réidentification est la dernière étape du processus de suivi de véhicules. Les caractéristiques (la signature par exemple) d'un véhicule sont comparées par rapport aux caractéristiques des véhicules candidats. L'algorithme de réidentification permet ou non l'association des caractéristiques avec celles d'un des véhicules candidats. Les résultats obtenus par la réidentification donnent le suivi des véhicules entre leur origine et leur destination. Puis, les applications telles que l'estimation des matrices origine – destination et temps de parcours peuvent être effectuées.

Ce manuscrit va s'intéresser aux étapes de détection, de prétraitement et de réidentification. Les prétraitements concernant la réduction spatiale et temporelle ne feront pas l'objet de propositions et ne serviront que pour certaines applications.

4.4 Évaluation et indicateurs

4.4.1 Introduction

Plusieurs expérimentations ont été menées afin de valider et évaluer les méthodes de suivi de véhicules proposées dans ce manuscrit. Lors des différentes expérimentations réalisées, plusieurs types de données sont acquises. Outre les données des capteurs étudiés, à savoir la boucle inductive et le magnétomètre, des vidéos sont réalisées. Les vidéos servent de référence pour confirmer ou non les réidentifications ainsi que pour vérifier les catégories des véhicules.

Boucle inductive

Les boucles inductives se divisent en deux catégories au cours des expérimentations : celles délivrant la signature et celles délivrant des données agrégées. Les boucles inductives délivrent les informations agrégées suivantes : la voie de circulation, l'heure de passage à la milliseconde, la vitesse (V) exprimée en kilomètre par heure, le temps intervéhiculaire (TI) exprimé en dixième de seconde, la longueur (L), exprimée en décimètre, le temps de présence exprimé en milliseconde et la distance intervéhiculaire exprimée en mètre.

Pour les boucles inductives relevant la signature, les informations recueillies en plus lors du passage d'un véhicule sont : le nombre de points de la signature, l'horaire de passage du véhicule, la catégorie, la signature, l'horaire dans le système heure minute seconde centième, ainsi que la date. Il est à noter que l'acquisition de la signature n'est pas effectuée avec une fréquence d'échantillonnage constante. Cette fréquence d'échantillonnage varie faiblement autour de 500 Hz.

Magnétomètre

Comme pour les boucles inductives, deux types de magnétomètres ont été utilisés : les magnétomètres de la société Sensys Network et les magnétomètres dits « signature » développés pour les besoins du projet MOCoPo par l'INRIA.

Les données acquises par les magnétomètres Sensys sont : le numéro de détection, la date et l'heure de passage à la seconde basée sur le référentiel GMT, la référence du capteur (point accès–voie), la vitesse (V) exprimée en miles par heure, la longueur (L) exprimée en pied, le temps intervéhiculaire (TI) exprimé en seconde. Les informations de mesures sont délivrées dans le système métrique américain.

Les données acquises par les magnétomètres enregistrant la signature magnétique sont : l'heure exprimée en milliseconde qui est réinitialisée quotidiennement, l'amplitude des signaux selon les trois axes x, y, z exprimée en millivolt. Il est à noter que l'acquisition du signal s'effectue à une fréquence d'échantillonnage constante de 128 Hz.

La vidéo

Chaque point de mesures du site d'expérimentation, c'est-à-dire l'emplacement des capteurs, est équipé de caméras numériques en nombre suffisant pour filmer correctement les plaques d'immatriculation (les plaques d'immatriculation sont encryptées pour des raisons d'anonymat) de tous les véhicules passant sur chaque voie. Le champ de la caméra est réglé de manière que les plaques soient enregistrées au droit des capteurs de mesures. La lecture des plaques se réalise manuellement. Un opérateur visionne les vidéos et recense tous les véhicules étant passés sur les zones instrumentées. Différents champs sont renseignés notamment la plaque d'immatriculation, le type de véhicule, l'horaire de passage et le nom du fichier contenant la signature liée à ce véhicule. Après le déchiffrement de toutes les vidéos, une base de données comprenant tous les véhicules passés sur un ou plusieurs points de mesures est obtenue. Tous les véhicules étant passés sur une origine et une destination

instrumentées sont connus. Cette base sert de référence et permet de vérifier les prédictions données par les différentes méthodes explicitées dans ce manuscrit.

FIGURE 4.6 – Enregistrement vidéo.

4.4.2 Méthode d'évaluation

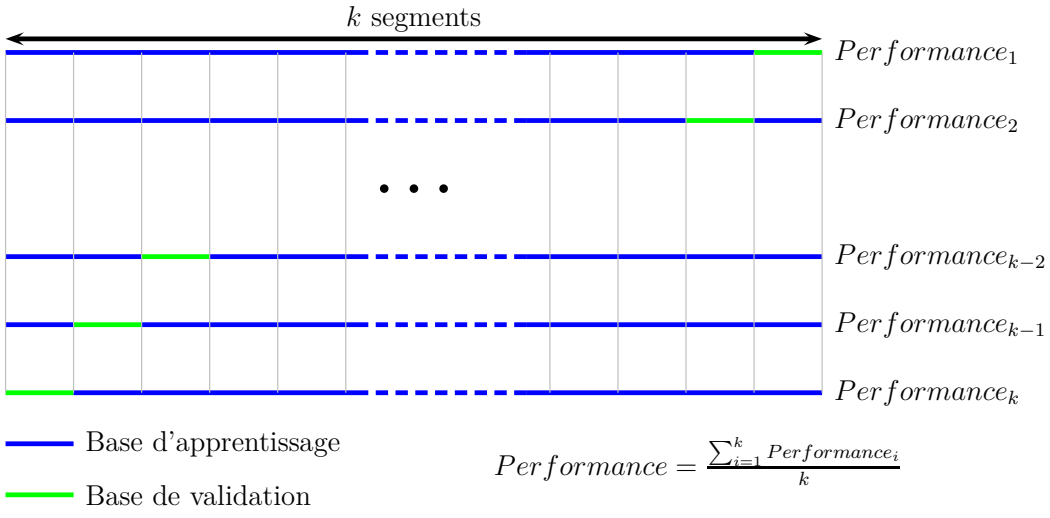


FIGURE 4.7 – Principe de la validation croisée pour k segments.

Pour évaluer les classifieurs, plusieurs méthodes existent : validation simple (ou hold out), leave one out, bootstrap... Dans ce manuscrit, nous avons choisi d'utiliser la validation croisée [35] : le classifieur est paramétré à partir d'un ensemble d'apprentissage, ensuite les performances du classifieur sont évaluées avec un ensemble de test. Afin de ne pas subir l'influence d'une base d'apprentissage qui favoriserait une méthode, il est proposé d'utiliser la k-validation croisée. La méthode de validation croisée demande de séparer la base entière de z véhicules en un ensemble d'apprentissage–validation (A–V) et un ensemble de test (T). Le premier ensemble (A–V) comprend $z/3$ véhicules et est subdivisé en k sous-ensembles disjoints de taille approximativement identique. Chaque sous-ensemble est

constitué aléatoirement. Chaque sous-ensemble est utilisé individuellement une et une seule fois pour la phase de validation, les $k-1$ autres sous-ensembles servent à l'apprentissage comme le montre la figure 4.7. Le classifieur retenu est celui présentant le meilleur score (erreur de généralisation minimum) de validation croisée, c'est-à-dire la moyenne des scores de chaque sous-ensemble de validation. En général k est pris entre 5 et 15. Dans un second temps, la base A–V sera utilisée pour l'apprentissage et les performances seront calculées sur la base de test (T).

Les performances utilisées seront évaluées par le taux de bonne réidentification (TBR), le taux de représentation (TR) et le taux de réidentification (TRi) définis comme suit :

$$\text{TBR} = \frac{\text{Nombre de bonnes réidentifications}}{\text{Nombre de réidentifications}} \times 100 \tag{4.3}$$

$$\text{TR} = \frac{\text{Nombre de bonnes réidentifications}}{\text{Nombre de réidentifications possibles}} \times 100 \tag{4.4}$$

$$\text{TRi} = \frac{\text{Nombre de réidentifications}}{\text{Nombre de réidentifications possibles}} \times 100 \tag{4.5}$$

Le taux de bonne réidentification (TBR) indique le pourcentage de couples corrects trouvés par rapport à l'ensemble des couples constitués par le classifieur. Le taux de représentation (TR) indique le pourcentage de couples corrects trouvés par rapport au nombre total de couples recensés grâce à la vidéo. Le taux réidentification (TRi) indique le pourcentage de couples trouvés par rapport au nombre total de couples recensés grâce à la vidéo. Le TBR est lié aux deux autres taux par l'équation suivante : $TBR = \frac{TR}{TRi} \times 100$.

4.5 Conclusion

Le processus de suivi anonyme par des capteurs magnétiques se décompose en trois étapes. La première étape consiste à détecter le véhicule et à segmenter le signal. La seconde étape cherche à minimiser les déformations du signal et à sélectionner les caractéristiques pertinentes. La dernière étape utilise les caractéristiques précédentes pour prendre la décision d'associer les signaux origine et destination.

La performance du processus est évaluée par le taux de bonne réidentification, le taux de représentation et le taux de réidentification. Ces différents taux sont estimés à l'aide de vidéos servant de référence lors d'expérimentations.

5

Détection

Sommaire

5.1 Réalisation de la détection 55
5.2 Problèmes rencontrés 61
5.3 Proposition d'un algorithme de détection 63
5.4 Évaluation des méthodes de détection 66
5.4.1 Évaluation 66
5.4.2 Conclusion 70
5.5 Conclusion 70

5.1 Réalisation de la détection

Le capteur réalise l'acquisition d'un signal brut échantillonné. Ce signal est acquis en permanence qu'un véhicule soit présent ou non. La première étape consiste à segmenter le signal, c'est-à-dire détecter le début et la fin de la présence du véhicule de manière à ne conserver que la partie du signal où le véhicule est présent. Pour évaluer les algorithmes de détection, deux indicateurs sont pris en compte : le taux de détection et le nombre de fausses détections. Le taux de détection est le nombre de bonnes détections sur le nombre de détections possibles. La fausse détection correspond à la détection erronée d'un véhicule quand celui-ci n'est pas présent.

Dans le cadre de la boucle inductive, la segmentation est simple à effectuer. Le signal $\mathbf{s_x}$ de la boucle inductive est représenté positivement. Une méthode consiste à comparer le signal à un seuil de détection α. Si $\mathbf{s_x} > \alpha$, alors la présence d'un véhicule est détectée et le signal est ensuite segmenté. La détermination du seuil α prend en compte le bruit du canal, pour éviter les fausses détections.

Il est à noter que Ndoye et coauteurs [36] ont développé une autre méthode de détection. En effet, d'après les auteurs, la mesure est entachée de bruit. Ils estiment que la méthode basée sur une seule valeur par rapport à un seuil pour déterminer la détection d'un véhicule n'est pas assez robuste. Dans [36], ils considèrent le signal comme un bruit blanc gaussien sans présence de véhicule. Lors de la présence du véhicule, le signal est constitué d'une information plus un bruit blanc gaussien additif. Les auteurs ont utilisé une règle de décision basée sur le test du rapport de vraisemblance. La détermination du seuil de décision est basée sur la probabilité de fausse détection qui suit une

loi de probabilité gamma inverse. En appliquant cette loi avec une fenêtre temporelle glissante sur le signal, la décision de la présence ou non du véhicule est prise. À partir de cette décision, le signal est segmenté pour obtenir la signature du véhicule.

De même Kwon [20] propose d'utiliser une autre méthode pour la segmentation du signal afin d'éviter un décalage temporel dans l'estimation des points de début et de fin de segmentation. Le début et la fin du signal sont dans ce cas estimés à partir de la dérivée première du signal. Ces différentes méthodes ne peuvent pas s'appliquer directement pour la segmentation avec un magnétomètre. En effet, le décalage d'origine de la mesure par le magnétomètre varie fortement d'un capteur à l'autre. De plus, cette valeur dérive au cours du temps en fonction de la température.

En 2004 [37] puis en 2005 [38], Cheung et coauteurs ont présenté des algorithmes pour détecter des véhicules à partir de magnétomètres. Les algorithmes développés obtiennent les taux de détection de 99 % lors des expérimentations décrites dans les rapports [37, 38]. La variation du décalage de l'origine n'est pas prise en compte dans ces algorithmes. Une des conséquences pourrait être une diminution du taux de détection au cours du temps. En 2007 [14], Cheung et coauteurs proposent une nouvelle méthode prenant en compte la variation du signal due à la température. Ils proposent ainsi la méthode : Adaptive Threshold Detection Algorithm (ATDA, algorithme de détection à seuil adaptatif).

Adaptive Threshold Detection Algorithm Cette méthode se décompose en plusieurs étapes. Pour réaliser la détection, le signal est lissé par une moyenne glissante (équation (5.1)).

$$a(k) = \begin{cases} \frac{r(k)+r(k-1)+\ldots r(1)}{k} & \text{pour } k < M \\ \frac{r(k)+r(k-1)+\ldots r(k-M+1)}{M} & \text{pour } k \geqslant M \end{cases} \tag{5.1}$$

où $a(k)$ représente les valeurs du signal lissé, $r(k)$ le signal brut et M la taille de la fenêtre glissante. Ce traitement est appliqué à chacun des signaux s_x, s_y et s_z. La variation du décalage de l'origine est négligeable sur une période courte. Par conséquent, cette variation n'intervient pas lors de la détection du passage d'un véhicule. Afin de tenir compte de la dérive à long terme, une ligne de base adaptative pour chacun des trois axes magnétiques est calculée par les équations suivantes :

$$B_i(k) = \begin{cases} B_i(k-1) \times (1-\alpha_i) + a_i(k) \times \alpha_i & \text{si } E_j,\ j \in \{1,2\} \\ B_i(k-1) & \text{autrement} \end{cases} \text{ pour } i \in \{x,y,z\} \tag{5.2}$$

où $B_i(k)$ est la ligne de base adaptative, α_i est un facteur de poids empirique, $a_i(k)$ est le signal lissé, E_j représente l'état du processus de détection et l'indice i représente l'un des trois axes magnétiques. La ligne de base adaptative est mise à jour uniquement lorsque aucun véhicule n'est en cours de détection et aucune fluctuation du signal n'est observée. $\alpha_i = 0{,}05$ est appelé le facteur d'oubli par Cheung et coauteurs.

À partir de la ligne de base, un indicateur de présence noté $T(k)$ est généré en fonction de l'équation (5.3). $h_z(k)$ et $h_x(k)$ sont les valeurs de seuils déterminées de manière empirique. Lorsqu'un véhicule effectue un arrêt–redémarrage sur le capteur, la valeur de l'axe z va décroître provoquant un passage en-dessous du seuil de $|a_z(k) - B_z(k)|$. Pour éviter un problème de non détection, les mesures de l'axe x et z sont prises en compte lorsque la détection d'un véhicule est en cours. En effet, il est peu probable d'après Cheung et coauteurs que les deux mesures soient en-dessous du seuil de détection en même temps avec la présence d'un véhicule. D'après Cheung et coauteurs, l'axe z est moins influencé par les véhicules passant sur les voies adjacentes. L'influence de l'axe z est prépondérant lors de la détection.

$$T(k) = \begin{cases} \begin{cases} \text{vrai} & \text{si } |a_z(k) - B_z(k)| > h_z(k) \\ \text{faux} & \text{autrement} \end{cases} & \text{pour } E_j(k-1),\ j \in \{3,4\} \\ \begin{cases} \text{vrai} & \text{si } |a_z(k) - B_z(k)| > h_z(k) \text{ ou } |a_x(k) - B_x(k)| > h_x(k) \\ \text{faux} & \text{autrement} \end{cases} & \text{pour } E_5(k-1) \end{cases} \tag{5.3}$$

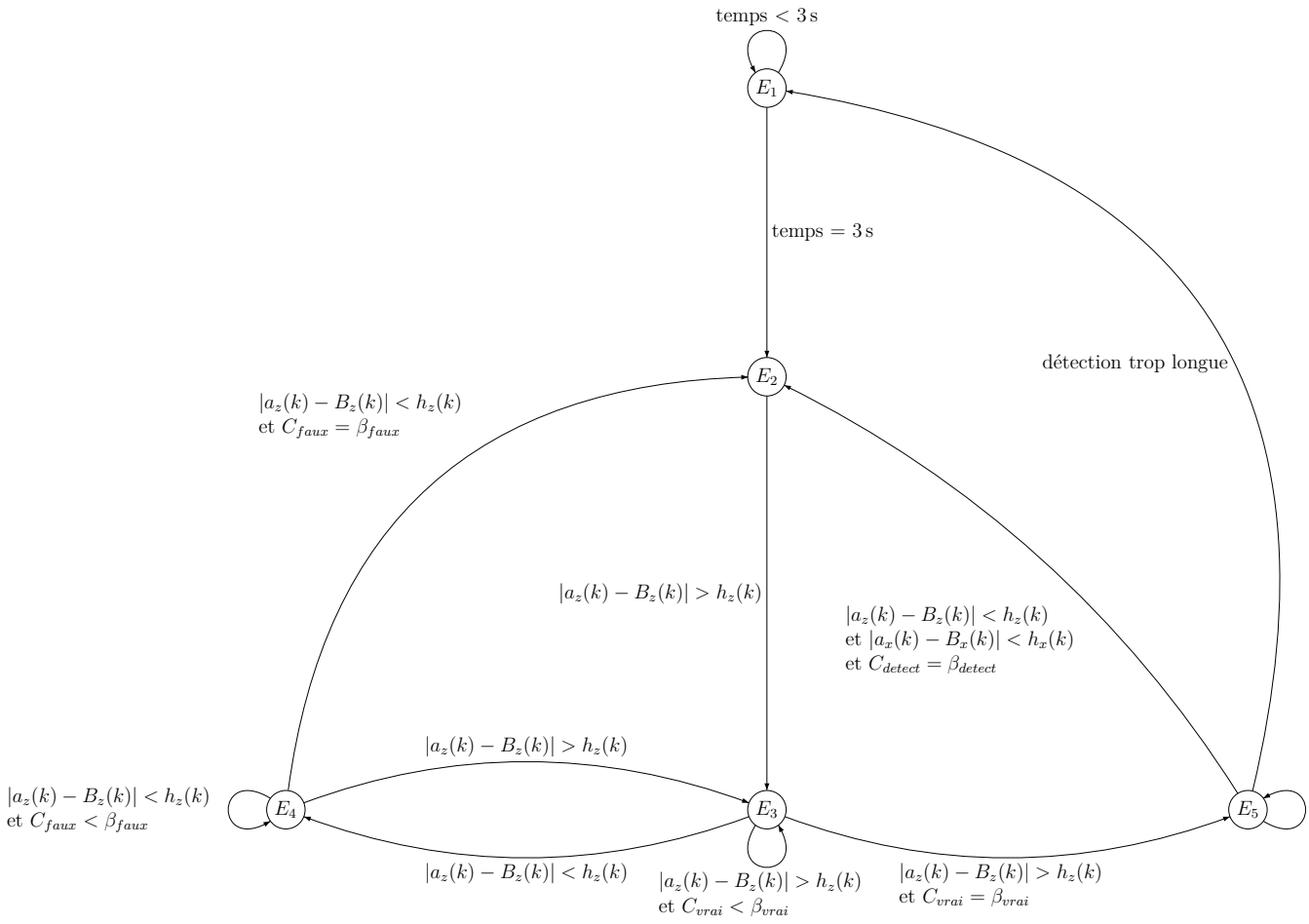


FIGURE 5.1 – Principe de l'algorithme ATDA (Adaptive Threshold Detection Algorithm) développé par Cheung et coauteurs [14].

Cet algorithme est un automate d'états et se décompose en cinq états comme le montre la figure 5.1. L'algorithme remplit deux objectifs : il filtre les signaux parasites qui ne sont pas causés par un véhicule ; il révèle le passage d'un véhicule par un indicateur afin de segmenter correctement le signal. Au démarrage du capteur, l'automate se situe dans l'état d'initialisation E_1. L'état d'initialisation consiste, à partir de la mesure, à établir la ligne de base. Cet état dure 3 s et suppose qu'aucun véhicule ne se trouve à proximité. Une fois la phase d'initialisation terminée, l'état E_2 est activé. L'algorithme met à jour la ligne de base adaptative en attendant le passage d'un véhicule. Le passage à l'état E_3 s'effectue lorsque la mesure de l'axe z est supérieure au seuil, c'est-à-dire $|a_z(k) - B_z(k)| > h_z(k)$. Lors du changement d'état, le temps est enregistré comme le début de la segmentation dans l'hypothèse où il s'agirait d'une détection valide.

L'état E_3 comporte un compteur C_{vrai}. Le passage d'un véhicule va générer une succession de valeurs T_k vraie, C_{vrai} sert à compter le nombre de T_k vrai. L'état E_3 reste actif tant que : $|a_z(k) - B_z(k)| > h_z(k)$ et $C_{vrai} < \beta_{vrai}$ avec β_{vrai} le seuil déterminant une séquence assez longue pour valider le passage d'un véhicule. Le passage à l'état E_4 s'effectue lorsque $|a_z(k) - B_z(k)| < h_z(k)$, et celui à E_5 lorsque $|a_z(k) - B_z(k)| > h_z(k)$ et $C_{vrai} = \beta_{vrai}$.

L'état E_4 possède aussi son propre compteur C_{faux}. Le compteur C_{faux} est incrémenté tant que $|a_z(k) - B_z(k)| < h_z(k)$. Si $C_{faux} = \beta_{faux}$ avec β_{faux} le seuil déterminant une fausse détection, alors l'état E_4 est désactivé au profit de l'état E_2. Le passage à l'état E_2 implique la réinitialisation à zéro des compteurs C_{faux} et C_{vrai}. Dans ce cas, les variations de $|a_z(k)|$ sont considérées comme provenant de perturbations et non du passage d'un véhicule. Si $|a_z(k) - B_z(k)| > h_z(k)$ et $C_{faux} < \beta_{faux}$, alors l'état E_3 est à nouveau activé. Dans ce cas, la détection du véhicule est toujours en cours.

L'état E_5 implique que le véhicule est toujours en état de détection par le capteur. Maintenant, la

génération de $T(k)$ est déterminée à la fois par l'axe magnétique x et z. À chaque fois qu'un nouveau $T(k)$ vrai arrive, il est enregistré comme la fin pour la segmentation du signal. Un compteur C_{detect} compte le nombre de fois successives où $|a_z(k) - B_z(k)| < h_z(k)$ et $|a_x(k) - B_x(k)| < h_x(k)$ en même temps. Lorsque $C_{detect} = \beta_{detect}$, un événement complet de détection a lieu. L'état E_5 est désactivé au profit de l'état E_2. Le passage d'un véhicule est détecté et le signal est segmenté pour extraire la signature du véhicule. Si l'état E_5 persiste durant un temps déraisonnablement long alors une situation de blocage est détectée et l'état E_1 est activé à la place de E_5.

Cheung et coauteurs ont évalué cette méthode [14] sur des expérimentations de deux heures, le taux de détection obtenu est d'environ 98 %. Elle a été développée conjointement avec la société Sensys Network qui commercialise des magnétomètres pour la gestion de trafic.

Adaptive Threshold Detection Algorithm amélioré En 2010, Kanathantip et coauteurs [39] ont proposé une amélioration de l'algorithme ATDA. Les évolutions sont représentées sur la figure 5.2 et concernent les états E_2 et E_5. Ils ont modifié la valeur du facteur de poids α_i ($\alpha_i = 0{,}02$) et des seuils ($\beta_{vrai} = 7$, $\beta_{faux} = 8$ et $\beta_{detect} = 16$). De plus au niveau de l'état E_5, différents tests ont été ajoutés pour prendre en compte des situations particulières. L'un des tests consiste à vérifier le cas d'une double détection. Si $|a_z(k) - B_z(k)| < h_z(k)$ et $|a_x(k) - B_x(k)| < h_x(k)$ et que la détection a duré moins d'une seconde, alors deux cas sont pris en compte : soit $|Max_z - Min_z| < 50$ dans ce cas la détection est ignorée, soit $|Max_z - Min_z| > 50$ alors la détection est prise en compte. Max_z représente le maximum du signal selon l'axe magnétique z, et Min_z le minimum aussi suivant z. De plus, trois autres cas sont considérés lorsque la détection du véhicule dure plus de 7,5 s : soit Moy_x (ou Moy_z) représente la moyenne du signal selon l'axe magnétique x (ou z).

- $|Moy_z - B_z(k)| < 200$ et $|Moy_x - B_x(k)| < 300$ et le temps de détection est inférieur à 120 s alors la détection est ignorée. Ce cas arrive lors d'une exposition soudaine du capteur au soleil d'après les auteurs.

- $|Moy_z - B_z(k)| > 200$ ou $|Moy_x - B_x(k)| > 300$, et le temps de détection est inférieur à 120 s alors le cas d'un arrêt–redémarrage du véhicule est considéré.

- $|Moy_z - B_z(k)| > 200$ ou $|Moy_x - B_x(k)| > 300$, et le temps de détection est supérieur à 120 s alors le véhicule est considéré comme arrêté.

Les auteurs ont ajouté aussi un état pour signaler l'enregistrement des données lors de la détection, et un état transitoire entre l'enregistrement des données et le passage à l'état E_2. Cet état permet de mettre à jour la ligne de base à partir des données dès la fin de détection, en reprenant l'historique des 16 données inférieures au seuil β_{detect}. Les auteurs ont obtenu des taux de détection de 99,33 % et 99,76 % lors de deux expérimentations comportant respectivement chacune 298 et 422 véhicules. Il est à noter que le taux de détection pour les deux-roues motorisées est de 88,89 % et de 95,45 % pour un nombre respectif de 54 et 66 deux-roues motorisés, respectivement. D'après les auteurs, cette méthode améliore la détection des petits véhicules et évite les fausses détections dues aux véhicules roulant à des vitesses faibles.

Lee et coauteurs [40, 41] se sont aussi basés sur la méthode de la ligne de base définie par l'équation (5.2) pour réaliser la détection de véhicule par magnétomètre. Ils ont utilisé seulement l'axe magnétique x pour la détection et ils ont modifié le facteur d'oubli ($\alpha_x = 0{,}1$) de l'équation (5.2). À la différence de Cheung et coauteurs qui effectuent la détection à partir de la norme $\ell 1$ ($|a_x - B_x|$), Lee et coauteurs utilisent la norme $\ell 2$ ($(a_x - B_x)^2$) ce qui a pour effet d'accentuer les différences entre les deux signaux. Les bruits impulsifs sont éliminés par un filtre passe-bas. La dernière étape consiste à lisser le signal avec une moyenne glissante suivant l'équation (5.1) avec une fenêtre de 40 points.

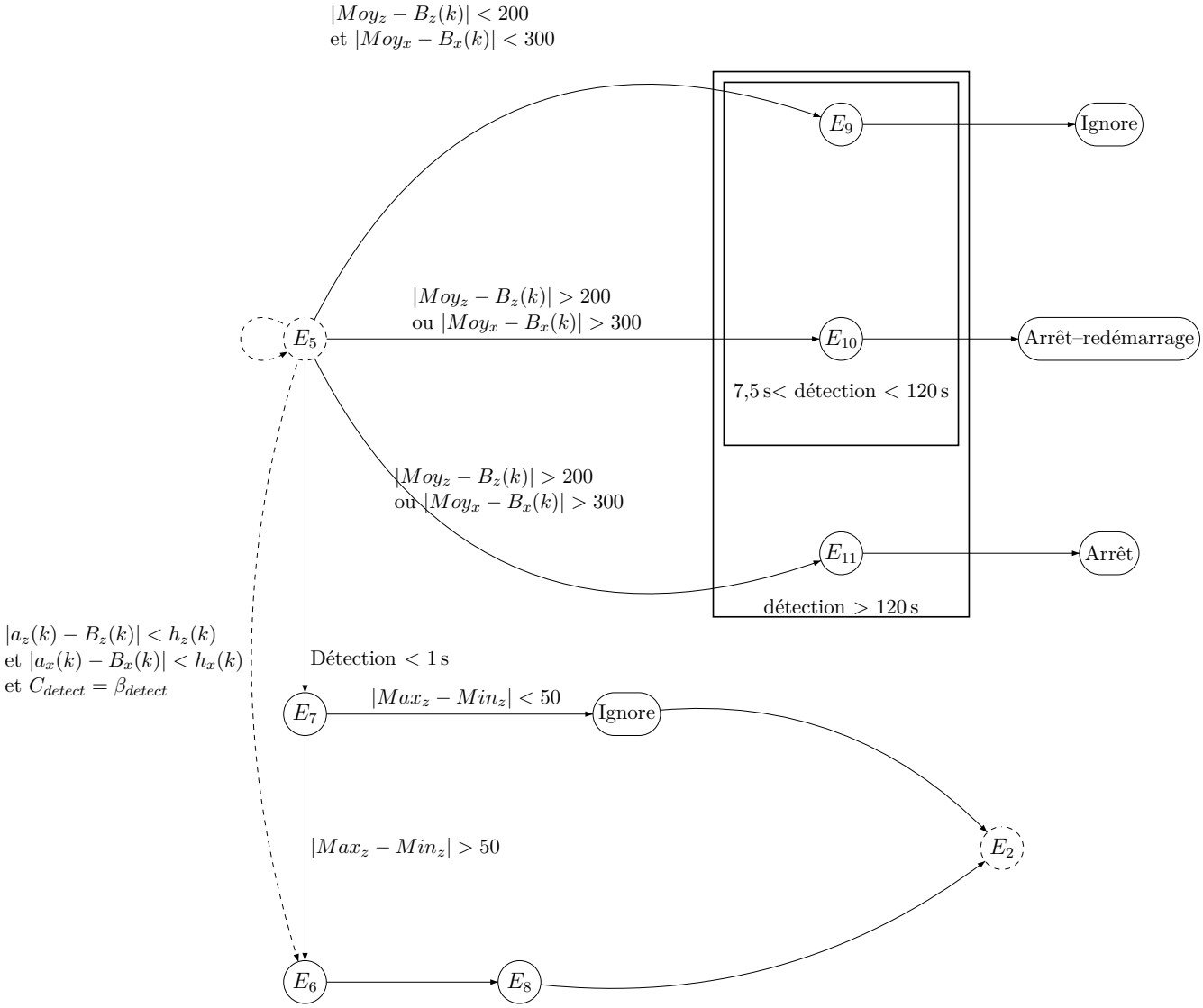


FIGURE 5.2 – Partie modifiée de l'algorithme ATDA par Kanathantip et coauteurs [39].

Algorithme de Chinrungrueng et coauteurs En 2009, Chinrungrueng et coauteurs [42] proposent un algorithme de détection de véhicules qui a l'avantage de ne pas dépendre de la ligne de base. Comme Cheung et coauteurs, ils commencent par lisser le signal par une moyenne glissante (5.1) pour obtenir le signal $a_i(k)$. Ensuite les auteurs proposent de calculer la différence d'amplitude entre les points adjacents telle que :

$$A(k) = \sqrt{(a_x(k) - a_x(k-1))^2 + (a_y(k) - a_y(k-1))^2 + (a_z(k) - a_z(k-1))^2} \tag{5.4}$$

Ce signal est à son tour lissé par une moyenne glissante (5.1) pour obtenir le signal $\bar{A}(k)$. La présence d'un véhicule est déterminée lorsque la $\bar{A}(k) \geqslant A_t$, avec A_t le seuil de détection. Afin de prévenir des fausses détections, un automate à quatre états est développé. Le principe de fonctionnement de cet automate est schématisé sur la figure 5.3.

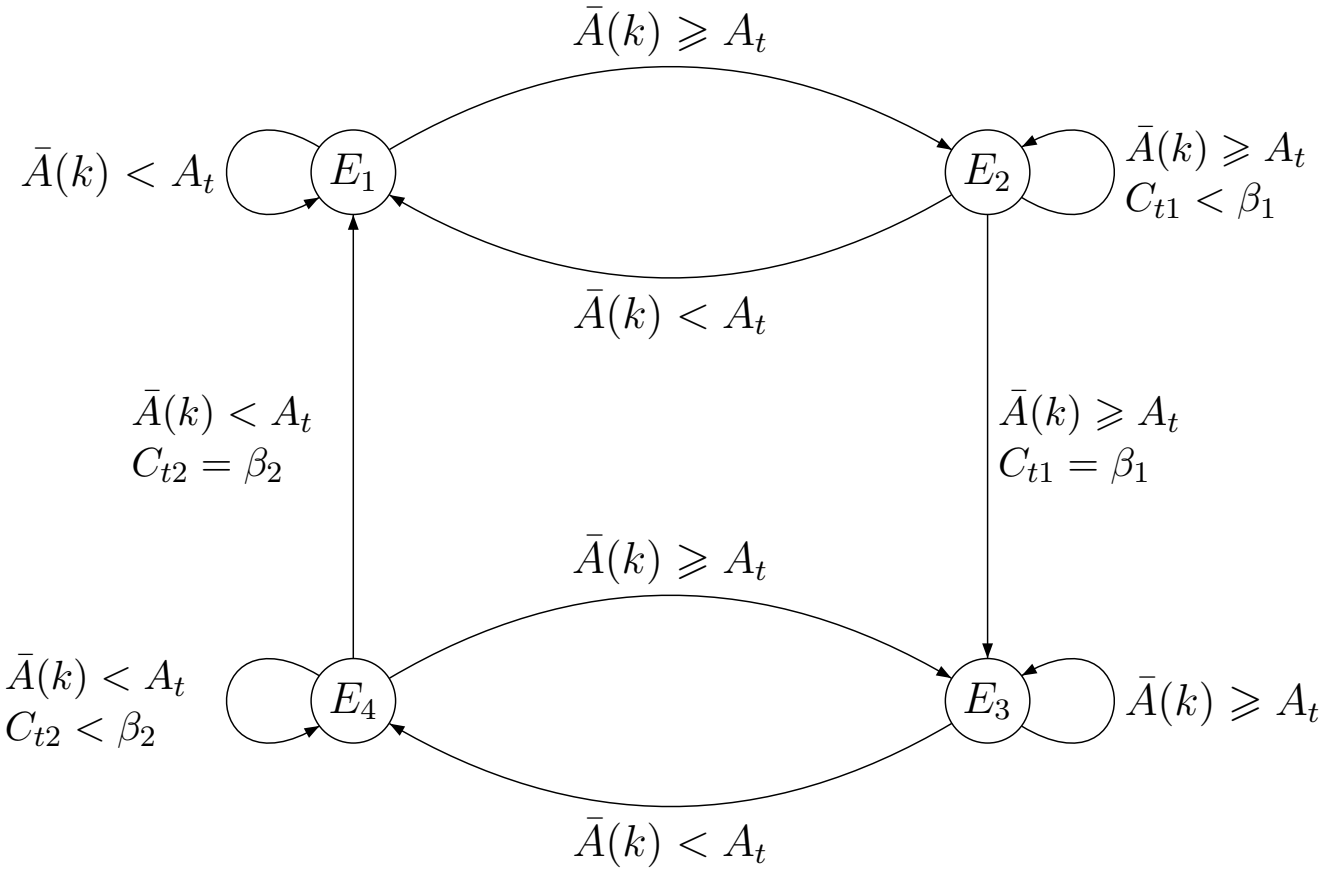


FIGURE 5.3 – Automate d'états pour la détection des véhicules proposé par Chinrungrueng et coauteurs [42].

L'état E_1 représente l'état d'attente du passage d'un véhicule. Dès que la valeur $\bar{A}(k)$ est supérieure ou égale au seuil A_t, l'état E_2 est activé. La détection du véhicule commence. Pour prévenir d'une fausse détection, si le seuil A_t n'est pas successivement dépassé durant le temps C_{t1} la détection est ignorée et l'automate passe à l'état E_1. Dans le cas contraire, la détection est validée et l'automate passe à l'état E_3. L'état E_3 correspond au passage du véhicule sur le capteur. Lorsque $\bar{A}(k) < A_t$, l'automate passe à l'état E_4. L'état E_4 détermine la fin du passage du véhicule. Le passage du véhicule est considéré comme terminé lorsque le signal $\bar{A}(k)$ est successivement inférieur au seuil A_t durant le temps C_{t2}. Cette méthode n'a pas été évaluée par les auteurs mais a néanmoins été utilisée à une des entrées de l'université de Thammasart en Thaïlande, après le poste de sécurité.

5.2 Problèmes rencontrés

La détection des véhicules par les boucles inductives s'est révélée fiable et élevée lors des différentes expérimentations que nous avons menées.

Par contre dans le cadre projet PREDIT MOCoPo (Section 8.3), les signatures issues des magnétomètres ne sont pas toutes segmentées correctement dans la base de données. La figure 5.4 montre le signal d'un même véhicule pour deux points de mesures différents. Comme attendu, la vitesse a une influence sur la longueur du signal et l'offset des capteurs n'est pas le même.

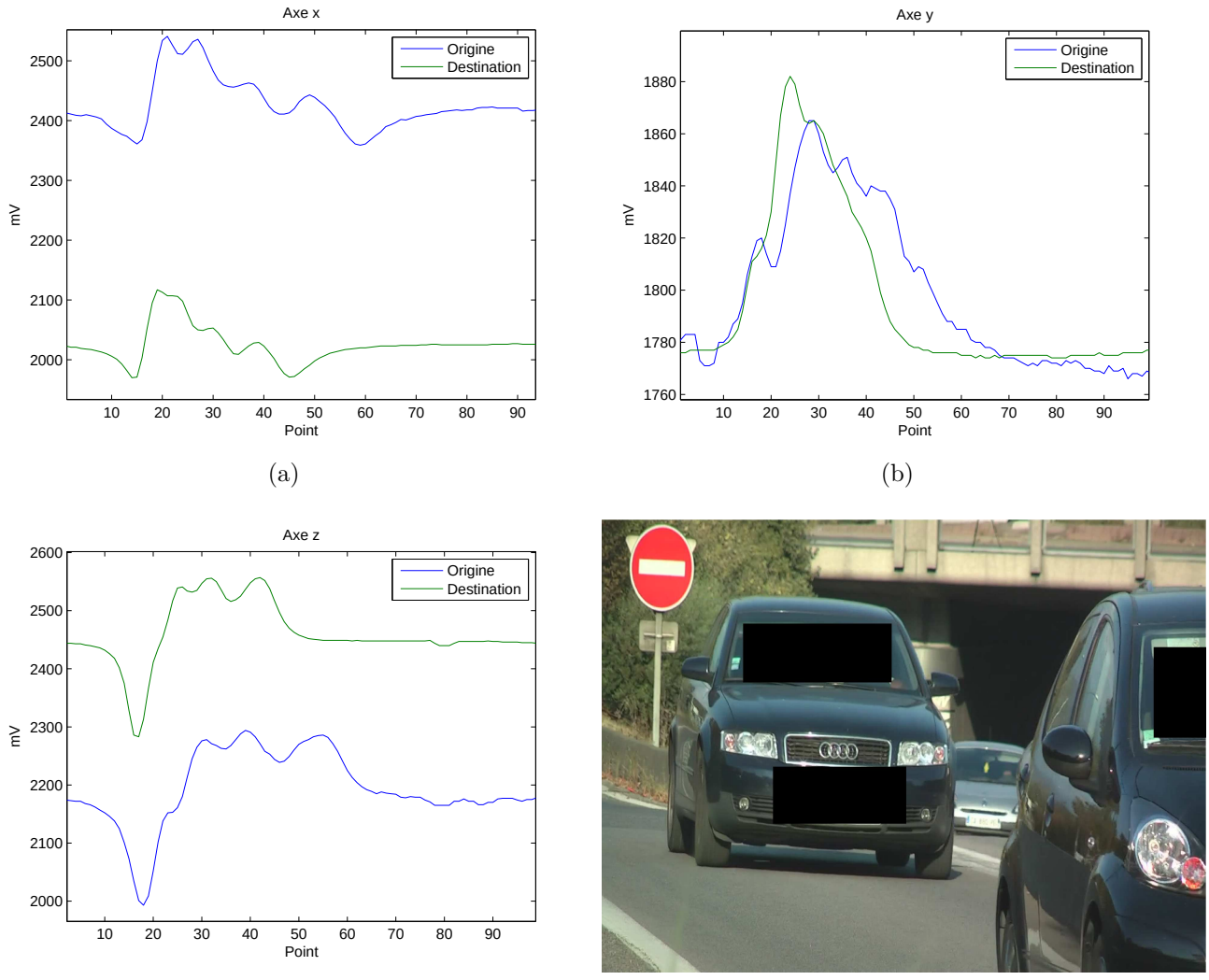
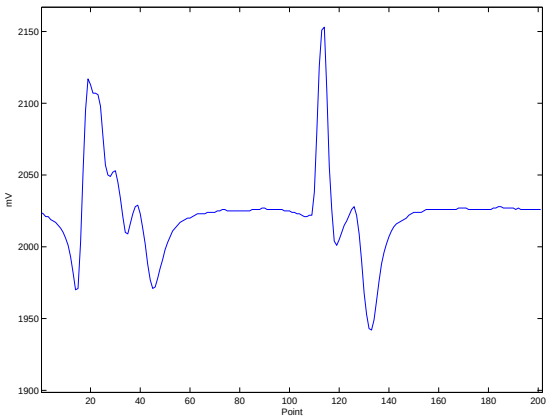


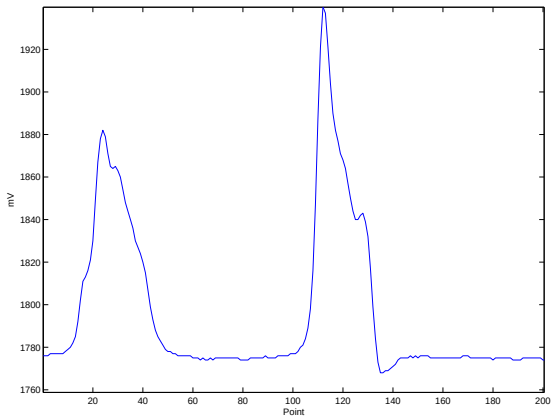
FIGURE 5.4 – Déformation d'une signature par la vitesse : (a) : axe x ; (b) : axe y ; (c) : axe z ; (d) : photo du véhicule.

La figure 5.5 montre une détection détériorée. La détection du véhicule est déjà effectuée mais celle-ci n'est pas toujours optimale. Lors de l'enregistrement, la détection n'a pas fonctionné correctement et les signatures de deux véhicules ont été considérées comme une seule signature.

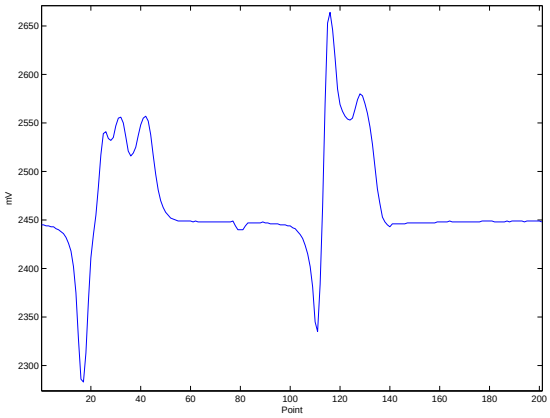
La figure 5.6 montre un autre cas de détection détériorée. En comparant la signature Origine et la signature Destination du véhicule, une différence importante apparaît. En effet le début de la signature Origine est supérieur à l'offset supposé du capteur. Le début de la signature Origine est tronqué. La segmentation des signatures n'est pas correcte, le début des signatures est parfois tronqué, la fin des signatures comporte un signal plus ou moins long sans la présence de véhicule... Ces types de données se retrouvent malheureusement dans la base de données et risquent de perturber le suivi



(a)



(b)

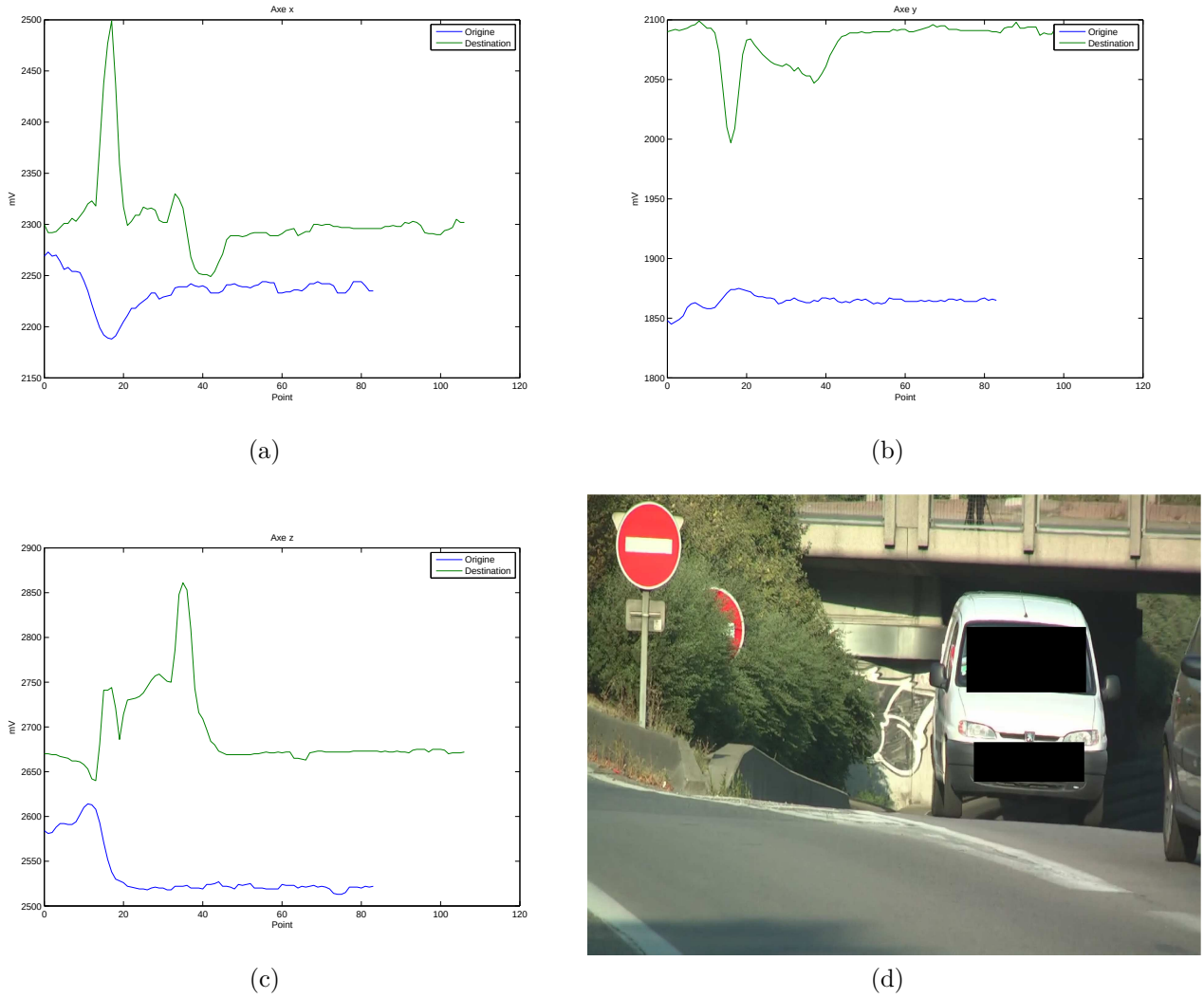


(c)

(d)

FIGURE 5.5 – Détection d'une signature au lieu de deux : (a) : axe x ; (b) : axe y ; (c) : axe z ; (d) : photo des véhicules.

de véhicules. Ainsi, une nouvelle méthode de détection est proposée pour segmenter à nouveau le signal.



(a) (b)

(c) (d)

FIGURE 5.6 – Signature tronquée : (a) : axe x ; (b) : axe y ; (c) : axe z ; (d) : photo du véhicule.

5.3 Proposition d'un algorithme de détection

L'utilisation des algorithmes [14, 39–41] présentés à la section précédente 5.1 ne peut pas s'appliquer pour les signaux acquis lors des expérimentations de MOCoPo (Section 8.3). En effet, ces algorithmes déterminent la ligne de base, et cette ligne de base est mise à jour en fonction des points précédents. Une plus grande importance est donnée aux points précédents en effectuant une moyenne pondérée. Ensuite, la ligne de base est utilisée pour effectuer la détection des véhicules suivant l'équation (5.3). Dans les expérimentations menées, le nombre de points au début du signal est insuffisant pour calculer la ligne de base. La phase d'initialisation pour calculer la ligne de base sans la présence de véhicules dure 3 s pour [14]. Dans nos expérimentations, la détection réalisée fait que cette durée est inférieure à trois secondes et est même parfois nulle.

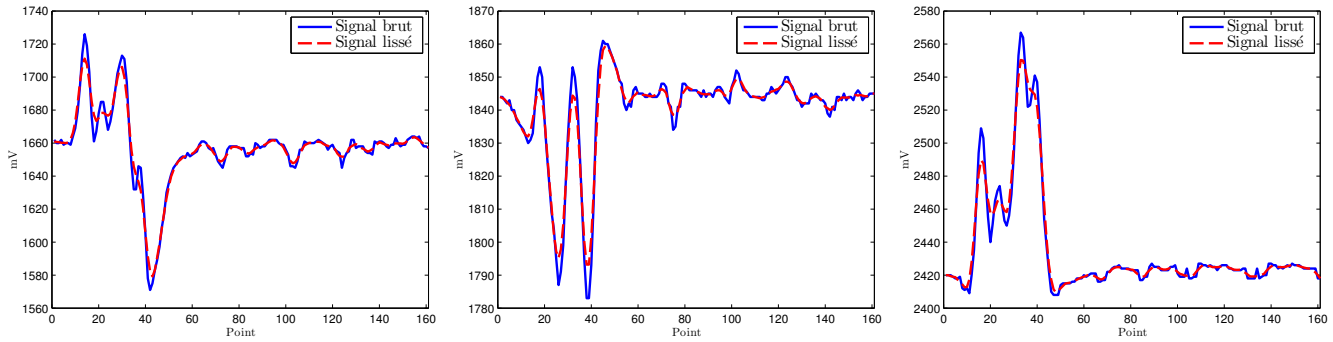
Par contre, la méthode proposée par Chinrungrueng et coauteurs (présentée précédemment à la page 60) n'utilise pas la ligne de base pour la détection. Cette méthode est reprise et évaluée. Le site d'expérimentation utilisé par Chinrungrueng et coauteurs n'a *a priori* pas de voies adjacentes. Dans notre situation, des voies adjacentes sont présentes et les changements de voies sont fréquents ce qui

risque de complexifier la détection. Nous présentons plusieurs modifications de l'algorithme proposé par Chinrungrueng et coauteurs, et nous évaluons ces modifications vis-à-vis de l'algorithme initial.

Comme proposé par Chinrungrueng et coauteurs, le signal est lissé comme suit :

$$a_i(k) = \frac{1}{M} \sum_{j=-m}^{m} r_i(k+j) \tag{5.5}$$

avec M la largeur de la fenêtre temporelle, $m = \lfloor \frac{M}{2} + 1 \rfloor$, $a_i(k)$ le point moyenné, $i \in [x,y,z]$. La fenêtre utilisée a une largeur $M = 3$ dans notre cas. La figure 5.7 montre le signal brut et le signal une fois lissé.



(a) (b) (c)

FIGURE 5.7 – Signal brut et lissé d'un véhicule (ordonnée en mV, abscisse en nombre de points, fréquence d'échantillonnage de 128 Hz) : (a) : selon l'axe x ; (b) : selon l'axe y ; (c) : selon l'axe z.

Une modification est apportée entre l'état E_1 et E_2 de l'automate d'états de Chinrungrueng et coauteurs. Nous ajoutons l'état E_{2bis}, cet état joue le même rôle que l'état E_4 de l'automate d'états de Cheung et coauteurs. Dans le cas précédent, dès que $\bar{A}(k)$ était inférieur au seuil A_t, nous revenions à l'état initial E_1. Si le passage de la valeur $\bar{A}(k)$ en-dessous du seuil était dû à un bruit, le début de la signature du véhicule était perdu. L'état E_{2bis} est introduit pour palier ce problème. L'état E_{2bis} localise les passages en-dessous du seuil et les considère comme du bruit si ceux-ci sont courts. En effet, lorsque l'automate d'états est dans l'état E_2, si $\bar{A}(k)$ est inférieur au seuil A_t, l'automate d'états active l'état E_{2bis} et non plus l'état initial. Un nouveau compteur C_{t1bis} est mis en place pour l'état E_{2bis}. Si ce compteur reste inférieur à la valeur β_{1bis} et que $\bar{A}(k)$ est à nouveau supérieur au seuil A_t, l'état E_2 est activé. L'activation de l'état initial E_1 se fait lorsque $C_{t1bis} \geqslant \beta_{1bis}$ et $\bar{A}(k) < A_t$.

Pour éviter les fausses détections, une condition supplémentaire est à valider entre l'état E_4 et l'état E_1. Le maximum de la signature détectée doit être supérieur au seuil β_3 sinon la détection n'est pas prise en compte, c'est-à-dire que la signature du véhicule n'est pas comptée ni enregistrée.

$$\begin{cases} \max \bar{A}(\text{début} : \text{fin}) \geqslant \beta_3 \Rightarrow \text{Véhicule détecté} \\ \max \bar{A}(\text{début} : \text{fin}) < \beta_3 \Rightarrow \text{Véhicule non détecté} \end{cases} \tag{5.6}$$

Le seuil β_3 est fixé après avoir analysé les fausses détections sur une base de données. Dans les deux cas (véhicule détecté et véhicule non détecté), la machine d'état passe de l'état E_4 à l'état E_1 avec la validation des conditions $\bar{A}(k) < A_t$ et $C_{t2} \geqslant \beta_2$.

Les modifications apportées à l'automate d'états sont représentées en trait plein sur la figure 5.8, les traits pointillés correspondent aux parties inchangées de l'automate d'états.

D'après Cheung et coauteurs [14], l'axe magnétique y subit trop l'influence des voies adjacentes pour permettre une segmentation du signal correcte. Une augmentation du nombre de fausses détections ainsi qu'une segmentation incorrecte au début et à la fin du signal risquent de se produire sous l'influence de cet axe selon Cheung et coauteurs. Chinrungrueng et coauteurs utilisent les trois

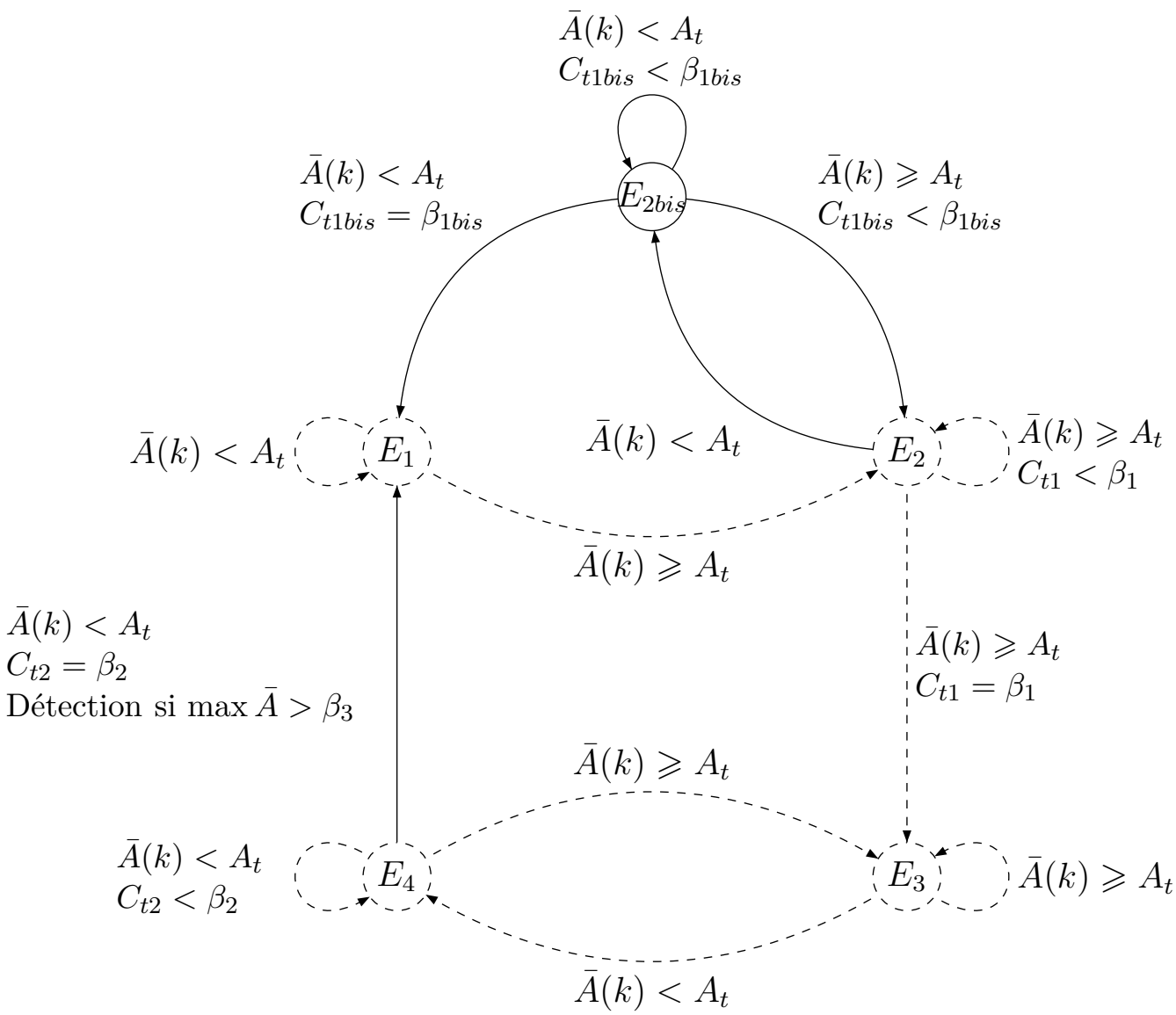


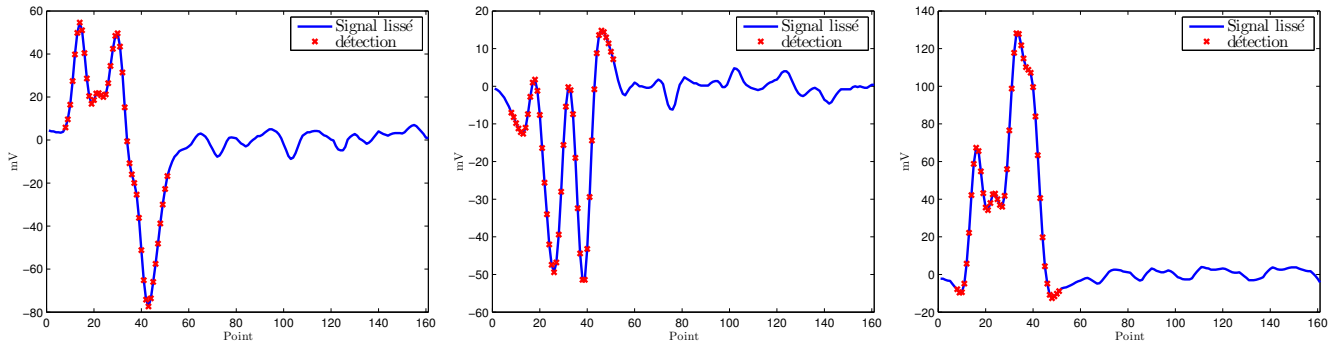
FIGURE 5.8 – Automate d'états proposé pour la détection des véhicules.

signaux selon l'équation (5.4). Pour s'affranchir de l'influence de l'axe magnétique y, celui-ci n'est pas pris en compte et l'équation (5.4) devient :

$$A(k) = \sqrt{(a_x(k) - a_x(k-1))^2 + (a_z(k) - a_z(k-1))^2} \tag{5.7}$$

Un lissage par la moyenne glissante (5.5) est effectué avec $M = 3$ pour calculer le signal $\bar{A}(k)$.

L'automate d'états proposé par Chinrungrueng et coauteurs (figure 5.3) est utilisé et évalué. Cet automate est nommé par la suite $Detect_{init}$. Il sera comparé à l'automate d'états modifié (nommé $Detect_{modif}$) et à l'automate d'états modifié ne prenant pas en compte l'axe magnétique y (nommé $Detect_{modifxz}$). Une fois la détection accomplie et la signature extraite du signal par l'un des automates d'états, nous proposons de calculer la ligne de base. Pour cela les points qui n'ont pas été retenus pour la signature sont considérés comme appartenant à la ligne de base. La moyenne de ces points est soustraite au signal $a_i(k)$. Ainsi, le calcul de la ligne de base permet de centrer les signatures des différents magnétomètres au niveau du zéro. La figure 5.9 représente la signature d'un véhicule détectée par un automate d'états et centrée sur zéro.



(a) (b) (c)

FIGURE 5.9 – Signal et signature extraite d'un véhicule (ordonnée en mV, abscisse en nombre de points, fréquence d'échantillonnage de 128 Hz) : (a) : selon l'axe x ; (b) : selon l'axe y ; (c) : selon l'axe z.

5.4 Évaluation des méthodes de détection

5.4.1 Évaluation

Les algorithmes proposés précédemment dans la section 5.3 pour la détection ont été utilisés sur la première session de mesures effectuées sur l'échangeur du Rondeau. Cette expérimentation et la configuration du site sont décrites à la section 8.3. Les taux de détection et le nombre de fausses détections des trois magnétomètres à la sortie de l'échangeur du Rondeau sont évalués et comparés à ceux des trois boucles inductives. Les taux de détection et le nombre de fausses détections des trois autres magnétomètres à l'entrée de l'échangeur du Rondeau sont évalués et comparés aux magnétomètres commercialisés par la société Sensys. Les tableaux 5.1 et 5.2 montrent les résultats de la détection par les différents capteurs ainsi que les résultats obtenus par les algorithmes proposés. Les nombres de détections réelles obtenues par l'analyse des vidéos servent de référence (Réf.). Pour la sortie de l'échangeur du Rondeau, les résultats de la détection sont établis sur la première demi-heure alors que pour l'entrée de l'échangeur du Rondeau les résultats sont basés sur l'ensemble de la durée de l'expérimentation. Le nombre de détections au niveau de l'entrée et de la sortie de l'échangeur est par conséquent différent. L'entrée de l'échangeur du Rondeau comporte trois voies appelées : Voie d'Insertion (VI), Voie Lente (VL) et Voie Rapide (VR). La sortie de l'échangeur

du Rondeau se compose de trois voies dénommées : Droite (D), Milieu (M) et Gauche (G). Les paramètres utilisés pour l'algorithme $Detect_{Init}$ (proposé par Chinrungrueng et coauteurs [42] et présenté précédemment à la section 5.1 page 60) ont été fixés de manière empirique en fonction de la sortie droite de l'échangeur du Rondeau. Ainsi la valeur A_t du seuil de détection est fixée à 2,5. Le seuil β_1 qui représente le nombre minimum de points pour que la détection puisse être valide est fixé à 10. Le seuil β_2 qui détermine le nombre de points inférieurs à A_t pour terminer la détection d'un véhicule est défini à 12. Pour l'algorithme $Detect_{Modif}$, ces paramètres sont conservés, β_{1bis} et β_3 sont égaux à 3 et 4,5, respectivement. Nous avons retenu exactement les mêmes paramètres pour l'algorithme $Detect_{Modifxz}$.

Nombre (%)			Signature			
	Vidéo	Sensys	Par défaut	$Detect_{Init}$	$Detect_{Modif}$	$Detect_{Modifxz}$
VI	429	386 (89,98)	405 (94,41)	415 (96,74)	416 (96,97)	415 (96,74)
VL	1594	1408 (88,33)	1337 (83,88)	1494 (93,73)	1499 (94,04)	1494 (93,73)
VR	878	582 (66,29)	693 (78,93)	803 (91,46)	801 (91,23)	800 (91,12)
Total	2901	2376 (81,90)	2435 (83,94)	2712 (93,49)	2716 (93,62)	2709 (93,38)

Tableau 5.1 – Détection en entrée de l'échangeur du Rondeau.

Nombre (%)			Signature			
	Vidéo	Boucle inductive	Par défaut	$Detect_{Init}$	$Detect_{Modif}$	$Detect_{Modifxz}$
D	745	722 (96,91)	608 (81,61)	706 (94,77)	704 (94,50)	703 (94,36)
M	679	664 (97,79)	587 (86,45)	649 (95,58)	650 (95,73)	649 (95,58)
G	329	319 (96,96)	299 (90,88)	308 (93,62)	309 (93,92)	308 (93,62)
Total	1753	1705 (97,26)	1494 (85,23)	1663 (94,87)	1663 (94,87)	1660 (94,69)

Tableau 5.2 – Détection en sortie de l'échangeur du Rondeau.

Les algorithmes proposés présentent des taux de détection similaires pour une même voie. Que se soit pour l'entrée (tableau 5.1) ou la sortie (tableau 5.2) du Rondeau, les algorithmes proposés permettent une amélioration de la détection des véhicules d'environ 10 points. Dans le cadre de l'entrée du Rondeau, le capteur « signature » par défaut a un taux de détection moindre que le capteur industriel Sensys pour la VL. Avec les différents algorithmes proposés, le taux de détection est devenu supérieur à celui du capteur Sensys. Pour les différentes voies en entrée du Rondeau, le taux de détection varie fortement de 66 % à 90 % pour le capteur Sensys, soit une variation d'environ 14 points. Le capteur signature avec la méthode de détection par défaut a une variation de son taux de détection d'environ 16 points. Cette variation diminue avec les algorithmes proposés. Le taux de détection oscille entre 91 % et 97 % et représente une amplitude de 6 points.

Au niveau de la sortie, le taux de détection se rapproche de celui de la boucle inductive en évoluant de 12 points d'écart à 3 points d'écart environ. La variation du taux de détection entre les différentes voies pour la boucle inductive est faible avec moins de 2 points d'écart. Cet écart est d'environ 10 points pour la détection par défaut. Les algorithmes proposés permettent de réduire cette amplitude à 3 points.

Les algorithmes proposés améliorent la détection des véhicules par rapport à la détection par défaut. De plus, la variabilité du taux de détection entre les différentes voies a fortement diminué avec les nouveaux algorithmes. Nous nous intéressons maintenant aux fausses détections, en effet l'augmentation du taux de détection ne doit pas se faire aux dépens des fausses détections. Les fausses détections sont identifiées lorsqu'une détection est produite alors qu'aucun véhicule n'est présent sur la vidéo de référence.

Nombre		Signature		
	Sensys	$Detect_{Init}$	$Detect_{Modif}$	$Detect_{Modifxz}$
VI	7	4	2	4
VL	2	23	10	15
VR	4	31	14	11
Total	13	58	26	30

Tableau 5.3 – Fausses détections en entrée de l'échangeur du Rondeau.

Nombre (%)		Signature		
	Boucle inductive	$Detect_{Init}$	$Detect_{Modif}$	$Detect_{Modifxz}$
D	3	7	0	0
M	0	10	0	2
G	0	1	0	1
Total	3	18	0	3

Tableau 5.4 – Fausses détections en sortie de l'échangeur du Rondeau.

Le tableau 5.3 indique que pour l'entrée de l'échangeur du Rondeau, le nombre de fausses détections par le capteur Sensys est faible (13). Sur ce même tableau, nous constatons pour l'algorithme $Detect_{Init}$ une forte augmentation des fausses détections, puisque celles-ci passent de 13 à 58. La modification des algorithmes de détection permet de réduire le nombre de fausses détections par deux, mais ce nombre reste supérieur à celui des capteurs Sensys. Les fausses détections, pour les algorithmes $Detect_{Modif}$ et $Detect_{Modifxz}$, reste acceptable. Le tableau 5.4 montre que pour la sortie de l'échangeur du Rondeau, le nombre de fausses détections pour la boucle inductive est très faible avec seulement trois fausses détections. Comme pour l'entrée de l'échangeur du Rondeau, les fausses détections avec l'algorithme $Detect_{Init}$ augmentent. L'intérêt de modifier cet algorithme et de proposer les algorithmes $Detect_{Modif}$ et $Detect_{Modifxz}$ se justifie en constatant que l'algorithme $Detect_{Modifxz}$ fait aussi bien que les boucles inductives quant aux fausses détections et que l'algorithme $Detect_{Modif}$ fait mieux en ne commettant aucune fausse détection.

Les taux de détection avec les algorithmes proposés ($Detect_{Init}$, $Detect_{Modif}$ et $Detect_{Modifxz}$) au niveau de la sortie de l'échangeur du Rondeau sont supérieurs d'environ un point à ceux de l'entrée de l'échangeur du Rondeau. De même, les nombres de fausses détections sont inférieurs pour la sortie de l'échangeur du Rondeau. Les meilleurs résultats pour la sortie de l'échangeur du Rondeau s'expliquent en partie par la topologie des lieux. La description et la topologie des lieux sont données à la section 8.3. Comme le montrent les figures 8.12 et 8.13a, chacune des voies à la sortie de l'échangeur a une direction spécifique et le marquage au sol au niveau de la position des capteurs interdit les changements de voies. Les véhicules sont, en général, centrés sur leur voie. Au niveau de l'entrée de l'échangeur du Rondeau (figure 8.12), des changements de voies se produisent au niveau des capteurs. Les usagers cherchent à se positionner vis-à-vis de la direction qu'ils souhaitent prendre. Nous avons pris trois exemples de situations complexes pour la détection au niveau de l'entrée de l'échangeur du Rondeau. La figure 5.10 montre un véhicule réalisant un changement de voie entre la VR et la VL au niveau de l'installation des magnétomètres dans le cas d'une circulation relativement fluide. Le véhicule désigné par la flèche se situe sur la séparation des deux voies. Cette situation rend la détection difficile car : soit le véhicule est détecté deux fois (une fois sur la VR et une fois sur la VL), soit le véhicule n'est détecté par aucun des capteurs, soit il est détecté une seule fois par l'un des magnétomètres.

La figure 5.11 représente une situation de ralentissement brusque par un ensemble de véhicules. Les véhicules qui freinent sont facilement identifiables sur la photographie par leurs feux stop allumés. Un véhicule effectue même un changement de voie pendant cette phase de « freinage ». Dans ce cas,

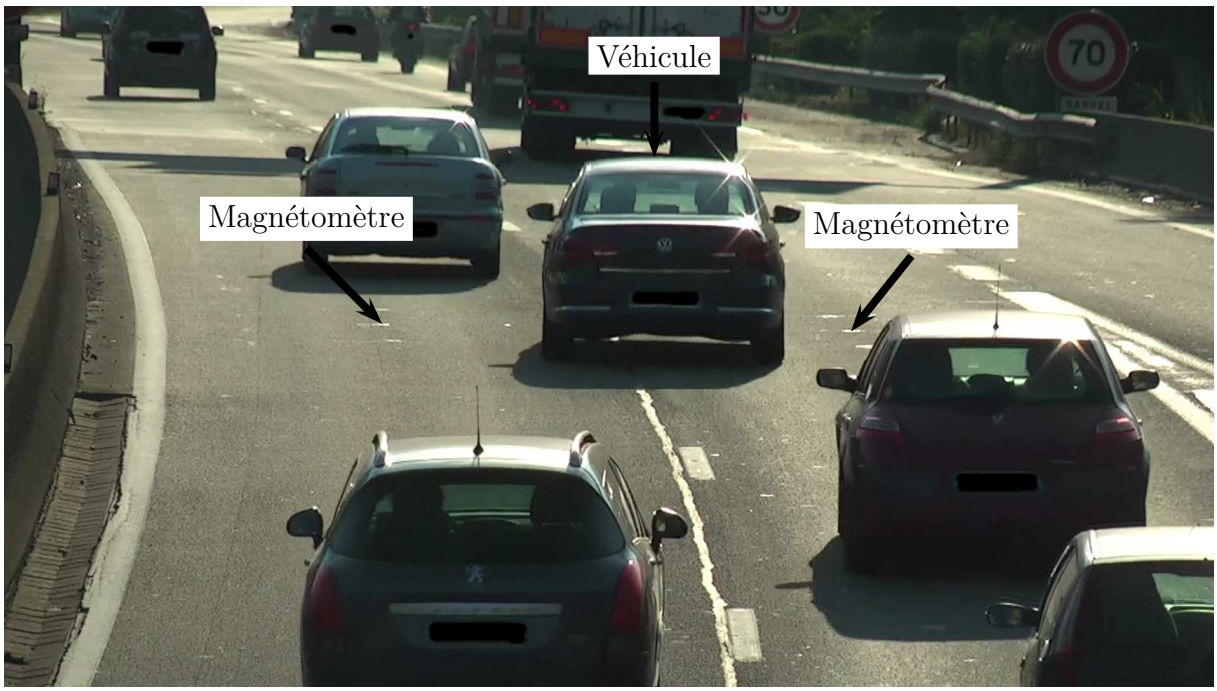


FIGURE 5.10 – Véhicule réalisant un changement de voie entre la VR et la VL. L'emplacement des magnétomètres des voies lente et rapide est indiqué par des flèches.

la segmentation des signatures se complique. Le véhicule de la VR qui change de voie pour la VL peut provoquer un allongement de la signature du véhicule de la voie lente.

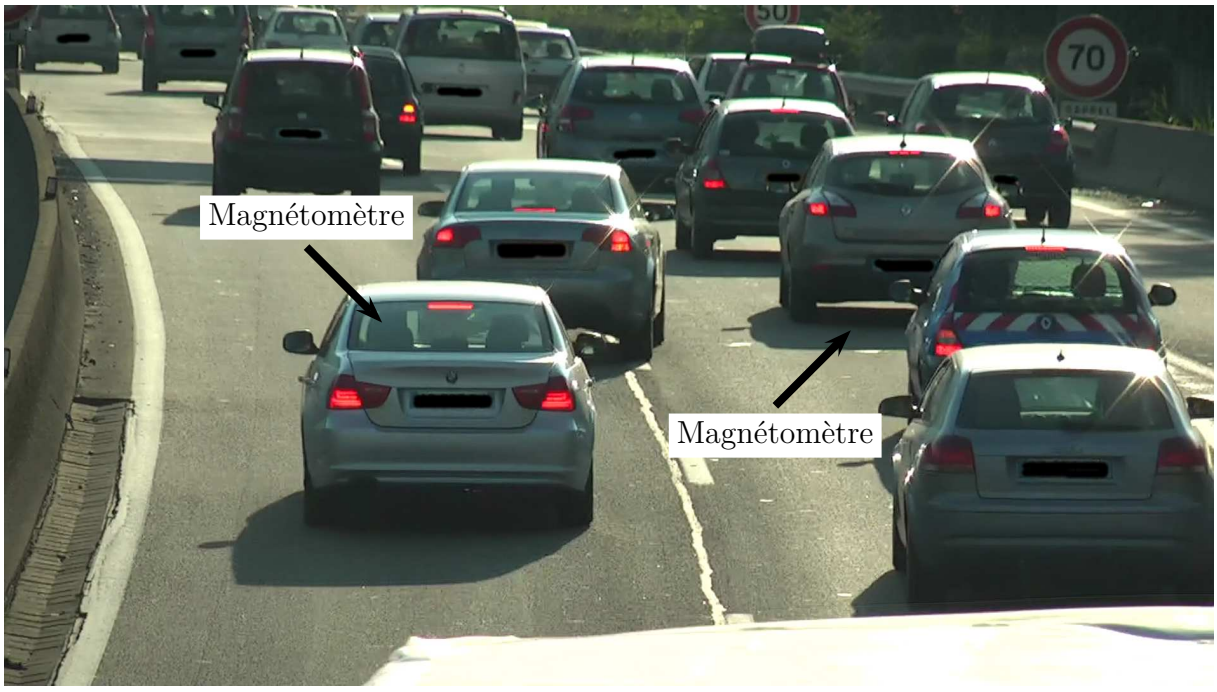


FIGURE 5.11 – Situation de freinage brusque sur les capteurs. L'emplacement des magnétomètres des voies lente et rapide est indiqué par des flèches.

La figure 5.12 correspond à trois véhicules de front sur la VR et la VL. Cette situation implique obligatoirement la non-détection du véhicule 2.

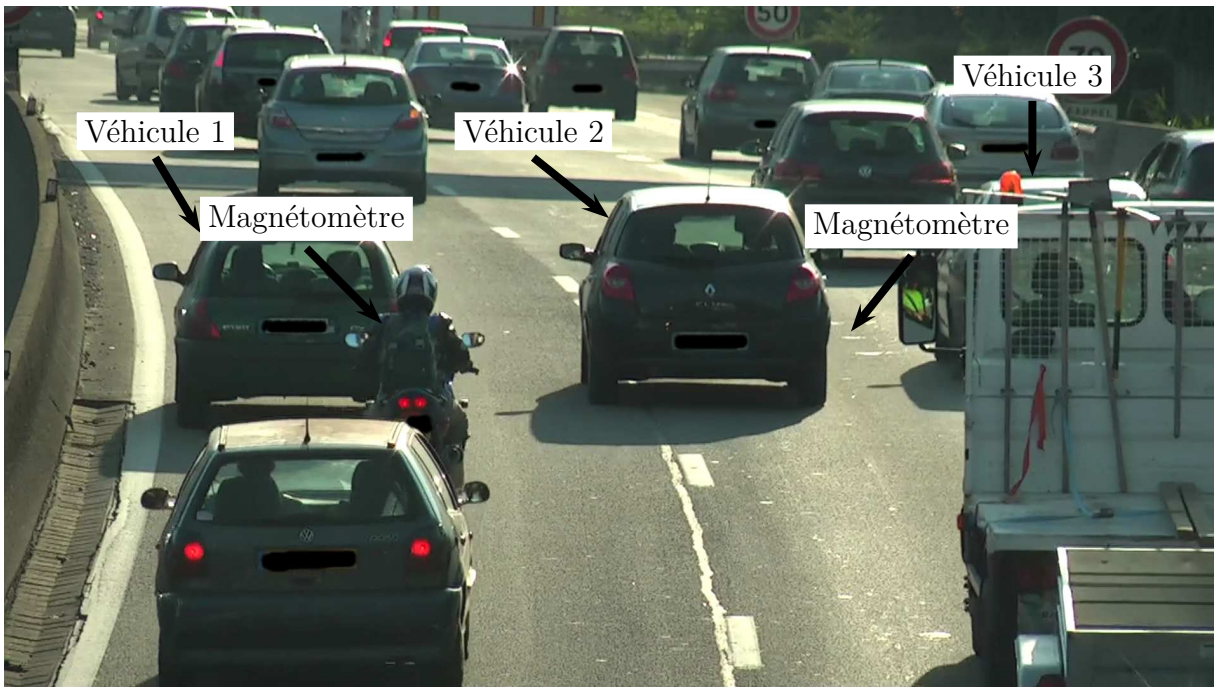


FIGURE 5.12 – Trois véhicules de front sur deux voies. L'emplacement des magnétomètres de la VL et VR est indiqué par des flèches.

5.4.2 Conclusion

L'évaluation de l'algorithme proposé ne peut pas être complète car le taux de détection dépend de la première détection réalisée par défaut. Les taux de détection des algorithmes proposés seraient peut être améliorés si les signatures manquantes pouvaient être analysées. En revanche, il est à noter que la première détection réalisée par défaut, évite peut-être aux algorithmes proposés de faire de fausses détections. Les différentes situations rencontrées sur le site de l'échangeur du Rondeau sont souvent complexes, nous pouvons donc considérer que les algorithmes proposés pour la détection obtiennent de bons taux de détection et de faibles taux de fausses détections. Néanmoins, de fortes déformations des signatures vont être engendrées par : les positions décentrées des véhicules par rapport à l'axe des voies ; les ralentissements des véhicules au niveau des capteurs ; le passage de plusieurs véhicules en simultané sur la même voie au niveau du magnétomètre... Ceci risque de rendre le suivi de véhicules difficile à réaliser. L'algorithme $Detect_{Modif}$ a obtenu les meilleurs résultats au niveau de la détection et des fausses détections, il est donc utilisé pour la segmentation des signatures dans la suite du manuscrit. Les signatures segmentées sont utilisées pour le suivi de véhicules.

5.5 Conclusion

La détection des véhicules par une boucle inductive est fiable et élevée. Nous avons constaté très peu de variation au niveau du taux de détection entre différentes boucles inductives au cours de nos expérimentations. Le taux de détection se situe à environ 96 %. La segmentation du signal est réalisée correctement.

Les algorithmes de détection des véhicules pour les magnétomètres sont plus complexes à réaliser. Ces algorithmes cherchent à être robuste en prenant en compte la variation de la température de la chaussée. Nous avons constaté lors de nos expérimentations dans le cadre du projet MOCoPo une variation des taux de détection des magnétomètres, ces taux de détection sont plus faibles que ceux des boucles inductives. L'un des principaux algorithmes de détection (ATDA) est conçu en prenant en compte une ligne de base. Cette ligne de base s'ajuste automatiquement en fonction des variations de

la température. La détection d'un véhicule est effectuée en fonction de la ligne de base. Cependant dans nos expérimentations, le signal est déjà segmenté nous empêchant d'utiliser cet algorithme pour la détection. Un nouvel algorithme de détection basé sur l'algorithme de Chinrungrueng et coauteurs est développé. Cet algorithme a permis d'améliorer significativement le taux de détection des magnétomètres.

6

Prétraitements

Sommaire

6.1 Différents prétraitements **74**
6.1.1 Normalisation 74
6.1.2 Réduction du nombre de données 74
6.1.3 Déconvolution dans le cadre de la boucle inductive 76
6.2 Propositions de sélection de données **81**
6.2.1 Informations recueillies & signatures 81
6.2.2 Étude réalisée 83
6.2.3 Population de VL 85
6.2.4 Population de PL 85
6.2.5 Population de VL–PL 85
6.2.6 Conclusion 85
6.3 Propositions d'algorithmes de déconvolution **86**
6.3.1 Contexte 86
6.3.2 Définition du problème 86
6.3.3 Algorithmes utilisés 89
6.3.4 Estimation de la signature « réelle » 94
6.3.5 Conclusion 96
6.4 Conclusion **96**

La boucle inductive est l'un des capteurs de trafic prédominants au niveau mondial. De ce fait, de nombreuses références sur la boucle inductive existent dans la littérature. Même si le magnétomètre existe depuis de nombreuses années (d'après [16] cette technologie a été introduite dans les années 60), il n'a été adopté par des gestionnaires de trafic que récemment. Les références concernant le magnétomètre comme capteur de trafic apparaissent plus tardivement et sont donc moins nombreuses. Que ce soit pour la boucle inductive ou le magnétomètre, des prétraitements sont effectués dans l'objectif d'améliorer les TBR (cf. la définition et l'équation (4.3) à la page 53). Pour rappel, le prétraitement se positionne après la phase de détection et s'attache à préparer les données pour la phase de réidentification.

La proposition de sélection de données est effectuée dans le cadre de la boucle inductive. Cette sélection peut facilement être adaptée pour le magnétomètre. La proposition de déconvolution est réalisée pour la boucle inductive, par contre une étude plus approfondie est nécessaire si nous souhaitons étendre cette méthode pour le magnétomètre.

6.1 Différents prétraitements

6.1.1 Normalisation

Lors de l'acquisition de la signature celle-ci peut être déformée pour diverses raisons. Le véhicule n'est pas passé dans l'axe du capteur, la vitesse du véhicule à l'origine et à la destination n'est pas la même... Afin de compenser les déformations, deux normalisations sont possibles, une au niveau de l'amplitude du signal et l'autre au niveau de la base temporelle.

Plusieurs références [20, 33, 34, 36, 43–62] évoquent la normalisation en amplitude pour le signal issu de la boucle inductive. Il en est de même pour les magnétomètres [14] . Les signatures sont normalisées en amplitude par rapport à leur maximum d'une part pour éliminer les variations de sensibilité entre les différents capteurs d'autre part pour compenser le déport des véhicules par rapport à l'axe central de la voie.

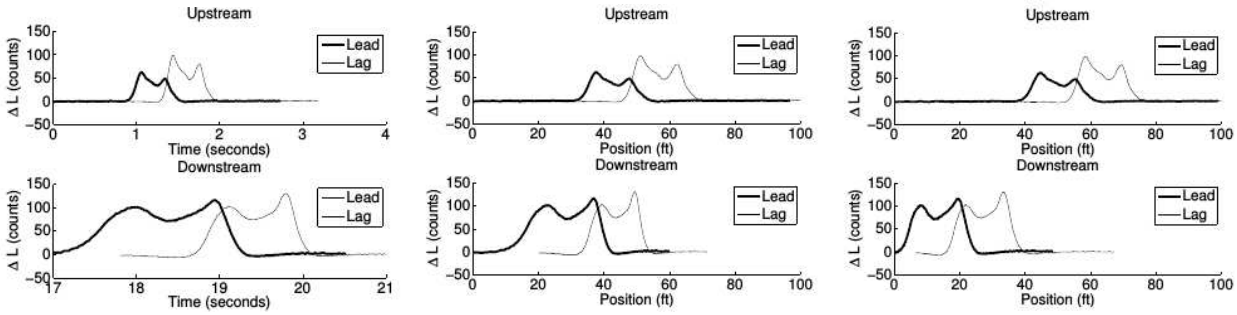
Pour la normalisation temporelle, plusieurs approches sont possibles. Elle est aussi bien utilisée pour les signatures issues de boucles inductives que de magnétomètres. La première approche consiste à utiliser la mesure de la vitesse individuelle du véhicule pour convertir la signature dans le domaine spatial. La seconde approche ne nécessite pas la connaissance de la vitesse du véhicule, et elle réside dans l'interpolation de la signature en un nombre de points prédéfini. L'interpolation permet aussi de limiter le nombre d'informations à transmettre et à rendre la taille de l'échantillon constant. Une dernière solution consiste à mixer les deux méthodes, c'est-à-dire de convertir le signal dans le domaine spatial puis ensuite d'interpoler la signature en un nombre de points prédéfini.

Ernst et coauteurs [62] ont analysé une méthode de réidentification avec différentes normalisations. En reprenant les signatures issues de paires de boucles inductives, ils ont comparé trois cas de réidentification : sans compensation du signal (figure 6.1a), en compensant la vitesse (figure 6.1b), en compensant l'accélération (figure 6.1c). La compensation en vitesse est similaire à la première approche citée précédemment, la mesure de vitesse étant effectuée par la paire de boucles inductives. La compensation en accélération des véhicules vise à comparer la première et la seconde signature d'une paire de boucles inductives. Lors d'une accélération la première signature est plus longue (support temporel plus important) que la deuxième. En émettant l'hypothèse que l'accélération est constante durant le passage du véhicule, il est alors possible de compenser cette déformation du signal.

Ernst et coauteurs ont comparé les taux de bonne réidentification en utilisant le coefficient de corrélation pour chacune des trois normalisations. Lors de leurs expérimentations, ils ont obtenu les taux de bonne réidentification suivants : 38,5 % sans compensation, 56,5 % avec une compensation en vitesse et 68,3 % avec une compensation en accélération. Une amélioration du taux de bonne réidentification est constatée en apportant une compensation du signal.

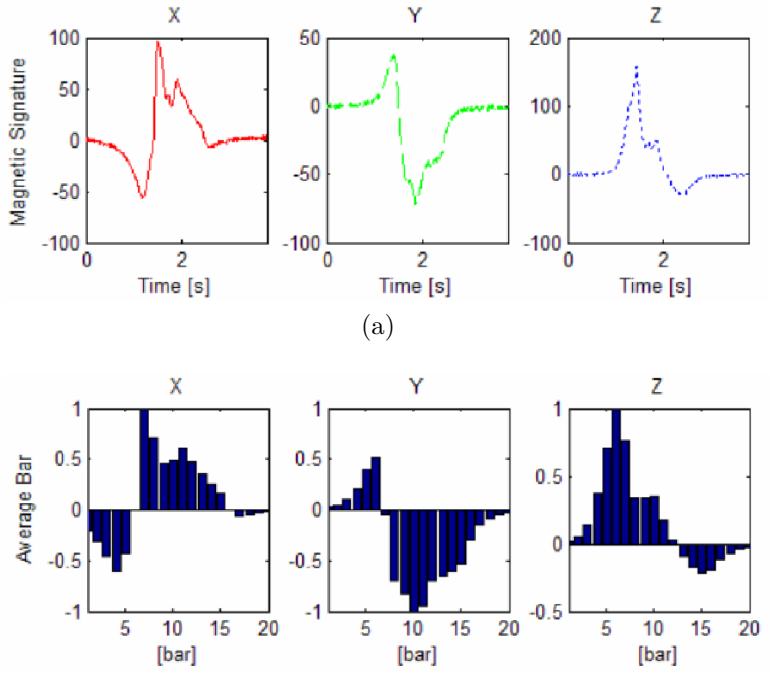
6.1.2 Réduction du nombre de données

Plusieurs approches ont été utilisées pour réduire le nombre de données en conservant le maximum d'informations possibles. Une sélection des données a été réalisée par Ritchie et coauteurs [50] : le maximum avant normalisation, le degré de symétrie, la longueur, le nombre de points supérieurs à 0,5, la somme des points supérieurs à 0,5. En 2007, Jeng [54] a proposé d'interpoler le signal par une spline cubique en 61 points. À partir de ces 61 points, la dérivée première du signal est calculée tous les deux points. Les valeurs des 30 pentes sont utilisées ensuite pour la réidentification.



(a) (b) (c)

FIGURE 6.1 – Les signaux des paires de boucles inductives (extrait de [62]) : (a) : sans compensation ; (b) : avec compensation en vitesse (c) : avec compensation en vitesse et accélération. Lead et Lag représentent les signatures de la première et de la seconde boucle inductive de la paire de boucle inductive, respectivement. Downstream et upstream sont les paires de boucle inductive de la destination et de l'origine, respectivement.



(a)

(b)

FIGURE 6.2 – Compression de la signature en 20 sections moyennées (extrait de [14]) : (a) : signal lissé ; (b) : signal moyenné en 20 points.

Dans le cadre des magnétomètres, les recherches menées par Cheung et coauteurs se sont portées sur des capteurs sans fil, ceci a orienté les recherches vers des méthodes peu coûteuses en calcul pour préserver la batterie et limiter les transmissions de données. De manière à lisser et compresser le signal, les auteurs de [14, 17] divisent la signature en 20 segments. L'amplitude moyenne est calculée pour chaque segment. La figure 6.2a représente la signature des trois axes et la figure 6.2b la compression correspondante en 20 points.

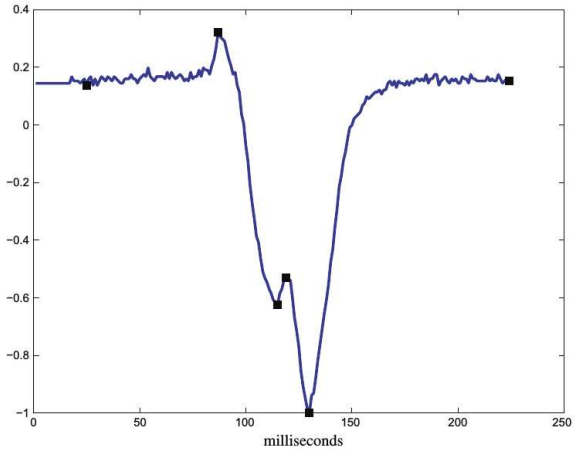


FIGURE 6.3 – Compression de la variation magnétique selon l'axe z ; les carrés représentent les extremums extraits (extrait de [63]).

Kwong et coauteurs [63, 64] proposent, pour chacun des axes x, y et z du magnétomètre, de conserver le premier et le dernier point du signal et d'extraire les extremums entre ces deux points comme représentés sur la figure 6.3.

Les différentes études développées par Ieng et coauteurs [59, 60], ont démontré l'intérêt de compresser les données pour la transmission de celles-ci et pour réduire les temps de calcul afin de réaliser des algorithmes de réidentification en temps réel. Ils ont d'une part repris des caractéristiques de la signature précédemment décrites dans l'algorithme lexicographique évoqué dans [33], et d'autre part ils ont aussi calculé d'autres données telles que la transformée de Fourier et des données statistiques comme la variance, la moyenne, le skewness, le kurtosis...

Sur cet ensemble de données, une analyse en composantes principales a été réalisée. Les données étant hétérogènes, elles ont été normalisées (moyenne nulle et écart-type égal à un). Les données sont stockées sous la forme d'une matrice $\mathbf{X}$ de taille $n \times p$. Chaque ligne représente les valeurs prises par l'individu i sur les p données. Chaque colonne représente les valeurs de la donnée j pour les n individus. La matrice $\mathbf{M}$ de corrélation est calculée comme suit : $\mathbf{M} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$. Ensuite les valeurs et vecteurs propres associés à $\mathbf{M}$ sont calculés. Les valeurs propres λ_j représentent la quantité d'informations, avec le taux d'informations égal à $\frac{\lambda_j}{\sum_{i=1}^{p}\lambda_i}$. Les composantes principales sont une combinaison linéaire des données d'origine. Les composantes principales sont deux à deux non corrélées, de variance maximale et d'importance décroissante. Les données retenues sont celles présentant une forte corrélation avec les deux premiers axes principaux et étant faiblement corrélées entre elles. Finalement, les auteurs de [59] ont retenu treize données.

6.1.3 Déconvolution dans le cadre de la boucle inductive

D'après Cheung et coauteurs [14], le magnétomètre mesure directement des variations locales et fournit plus d'informations que la boucle inductive de par ce fait. Kwon a démontré dans [20, 65] qu'il est possible d'obtenir une mesure de la variation locale du champ magnétique à partir de la boucle inductive. Pour cela, il considère que la surface de la boucle inductive a une influence sur le signal mesuré. Le signal serait lissé. Il fait l'hypothèse dans [65] que le signal mesuré est le résultat

d'une moyenne glissante de l'inductance dont la fenêtre temporelle varie selon la surface de la boucle inductive. Kwon propose de supprimer l'effet de la moyenne glissante en utilisant la déconvolution avant d'effectuer la réidentification. Ce processus vise à restaurer les informations d'inductance perdues dues à la moyenne glissante. Il vise aussi à accentuer l'unicité de chaque signature. Ce processus a pour but d'améliorer le taux de réidentification. La déconvolution du signal de la boucle inductive permet de se rapprocher d'une mesure de variation locale du champ magnétique.

Mathématiquement, le passage d'un véhicule au-dessus de la boucle inductive peut être modélisé par le produit de convolution de deux fonctions compactes comme suit :

$$s(t) = h(t) * e(t) \tag{6.1}$$

La première fonction $e(t)$ caractérise le véhicule en fonction du temps, elle est considérée comme le signal d'entrée. La seconde fonction $h(t)$ représente le capteur en fonction du temps, et plus précisément dans notre cas la boucle inductive, elle est appelée la fonction de transfert. Le signal observé en fonction du temps, noté $s(t)$, n'est pas le signal « réel » du véhicule mais le résultat de la convolution des deux signaux $e(t)$ et $h(t)$.

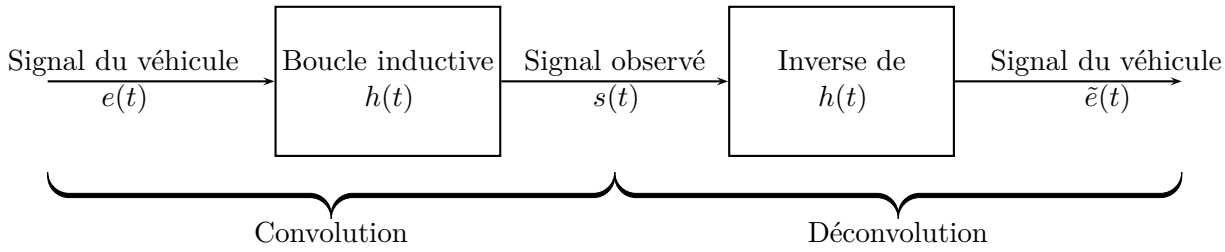


FIGURE 6.4 – Principe de la convolution et de la déconvolution.

Si la fonction de transfert $h(t)$ de la boucle inductive était une impulsion de Dirac alors le signal observé $s(t)$ et le signal du véhicule $e(t)$ seraient identiques. Cependant, la surface de la boucle inductive n'est pas ponctuelle et le support, au sens mathématique de la fonction de transfert $h(t)$ ne l'est donc pas. $h(t)$ est différente d'une impulsion de Dirac. Pour cette raison, la fonction de transfert $h(t)$ a donc l'effet d'une moyenne glissante pondérée sur le signal d'entrée $e(t)$ masquant ainsi des détails du signal. Le problème est donc d'estimer le signal d'entrée $e(t)$, tout en sachant que la fonction de transfert $h(t)$ est inconnue comme le montre la figure 6.4.

Une méthode pour évaluer la fonction de transfert $h(t)$ est de mesurer la réponse impulsionnelle. Cependant, dans le cas de la boucle inductive, il n'est pas évident de créer une impulsion en entrée et ainsi de déterminer cette fonction de transfert $h(t)$.

Cette constatation motive l'utilisation d'algorithme de déconvolution pour extraire les caractéristiques cachées du signal. Quand un véhicule est détecté par une boucle inductive, le problème est difficile car la seule information disponible est le signal $s(t)$. Kwon a proposé deux algorithmes de déconvolution pour résoudre ce problème : une approche par les moindres carrés et l'algorithme de déconvolution aveugle de Godard [20, 65].

Approche par les moindres carrés

Kwon dans [20, 65] propose de modéliser le système de la boucle inductive comme suit :

$$\mathbf{s} = \mathbf{He} + \mathbf{n} \tag{6.2}$$

avec $\mathbf{s}$, $\mathbf{e}$ et $\mathbf{n}$ des vecteurs de dimension $m \times 1$ qui désignent le signal observé, la vraie signature du véhicule et le bruit aléatoire gaussien, respectivement. m représente le nombre de points de la

signature. **H** est une matrice circulante de dimension $m \times m$ (matrice carrée dont le passage d'une ligne à la suivante s'effectue par la permutation circulaire des éléments cf. [66]) et s'écrit comme suit :

$$\mathbf{H} = \begin{bmatrix} h_0 & h_{m-1} & h_{m-2} & \cdots & h_1 \\ h_1 & h_0 & h_{m-1} & \cdots & h_2 \\ h_2 & h_1 & h_0 & \cdots & h_3 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ h_{m-1} & h_{m-2} & h_{m-3} & \cdots & h_0 \end{bmatrix} \tag{6.3}$$

Pour réaliser la déconvolution, une estimation précise de la réponse impulsionnelle est nécessaire. Kwon fait l'hypothèse que le flux magnétique produit par la boucle inductive est croissant dans un premier temps. Puis dans un second temps, une fois sa valeur maximale atteinte, le flux magnétique est constant dans la partie centrale. Enfin, le flux magnétique est décroissant à l'autre extrémité de la boucle inductive. Nous pouvons considérer que le flux magnétique ressemble à une fonction trapèze. La durée pendant laquelle les courants de Foucault sont induits dans le véhicule par la boucle inductive, est fonction du temps nécessaire au véhicule pour passer au-dessus de la boucle inductive. Cette durée dépend de la vitesse et de la longueur du véhicule. Ainsi, la longueur de la réponse de la boucle inductive doit être modélisée pour chaque véhicule en fonction de sa vitesse et de la durée de passage au-dessus de la boucle. La durée de la signature donne le temps nécessaire pour parcourir la longueur de la boucle inductive (l'un des formats de la boucle inductive américaine fait 6 pieds $\simeq$ 1,8288 m). À partir d'une évaluation de la longueur du véhicule, Kwon propose l'équation suivante pour estimer la vitesse :

$$V = \frac{(6 + l) \times 0{,}6818}{\Delta t} \tag{6.4}$$

avec V la vitesse estimée en miles par heure, l la longueur estimée du véhicule, Δt la durée de la signature. Le terme 0,6818 est utilisé pour convertir les pieds en miles et les secondes en heure. L'estimation de la longueur du véhicule proposée par Kwon se base sur le nombre de pics dans la signature et est décrite dans le tableau 6.1.

Nombre de pics	Type de véhicule	Longueur (en pieds)
1	Voiture de tourisme	21
2	Camionnette	23
3	Camion de livraison	30
4	Camion	40
5	Semi-remorque	58
$\geqslant 6$	Camion avec remorque	66

Tableau 6.1 – Estimation de la longueur des véhicules en fonction du nombre de pics dans la signature (extrait de [20]).

Le temps nécessaire au véhicule pour traverser la boucle inductive est calculé à partir de la longueur et de la vitesse estimée. Cette durée est d'autant plus courte que la vitesse du véhicule est élevée. Bien que cette durée dépende de la vitesse du véhicule, la forme de la fonction représentant **h** reste la même. Kwon propose de modéliser **h** par une somme de gaussiennes en se basant sur le fait que le signal observé est toujours lisse. La fonction doit être symétrique et elle est modélisée par :

$$\mathbf{h} = \sum_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} e^{\frac{-(\mathbf{t}-\mu_i)^2}{2\sigma_i^2}} \tag{6.5}$$

avec μ_i la moyenne et σ_i l'écart-type de la i$^{\text{e}}$ fonction gaussienne. **t** représente les valeurs temporelles et N le nombre de fonctions gaussiennes. Ces valeurs varient de manière à correspondre aux caractéristiques physiques de la boucle inductive. Cependant Kwon ne précise pas comment estimer les paramètres μ_i, σ_i et N.

En assumant maintenant que **H** est connu, l'objectif est de trouver **e** en utilisant **s** et **H**. L'approche des moindres carrés avec régularisation permet d'obtenir la fonction de coût suivante :

$$J(\mathbf{e}) = \frac{1}{2}(\| \mathbf{s} - \mathbf{He} \|^2) + \frac{1}{2}\alpha \| \mathbf{Qe} \|^2 \tag{6.6}$$

avec α le paramètre de régularisation et **Q** un opérateur linéaire agissant comme un stabilisateur. La signature « réelle » est obtenue en minimisant la fonction de coût, c'est-à-dire en calculant la dérivée première :

$$\frac{\partial J(\mathbf{e})}{\partial \mathbf{e}} = 0 \tag{6.7}$$

L'estimation $\tilde{\mathbf{e}}$ est donnée par :

$$\tilde{\mathbf{e}} = \left(\mathbf{H}^T\mathbf{H} + \alpha\mathbf{Q}^*\mathbf{Q}\right)^{-1}\mathbf{H}^T\mathbf{s} \tag{6.8}$$

avec $\mathbf{Q}^*$ le complexe conjugué transposé de **Q**. Kwon définit cette approche comme un filtre des moindres carrés avec contrainte. Nous préférons faire référence, en présence de l'équation (6.6), à une régularisation de Tikhonov. Kwon précise simplement que l'opérateur linéaire **Q** est généralement considéré comme un filtre passe-haut. Les calculs sont effectués dans le domaine fréquentiel en utilisant la transformée de Fourier.

Algorithme de déconvolution aveugle

La déconvolution aveugle consiste à déterminer un filtre modélisant au mieux le comportement d'un processus inconnu avec comme seule connaissance le signal de sortie pour estimer le signal d'entrée. Contrairement à la méthode précédente, cet algorithme ne nécessite pas d'hypothèse sur la fonction de transfert. Le principe de fonctionnement de la déconvolution aveugle est représenté par le schéma 6.5. Cet algorithme est constitué d'un filtre impulsionnel et d'une fonction non linéaire. La fonction non linéaire permet la génération d'un signal d'erreur traduisant l'erreur du filtre.

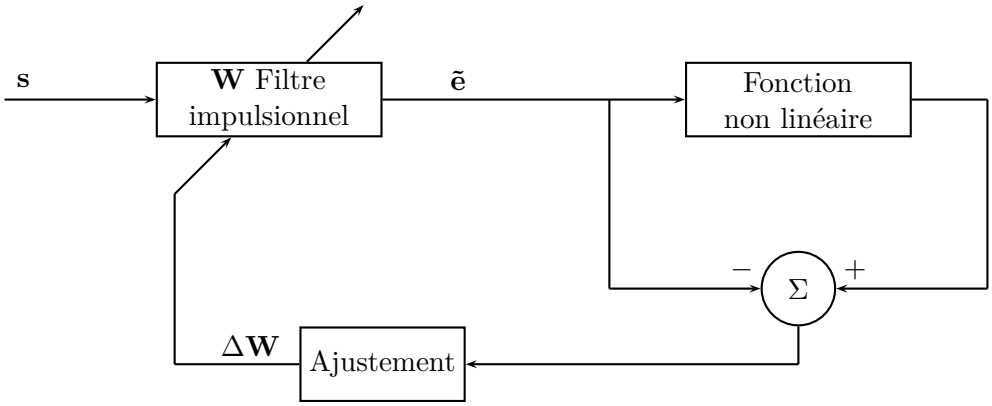


FIGURE 6.5 – Déconvolution de Godard.

Kwon dans [20, 65] a choisi d'utiliser un modèle développé par Godard pour réaliser la déconvolution aveugle. D'après la figure 6.5, $\tilde{\mathbf{e}}$ est calculé comme la convolution de la signature observée **s** avec un filtre inverse caractérisé par **W** ayant m paramètres, à savoir,

$$\tilde{\mathbf{e}} = \mathbf{Ws} \tag{6.9}$$

avec $\tilde{\mathbf{e}}$ et **s** deux vecteurs de dimension m et **W** une matrice circulante de dimension $m \times m$. La fonction de non-linéarité $\gamma(.)$ définie par Godard dans l'article [67] est :

$$\gamma_p(\mathbf{e}) = \frac{E\left\{|\mathbf{e}|^{2p}\right\}}{E\left\{|\mathbf{e}|^{p}\right\}} \tag{6.10}$$

où $E\{\cdot\}$ est l'espérance mathématique, p un entier positif et $|\cdot|$ le module. L'ajustement de ΔW consiste selon l'article [67] à minimiser le critère :

$$J_p = E\left[\left(|\tilde{\mathbf{e}}|^p - \gamma_p(\mathbf{e})\right)^2\right] \tag{6.11}$$

En utilisant l'algorithme du gradient stochastique, la problématique résulte en l'équation de mise à jour suivante :

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu \nabla_{\mathbf{W}} J_p \tag{6.12}$$

avec μ un réel positif qui représente le pas d'adaptation. $\mathbf{W}$ est constitué de $\mathbf{w}_i$ avec $i = 0, \ldots, m-1$. Le gradient pour $\mathbf{w}_i$ est donné par l'équation suivante :

$$\nabla_{\mathbf{w}_i} J_p = 2p(|\tilde{e}_i|^p - \gamma_p(\mathbf{e}))|\tilde{e}_i|^{p-1}\mathbf{s}^* \tag{6.13}$$

avec $\mathbf{s}^*$ le complexe conjugué transposé de $\mathbf{s}$.

Kwon a choisi $p = 2$ qui est un cas particulier de l'algorithme de Godard appelé algorithme à module constant (CMA : Constant Modulus Algorithm). Ainsi l'équation (6.13) s'écrit :

$$\nabla_{\mathbf{w}_i} J_2 = (|\tilde{e}_i|^2 - \gamma_2(\mathbf{e}))|\tilde{e}_i|\mathbf{s}^* = \Delta_i \mathbf{s}^* \tag{6.14}$$

et $\gamma_2(\mathbf{e})$ est égal à 1. L'équation de mise à jour devient :

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu[\Delta_1, \ldots, \Delta_m]\mathbf{s}^* \tag{6.15}$$

Le processus de déconvolution ne nécessite aucune connaissance sur la fonction de transfert $\mathbf{h}$. La figure 6.5 illustre ce processus de déconvolution aveugle de Godard. La déconvolution aveugle est alors obtenue par des étapes itératives de réglage des coefficients du filtre inverse $\mathbf{W}$. Ces coefficients sont obtenus à partir de la différence entre la fonction d'erreur et l'estimation de la signature.

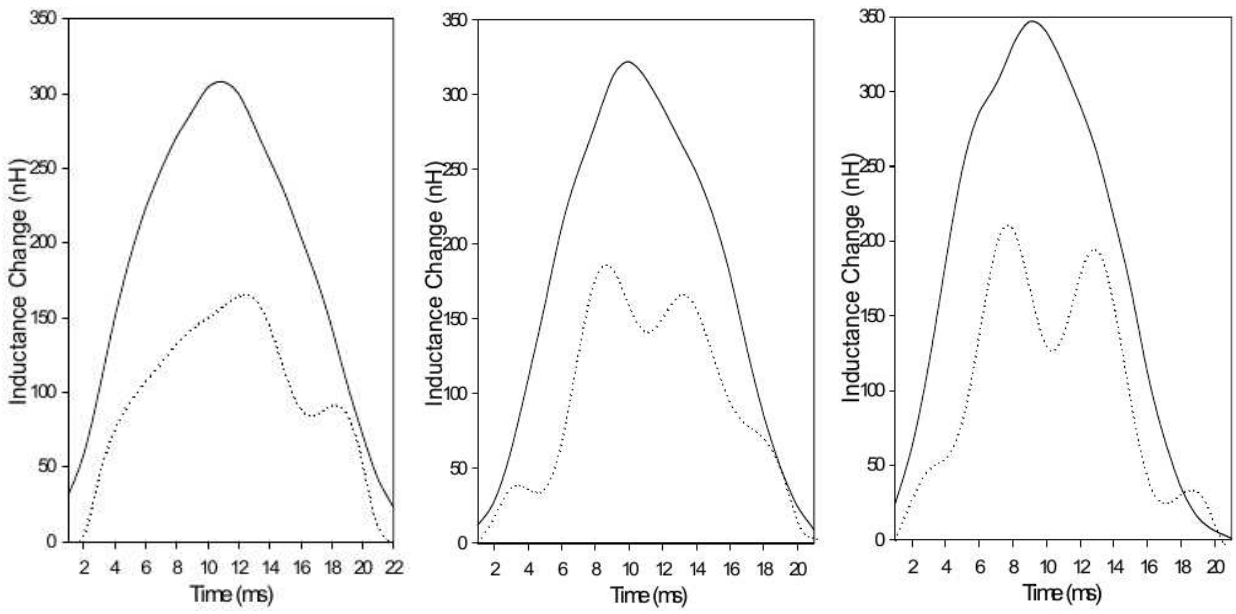
Lorsque le système converge, le filtre inverse obtenu $\mathbf{W}$ est proche de la fonction de transfert $\mathbf{H}$ et la signature réelle estimée $\tilde{\mathbf{e}}$ est proche de la signature réelle $\mathbf{e}$. Il est à noter que ce processus implique seulement la connaissance du signal observé $\mathbf{s}$. La mise en œuvre de cet algorithme par Kwon est effectuée dans le domaine fréquentiel après une transformée de Fourier. Ensuite le résultat est converti dans le domaine temporel.

Résultats et discussions

Kwon dans [20, 65] a testé les deux algorithmes de déconvolution sur une base de données de 563 véhicules. Les figures 6.6 montrent les résultats obtenus après la déconvolution par les moindres carrés pour différents véhicules, et les figures 6.7 par CMA appelé méthode de Godard par Kwon. Pour les figures 6.6a, 6.6b et 6.6c, la différence entre les signatures sans déconvolution est plus difficile à faire que pour les signatures déconvoluées. Cette différence devrait faciliter la réidentification. De même pour les figures 6.7a, 6.7b et 6.7c, la différence entre les signatures est accentuée par la déconvolution. Les signatures de la figure 6.6a (ou 6.6b, ou 6.6c) sont issues du même véhicule que celles de la figure 6.7a (ou 6.7b, ou 6.7c). En comparant les figures, nous nous apercevons que le résultat de la déconvolution par les moindres carrés n'est pas identique à celui de la déconvolution par la méthode de Godard. La méthode employée a une influence sur l'amplitude du signal déconvolué. Les hypothèses émises influencent la déconvolution. Pour réaliser la déconvolution par les moindres carrés, Kwon fait une hypothèse forte sur la forme de la fonction de transfert $\mathbf{H}$. L'hypothèse que la fonction de transfert est une somme de gaussiennes n'est pas certaine, et introduit des questions : comment déterminer les paramètres des fonctions gaussiennes ? combien de fonctions gaussiennes faut-il sommer ? La détermination des paramètres de l'équation (6.8) n'est pas expliquée par Kwon. Il est légitime de se poser des questions sur les valeurs prises par ces paramètres : $\mathbf{Q}$ est-il toujours le même pour les différents véhicules ? Comment est calculé le paramètre α ?

L'algorithme de Godard a un temps de calcul très supérieur à celui des moindres carrés pour effectuer la déconvolution d'après le rapport [20] de Kwon. Par contre, cette méthode n'a pas d'*a priori* sur la forme de la fonction de transfert et elle ne nécessite pas d'inversion de matrice. Mais il est impossible de connaître le nombre d'itérations nécessaire pour que l'algorithme converge.

La déconvolution par les moindres carrés proposée par Kwon oblige le passage de chaque véhicule pour calculer les paramètres α, la fonction de transfert de la boucle inductive. La déconvolution aveugle contraint de minimiser le critère J_2 par l'algorithme du gradient stochastique pour chaque véhicule. Le calcul des paramètres à chaque passage de véhicule pour les deux algorithmes proposés par Kwon, est un inconvénient et représente un temps de calcul élevé. Une estimation fiable et robuste de la fonction de transfert de la boucle inductive utilisable directement pour tous les véhicules permettrait d'améliorer les temps de calcul.



(a) (b) (c)

FIGURE 6.6 – Déconvolution par la méthode des moindres carrés développée par Kwon (extrait de [20]). En trait continu la signature sans déconvolution et en pointillé la signature déconvoluée : (a) : pour une voiture ; (b) : pour un van ; (c) : pour un petit camion.

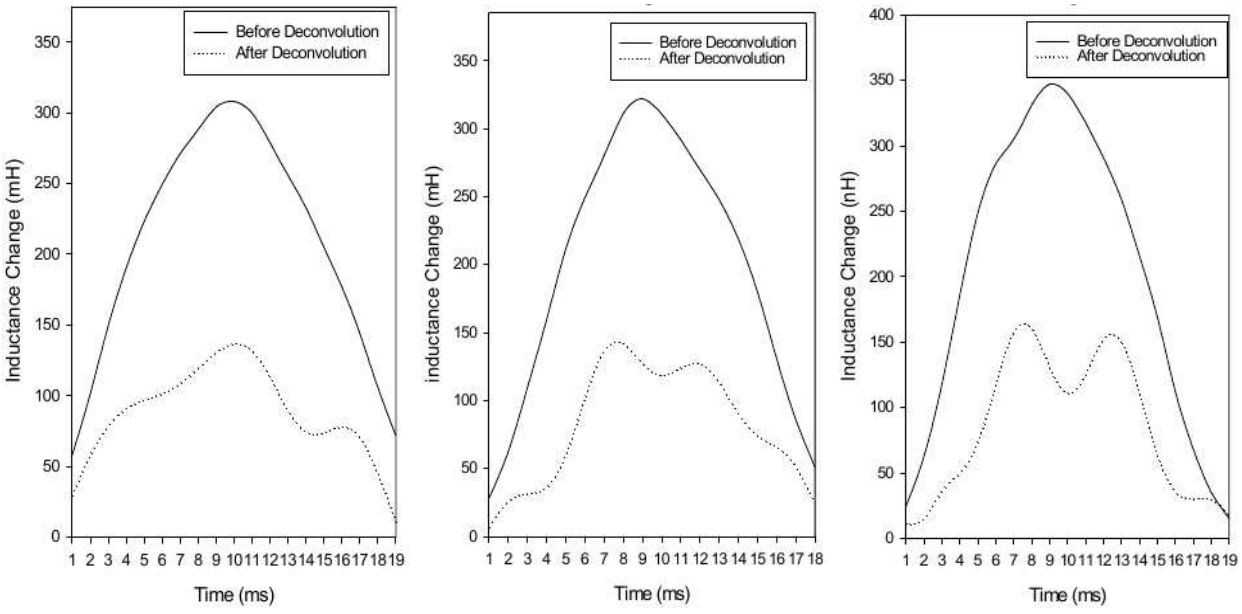
6.2 Propositions de sélection de données en fonction de la catégorie des véhicules

La sélection de données est réalisée avec des signatures issues de la boucle inductive, mais cette analyse peut être effectuée avec des signatures issues du magnétomètre.

6.2.1 Informations recueillies & signatures

Les figures 6.8 et 6.9 montrent les signatures inductives d'un véhicule léger (signature avec un maximum) et d'un poids lourd (signature avec plusieurs maximums).

Dans les études présentées dans [47, 59, 68, 69], afin de pouvoir comparer au mieux les deux signaux d'un même véhicule au passage sur deux boucles inductives différentes, divers prétraitements ont été réalisés :



(a) (b) (c)

FIGURE 6.7 – Déconvolution par la méthode de Godard développée par Kwon (extrait de [20]). En trait continu la signature sans déconvolution et en pointillé la signature déconvoluée : (a) : pour une voiture ; (b) : pour un van ; (c) : pour un petit camion.

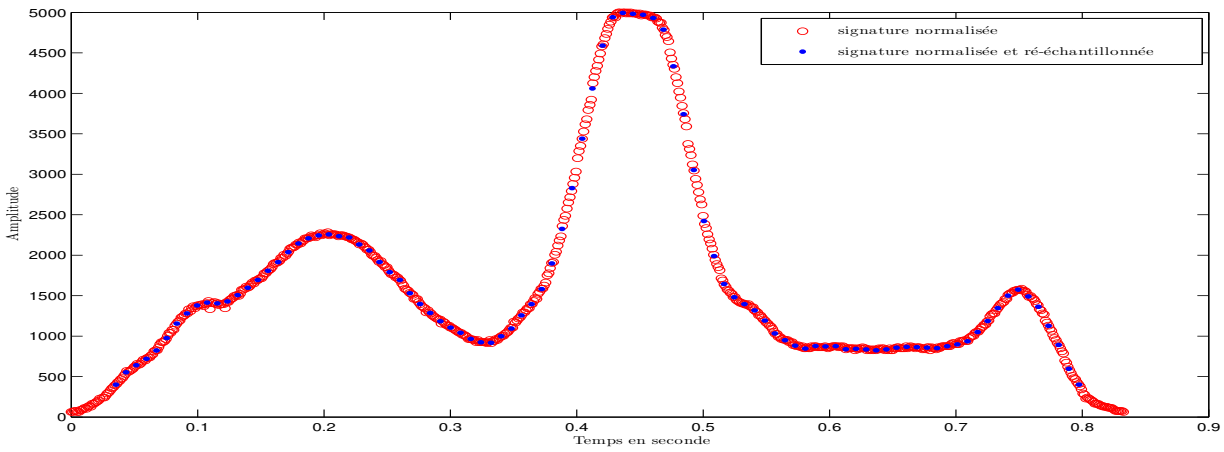


FIGURE 6.8 – Exemple de signature d'un véhicule de classe 10.

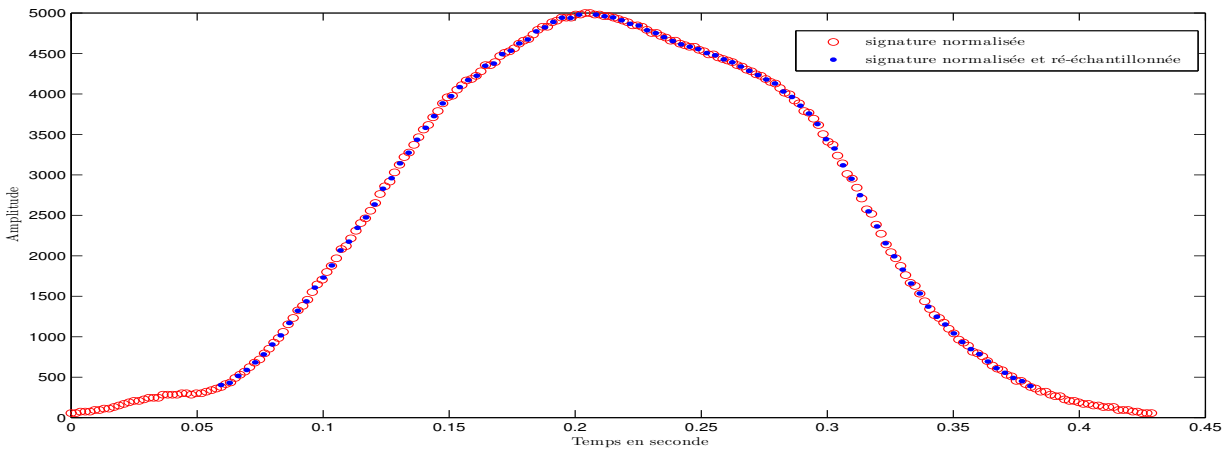


FIGURE 6.9 – Exemple de signature d'un véhicule de classe 1.

- pour s'affranchir des problèmes liés aux temps d'acquisition des données, le signal a été interpolé en fonction d'un temps moyen entre chaque acquisition. Le signal a été normalisé en abscisse à 96 points.

- l'amplitude du signal est fonction du type de véhicule. Elle dépend en particulier : de la hauteur de la masse métallique du véhicule par rapport au capteur, de la position du véhicule sur la boucle inductive. Comme il s'agit de comparer deux signaux d'un même véhicule passant sur deux capteurs différents et éloignés l'un de l'autre, les trajectoires peuvent différer et par conséquent leur amplitude aussi. Pour remédier à ce problème, chaque signal a été normalisé en amplitude [68]. À noter qu'un véhicule qui ne passe pas au même niveau sur les deux boucles inductives, ne possède pas exactement la même signature, même après normalisation [70].

- Afin d'éliminer le bruit de fond électronique du signal, seules les valeurs au-dessus d'un certain seuil ont été enregistrées. Pour information, le lecteur trouvera dans [68] plus de détails sur cette normalisation.

Les figures 6.8 et 6.9 mettent en lumière une partie du prétraitement. Par exemple sur les figures, les ronds rouges correspondent au relevé de la boucle normalisée à la valeur 5000. La normalisation aurait pu être effectuée avec une valeur de un. Les points bleus correspondent au signal rééchantillonné à pas constant. A l'issue de ces divers traitements, un signal normalisé en 96 points est créé. Dans un second temps, différentes données globales sont calculées.

6.2.2 Étude réalisée

Les signatures inductives sont issues des boucles inductives des différents sites expérimentaux. La base de données est présentée à la section 8.1. Chaque signature est rééchantillonnée sur 96 points, quel que soit le type de véhicule (Véhicules Légers ou Poids Lourds). Cette valeur a été choisie afin de ne pas perdre trop d'information (notamment pour les poids lourds), mais aussi pour conserver une certaine rapidité de calcul en ne surchargeant pas le calculateur de données. Dans [68], 206 données sont calculées à partir de la signature : 96 valeurs liées à la signature traitée, 47 données fréquentielles obtenues à partir de la transformée de Fourrier et 63 données dites globales qui sont obtenues par calcul. Ainsi, chaque signature est caractérisée par un nombre élevé de données, et il existe certainement de la redondance. Cela conduit aussi à des temps de calcul trop longs et des traitements inutiles. Le nombre de données doit être réduit tout en conservant au mieux l'information contenue dans ces données. Pour ce faire, comme dans [71], une analyse en composantes principales a été réalisée. Elle permettra de déterminer quelles sont les données à conserver et celles à rejeter. La procédure adoptée est celle présentée dans [71]. Le but est de sélectionner les données les plus représentatives et les moins corrélées entre elles. Contrairement à [71], les Véhicules Légers (VL) et les Poids Lourds (PL) sont différenciés. Cette analyse, présentée ici, permet de sélectionner les données les plus pertinentes pour chaque type de population. Le tableau analysable est constitué de z véhicules caractérisés chacun par 204 données ; z dépend de la population analysée dans ce manuscrit. Trois types de populations sont analysées :

- Véhicules Légers (VL : classe 1 selon la norme NF P 99-300) ;

- Poids Lourds (PL : classe 2 à 10 selon la norme NF P 99-300) ;

- VL–PL (classe 1 à 10 selon la norme NF P 99-300).

Chaque véhicule a la même importance dans l'analyse. Les données étant hétérogènes, elles ont été normalisées (moyenne nulle et écart-type égal à un). Ainsi, après l'Analyse en Composantes Principales (ACP), les valeurs propres, les taux d'inertie et les coordonnées des données dans le plan 1–2 (1^{re} et 2^{e} composantes principales) sont obtenues. Dans le cas où les variables sont centrées et

réduites, la variance de chaque variable vaut 1. L'inertie totale est alors égale au nombre de variables. Le taux d'inertie d'un vecteur propre est calculé en divisant la valeur propre associée par le nombre de variables.

L'ensemble des données est projeté dans le plan 1–2 afin de rechercher les corrélations ou non corrélations entre elles comme l'illustre la figure 6.10.

L'objectif est de choisir un nombre minimum de données qui soient à la fois bien représentées dans le plan 1–2, non corrélées entre elles et qui conservent l'ensemble des caractéristiques du jeu de données.

À partir du taux d'inertie, neuf données peuvent fournir une bonne représentation approximative des signatures étudiées pour une population de **VL**. L'inertie des neuf premiers axes représente 84 % de l'inertie totale. Pour les **PL**, neuf données donnent aussi une bonne représentation des signatures ; l'inertie totalisée par les neuf premiers axes représente 83 % de l'inertie totale.

Les données choisies pour représenter les signatures ne sont pas les composantes principales proposées par l'ACP. Aussi, pour ne pas générer de calculs supplémentaires et représenter les axes, les données les plus proches de ces axes sont sélectionnées. C'est pourquoi dans les exemples suivants, un nombre plus important de données que celui déterminé par l'ACP est sélectionné.

Pour sélectionner les données appropriées, la projection des anciennes données sur la nouvelle base est utilisée, comme il est visible sur la figure 6.10. Celle-ci représente la projection des données sur les deux premiers axes de la base. Ces deux axes possèdent la majorité de l'inertie totale. Cependant, pour le choix des données, la projection sur les autres axes est aussi utilisée.

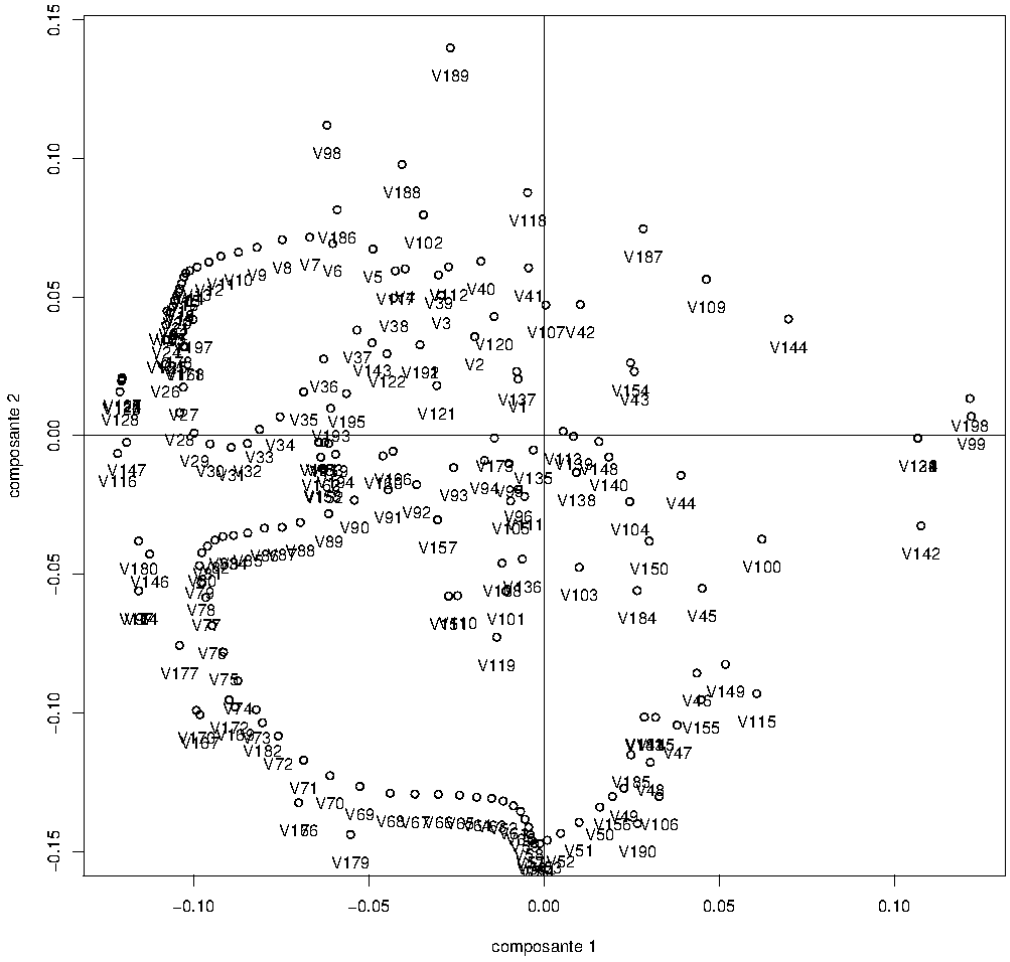


FIGURE 6.10 – Représentation des 204 données sur les deux axes principaux de la nouvelle base pour les Poids Lourds.

6.2.3 Population de VL

Dans ce cas, il a été retenu onze caractéristiques. Trois de ces caractéristiques sont issues du signal échantillonné : ce sont les valeurs du 42e, 56e et 90e point de la signature. Cinq de ces caractéristiques sont issues de la transformée de Fourrier : la partie réelle de la 9e composante, la partie imaginaire des 4e, 5e et 6e composantes ainsi que le module de la 3e composante. Enfin, trois caractéristiques temporelles globales sont aussi retenues : la bande passante, l'écart-type total et le coefficient de Skewness. La bande passante représente le nombre de valeurs où l'ordonnée est supérieure au maximum de la signature divisé par deux. Le coefficient de Skewness est le moment d'ordre trois. Il caractérise l'obliquité ou l'asymétrie du signal. Un coefficient nul indique une distribution symétrique. Une valeur différente indique un décalage de la distribution par rapport à la médiane.

6.2.4 Population de PL

Les classes 2 à 10 correspondent aux véhicules « poids lourds ». Lors de cette étude, neuf caractéristiques ont été retenues. Cinq caractéristiques sont issues de la transformée de Fourier : la partie réelle des 1re et 2e composantes, la partie imaginaire des 2e et 3e composantes et le module de la 3e composante. De plus, il y a quatre caractéristiques temporelles globales : la valeur du maximum secondaire, la valeur quadratique moyenne totale (cf. équation (6.16)), la valeur quadratique moyenne gauche (cf. équation (6.17)) et le coefficient de Skewness.

$$\frac{\sum_{k=1}^{longueur}[signature(k)]^2}{longueur} \tag{6.16}$$

$$\frac{\sum_{k=1}^{\frac{longueur}{2}}[signature(k)]^2}{\frac{longueur}{2}} \tag{6.17}$$

6.2.5 Population de VL–PL

La population totale comprend 90 % de VL et 10 % de PL. Au cours de cette analyse de données, douze caractéristiques ont été retenues. Trois caractéristiques viennent du signal échantillonné : ce sont les valeurs des 42e, 43e et 44e points de la signature. Huit caractéristiques proviennent de la transformée de Fourrier : la partie réelle des 2e, 3e et 7e composantes, la partie imaginaire des 2e, 3e et 4e composantes, le module de la 2e composante ainsi que le taux de distorsion harmonique (THD cf. équation (6.18)).

$$THD = \frac{\sqrt{A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2 + A_6^2 + A_7^2 + A_8^2}}{A_0} \tag{6.18}$$

avec A_i le module de la ie composante de la transformée de Fourier. Il y a aussi une caractéristique temporelle qui est la somme de la partie droite de la signature divisée par la somme générale.

6.2.6 Conclusion

Cette section a permis de réduire le nombre de données par Analyse en Composantes Principales (ACP) pour les signaux issus de la boucle inductive et a déterminé les p caractéristiques les plus pertinentes. Ainsi, seules les p[1] données (ou identifiants) les plus pertinentes seront utilisées. Alors, chaque véhicule est caractérisé par p données $(x_1,...,x_p)$.

[1] 11 pour les VL ; 9 pour les PL et 12 pour les VL/PL

6.3 Propositions d'algorithmes de déconvolution

La déconvolution est réalisée pour les signatures issues de la boucle inductive.

6.3.1 Contexte

Kwon dans [20, 65] a émis l'hypothèse, dans le cas de la déconvolution par les moindres carrés que la fonction de transfert $\mathbf{h}$ est la somme de gaussiennes. Ainsi, cette somme de fonctions gaussiennes représente l'estimation de la fonction de transfert $\tilde{\mathbf{h}}$. Cependant, l'hypothèse formulée par Kwon sur l'estimation de la fonction de transfert $\tilde{\mathbf{h}}$ est forte. De plus, les paramètres (nombre de gaussiennes, moyenne, écart-type) pour établir la somme de fonctions gaussiennes ne sont pas évidents à estimer. Il est à noter aussi que la fonction de déconvolution est recalculée à chaque détection de véhicules. Ce calcul est encore plus pénalisant en utilisant l'algorithme de Godard pour la déconvolution.

Dans cette section, un nouvel algorithme de déconvolution aveugle est proposé. Cette méthode ne nécessite pas d'*a priori* sur la fonction de transfert. Seules les propriétés mathématiques de régularité sont utilisées. L'idée principale est d'estimer $\tilde{\mathbf{h}}$ et $\tilde{\mathbf{e}}$ de manière itérative comme le montre la figure 6.11. L'hypothèse émise se situe sur la forme du signal d'entrée $\mathbf{e_0}$, afin d'estimer la fonction de transfert par la déconvolution du signal $\mathbf{s}$ par $\mathbf{e_0}$. Le signal d'entrée $\mathbf{e_0}$ est choisi arbitrairement. Ensuite, en appliquant la déconvolution à la fonction de transfert estimée $\tilde{\mathbf{h}}$ par le signal observé $\mathbf{s}$, l'estimation du signal d'entrée est obtenue. Le signal d'entrée estimé $\tilde{\mathbf{e}}$ est utilisé pour déterminer une nouvelle fonction de transfert $\tilde{\mathbf{h}}$ et ainsi de suite. De manière à valider le processus, le signal d'entrée estimé $\tilde{\mathbf{e}}$ est convolué avec la fonction de transfert estimée $\tilde{\mathbf{h}}$ pour obtenir l'estimation du signal observé $\tilde{\mathbf{s}}$. L'erreur est calculée, en utilisant la norme $\ell 2$, entre le signal observé et l'estimation de celui-ci. L'erreur sert de critère d'arrêt pour l'algorithme.

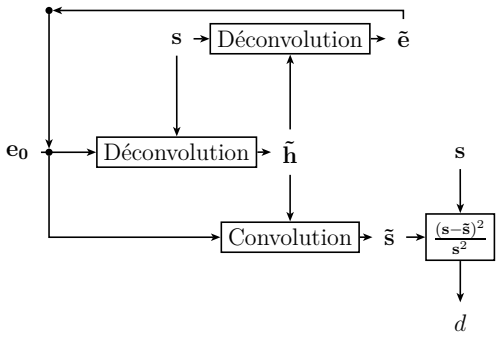


FIGURE 6.11 – Diagramme pour estimer $\tilde{\mathbf{h}}$ et $\tilde{\mathbf{e}}$.

6.3.2 Définition du problème

Intuitivement pour des boucles inductives définies selon la même norme (SIREDO par exemple), la fonction de transfert $\mathbf{h}$ doit être identique quel que soit le véhicule qui passe. La première étape va donc consister à estimer la fonction de transfert $\tilde{\mathbf{h}}$ pour un format de boucle inductive donné. Cependant un même véhicule va générer un signal observé plus ou moins long suivant sa vitesse. Cette durée du signal est appelée temps de présence, notée par la suite TI. En restant dans l'espace temporel, la durée de la fonction de transfert va aussi varier suivant la vitesse du véhicule qui passe. Pour s'affranchir de ce problème, il est proposé de passer du domaine temporel au domaine spatial. Pour réussir le changement de représentation, il faut que la vitesse individuelle v de chaque véhicule soit mesurée. Pour mesurer cette vitesse, plusieurs possibilités peuvent être utilisées :

- dans le cas d'une paire de boucles inductives, la vitesse est mesurée suivant le temps de parcours entre les deux boucles inductives,

- dans le cas d'une seule boucle inductive, la vitesse peut être estimée à partir de l'analyse de la signature par les méthodes développées dans [55, 72],
- autrement, un autre type de capteur de vitesse peut être employé comme par exemple un radar. Ensuite, les mesures de vitesse relevées par le capteur sont associées de manière précise avec les signatures.

Lors des expérimentations que nous avons menées, les boucles inductives sont par paire et nous permettent ainsi de mesurer la vitesse. Le nouveau référentiel considère comme origine spatiale le centre de la largeur de la boucle inductive. Nous faisons l'hypothèse que la moitié du temps d'occupation ($\frac{TI}{2}$) correspond à l'instant duquel le centre longitudinal du véhicule est placé au niveau du centre longitudinal de la boucle inductive. Nous considérons aussi que le centre de la boucle inductive correspond au centre du véhicule. Le changement de repère s'effectue en appliquant l'équation suivante :

$$x = v \times (-t + \frac{TI}{2}) \tag{6.19}$$

Grâce à cette équation, l'ensemble des signaux s'exprime en fonction de la position spatiale. La figure 6.12 illustre pour une signature la relation entre les repères temporel et spatial.

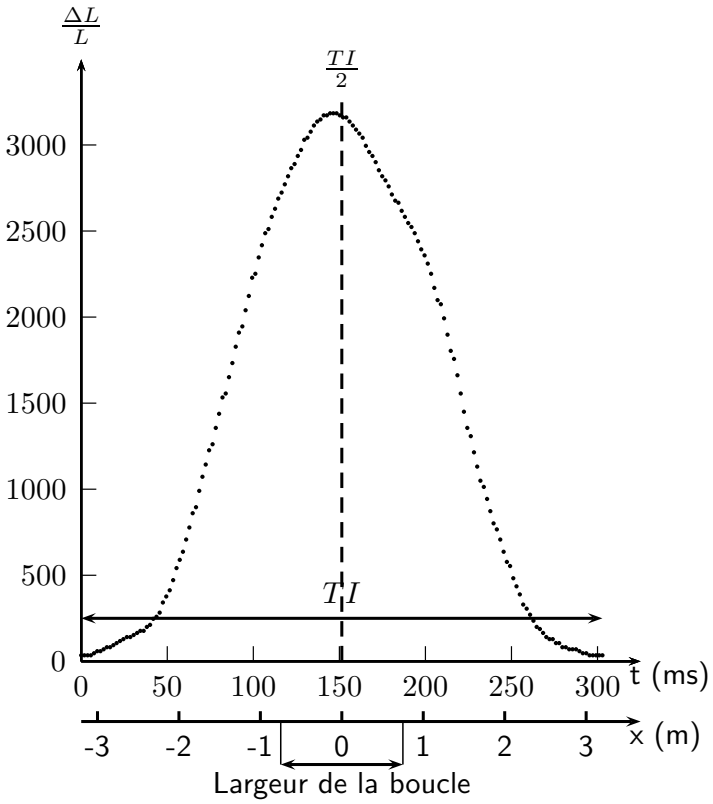


FIGURE 6.12 – Changement de repère.

Le signal observé devient alors $s(x)$. L'équation (6.1) dans le domaine temporel s'écrit dans le domaine spatial comme suit :

$$s(x) = h(x) * e(x) \tag{6.20}$$

Les signaux $s(x)$, $e(x)$ et $h(x)$ sont à support compact :

$$\exists (a,b) \in \mathbf{R}^2, a < b, s(x) = \int_{-\infty}^{+\infty} h(x-l)e(l)dl = \int_a^b h(x-l)e(l)dl \tag{6.21}$$

L'équation (6.21) représente une équation intégrale de Fredholm de première espèce qui est généralement un problème mal posé. Les fonctions de l'équation dans un espace Hilbertien sont de carré intégrable puisque leur norme est finie. Cette condition assure que les moindres carrés peuvent être appliqués à l'équation (6.20). L'existence d'une approximation de la solution par les moindres carrés est conditionnée par le théorème de Picard [73].

Théorème 1 (Théorème de Picard)
Soit K un opérateur compact linéaire défini par $Ke(x) = \int_a^b k(x,l)e(l)dl = s(x)$ sur l'espace de Hilbert $L^2[a,b]$ et $k(x,l) = \sum_{i=1}^{\infty} \mu_i u_i(x) v_i(l)$ le développement en valeurs singulières du noyau k. μ_i représente les valeurs singulières et le couple (u_i,v_i) les vecteurs singuliers associés gauche et droit. L'équation $Ke = s$ a une solution si et seulement si la condition de Picard est vérifiée.

La condition de Picard est la suivante :

$$\sum_{i=1}^{\infty} \left(\frac{\langle u_i,s \rangle}{\mu_i} \right)^2 < \infty \quad \mu_i \neq 0 \tag{6.22}$$

avec $\langle u_i,s \rangle$ le produit scalaire dans $L^2[a,b]$.

Dans le cas présent, les mesures discrètes sont considérées comme des vecteurs ou des matrices. Le signal observé est représenté par le vecteur $\mathbf{s} = [s_0,s_1,\ldots,s_{n-1}]^T$ de dimension n, les valeurs $s_i, i = 0 \cdots n-1$ sont calculées par une interpolation B-spline pour palier la fréquence d'échantillonnage variable de la boucle inductive. La fonction de transfert associée à la boucle inductive est représentée par le vecteur $\mathbf{h} = [h_0,h_1,\ldots,h_{m-1}]^T$ de dimension m. Le signal « réel » d'entrée est symbolisé par le vecteur $\mathbf{e} = [e_0,e_1,\ldots,e_{l-1}]$ de dimension l. La relation entre les différentes dimensions est : $n = m + l - 1$. Soit $\mathbf{E}$ la matrice de Toeplitz de dimension $n \times m$ construite à partir du vecteur $\mathbf{e}$ tel que :

$$\mathbf{E} = \begin{bmatrix} e_0 & \cdots & e_{l-1} & 0 & \cdots & \cdots \\ 0 & e_0 & \cdots & e_{l-1} & 0 & \cdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & e_0 \end{bmatrix} \tag{6.23}$$

Soit $\mathbf{H}$ la matrice de Toeplitz de dimension $n \times l$ construite à partir du vecteur $\mathbf{e}$ de la même façon que la matrice $\mathbf{E}$. L'équation de convolution (6.20) s'écrit maintenant sous forme matricielle de la façon suivante :

$$\mathbf{Eh} = \mathbf{He} = \mathbf{s} \tag{6.24}$$

Cependant, dans notre problème, la seule information disponible est $\mathbf{s}$. Les vecteurs $\mathbf{e}$ et $\mathbf{h}$ sont inconnus. Pour estimer ce couple de vecteurs, une approche basée sur la déconvolution aveugle est proposée à partir de la résolution de deux problèmes par la méthode des moindres carrés :

$$\tilde{\mathbf{h}}_\mathbf{n} = \arg\min_{\mathbf{h_n}} \|\mathbf{s} - \tilde{\mathbf{E}}_{\mathbf{n-1}} \mathbf{h_n}\|_2^2 \tag{6.25a}$$

$$\tilde{\mathbf{e}}_\mathbf{n} = \arg\min_{\mathbf{e_n}} \|\mathbf{s} - \tilde{\mathbf{H}}_\mathbf{n} \mathbf{e_n}\|_2^2 \tag{6.25b}$$

avec $\tilde{\mathbf{h}}_\mathbf{n}$ et $\tilde{\mathbf{e}}_\mathbf{n}$ les solutions du problème à la n^e itération, $\tilde{\mathbf{H}}_\mathbf{n}$ et $\tilde{\mathbf{E}}_\mathbf{n}$ les matrices de Toeplitz des vecteurs $\tilde{\mathbf{h}}_\mathbf{n}$ et $\tilde{\mathbf{e}}_\mathbf{n}$ respectivement. Afin d'initialiser l'algorithme, nous émettons une hypothèse *a priori* sur $\mathbf{e_0}$. L'approche proposée est basée sur la résolution des problèmes (6.25a) et (6.25b) alternativement. La première étape consiste à estimer $\tilde{\mathbf{h}}_\mathbf{1}$ à partir de l'hypothèse $\mathbf{e_0}$ avec l'équation (6.25a). Ensuite en utilisant $\tilde{\mathbf{h}}_\mathbf{1}$ dans l'équation (6.25b), c'est $\mathbf{e_1}$ qui est estimé. Ces deux étapes alternatives sont réitérées jusqu'à la convergence. Toutefois, il est admis que l'équation de Fredholm conduit généralement à un problème inverse discret mal posé comme il est défini dans [74]. En utilisant le théorème de Picard et la condition de Picard discrète introduite dans [74], la vérification d'une approximation numérique sans oscillation pour les signaux $\tilde{\mathbf{h}}$ et $\tilde{\mathbf{e}}$ peut être obtenue.

Dans la forme discrète, $\mathbf{s}$ satisfait la condition de Picard discrète si pour toutes les valeurs singulières non nulles μ_i, le produit scalaire correspondant $\langle u_i,s\rangle$ décroît en moyenne plus vite que μ_i. Pour vérifier les conditions discrètes de Picard, Hansen dans [74] a proposé de calculer la moyenne géométrique mobile (MGM). Pour cela, la matrice $\mathbf{E}$ est décomposée en valeurs singulières (SVD). $\mathbf{E} = \mathbf{UAV^T} = \sum_{i=1}^{m} \alpha_i \mathbf{u}_i \mathbf{v}_i^T$, avec $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ et $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ les matrices gauche et droite orthonormées. Les vecteurs singuliers à gauche et à droite sont associés aux valeurs singulières correspondantes α_i dans l'ordre non décroissant. Les coefficients diagonaux de la matrice diagonale $\mathbf{A}$ sont égaux aux valeurs singulières α_i de $\mathbf{E}$. De la même manière, la décomposition en valeurs singulières est appliquée à $\mathbf{H} = \mathbf{GBW^T} = \sum_{i=1}^{l} \beta_i \mathbf{g}_i \mathbf{w}_i^T$, avec $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_n)$ et $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ les matrices gauche et droite orthonormées. Les vecteurs singuliers à gauche et à droite sont associés aux valeurs singulières correspondantes β_i dans l'ordre non décroissant. Les coefficients diagonaux de la matrice diagonale $\mathbf{B}$ sont égaux aux valeurs singulières α_i de $\mathbf{H}$. Ainsi, la condition de Picard discrète affirme que le signal observé $\mathbf{s}$ est la sortie de droite de l'équation (6.24) si pour toutes les valeurs singulières différentes de zéro α_i (ou β_i), le produit scalaire correspondant $|\mathbf{u}_i^T\mathbf{s}|$ (ou $|\mathbf{g}_i^T\mathbf{s}|$) décroît vers zéro en moyenne plus vite que α_i (ou β_i).

Les sections suivantes aborderont l'algorithme avec la méthode des moindres carrés. La condition de Picard montre que le problème est mal posé au sens de Hadamard, c'est-à-dire que la solution n'existe pas, n'est pas unique ou n'est pas stable. Pour résoudre ce problème, la théorie de régularisation de Tikhonov est utilisée.

6.3.3 Algorithmes utilisés

Approche par les moindres carrés

La méthode est d'utiliser l'approche par les moindres carrés pour résoudre les équations (6.25a) et (6.25b). La première étape est d'estimer $\tilde{\mathbf{h}}$ à partir de l'équation (6.25a). Le critère à minimiser s'écrit :

$$\begin{aligned} \|\mathbf{s} - \mathbf{Eh}\|^2 &= (\mathbf{s} - \mathbf{Eh})^T(\mathbf{s} - \mathbf{Eh}) \\ &= \mathbf{s}^T\mathbf{s} - \mathbf{h}^T\mathbf{E}^T\mathbf{s} - \mathbf{s}^T\mathbf{Eh} + \mathbf{h}^T\mathbf{E}^T\mathbf{Eh} \end{aligned} \tag{6.26}$$

or :

$$\begin{aligned} \mathbf{h}^T\mathbf{E}^T\mathbf{s} &= \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} e_{i,j} h_i s_j \\ &= \mathbf{s}^T\mathbf{Eh} \end{aligned} \tag{6.27}$$

En remplaçant, (6.27) dans (6.26), nous obtenons :

$$\|\mathbf{s} - \mathbf{Eh}\|^2 = \mathbf{h}^T\mathbf{E}^T\mathbf{Eh} - 2\mathbf{h}^T\mathbf{E}^T\mathbf{s} + \mathbf{s}^T\mathbf{s} \tag{6.28}$$

Le terme $\mathbf{s}^T\mathbf{s}$ étant constant, la minimisation est équivalente à :

$$J(\mathbf{h}) = \mathbf{h}^T\mathbf{E}^T\mathbf{Eh} - 2\mathbf{h}^T\mathbf{E}^T\mathbf{s} \tag{6.29}$$

En calculant le gradient de $J(\mathbf{h})$:

$$\begin{aligned} \nabla J(\mathbf{h}) &= 2\mathbf{E}^T\mathbf{Eh} - 2\mathbf{E}^T\mathbf{s} \\ &= 2[\mathbf{E}^T\mathbf{Eh} - \mathbf{E}^T\mathbf{s}] \end{aligned} \tag{6.30}$$

La condition d'annulation du gradient de $J(\mathbf{h})$ est donnée par $\mathbf{E}^T\mathbf{Eh} - \mathbf{E}^T\mathbf{s} = 0$. Ce qui donne :

$$\tilde{\mathbf{h}} = (\mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T\mathbf{s} \tag{6.31}$$

La deuxième étape consiste à estimer $\tilde{\mathbf{e}}$ à partir de l'équation (6.25b). De la même façon que précédemment l'équation s'écrit :

$$\tilde{\mathbf{e}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{s} \tag{6.32}$$

Les deux équations vont être résolues alternativement plusieurs fois (cf. figure 6.11) ce qui conduit à écrire les équations sous la forme :

$$\tilde{\mathbf{h}}_n = (\tilde{\mathbf{E}}_{\mathbf{n-1}}^{\mathbf{T}}\tilde{\mathbf{E}}_{\mathbf{n-1}})^{-1}\tilde{\mathbf{E}}_{\mathbf{n-1}}^{\mathbf{T}}\mathbf{s} \tag{6.33a}$$

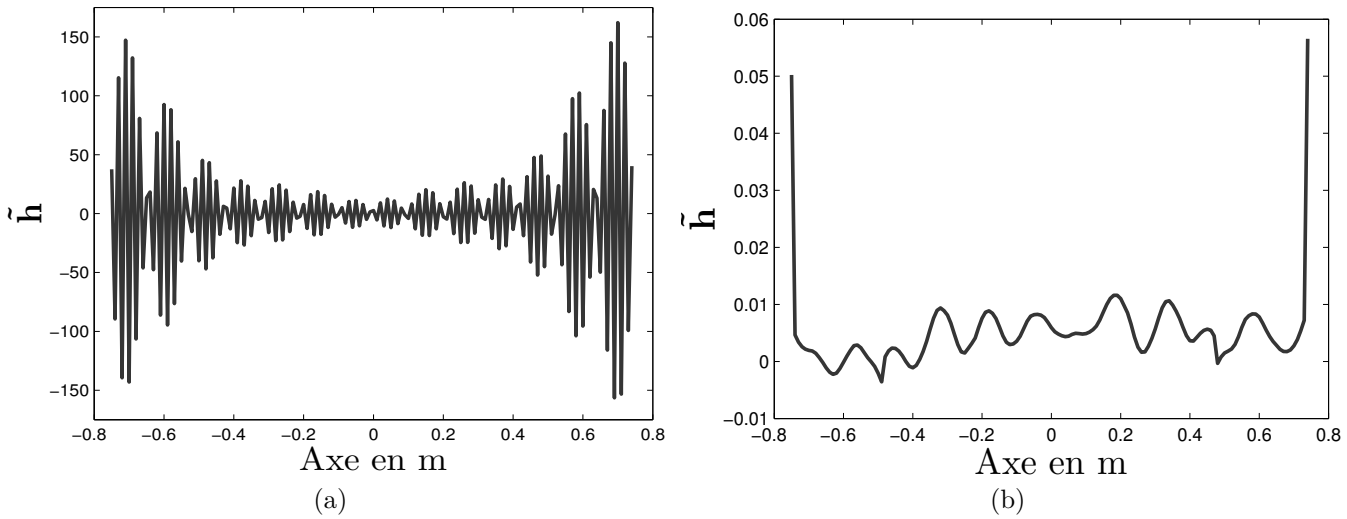
$$\tilde{\mathbf{e}}_n = (\tilde{\mathbf{H}}_{\mathbf{n}}^{\mathbf{T}}\tilde{\mathbf{H}}_{\mathbf{n}})^{-1}\tilde{\mathbf{H}}_{\mathbf{n}}^{\mathbf{T}}\mathbf{s} \tag{6.33b}$$

L'algorithme est appliqué sur un signal observé $\mathbf{s}$, ensuite une estimation $\tilde{\mathbf{s}}_n$ du signal observé est calculée à partir de $\tilde{\mathbf{e}}_n$ et $\tilde{\mathbf{h}}_n$: $\tilde{\mathbf{s}}_\mathbf{n} = \tilde{\mathbf{H}}_\mathbf{n}\tilde{\mathbf{e}}_\mathbf{n}$. Pour l'initialisation de $\mathbf{e_0}$, deux hypothèses ont été testées :

- la somme de deux fonctions gaussiennes,
- une fonction rectangle.

L'algorithme est considéré comme convergent si et seulement si l'équation (6.34) est vérifiée : c'est-à-dire que l'erreur converge vers zéro, et que la condition de Picard discrète est confirmée pour chaque itération n.

$$\lim_{n\to+\infty} \frac{\|\mathbf{s} - \tilde{\mathbf{s}}_\mathbf{n}\|_2^2}{\|\mathbf{s}\|_2^2} = 0 \tag{6.34}$$



(a) (b)

FIGURE 6.13 – Estimation de $\mathbf{h}$ avec les différentes hypothèses (a) : somme de gaussiennes, (b) : fonction rectangle.

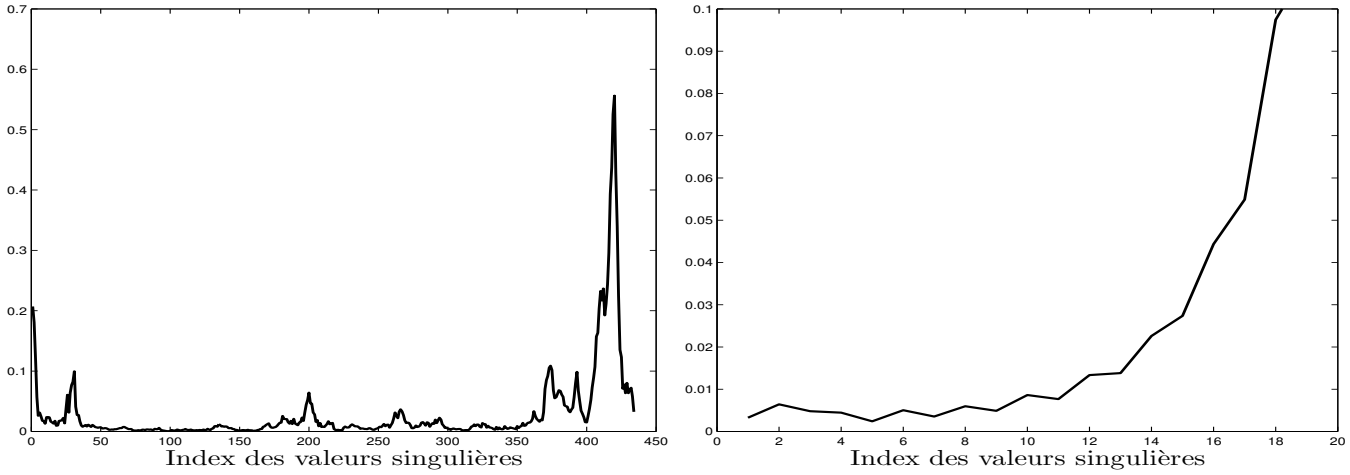
La figure 6.13 montre que les solutions sont vraiment très différentes même si l'équation (6.34) converge vers zéro. Pour l'hypothèse de la somme de deux gaussiennes, la figure 6.13a présente une solution avec des valeurs oscillantes. De plus, comme le montre la figure 6.14 la condition de Picard discrète n'est pas vérifiée. Il s'agit généralement d'une solution calculée pour un problème mal posé. Pour cette raison, l'algorithme ne donne pas de solution fiable.

Approche par la régularisation de Tikhonov

L'approche classique par les moindres carrés n'offre pas de solution fiable, il est donc envisagé d'utiliser une autre approche pour résoudre le problème (6.25). L'approche de Tikhonov consiste à introduire un terme de régularisation dans la minimisation de manière à privilégier une solution particulière dont les propriétés semblent pertinentes :

$$\tilde{\mathbf{h}}_\mathbf{n} = \arg\min_{\mathbf{h_n}} \|\mathbf{s} - \tilde{\mathbf{E}}_{\mathbf{n-1}}\mathbf{h}_\mathbf{n}\|_2^2 + \lambda\|\mathbf{L}\mathbf{h}_\mathbf{n}\|_2^2 \tag{6.35a}$$

$$\tilde{\mathbf{e}}_\mathbf{n} = \arg\min_{\mathbf{e_n}} \|\mathbf{s} - \tilde{\mathbf{H}}_\mathbf{n}\mathbf{e}_\mathbf{n}\|_2^2 + \mu\|\mathbf{M}\mathbf{e}_\mathbf{n}\|_2^2 \tag{6.35b}$$



(a) (b)

FIGURE 6.14 – (a) : La moyenne géométrique mobile pour l'estimation de $\tilde{\mathbf{e}}$ varie très peu pour les premières valeurs singulières. (b) : La moyenne géométrique mobile pour l'estimation de $\tilde{\mathbf{h}}$ est croissante. La condition de Picard discrète n'est pas vérifiée.

où $\mathbf{L}$ et $\mathbf{M}$ sont des matrices bien conditionnées. Dans le cas présent, elles sont égales à la matrice identité : $\mathbf{L} = \mathbf{M} = \mathbf{I}$. Le second terme $\lambda\|\mathbf{L}\mathbf{h_n}\|_2^2$ ($\mu\|\mathbf{M}\mathbf{e_n}\|_2^2$) représente la norme de la fonction de transfert (ou du signal d'entrée) pondérée par un paramètre de régularisation. Les équations (6.35) impliquent un compromis entre la norme $\ell 2$ de la solution régularisée et la qualité de l'ajustement prévue par la solution. La régularisation permet d'éviter de très grandes solutions qui ne sont pas physiquement réalisables. Ce problème est bien conditionné si les paramètres de régularisation λ et μ ne sont pas trop petits [74]. Il est intéressant de vérifier la condition de Picard discrète pour l'équation régularisée pour les mêmes données utilisées dans l'approche des moindres carrés. Pour résoudre le problème (6.35), les hypothèses décrites précédemment pour $\mathbf{e_0}$ sont à nouveau utilisées.

Les solutions numériques des deux hypothèses semblent avoir une forme en U similaire sur les figures 6.15a et 6.15c. Les figures 6.15b et 6.15d montrent que l'erreur diminue à mesure que le nombre d'itérations augmente, c'est-à-dire que l'erreur converge vers 0.

L'analyse de la condition de Picard discrète révèle que les produits scalaires $|\mathbf{u}_i^T\mathbf{s}|$ ou $|\mathbf{g}_i^T\mathbf{s}|$ décroissent plus rapidement que les valeurs singulières respectives α_i ou β_i comme le montre la figure 6.16b. Ainsi la condition de Picard discrète est vérifiée dans le cadre du problème régularisé. En effet, l'estimation de $\tilde{\mathbf{e}}$ semble plus simple. Le problème (6.35b) est en fait bien conditionné et le produit scalaire $|\mathbf{g}_i^T\mathbf{s}|$ décroît plus vite que β_i pour chaque indice $i = 1 \cdots 150$. Au contraire, le problème (6.35a) semble être plus difficile, surtout pour l'hypothèse initiale d'une fonction rectangle. En effet, la figure 6.16d montre que la moyenne géométrique mobile décroît globalement jusqu'à la 35$^{\text{e}}$ valeur singulière et ensuite elle semble constante. Dans ce cas, la condition de Picard discrète est remplie et les valeurs singulières au delà de la 35$^{\text{e}}$ valeur sont considérées comme dominées par le bruit. La régularisation est donc utile dans ce cas.

La condition de Picard discrète est un critère efficace pour vérifier si les équations (6.35) ont des solutions numériques, mais la moyenne géométrique mobile ne détermine pas efficacement les paramètres de régularisation λ et μ. À partir de [75] et à l'aide de la décomposition en valeurs singulières, les solutions de (6.35) sont les suivantes :

$$\tilde{\mathbf{h}}_\lambda = \sum_i \phi_i \frac{\mathbf{u}_i^T\mathbf{s}}{\alpha_i}\mathbf{v}_i \tag{6.36a}$$

$$\tilde{\mathbf{e}}_\mu = \sum_i \gamma_i \frac{\mathbf{g}_i^T\mathbf{s}}{\beta_i}\mathbf{w}_i \tag{6.36b}$$

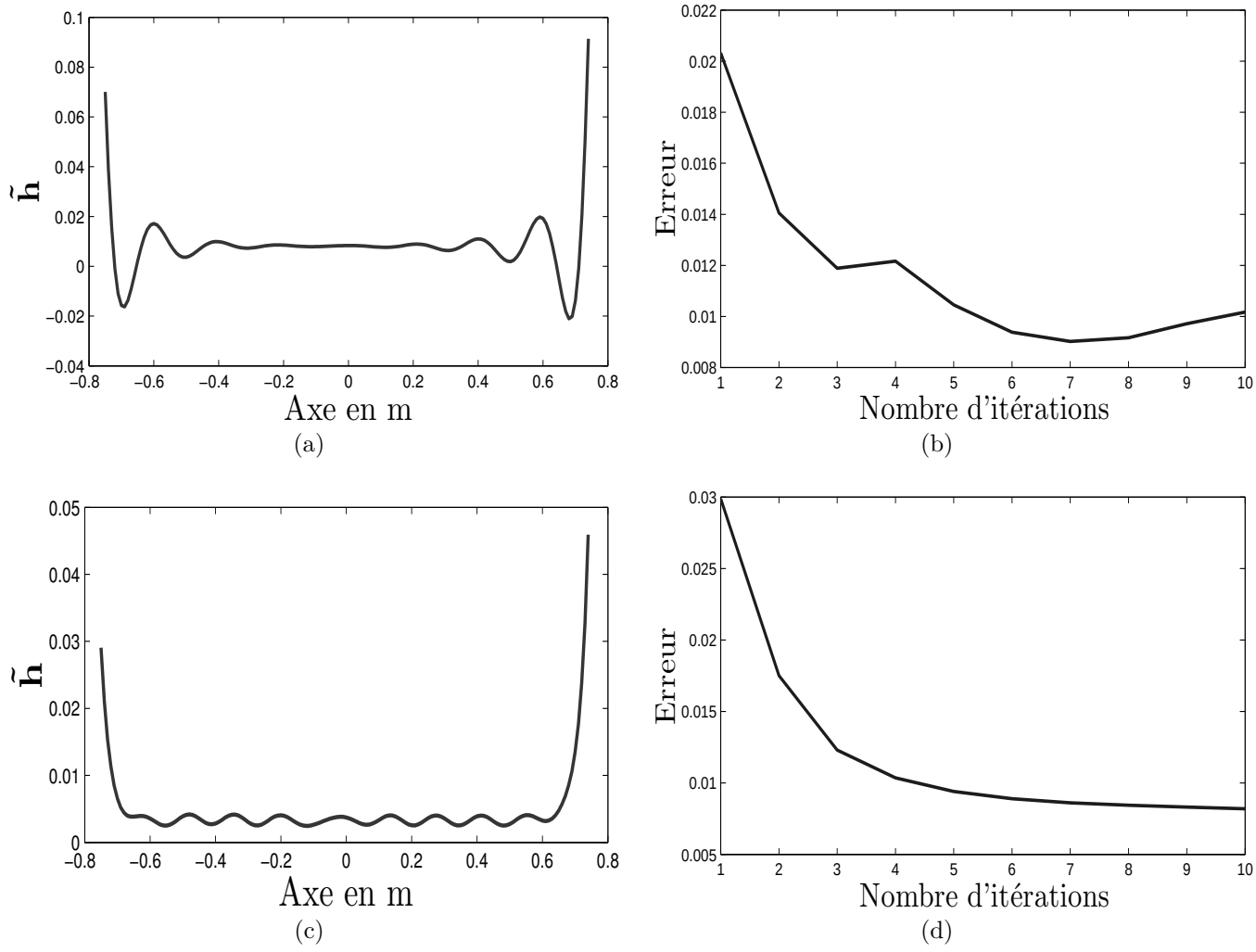


FIGURE 6.15 – Les solutions numériques de $\tilde{\mathbf{h}}$ pour la somme de gaussiennes (a) ou la fonction rectangle (c) comme hypothèse et respectivement (b), (d) l'erreur entre $\tilde{\mathbf{s}}$ et $\mathbf{s}$.

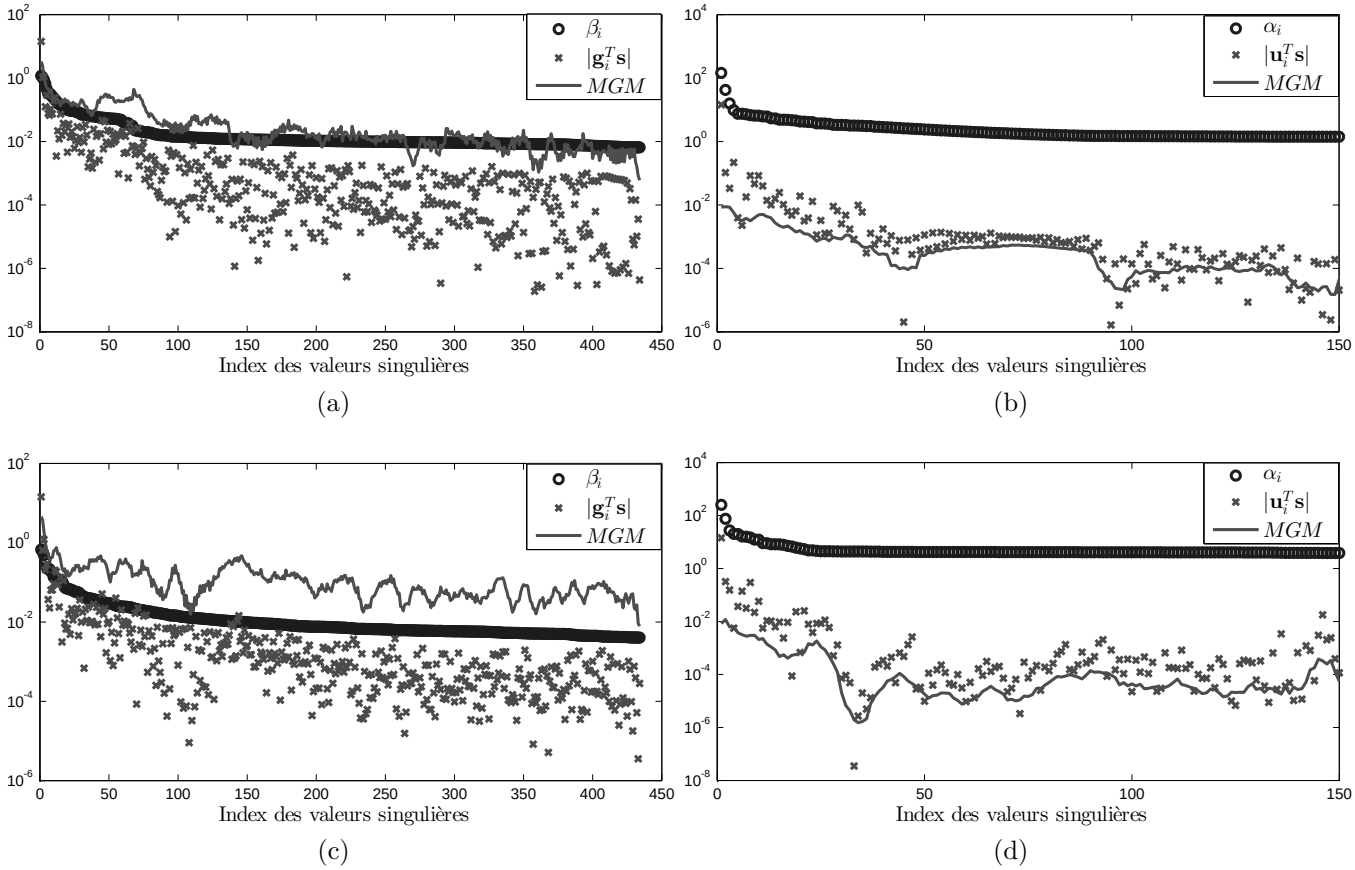


FIGURE 6.16 – Les courbes de Picard : $\tilde{\mathbf{e}}$ (a) et $\tilde{\mathbf{h}}$ (b) avec l'hypothèse d'une somme de deux gaussiennes ; $\tilde{\mathbf{e}}$ (c) et $\tilde{\mathbf{h}}$ (d) avec l'hypothèse d'une fonction rectangle.

D'après Hansen [76], quand la régularisation de Tikhonov est utilisée, les facteurs de filtrage ϕ_i et γ_i sont définis par :

$$\phi_i = \frac{\alpha_i^2}{\alpha_i^2 + \lambda^2} \simeq \begin{cases} 1, & \alpha_i \gg \lambda \\ \frac{\alpha_i^2}{\lambda^2}, & \alpha_i \ll \lambda \end{cases} \tag{6.37}$$

$$\gamma_i = \frac{\beta_i^2}{\beta_i^2 + \mu^2} \simeq \begin{cases} 1, & \beta_i \gg \mu \\ \frac{\beta_i^2}{\mu^2}, & \beta_i \ll \mu \end{cases} \tag{6.38}$$

D'autres méthodes de régularisation existent comme la troncature spectrale (souvent notée TSVD pour Truncated Singular Value Decomposition) qui consiste à tronquer le développement à un certain rang. Ce rang est l'analogue du paramètre de régularisation dans la méthode de Tikhonov. À la différence de l'approche de Tikhonov, la TSVD enlève la contribution des valeurs singulières au-delà d'un certain rang en les forçant à zéro. Une solution est de déterminer le rang en fonction de la condition de Picard discrète. Le rang est fixé pour ne conserver que les valeurs singulières respectant la condition de Picard discrète. Cependant, le choix des paramètres de régularisation est très difficile quand il est basé uniquement sur la condition de Picard discrète. Il semble que le graphique de Picard donne juste une valeur très qualitative de la plus petite valeur singulière qui n'est pas dominée par le bruit.

L'approche de Tikhonov fait varier l'importance de la contribution de chaque valeur singulière. Lorsque les valeurs singulières diminuent, les facteurs de filtrage correspondant convergent vers zéro. Ainsi les valeurs singulières dont la valeur numérique est faible sont filtrées. Par contre, les facteurs de filtrage correspondant aux valeurs singulières dont la valeur numérique est grande sont faiblement modifiées, ce qui est souhaitable étant donné que ce sont les valeurs les plus représentatives. Ainsi les valeurs λ (ou μ) doivent être choisies avec soin en fonction des valeurs singulières de $\mathbf{\tilde{E}_{n-1}}$ (ou $\mathbf{\tilde{H}_n}$). De plus, les paramètres de régularisation doivent être déterminés de telle sorte que les valeurs singulières respectant la condition de Picard discrète ne soit pas filtrées. C'est-à-dire que λ et μ sont très petits vis-à-vis des valeurs singulières respectant les conditions de Picard discrètes et très grand vis-à-vis des valeurs singulières ne respectant les conditions de Picard discrètes. En ce qui concerne la condition de Picard discrète, les paramètres de régularisation doivent être choisis de telle sorte que chaque valeur unique qui répond à la condition de Picard discrète soit plus grande que les paramètres de régularisation.

Une méthode d'analyse pratique et efficace pour déterminer les paramètres de régularisation est l'utilisation de la courbe en L [76] qui est une courbe log–log de la norme $\ell 2$ de $\|\mathbf{\tilde{h}}\|_\mathbf{2}$ (ou $\|\mathbf{\tilde{e}}\|_\mathbf{2}$) par rapport à la norme résiduelle correspondante $\|\mathbf{s} - \mathbf{\tilde{E}_{n-1}\tilde{h}_n}\|_2$ (ou $\|\mathbf{s} - \mathbf{\tilde{H}_{n-1}\tilde{e}_n}\|_2$). Cette courbe a en général l'allure de la figure 6.17, ce qui explique son nom de courbe en L.

La partie horizontale correspond à la domination des erreurs de régularisation. Si λ (ou μ) est trop grand, la solution est aplanie. La régularisation est devenue trop importante. La partie non horizontale correspond à la domination de l'erreur de perturbation. Si λ (ou μ) est trop faible, la solution est oscillante. L'erreur ne diminue plus et la solution est instable. Le coin de la courbe en L définit la valeur optimale de λ (ou μ) et semble être un bon compromis entre la norme de la solution régularisée et la norme résiduelle. Pour construire cette courbe, l'estimation doit être faite pour tous les paramètres de régularisation valides.

6.3.4 Estimation de la signature « réelle »

Pour estimer la signature « réelle », deux hypothèses ont été effectuées sur le signal d'entrée initial. Dans le premier cas, le véhicule est considéré comme une masse métallique rectangulaire uniforme. Ainsi, le signal d'entrée est un signal rectangulaire unitaire de la longueur du véhicule. Dans le second cas, J.P. Sebastiá et coauteurs [77] énoncent l'hypothèse d'une concentration de la masse métallique au niveau des essieux avant et arrière du véhicule. Le signal d'entrée est représenté par la somme

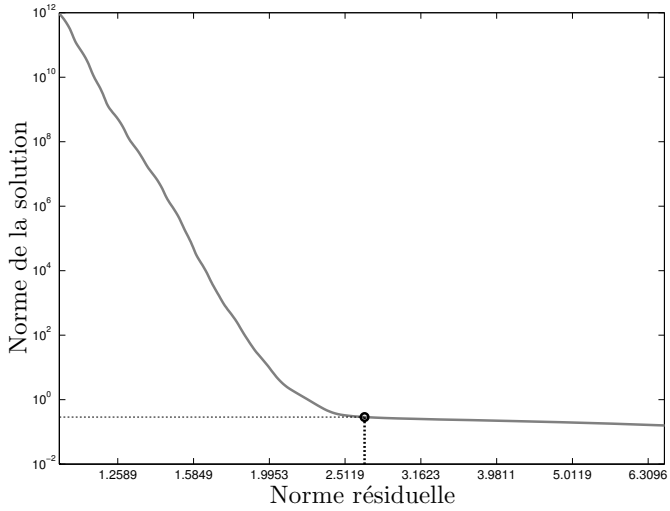


FIGURE 6.17 – Courbe en L : $\|\tilde{\mathbf{h}}\|_{\mathbf{2}}$ versus $\|\mathbf{s} - \tilde{\mathbf{E}}_{\mathbf{n-1}}\tilde{\mathbf{h}}_{\mathbf{n}}\|_2$.

de deux courbes gaussiennes et défini par l'équation (6.39). Pour différents types de véhicules, la concentration métallique n'est pas toujours la même et n'est pas toujours à la même place. Pour prendre en compte ces différences, l'équation comporte quatre paramètres. x_1 et x_2 représentent les deux emplacements de la concentration métallique, ϵ_1 et ϵ_2 représentent la largeur d'influence de la concentration métallique.

$$e(x) = e^{\frac{(x-x_1)^2}{(\epsilon_1)^2}} + e^{\frac{(x-x_2)^2}{(\epsilon_2)^2}} \tag{6.39}$$

L'ensemble du processus itératif a été appliqué pour les deux hypothèses sur le signal d'entrée. Comme l'hypothèse de la somme de deux distributions gaussiennes a donné de meilleurs résultats, seule cette hypothèse pour le signal d'entrée est retenue pour le reste de l'expérimentation.

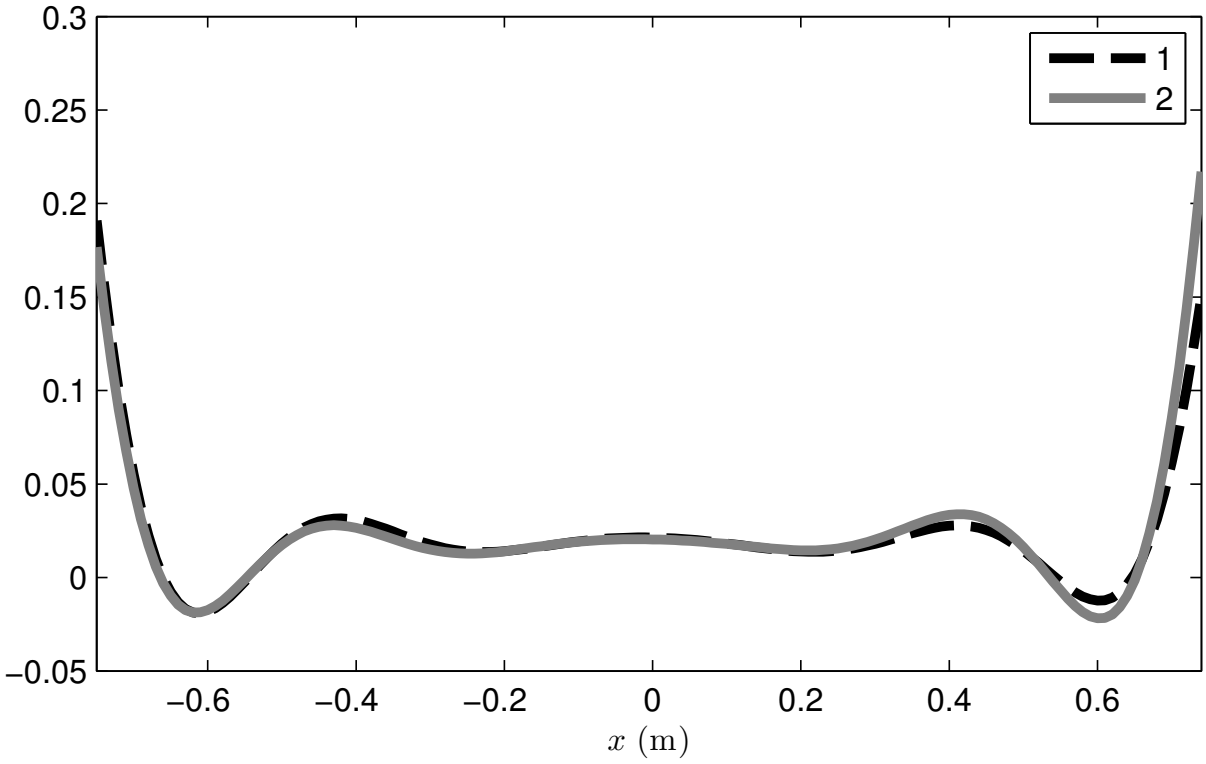


FIGURE 6.18 – Fonctions transferts estimées pour 2 boucles différentes et 2 véhicules différents.

À partir de l'équation (6.35), la fonction de transfert $\tilde{\mathbf{h}}$ est évaluée. La figure 6.18 montre deux

fonctions de transfert estimées par l'algorithme de déconvolution proposé. La courbe numéro 1 représente la fonction de transfert estimée $\tilde{\mathbf{h}}$ d'une boucle inductive pour la voiture de tourisme avec remorque de la figure 6.19. La courbe numéro 2 représente la fonction de transfert estimée $\tilde{\mathbf{h}}$ d'une autre boucle inductive pour la voiture de tourisme de la figure 6.20. La forme des fonctions de transfert estimées pour les deux capteurs est quasiment similaire alors que les véhicules sont différents. En effet, pour des vitesses de passage différentes, des véhicules différents et des boucles inductives différentes mais répondant aux mêmes caractéristiques techniques, la fonction de transfert estimée reste toujours approximativement la même. En conséquence, la moyenne des fonctions de transfert estimées a été calculée à partir d'une base de données d'apprentissage et sera comparée vis-à-vis des fonctions de transfert estimées à chaque passage de véhicules.

FIGURE 6.19 – Voiture de tourisme avec remorque « 1 » sur le capteur UD1J (vitesse estimée $76\,\mathrm{km\,h^{-1}}$).

Sur la figure 6.21, les signatures des véhicules « 1 » et « 2 » sans déconvolution sont représentées par les courbes en discontinu. Pour le véhicule « 2 », la signature sans déconvolution ne comporte qu'un seul pic alors que sur la signature estimée deux pics sont clairement visibles. Les deux signatures estimées après déconvolution des véhicules semblent offrir plus de détails.

6.3.5 Conclusion

Nous avons démontré que la déconvolution aveugle proposée converge vers une solution pour l'estimation de la fonction de transfert de la boucle inductive. Cette convergence nous permet d'estimer lors d'une phase d'apprentissage la fonction de transfert et ainsi ne plus avoir à calculer cette fonction pour chaque type de véhicules. La signature déconvoluée semble comporter des informations supplémentaires par rapport à la signature non déconvoluée. Pour valider cette hypothèse, une expérimentation est menée et décrite dans la section 8.2.3 à la page 147.

6.4 Conclusion

Parmi les prétraitements, nous avons identifié ceux concernant la compensation de la déformation temporelle et en amplitude des signaux, ceux pour la réduction du nombre de données et plus

FIGURE 6.20 – Voiture de tourisme « 2 » sur le capteur UD1k (vitesse estimée 116 km h^{-1}).

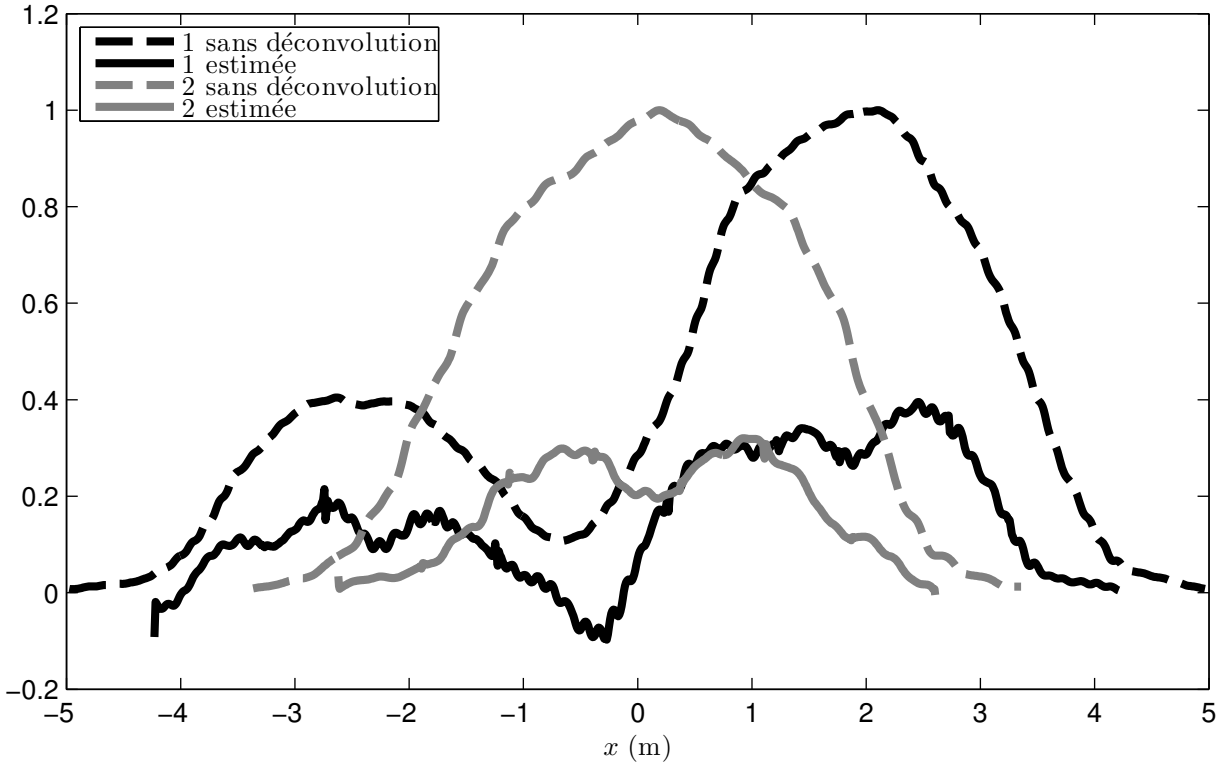


FIGURE 6.21 – Signatures sans et avec déconvolution pour les voitures de tourisme « 1 » et « 2 ».

spécifiquement pour la boucle inductive la déconvolution. Nous avons proposé lors de la réduction du nombre de données de prendre en considération le type de population. La prise en compte du type de population donne des sélections de variables différentes pour les différentes populations. Les variables ainsi sélectionnées ont pour objectif de mieux caractériser les véhicules de chaque population.

La déconvolution élaborée par Kwon nécessite de calculer les paramètres pour déconvoluer le signal au passage de chaque véhicule. Nous avons proposé une nouvelle méthode permettant de calculer les paramètres lors d'une phase d'apprentissage et d'utiliser directement ces paramètres lors de chaque passage de véhicules pour obtenir le signal déconvolué. La déconvolution du signal devrait améliorer la réidentification des véhicules.

7

Réidentification

Sommaire

7.1 Présentation **99**
7.1.1 Méthodes sans apprentissage 100
7.1.2 Méthodes avec apprentissage 102
7.2 Propositions de classifieurs **105**
7.2.1 Distance de Canberra 105
7.2.2 Séparateur à vaste marge 105
7.3 Détermination d'un seuil et utilisation de plusieurs méthodes **111**
7.3.1 Introduction 111
7.3.2 Étude sur une base idéale 111
7.3.3 Étude sur une base perturbée 117
7.3.4 Conclusion 125
7.4 Nouvelle méthode d'évaluation **125**
7.4.1 Contexte 125
7.4.2 Classification et comparaisons 125
7.4.3 Réidentification et comparaisons 131
7.5 Conclusion **137**

7.1 Présentation

La réidentification de véhicule consiste à comparer des caractéristiques (la signature par exemple) d'un véhicule destination à celles d'un véhicule origine. C'est un problème de classification représenté par deux classes. Si la signature destination et la signature origine appartiennent au même véhicule, la classe est considérée comme identique. Lorsque les deux signatures n'appartiennent pas au même véhicule, la classe est considérée comme différente. La prise de décision pour établir si le véhicule origine est identique à celui de destination est déterminée par un classifieur. En réalité, la signature d'un véhicule pour la comparaison n'est pas forcément l'ensemble des points de mesure mais un

ensemble de variables soit issu directement de la mesure, soit calculé à partir de la mesure. Le vecteur $\mathbf{c}$ est la variable classe avec les deux valeurs {identique, différent}. Les variables d'un véhicule origine k sont représentées par le vecteur $\mathbf{x}_{o,k}$, celles d'un véhicule destination l sont représentées par le vecteur $\mathbf{x}_{d,l}$. La réidentification se différencie de la classification dans le sens ou au plus un seul candidat est identique. En effet, lorsqu'il s'agit de classifier une signature destination par rapport à plusieurs signatures origine, la classification pourra classer plusieurs candidats comme identiques alors que la réidentification ne donnera au plus qu'un candidat comme identique. Pour la réidentification, deux approches sont couramment utilisées : sans apprentissage et avec apprentissage.

7.1.1 Méthodes sans apprentissage

Les méthodes sans apprentissage ont l'avantage d'être utilisables immédiatement sans connaissance *a priori*. L'une des méthodes les plus simples, d'après [33, 43, 44] est l'utilisation d'un classifieur basé sur la distance minimale. La référence à la distance est un abus de langage. Elle ne représente pas forcément la distance au sens mathématique du terme mais la mesure de similarité. Ce classifieur a l'avantage de s'adapter facilement à la réidentification. Dans le cas des boucles inductives, plusieurs distances ont été utilisées par [33]. En reprenant les notations $\mathbf{x}_o$, $\mathbf{x}_d$, pour les variables origine et destination, respectivement, et p représentant le nombre de variables, les différentes distances sont :

- La distance de Manhattan (appelée distance euclidienne dans [33, 34]) $d_M(\mathbf{x}_d,\mathbf{x}_o)$:

$$d_M(\mathbf{x}_d,\mathbf{x}_o) = \sum_{i=1}^{p} |x_{d,i} - x_{o,i}| \tag{7.1}$$

  La distance de Manhattan est la distance associée à la norme $\ell 1$.

- Le coefficient de corrélation $r(\mathbf{x}_d,\mathbf{x}_o)$:

$$r(\mathbf{x}_d,\mathbf{x}_o) = \frac{\sum_{i=1}^{p}(x_{o,i} - \bar{x}_o)(x_{d,i} - \bar{x}_d)}{\sqrt{\sum_{i=1}^{p}(x_{o,i} - \bar{x}_o)^2}\sqrt{\sum_{i=1}^{p}(x_{d,i} - \bar{x}_d)^2}} \tag{7.2}$$

  avec $\bar{x}_o$ et $\bar{x}_d$ les moyennes des vecteurs $\mathbf{x}_o$ et $\mathbf{x}_d$, respectivement. La dénomination coefficient de corrélation est employée parfois pour désigner la valeur $1 - r(\mathbf{x}_d,\mathbf{x}_o)$: cette opération a pour objectif d'obtenir une valeur proche de zéro pour une grande similarité.

- La distance de Tanimoto (appelée distance de similarité dans [33, 34]) $d_T(\mathbf{x}_d,\mathbf{x}_o)$:

$$d_T(\mathbf{x}_d,\mathbf{x}_o) = 1 - \frac{\langle \mathbf{x}_d,\mathbf{x}_o \rangle}{||\mathbf{x}_d||_2^2 + ||\mathbf{x}_o||_2^2 - \langle \mathbf{x}_d,\mathbf{x}_o \rangle} \tag{7.3}$$

  avec $\langle \cdot,\cdot \rangle$ le produit scalaire. Elle ressemble à un produit scalaire de vecteur normalisé. Elle représente une mesure analogue à celle d'un déplacement angulaire entre les variables origine et destination.

- La distance de Lebesgue $d_L(\mathbf{x_d},\mathbf{x_o})$: cette distance nécessite d'interpoler et de rééchantillonner le signal selon l'axe des amplitudes nommé y.

$$d_L(\mathbf{x_d},\mathbf{x_o}) = \sum_{k=1}^{N_1}\left(\delta y_k - \overline{\Delta y}\right)^2 + \cdots + \sum_{k=N_{q-1}}^{N_q}\left(\delta y_k - \overline{\Delta y}\right)^2 \tag{7.4}$$

  avec $\delta y_k = y_{d,k} - y_{o,k}$, $y_{d,k}$ (ou $y_{o,k}$) les amplitudes de x_d (ou x_o) selon l'axe des ordonnées y, N est le nombre de points de chaque segment, q est le nombre de segments, et $\overline{\Delta y}$ la distance horizontale moyenne entre les deux signaux. Cette technique permet de mesurer un déplacement horizontal.

En plus de ces distances, Tabib [34] a proposé la distance euclidienne :

- La distance Euclidienne $d_E(\mathbf{x_d},\mathbf{x_o})$:

$$d_E(\mathbf{x_d},\mathbf{x_o}) = \sqrt{\frac{1}{p}\sum_{i=1}^{p}(x_{d,i} - x_{o,i})^2} \tag{7.5}$$

La distance euclidienne est la distance associée à la norme $\ell 2$.

Ritchie et coauteurs dans [33] ont utilisé les distances de Manhattan, de Tanimoto, du coefficient de corrélation et de Lebesgue sur la signature du véhicule. Tabib dans [34] a utilisé les mêmes distances avec en plus la distance euclidienne. Ces distances ont été calculées à partir de la signature. Tabib a aussi calculé ces distances sur la dérivée première. L'utilisation de la dérivée première met en évidence les minima et les maxima d'une signature. Il a également calculé ces distances sur la représentation fréquentielle de la signature après avoir effectué une transformée de Fourier. Selon [34], la distance de Manhattan sur la dérivée première selon l'axe des ordonnées et la distance de Lebesgue donnent les TBR (cf. définition page 53) de 56,6 % et 53,3 %, respectivement. En multipliant l'ensemble des distances qu'il a calculées entre-elles, le TBR est de 55 %. En sélectionnant les distances de Manhattan pour la dérivée première selon l'axe des abscisses et des ordonnées ainsi que leurs variances, et la distance de Lebesgue, puis en les multipliant entre-elles, le TBR est de 62,3 %.

Jeng et coauteurs dans [53, 54, 57, 58, 78] ont utilisé la distance de Manhattan sur la dérivée première. Dans [58] le taux de bonne réidentification varie entre 75 % et 79 % pour des conditions de trafic fluide et entre 52 % et 79 % pour des conditions de trafic congestionné.

Dans le cadre des magnétomètres, les variables sont issues des trois axes du capteur. De plus lors d'expérimentations, il est possible de placer plusieurs magnétomètres en ligne au travers de la voie selon l'axe y (comme le montre la figure 3.14). Ainsi Cheung et coauteurs dans [14] proposent d'utiliser le produit des coefficients de corrélation, qui est calculé comme suit :

$$r(\mathbf{S_d},\mathbf{S_o}) = \prod_{j\in\{x,y,z\}} \left| \frac{\sum_{i=1}^{p}(S_{o,j,i} - \bar{S}_{o,j})(S_{d,j,i} - \bar{S}_{d,j})}{\sqrt{\sum_{i=1}^{p}(S_{o,j,i} - \bar{S}_{o,j})^2}\sqrt{\sum_{i=1}^{p}(S_{d,j,i} - \bar{S}_{d,j})^2}} \right| \tag{7.6}$$

avec $i \in [1;p]$, p étant le nombre de variables de l'axe, $j \in \{x,y,z\}$ et o précisant que le capteur est à l'origine, un capteur à la destination sera noté avec d. Étant donné que plusieurs capteurs sont placés à l'origine comme à la destination, seul le coefficient de corrélation maximal est conservé :

$$r_{max}(\mathbf{S_d},\mathbf{S_o}) = \max\{r(\mathbf{S_{d,m}},\mathbf{S_{o,n}}) \text{ avec } \forall m \in [1;k] \text{ et } \forall n \in [1;k]\} \tag{7.7}$$

avec k le nombre de capteurs employés. Dans le cas de [14], sept capteurs par site sont utilisés. $\mathbf{S_{d,m}}$ (ou $\mathbf{S_{o,n}}$) représente l'un des capteurs à la destination (à l'origine, respectivement). Si le coefficient $r_{max}(\mathbf{S_d},\mathbf{S_o})$ est supérieur à un seuil r_{seuil} alors le véhicule est considéré comme identique entre l'origine et la destination. Si deux signatures origine ou plus sont considérées comme identiques à une même signature destination, alors celle dont la somme des coefficients de corrélation (7.8) est maximale, est retenue.

$$r_{max_s} = \sum_{m=1}^{k}\sum_{n=1}^{k} r(\mathbf{S_{d,m}},\mathbf{S_{o,n}}) \tag{7.8}$$

Cheung et coauteurs [14] décrivent deux expérimentations utilisant ce processus. Une première expérimentation est réalisée avec sept véhicules. Chaque véhicule a effectué cinq passages et certains passages ont été volontairement décalés du centre de la voie. Les passages ont tous été effectués sur la même ligne de capteurs. Les signatures relevées ont été compressées selon 20 sections moyennées comme le montre la figure 6.2. Le taux de réidentification obtenu est de 98,9 %, ce résultat démontre la faisabilité du suivi de véhicules par magnétomètre. La deuxième expérimentation est réalisée sur

quatre heures et 80 véhicules sont passés sur un réseau de sept capteurs à l'origine et un autre réseau de sept capteurs à la destination. Les signatures relevées ont été compressées, cette fois-ci, selon 10 sections moyennées. Nous supposons que le choix des auteurs de compresser plus fortement la signature est réalisé pour limiter les calculs. Dans cette deuxième expérimentation, le taux de réidentification obtenu est de 72,5 %.

En 2012, Sanchez [79] ainsi que Charbonnier et coauteurs [80] utilisent la déformation temporelle dynamique (DTW : Dynamic Time Warping en anglais) comme mesure de dissimilarité. Lors de leurs passages sur les différents capteurs, les véhicules peuvent avoir des vitesses différentes conduisant à des distorsions temporelles entre deux signatures du même véhicule. L'algorithme de DTW vise à aligner les déformations de deux signaux en fonction de l'axe temporel, il est donc parfaitement adapté pour la comparaison des signatures. Pour aligner une signature origine de longueur n avec une signature destination de longueur m, une matrice $n \times m$ est construite. L'élément (i,j) de la matrice correspond à la distance entre le i^{e} point de la signature origine et j^{e} point de la signature destination. Pour trouver la meilleure adéquation entre les deux signatures, un chemin à travers la matrice qui minimise la distance totale cumulée entre les deux signatures est calculé. Charbonnier et coauteurs [80] ont considéré le cas unidimensionnel selon chaque axe x, y et z avec la distance euclidienne :

$$D^2(i,j) = \left(S_{d,a}(i) - S_{o,a}(j)\right)^2 \text{ avec } a \in \{x,y,z\},\ \forall i \in [1,n],\ \forall j \in [1,m] \tag{7.9}$$

ainsi que le cas tridimensionnel :

$$D^2(i,j) = \frac{\left(S_{d,x}(i) - S_{o,x}(j)\right)^2 + \left(S_{d,y}(i) - S_{o,y}(j)\right)^2 + \left(S_{d,z}(i) - S_{o,z}(j)\right)^2}{3} \ \forall i \in [1,n],\ \forall j \in [1,m] \tag{7.10}$$

Le cas tridimensionnel permet de prendre en compte les liens entre les trois champs magnétiques dans le calcul de la distance. Finalement, chaque distance est normalisée en divisant la distance obtenue par l'amplitude maximale crête à crête des deux signatures. Sanchez [79] contrairement à Charbonnier et coauteurs ne donne aucune information sur la distance calculée. La DTW nécessite un temps de calcul important. Ainsi, Charbonnier et coauteurs sous-échantillonnent le signal en prenant la moyenne de dix points consécutifs, alors que Sanchez [79] ne prend que le premier et le dernier point ainsi que l'ensemble des extremums de la signature. Charbonnier et coauteurs démontrent que l'utilisation du signal tridimensionnel permet d'avoir un taux réidentification de 78 % sans commettre d'erreur de réidentification alors qu'en unidimensionnel le taux de réidentification est de 50 %. En analysant les courbes ROC publiées dans [80], les meilleurs résultats en unidimensionnel sont obtenus selon l'axe y alors que [14, 63, 64, 79, 81] accordent moins d'importance à cet axe et le jugent plutôt pénalisant pour la réidentification. D'après Charbonnier et coauteurs, le principal inconvénient de la DTW reste le temps de calcul. Pour Sanchez [79] le calcul de la DTW est une étape intermédiaire avant l'utilisation d'un classifieur.

7.1.2 Méthodes avec apprentissage

Dans le cadre des méthodes avec apprentissage, une mesure de dissimilarité est calculée entre les caractéristiques origine et les caractéristiques destination. Les mesures de dissimilarité sont définies dans la section précédente 7.1.1 et sont par exemple : la distance de Manhattan, la distance de Tanimoto, la distance de Lebesgue, la distance Euclidienne, le coefficient de corrélation... La mesure de dissimilarité est représentée par $\mathbf{v_i}$. $\mathbf{v_i} = (v_1,...,v_p)$ est le i^{e} vecteur des variables attributs pour le couple origine – destination (l,k) avec p le nombre de variables utilisées et pourra aussi s'écrire $\mathbf{v}(l,k)$. Ces algorithmes nécessitent une base de données d'apprentissage pour déterminer les paramètres intrinsèques de l'algorithme. Ensuite les algorithmes pourront prédire la classe d'une nouvelle donnée observée.

Méthode de logique floue

Le passage des véhicules sur les capteurs engendre des imprécisions dans le recueil des données. En effet, les véhicules ne passent pas tous au même endroit sur les boucles inductives, ce qui a une incidence sur l'amplitude des signatures. De plus, les capteurs mesurent la vitesse avec 10 % d'erreur, ce qui modifie la longueur des signatures. Pour tenir compte de ces incertitudes et imprécisions des données recueillies, une méthode basée sur la logique floue est développée dans [71]. De cette façon, aucune hypothèse sur la nature de la distribution des variables aléatoires n'est nécessaire, cependant des sous-ensembles flous pour ce problème doivent être définis.

Une fonction d'appartenance trapèze est choisie car elle est fréquemment utilisée dans la littérature [71]. Soit un sous-ensemble $E_{l,k}$ flou, cette fonction d'appartenance $f_{E_{l,k}}$ s'écrit :

$$f_{E_{l,k}}(v_{i,j}) = \begin{cases} 0 & \text{si } v_{i,j} \notin [a_j - \alpha_j, b_j + \alpha_j], \\ 1 & \text{si } v_{i,j} \in [a_j, b_j], \\ 1 + \frac{v_{i,j} - a_j}{\alpha_j} & \text{si } v_{i,j} \in]a_j - \alpha_j, a_j[, \\ 1 + \frac{b_j - v_{i,j}}{\alpha_j} & \text{si } v_{i,j} \in]b_j, b_j + \alpha_j[. \end{cases} \tag{7.11}$$

avec :

$$\begin{aligned} v_{i,j} \quad & j \in [1,p] \\ a_j &= m_k - p_1 \times \sigma_j \\ b_j &= m_k + p_1 \times \sigma_j \\ \alpha_j &= p_2 \times \sigma_j \end{aligned} \tag{7.12}$$

Les paramètres m_k et σ_k sont obtenus à l'aide d'un apprentissage supervisé, les paramètres p_1 et p_2 sont calculés empiriquement dans [71] et sont respectivement égaux à 0 et 15. Dans notre cas, et toujours selon [71], $a_k = b_k$ donc la fonction d'appartenance est triangulaire centrée sur la moyenne. La fonction discriminante est définie par :

$$F_i(\mathbf{v_i}) = \sum_{j=1}^{p} f_{E_{l,k}}(v_{i,j}) \tag{7.13}$$

Une valeur qui tend vers p, le nombre de variables d'entrée, traduit une grande similitude entre la signature origine et destination. Une expérimentation menée par Ieng et coauteurs [59, 60] a été réalisée sur des boucles inductives. Les boucles inductives à l'origine et celles à destination sont séparées de 15 m. Le taux de bonne réidentification obtenu, à partir d'une sélection de 12 variables, est de 94,8 %.

Approche Bayésienne

Ce classifieur est utilisé par Ieng et coauteurs [59, 60] pour le capteur boucle inductive, et par Kwong et coauteurs [63, 64], Sanchez et coauteurs [79, 81, 82] et par Charbonnier et coauteurs [80] avec le capteur magnétomètre.

Ieng et coauteurs [59, 60] ont émis les hypothèses suivantes : pour chaque classe c_j (la classe « identique » et la classe « différent »), la loi de probabilité est gaussienne ; chaque variable sachant la classe correspondante est indépendante. Nous avons alors $P(\mathbf{v}_i|c_j)$ qui suit une loi gaussienne de moyenne $\mu_\mathbf{j}$ et d'écart-type $\sigma_\mathbf{j}$. La matrice de covariance est $\mathbf{\Sigma_j}$. Donc $P(\mathbf{v}_i|c_j)$ s'écrit comme suit :

$$P(\mathbf{v}_i|c_j) = \frac{1}{\sqrt{2\pi}|\mathbf{\Sigma_j}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{v}_i - \boldsymbol{\mu_j})^T \mathbf{\Sigma_j}^{-1}(\mathbf{v}_i - \boldsymbol{\mu_j})\right) \tag{7.14}$$

Les paramètres sont estimés par un apprentissage supervisé. L'objectif est de déterminer la meilleure classification possible des variables $\mathbf{v}_i$. Ce qui revient à :

$$c_{opt}(\mathbf{v}_i) = \underset{j}{\text{argmax}}\,(P(c_j|\mathbf{v}_i)) \tag{7.15}$$

D'après le théorème de Bayes, nous avons :

$$P(c_j|\mathbf{v}_i) = \frac{P(\mathbf{v}_i|c_j)P(c_j)}{P(\mathbf{v}_i)} \tag{7.16}$$

$P(\mathbf{v}_i|c_j)$ représente la vraisemblance, $P(c_j|\mathbf{v}_i)$ la connaissance *a posteriori*, $P(c_j)$ et $P(\mathbf{v}_i)$ les connaissances *a priori*.

$$c_{opt}(\mathbf{v}_i) = \underset{j}{\operatorname{argmax}} \left(\frac{P(\mathbf{v}_i|c_j)P(c_j)}{P(\mathbf{v}_i)} \right) \tag{7.17}$$

Par déduction, il faut maximiser $P(\mathbf{v}_i|c_j)P(c_j)$ car $P(\mathbf{v}_i)$ est une constante de normalisation et n'intervient pas dans la décision.

$$c_{opt}(\mathbf{v}_i) = \underset{j}{\operatorname{argmax}} \left(P(\mathbf{v}_i|c_j)P(c_j) \right) \tag{7.18}$$

Suite à l'hypothèse d'indépendance des variables et sous l'hypothèse gaussienne et le fait d'avoir deux classes, l'équation (7.18) devient :

$$c_{opt}(\mathbf{v}_i) = \underset{j}{\operatorname{argmax}} \left(\frac{exp\left(-\frac{1}{2}\left(\mathbf{v}_i - \boldsymbol{\mu_j}\right)^T \boldsymbol{\Sigma_j}^{-1} \left(\mathbf{v}_i - \boldsymbol{\mu_j}\right)\right) P(c_j)}{\sqrt{2\pi}|\boldsymbol{\Sigma_j}|^{1/2}} \right) \tag{7.19}$$

Pour des raisons pratiques, nous préférons calculer le logarithme de l'expression (7.19) soit :

$$\ln\left(c_{opt}(\mathbf{v}_i)\right) = \underset{j}{\operatorname{argmax}} \left(-\frac{1}{2}\left(\mathbf{v}_i - \boldsymbol{\mu_j}\right)^T \boldsymbol{\Sigma_j}^{-1} \left(\mathbf{v}_i - \boldsymbol{\mu_j}\right) + \ln\left(P(c_j)\right) - \frac{1}{2}\left(\ln(2\pi) + \ln(|\boldsymbol{\Sigma_j}|)\right) \right) \tag{7.20}$$

Ieng et coauteurs [59, 60] ne donnent aucune information sur les connaissances *a priori* des probabilités relatives d'occurrences des classes c_j. Une première approche consiste à considérer que les probabilités d'occurrences des classes sont équiprobables. La fonction coût pour la classe « identique » est la suivante :

$$g_i(\mathbf{v}_i) = -\frac{1}{2}\left(\mathbf{v}_i - \boldsymbol{\mu_j}\right)^T \boldsymbol{\Sigma_j}^{-1} \left(\mathbf{v}_i - \boldsymbol{\mu_j}\right) \tag{7.21}$$

Cette approche est appelée maximum de vraisemblance (MV) par la suite.

La seconde approche consiste à considérer que la probabilité d'occurrence de la classe « identique » n'est pas équiprobable. Cette probabilité est déterminée par les conditions de circulation et la typologie des lieux. Un temps moyen t_m est estimé pour parcourir la distance entre l'origine et la destination. Si la différence temporelle t_{diff} entre l'heure d'acquisition de la signature origine et l'heure d'acquisition de la signature destination est proche de t_m alors la probabilité d'occurrence de la classe « identique » est importante. Par contre, plus la différence temporelle est éloignée de la valeur de t, plus la probabilité d'occurrence de la classe « identique » est faible. Nous émettons l'hypothèse que la probabilité d'occurrence de la classe « identique » suit une loi gaussienne :

$$P(c_j) = \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{t_{diff} - t_m}{\sigma_t} \right)^2 \right) \tag{7.22}$$

avec σ_t représentant l'écart-type des temps de parcours entre l'origine et la destination. La fonction coût pour la classe « identique » est la suivante :

$$g_i(\mathbf{v}_i) = -\frac{1}{2}\left(\mathbf{v}_i - \boldsymbol{\mu_j}\right)^T \boldsymbol{\Sigma_j}^{-1} \left(\mathbf{v}_i - \boldsymbol{\mu_j}\right) + \ln\left(P(c_j)\right) \tag{7.23}$$

Cette approche est appelée MV2 par la suite.

Que ce soit pour la première ou la seconde approche, la fonction coût est calculée entre une signature destination et m signatures origine. En reprenant la notation $\mathbf{v}_i = \mathbf{v}(l,k)$ pour désigner la

dissimilarité entre les caractéristiques du véhicule origine l et celles du véhicule destination k, nous obtenons l'équation suivante :

$$g_{opt}(\mathbf{v}(l,k)) = \underset{l\in[0,m-1]}{\operatorname{argmax}}\, g_{(l,k}(\mathbf{v}(l,k)) \tag{7.24}$$

Ainsi, la réidentification revient à chercher le maximum de la fonction coût ou le minimum de la fonction risque $r_{opt} = -g_{opt}$. Une expérimentation menée par Ieng et coauteurs [59, 60] est réalisée sur des boucles inductives. Les boucles inductives à l'origine et celles à destination sont séparées de 15 m. Ieng et coauteurs ne donnent aucune information sur les connaissances *a priori* des probabilités relatives d'occurrences de cette expérimentation. Le taux de bonne réidentification obtenu, à partir d'une sélection de 12 variables est de 93,3 %.

Charbonnier et coauteurs [80] ont extrait plusieurs variables indépendantes du temps à partir de chaque signal unidimensionnel x, y et z. Les trois variables suivantes ont été sélectionnées parmi les quinze variables initiales : l'asymétrie selon l'axe x, le maximum et l'entropie selon l'axe y. v_i correspond à la distance de Manhattan de ces trois variables. Les auteurs ont calculé les fonctions de densité de probabilité des classes « identique » et « différent ». Ils ont considéré ces fonctions comme des gaussiennes. Ils ont ainsi pu calculer la probabilité pour chaque nouveau couple origine – destination d'appartenir à la classe « identique ». Le taux de réidentification obtenu est de 30 % avec 15 % d'erreur de réidentification. Ce classifieur est utilisé ensuite pour filtrer le nombre de candidats avant l'utilisation de la mesure de dissimilarité par la déformation temporelle dynamique tridimensionnelle. 80 % des candidats sont filtrés engendrant un temps de calcul moindre pour la DTW et le taux de réidentification est de 80 % avec 2 % d'erreur de réidentification.

7.2 Propositions de classifieurs

7.2.1 Distance de Canberra

Une nouvelle mesure de dissimilarité est proposée avec la distance de Canberra $d_C(\mathbf{x_d},\mathbf{x_o})$:

$$d_C(\mathbf{x}_d,\mathbf{x}_o) = \sum_{i=1}^{p} \frac{|x_{d,i} - x_{o,i}|}{|x_{d,i}| + |x_{o,i}|} \tag{7.25}$$

avec $\mathbf{x}_d$ et $\mathbf{x}_o$ les caractéristiques des véhicules destination et origine, respectivement et p le nombre de caractéristiques représentant les véhicules.

7.2.2 Séparateur à vaste marge

Les Séparateurs à Vaste Marge (ou en anglais Support Vector Machine, SVM) sont des méthodes de classification qui furent introduites par V.Vapnik en 1995 dans son ouvrage « The nature of statistical learning theory » [83]. Ces méthodes sont initialement définies comme des algorithmes d'apprentissage pour la discrimination et ont ensuite été généralisées à la prévision d'une variable quantitative. Dans le cas d'une variable binaire, elles sont basées sur la recherche de l'hyperplan de marge optimale qui, lorsque c'est possible, classe ou sépare correctement les données tout en étant le plus éloigné possible de toutes les observations. Comme pour la présentation des SVM par Boubezoul [84] et par les auteurs du cours de l'Institut National des Sciences Appliquées de Toulouse [**baccini_wikistat_2014**], nous commençons par aborder les SVM dans le cadre de données linéairement séparables pour ensuite généraliser progressivement son application aux cas des données non linéaires et non séparables.

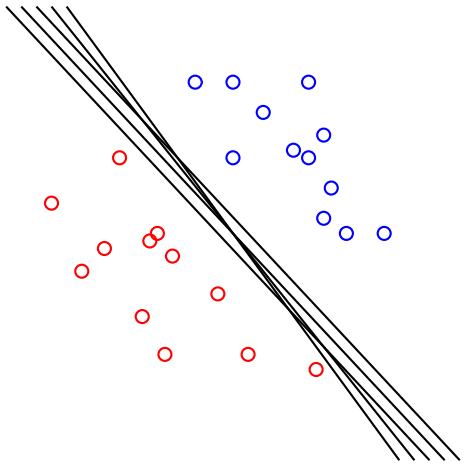


FIGURE 7.1 – Exemple d'un problème de discrimination binaire linéairement séparable où il s'agit de séparer les points bleus des points rouges. La frontière de décision est représentée en noir, cependant plusieurs solutions sont possibles.

Problème de discrimination binaire et classifieur linéaire pour les données séparables

La discrimination consiste à trouver une fonction de décision associant à chaque observation sa classe. Soient les vecteurs $\mathbf{x} \in \mathbb{R}^p$ les observations à discriminer et y la classe à laquelle appartient l'observation, nous définissons l'ensemble S de N éléments tel que :

$$\mathcal{S} = \{(\mathbf{x}_i,y_i) \in \mathbb{R}^p \times \{-1,1\}\}_{i=1}^{N} \tag{7.26}$$

Un problème de discrimination est dit linéairement séparable lorsqu'il existe une fonction de décision linéaire de la forme $D(\mathbf{x}) = signe(f(\mathbf{x}))$ avec $f(\mathbf{x}) = \mathbf{v}^T\mathbf{x} + a$, $\mathbf{v} \in \mathbb{R}^p$ et $a \in \mathbb{R}$, classant correctement toutes les observations de l'ensemble d'apprentissage ($D(\mathbf{x}_i) = y_i, i \in [1,N]$). La définition consiste à dire qu'il doit exister un hyperplan permettant de séparer les exemples positifs des exemples négatifs. Cette frontière de décision est déterminée, à un terme multiplicatif près, par :

$$\Delta(\mathbf{v},a) = \{\mathbf{x} \in \mathbb{R}^p | \mathbf{v}^T\mathbf{x} + a = 0\} \tag{7.27}$$

La figure 7.1 montre l'exemple d'un problème binaire de discrimination linéairement séparable. Comme le représente la figure, il existe généralement une infinité de fonctions de décision permettant la séparation des deux classes.

Pour garantir l'unicité de la solution nous considérons soit l'hyperplan standard (tel que $||\mathbf{v}|| = 1$), soit l'hyperplan canonique par rapport à un point $\mathbf{x}$ (tel que $\mathbf{v}^T\mathbf{x} + a = 1$).

Pour un classifieur linéaire, deux notions de marge sont définies : sa marge numérique et sa marge géométrique. Elles sont représentées sur la figure 7.2. La marge géométrique m d'un échantillon x_i est la plus petite distance d'un point de l'échantillon à la frontière de décision.

$$m = \min_{i \in [1,N]} dist(\mathbf{x_i},\Delta(\mathbf{v},a)) \tag{7.28}$$

La marge numérique μ est donnée par la plus petite valeur de la fonction de décision atteinte sur un point de l'échantillon.

$$\mu = \min_{i \in [1,N]} |\mathbf{v}^T\mathbf{x}_i + a| \tag{7.29}$$

La recherche d'un hyperplan optimal consiste à maximiser la marge. Cette recherche s'écrit comme un problème d'optimisation sous contrainte. La formulation « classique » des SVM s'obtient alors en minimisant $||\mathbf{w}||^2$ après avoir effectué les changements de variables suivants : $\mathbf{w} = \frac{\mathbf{v}}{m||\mathbf{v}||}$ et $b = \frac{a}{m||\mathbf{v}||}$.

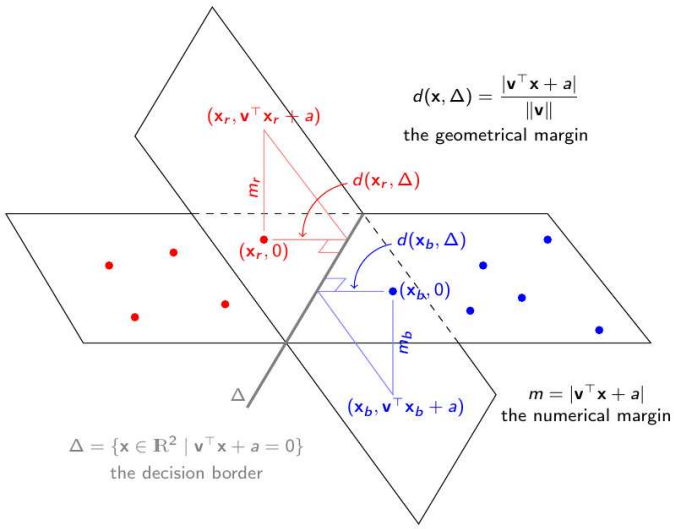


FIGURE 7.2 – Représentation des deux notions de marge sur un exemple de discrimination linéaire séparable en deux dimensions (extrait de [**baccini_wikistat_2014**]).

Soit $S = \{(\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \{-1,1\}\}_{i=1}^N$, un séparateur à vaste marge linéaire est un discriminateur linéaire de la forme $D(x) = signe(\mathbf{w}^T \mathbf{x}_i + b)$ où $\mathbf{w} \in \mathbb{R}^p$ et $b \in \mathbb{R}$ sont donnés par la résolution du problème suivant :

$$\begin{cases} \min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 \\ \text{avec } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geqslant 1 \,;\, i = 1,N \end{cases} \tag{7.30}$$

Il s'agit d'un problème quadratique sous contraintes linéaires dont la fonction objectif est à minimiser. Cette fonction objectif est le carré de l'inverse de la marge. L'unique contrainte se traduit par le fait que les exemples doivent être bien classés et qu'ils ne doivent pas dépasser les hyperplans canoniques. Ce problème d'optimisation sous contrainte est reformulé sous le formalisme lagrangien :

$$\mathcal{L}(\mathbf{w},b,\alpha) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i \left(y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \right) \tag{7.31}$$

où les $\alpha_i \geqslant 0$ sont les multiplicateurs de Lagrange associés aux contraintes.

Parmi les observations disponibles, les vecteurs supports sont recherchés. Les vecteurs supports sont les exemples les plus proches de la frontière de décision. La figure 7.3 illustre les observations qui sont appelées vecteur support.

Classifieur linéaire pour des données « non séparables »

Dans le cas pratique, les données utilisées sont souvent entachées de bruit (erreur d'acquisition, d'étiquetage...). Le classifieur linéaire est alors incapable de classer correctement toutes les données même si une relation linéaire existe entre les données et leurs classes. Il est quand même possible en suivant une démarche semblable au cas des données séparables d'établir une fonction de décision en relâchant les contraintes. Pour cela des variables d'écart, de ressort ou de relaxation notée ξ_i associées à chacune des observations sont introduites. La figure 7.4 représente la notion d'écart.

Nous avons donc les deux cas suivants :

$$\text{pas d'erreur : } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geqslant 1 \qquad \Rightarrow \xi_i = 0 \tag{7.32}$$

$$\text{erreur : } y_i(\mathbf{w}^T \mathbf{x}_i + b) < 1 \qquad \Rightarrow \xi_i = 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0 \tag{7.33}$$

Une fonction appelée coût charnière est définie comme suit à partir des deux équations précédentes :

$$\xi_i = max(0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)) \tag{7.34}$$

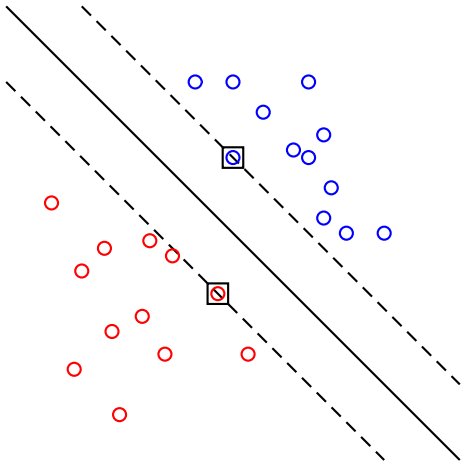


FIGURE 7.3 – Représentation des vecteurs supports dans le cas d'un problème binaire linéairement séparable. Les vecteurs supports des deux classes (points rouges et points bleus) sont encadrés. Les marges géométriques sont représentées en pointillé noir et la frontière de décision par un trait noir.

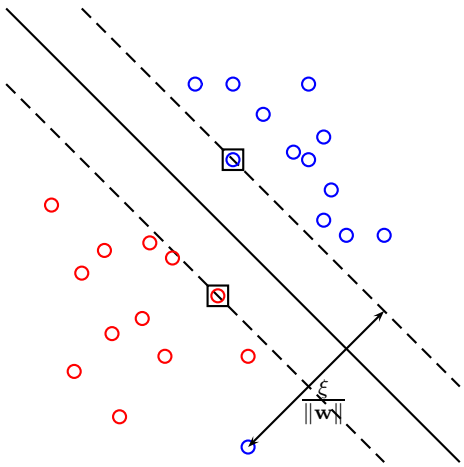


FIGURE 7.4 – Représentation de la notion d'écart : les points correctement classés ont un écart nul ; l'écart du point bleu mal classé se mesure entre lui et la marge numérique de l'hyperplan séparateur.

Le problème consiste à maximiser la marge et à minimiser la somme des termes d'erreur à la puissance d (avec typiquement $d = 1$ ou 2). Il s'écrit sous la forme suivante :

$$\begin{cases} \min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{d}\sum_{i=1}^{N}\xi_i^d \\ \text{avec : } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geqslant 1 - \xi_i \ , \ i = 1,\ldots,N \\ \xi_i \geqslant 0 \ , \ i = 1,\ldots,N \end{cases} \tag{7.35}$$

avec $C > 0$ une constante appelée terme d'équilibrage. Cette constante permet d'indiquer l'importance accordée aux erreurs commises. Le lagrangien du problème de minimisation est :

$$\mathcal{L}(\mathbf{w},b,\alpha,\beta) = \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{d}\sum_{i=1}^{N}\xi_i^d - \sum_{i=1}^{N}\alpha_i\left(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i\right) - \sum_{i=1}^{N}\beta_i\xi_i \tag{7.36}$$

avec des multiplicateurs de Lagrange $\alpha_i \geqslant 0$ et $\beta_i \geqslant 0$. Le calcul du gradient est identique par rapport à $\mathbf{w}$ et b. Dans le cas où $d = 1$, la dérivée partielle du lagrangien par rapport aux variables d'écart s'écrit : $\frac{\partial\mathcal{L}(\mathbf{w},b,\alpha,\beta)}{\partial_{\xi_i}} = C - \alpha_i - \beta_i = 0$. Cette condition de stationnarité permet d'éliminer les β_i car $\alpha_i \leqslant C$. Le problème dual s'exprime selon :

$$\begin{cases} \min_{\alpha\in\mathbb{R}^N} \frac{1}{2} & \alpha^T\mathbf{G}\alpha - \mathbf{e}^T\alpha \\ \text{avec :} & \mathbf{y}^T\alpha = 0 \\ & 0 \leqslant \alpha_i \leqslant C \ , \ i = 1,\ldots,N \end{cases} \tag{7.37}$$

avec $\mathbf{G}$ la matrice carrée de dimension N et de terme général $G_{ij} = y_i y_j \mathbf{x}_j^T\mathbf{x}_i$. Nous pouvons déterminer le statut d'une observation $\mathbf{x}_i$ en analysant sa variable duale α_i. Celle-ci possède trois statuts différents :

- $\alpha_i = 0$: l'exemple est bien classé et n'est pas sur l'un des deux hyperplans canoniques ;
- $0 \leqslant \alpha_i \leqslant C$: l'exemple est bien classé et appartient à un hyperplan canonique, c'est un vecteur support.
- $\alpha_i = C$: l'exemple est mal classé. Il est quand même considéré comme un vecteur support puisque $\alpha_i > 0$.

Cas non linéaire : les noyaux

En pratique, un grand nombre de jeux de données à classer sont non linéairement séparables, ainsi l'originalité des SVM est de transformer l'espace des observations des données d'entrée en un espace de plus grande dimension, dans lequel ces données sont linéairement séparables. Cette transformation est réalisée par l'intermédiaire des fonctions « noyau », comme l'illustre la figure 7.5. Un noyau k est défini de manière générale comme une fonction de deux variables sur $\mathbb{R}$:

$$k : \mathcal{X},\mathcal{X} \to \mathbb{R} \quad (\mathbf{x},\mathbf{x}') \to k(\mathbf{x},\mathbf{x}') \tag{7.38}$$

avec $\mathcal{X}$ le domaine des observations.

Il est possible de représenter des observations $\mathbf{x}$ de l'espace $\mathcal{X}$ à partir d'une transformation $\phi : \mathcal{X} \to \mathbb{R}$ définie telle que $\phi(\mathbf{x}) = k(\cdot,\mathbf{x})$ dans l'espace $\mathcal{H}$. L'espace $\mathcal{H}$ admet une structure d'espace de Hilbert à noyau reproduisant. Le cas général pour les SVM de données non linéairement séparables consiste en un classifieur à marge maximale dans lequel le produit scalaire a été remplacé par le noyau. Nous reprenons la définition donnée dans [**baccini_wikistat_2014**] pour le cas général des SVM :

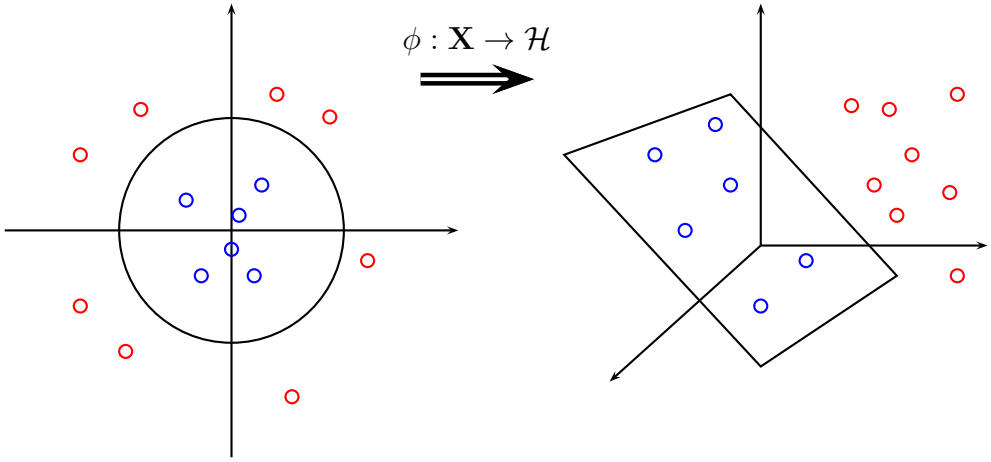


FIGURE 7.5 – Projection des données de l'espace non linéairement séparable vers un nouvel espace où les données sont linéairement séparables.

Définition 1 *(SVM cas général) Soit* $\{(\mathbf{x}_i,y_i); i = 1,\ldots,N\}$ *un ensemble de vecteurs d'observations avec* $\mathbf{x}_i \in \mathcal{X}$ *et* $y_i \in \{-1;1\}$. *Soit* $\mathcal{H}$ *un espace de Hilbert à noyau reproduisant de noyau* k. *Un séparateur à vaste marge est un discriminateur de la forme :* $D(x) = signe(f(\mathbf{x}) + b)$ *où* $f \in \mathcal{H}$ *et* $b \in \mathbb{R}$ *sont données par la résolution du problème suivant pour* $C \geqslant 0$ *donné :*

$$\begin{cases} min_{f\in\mathcal{H},b\in\mathbb{R}} & \frac{1}{2}||f||^2_{\mathcal{H}} + C\sum_{i=1}^{N}\xi_i \\ avec & y_i(f(\mathbf{x}_i) + b) \geqslant 1 - \xi_i \ , \ i = 1,\ldots,N \\ & 0 \leqslant \xi_i \ , \ i = 1,\ldots,N \end{cases} \tag{7.39}$$

La fonction de décision dans le cas non linéaire s'écrit :

$$f(\mathbf{x}) = \sum_{i\in\mathcal{A}} \alpha_i y_i k(\mathbf{x}_i,\mathbf{x}) \tag{7.40}$$

Le problème d'optimisation quadratique est le même que celui de l'équation (7.37), la différence se situe au niveau de la matrice $\mathbf{G}$ dont le terme général est maintenant défini comme suit : $G_{ij} = y_i y_j k(\mathbf{x}_i,\mathbf{x}_j)$.

Réidentification de véhicules

Un grand nombre de recherches a déjà été effectué avec les SVM (bio-informatique, finance, médicale, télécommunication, microélectronique...). Dans cette section, il est proposé d'utiliser cette méthode pour réidentifier les véhicules. Ainsi, notre problème de classification est binaire : réidentification du véhicule ou non. Soit l'ensemble d'apprentissage $\mathcal{S} = \{(\mathbf{v}_i,\mathbf{c}_i) \in \mathbb{R}^p \times \{-1;1\}\}_{i=1}^N$. Dans cette situation, l'étiquette c_i correspond à l'identification ou non-identification du véhicule. Pour passer les données de l'espace de description (non linéairement séparable) dans un espace de représentation pour les rendre séparables, le choix du noyau Radial Basis Function (RBF) est effectué. D'après Boubezoul [84], ce noyau est l'un des plus couramment utilisés dans la littérature. La forme générique de ce noyau est :

$$k(\mathbf{v_i},\mathbf{v}) = exp\left(-\frac{||\mathbf{v}_i - \mathbf{v}||_2^2}{2\sigma^2}\right) \tag{7.41}$$

où $||\cdot||_2$ désigne la norme euclidienne. Ce noyau permet de projeter une observation sur une fonction gaussienne représentant la similarité de l'observation avec tous les points de $\mathcal{V}$. Le paramètre σ représente la largeur de la gaussienne. Lorsque σ tend vers zéro, l'observation considérée ne sera similaire à aucune autre observation. Par contre plus σ sera grand, plus la similarité avec les observations aux

alentours augmentera. Nous aurons les paramètres C et σ à déterminer lors de la phase d'expérimentation pour la résolution du problème dual. Afin d'utiliser cette méthode SVM dans un problème de réidentification, nous avons ajouté une contrainte supplémentaire : seuls les couples dont l'origine n'est pas citée à de multiples reprises ne sont pas filtrés. Nous avons expérimenté la SVM sur les bases de données de la section 8.1. Pour l'ensemble des véhicules, les paramètres C et σ sont à fixer lors de la phase d'apprentissage. Le TBR est de 34,3 % sur la base de données d'Angers et de 7,8 % sur la base de données de Rennes en employant la SVM comme méthode de réidentification. La méthode SVM recueille des performances réduites. En quête d'un perfectionnement des performances de la SVM, une contrainte supplémentaire est prise en compte : la distance entre l'échantillon et l'hyperplan. Le couple valide correspondra à celui ayant la distance échantillon–hyperplan maximale dans la bonne classe. Cette version se nommera dans la suite du rapport SVM-d. Les résultats produits par cette amélioration sont détaillés dans la section 8.1 à la page 140.

7.3 Détermination d'un seuil et utilisation de plusieurs méthodes

7.3.1 Introduction

Dans cette section, les méthodes étudiées sont le maximum de vraisemblance présenté à la section 7.1.2 (MV), le maximum de vraisemblance avec un modèle de temps de parcours gaussien présenté aussi dans 7.1.2 (MV2), la méthode de logique floue présentée à la section 7.1.2 (AF) et la méthode SVM prenant en compte la distance échantillon–hyperplan présentée à la section 7.2.2. Les méthodes seront comparées par la méthode de validation croisée [35] afin de déterminer celle fournissant les meilleurs résultats. Les algorithmes développés sont testés sur plusieurs bases de données. Les différentes bases de données sont décrites aux chapitres 8.1, 8.2. Les multiples bases de données permettent d'avoir des situations et des typologies diversifiées, démontrant ainsi une convergence des résultats pour des situations disparates.

7.3.2 Étude sur une base idéale

La base idéale est constituée uniquement de paires de signatures, donc chaque destination a une origine et *vice versa*. L'élimination des perturbations a été réalisée grâce aux vidéos. Cette étude a pour but d'évaluer les performances des méthodes, afin de sélectionner les approches les plus efficaces.

Performances en validation croisée

Nous utilisons la méthode de validation croisée décrite à la section 4.4.2. Nous avons trois bases de données. La première base de donnée sera nommée SAROT1, la seconde Angers et la dernière Rennes (les noms faisant références aux sites d'expérimentations). L'expérimentation SAROT1 est décrite à la section 8.2.2, celle d'Angers à la section 8.1.2 et celle de Rennes à la section 8.1.2. La base de données SAROT1 utilise la sélection de variables effectuée par Ieng et coauteurs. Les deux autres bases de données sont établies en prenant en considération la population avec la sélection de variables, cette sélection est présentée à la section 6.2.

SAROT1 La k validation croisée, avec $k = 10$, est utilisée sur cette base de données. La base d'apprentissage comprend 460 véhicules et la base de test 936 véhicules. En se basant sur [59], seulement 9 caractéristiques par signature sont retenues. Les caractéristiques sont les suivantes : la longueur sur l'indice du maximum global de la signature ; le premier et le second coefficient de la partie réelle de la transformée de Fourier ; le premier, le second et le troisième coefficients de la partie imaginaire ; le module des quatrième, cinquième et sixième coefficients de la transformée de

Fourier. Sur cette base, l'algorithme du maximum de vraisemblance et celui de logique floue sont testés. Nous cherchons d'abord à établir les paramètres de l'algorithme de logique floue. La valeur de

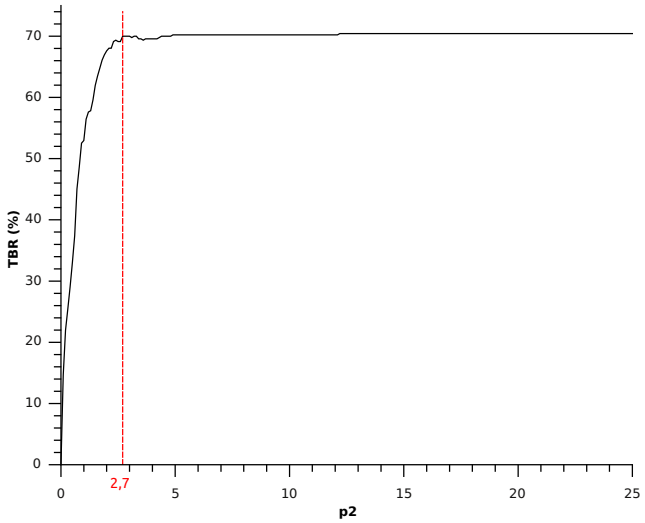


FIGURE 7.6 – TBR en fonction de p_2 pour l'approche floue.

p_2 est déterminée à partir d'une validation croisée afin d'obtenir une valeur optimale. Les paramètres m_k et σ_k sont calculés à partir de la base d'apprentissage. Après apprentissage, l'approche floue est appliquée sur la base de validation pour la réidentification. Les résultats sont calculés pour des valeurs de p_2 variant de 0 à 50 et pour chaque sous-ensemble de validation. La figure 7.6 présente le TBR (cf. (4.3)) moyen sur les sous-ensembles de validation en fonction de la valeur de p_2. À partir de $p_2 = 2{,}7$, le TBR atteint 70 %. Pour les valeurs de p_2 comprises entre 2,7 et 12,3, le TBR varie faiblement autour de 70 %. Ensuite pour les valeurs supérieures de p_2, le TBR reste constant ce qui a conduit à un choix de $p_2 = 12{,}5$.

Dans le cadre du maximum de vraisemblance, la moyenne $\boldsymbol{\mu}$ et la matrice de covariance $\boldsymbol{\Sigma}$ sont calculées à partir de chaque ensemble d'apprentissage.

TBR (%)	MV	AF
Sous-ensemble 1	58,7	54,35
Sous-ensemble 2	54,35	56,52
Sous-ensemble 3	76,09	67,39
Sous-ensemble 4	78,26	71,74
Sous-ensemble 5	82,61	73,91
Sous-ensemble 6	76,09	76,09
Sous-ensemble 7	63,04	65,22
Sous-ensemble 8	84,78	89,13
Sous-ensemble 9	78,26	73,91
Sous-ensemble 10	86,96	76,09
Moyenne	73,91	70,43

Tableau 7.1 – TBR par validation croisée pour le maximum de vraisemblance et l'approche floue.

L'analyse du tableau 7.1 met en évidence une meilleure performance pour le maximum de vraisemblance avec un TBR moyen proche de 74 %. En regardant plus en détail le tableau, certains sous-ensembles ont des TBR faibles par rapport à la moyenne. L'apprentissage de la moyenne $\boldsymbol{\mu}$ et de la matrice de covariance $\boldsymbol{\Sigma}$, est effectué sur la base de données d'apprentissage–validation. 28,74 % et 26,71 % de TBR pour respectivement MV et AF sont obtenus sur la base de données de test.

Distinction de la population Les bases de données d'Angers et de Rennes sont divisées chacune en un ensemble d'apprentissage–validation et un ensemble de test. L'ensemble d'apprentissage–validation est subdivisé en $k = 5$ sous-ensembles disjoints. Le nombre de PL étant très faible, la validation croisée a été utilisée sur la totalité de la base idéale divisée en $N = 5$ segments. Les méthodes de AF, MV, MV2 et SVM-d sont testées. Pour utiliser la méthode du maximum de vraisemblance avec un modèle de temps de parcours gaussien (MV2), les paramètres présentés dans la section 8.1.2 (vitesse moyenne : 60 $\mathrm{km\,h^{-1}}$ écart-type : 20 $\mathrm{km\,h^{-1}}$) pour la base d'Angers et les paramètres présentés dans la section 8.1.2 (vitesse moyenne : 50 $\mathrm{km\,h^{-1}}$, écart-type : 15 $\mathrm{km\,h^{-1}}$) pour la base de Rennes ont été pris en compte. Pour le classifieur SVM-d à noyau RBF, deux hyperparamètres sont à optimiser C et σ. Ces paramètres ont été estimés par validation croisée pour les trois types de population (VL, PL et mixte). Les paramètres de la méthode SVM-d sont présentés dans le tableau 7.2.

	SVM-d	
	C	σ
PL	2	11,314
VL	16	1
PL & VL	2	64

Tableau 7.2 – Paramètres établis pour la méthode SVM en fonction des différentes populations.

	Floue Triangulaire	MAP Uniforme	MAP Gaussien	SVM (Noyau RBF) Distance Hyperplan
Abréviation	Floue	MV	MV2	SVM-d
PL (%)	81,0	89,3	92,7	87,3
VL (%)	68,0	70,0	82,9	64,1
PL & VL (%)	51,6	69,6	87,1	62,2

Tableau 7.3 – TBR en validation croisée sur la base d'Angers.

	Floue	MV	MV2	SVM-d
PL (%)	90,0	92,5	93,7	86,2
VL (%)	51,2	55,3	62,4	52,0
PL & VL (%)	34,9	52,8	61,5	39,1

Tableau 7.4 – TBR en validation croisée sur la base de Rennes.

Le tableau 7.3 permet de remarquer que quelle que soit la méthode, il existe une différence de performances selon le type de la population. En effet, le TBR est plus grand pour les PL, suivi du taux pour les VL. Le taux le plus faible est obtenu par la population mixte.

L'analyse du tableau 7.4 montre que de fortes différences de TBR sont observables en fonction de la population comme pour le tableau 7.3. Les différences de TBR selon la population pour les tableaux 7.3 et 7.4 s'expliquent par la taille de la population étudiée, mais aussi par les variables choisies pour chaque type de population.

Les performances sont calculées dans cette section sur la base totale. Le premier tiers de la base est utilisé pour l'apprentissage et les deux autres sont utilisés pour l'évaluation.

Le tableau 7.5 montre sur la base totale que l'utilisation de la fenêtre temporelle dans le cadre de l'algorithme MV2 améliore les performances comparées aux autres méthodes qui en sont dépourvues.

	Floue	MV	MV2	SVM-d
PL (%)	69,6	71,7	84,1	79,7
VL (%)	36,7	38,8	67,1	27,9
PL & VL (%)	17,0	41,4	82,1	33,6

Tableau 7.5 – Angers - TBR sans fenêtrage dans le cadre d'une mesure sur la base totale.

Ainsi les TBR de l'algorithme MV2 sont supérieurs à ceux des autres algorithmes. De meilleurs résultats sont constatés pour la population de PL ainsi que de moins bons résultats pour la population mixte. Dans le tableau 7.6 de meilleurs résultats sont obtenus pour la méthode MV2, ainsi que pour les poids lourds.

	Floue	MV	MV2	SVM-d
PL (%)	79,6	81,5	87,0	75,9
VL (%)	37,7	41,0	57,4	31,8
PL & VL (%)	18,5	41,2	56,4	27,2

Tableau 7.6 – Rennes - TBR sans fenêtrage dans le cadre d'une mesure sur la base totale.

Les meilleurs TBR obtenus par la méthode MV2 sur les deux sites (tableaux 7.5 et 7.6) s'expliquent par la prise en compte d'une information *a priori* supplémentaire par rapport aux autres méthodes. Une pondération est effectuée avec la méthode MV2. Cette pondération est établie en fonction du temps de parcours entre l'origine et la destination. L'information *a priori* utilisée par cette méthode est le temps de parcours moyen des véhicules. Un temps parcours individuel éloigné du temps de parcours moyen pénalisera d'autant plus l'association.

La population réduite des PL favorise leur réidentification sur les deux sites.

Le fenêtrage temporel

Les TBR estimés lors de la phase d'apprentissage sont largement supérieurs à ceux de la phase de test pour l'ensemble des bases de données. La variation du TBR s'explique par le nombre de candidats lors de la réidentification. La base de données de validation contient du nombre total de signatures. Lors de la réidentification, une signature destination est comparée pour la base de données de validation à $\frac{1}{3} \times \frac{1}{k}$ signatures origine de l'ensemble de la base de données. Pour la base de données de test, la signature destination est comparée aux deux tiers des signatures de la base de données. Avec l'accroissement du nombre de signatures origine, le risque de procéder à une mauvaise réidentification augmente. Pour limiter ce risque, nous avons appliqué un fenêtrage temporel. Le fenêtrage temporel a pour objectif de réduire le nombre de signatures origine susceptible d'être réidentifié avec la signature destination. Cela consiste à chercher une signature origine qui appartient à la fenêtre temporelle. Nous étudions cette méthode sur les bases de données d'Angers et de Rennes.

Le positionnement de la fenêtre est déterminé avec l'horaire de passage sur le point de destination du véhicule. Les fenêtres temporelles sont estimées en fonction des vitesses constatées sur les sites d'Angers et de Rennes ainsi que des populations. Ces paramètres sont un facteur limitant, les TBR maximaux possibles sont explicités dans le tableau 7.7. En effet, certains véhicules sont éliminés de la base de données par ce prétraitement.

Population	VL	PL	VL & PL
Angers (%)	98,6	99,6	99,5
Rennes (%)	98,1	96,3	96,4

Tableau 7.7 – TBR maximaux potentiels avec le fenêtrage utilisé.

	Floue	MV	MV2	SVM-d
PL (%)	97,8	98,6	98,6	95,7
VL (%)	92,1	93,5	92,5	88,2
PL & VL (%)	81,2	82,4	86,4	83,0

Tableau 7.8 – Angers - TBR avec fenêtrage dans le cadre d'une mesure sur la base totale.

	Floue	MV	MV2	SVM-d
PL (%)	92,6	90,7	90,7	79,6
VL (%)	65,0	69,4	62,9	59,8
PL & VL (%)	47,9	65,1	55,8	51,3

Tableau 7.9 – Rennes - TBR avec fenêtrage dans le cadre d'une mesure sur la base totale.

Dans l'ensemble, une nette augmentation des performances est constatée dans le tableau 7.8 (ou 7.9) par comparaison avec le tableau 7.5 (ou 7.6). Seulement, pour le cas de la méthode MV2, ces augmentations observées sont plus faibles. La différence de performance entre les populations est moins importante que dans le cas où aucun fenêtrage temporel n'est utilisé. Les tableaux 7.8 et 7.9 confirment l'hypothèse que la pondération utilisée dans la méthode MV2 était une information *a priori* importante. Mais ces observations mettent aussi en valeur le bénéfice apporté par la réduction du nombre de signatures lors du fenêtrage.

Une nette amélioration des TBR est observée avec la mise en œuvre du fenêtrage. Par ailleurs, les résultats liés à la population mixte sont inférieurs à ceux des autres populations. Enfin, les performances sur le site de Rennes sont moins grandes par comparaison de celles obtenues sur le site d'Angers. Cette différence peut s'expliquer par le nombre de candidats possibles qui est plus important sur le site de Rennes.

Pour démontrer graphiquement la réduction du nombre de signatures par la fenêtre temporelle, nous avons constitué les matrices origine – destination pour le site d'Angers et de Rennes dans le cadre des VL. Elles sont représentées sur les figures 7.7 et 7.8. Dans cet exemple, l'ensemble des distances entre les signatures destination et les signatures origine est obtenu par la méthode de logique floue. Chaque ligne de la matrice représente les véhicules destinations et chaque colonne les véhicules origines. Il est à noter que les véhicules sont triés par ordre de passage et que dans le cas de la logique floue, une valeur proche de 11 (le nombre de variables utilisées) traduit une grande similarité. L'utilisation de la fenêtre temporelle limite les calculs à effectuer. En effet, au lieu de calculer la distance pour tous les couples possibles, nous attribuons systématiquement la distance zéro pour les couples origine – destination n'appartenant pas à la fenêtre temporelle. Suivant la restriction apportée par la fenêtre temporelle, une diagonale plus ou moins large variant du jaune au rouge est visible sur les figures 7.7 et 7.8. Dans le cas de Rennes (7.8), la diagonale observée est très large par rapport au cas d'Angers (figure 7.7). Ceci peut s'expliquer par la circulation importante, mais aussi par la grande distance entre les deux zones instrumentées à laquelle s'ajoute la vitesse relativement lente des véhicules. Dans la suite de cette étude, il a donc été décidé de se concentrer uniquement sur la base d'Angers.

Le fenêtrage temporel apporte un gain de TBR important grâce à la réduction de données effectuée. Par conséquent, la méthode MV2 n'est pas retenue pour la suite des analyses. En effet, cette méthode perd de l'intérêt au profit de la méthode MV lorsqu'un fenêtrage temporel est utilisé. Sans fenêtrage, les TBR de la méthode MV2 sont supérieurs à ceux de la méthode MV dans les tableaux 7.5 et 7.6. Avec le fenêtrage, les TBR des méthodes MV et MV2 sont identiques pour la population de PL dans les tableaux 7.8 et 7.9. Pour ces deux tableaux, les TBR de la méthode MV sont supérieurs à ceux de la méthode MV2 pour la population de VL. Pour la population mixte de PL & VL, le meilleur TBR est obtenu avec la méthode MV2 dans le tableau 7.5, alors que pour le tableau 7.9, le meilleur

TBR est donné par la méthode MV. Avec le fenêtrage, la méthode MV obtient un TBR équivalent ou supérieur à la MV2 dans la majorité des cas.

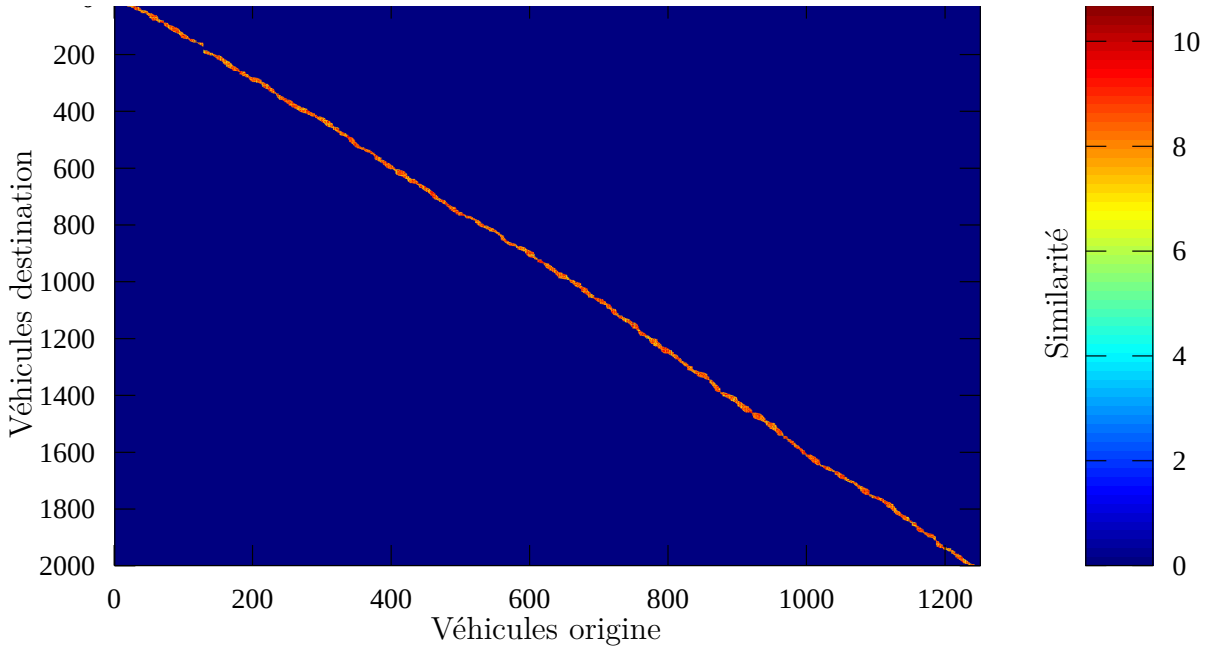


FIGURE 7.7 – Matrice des distances réalisée à partir des données d'Angers.

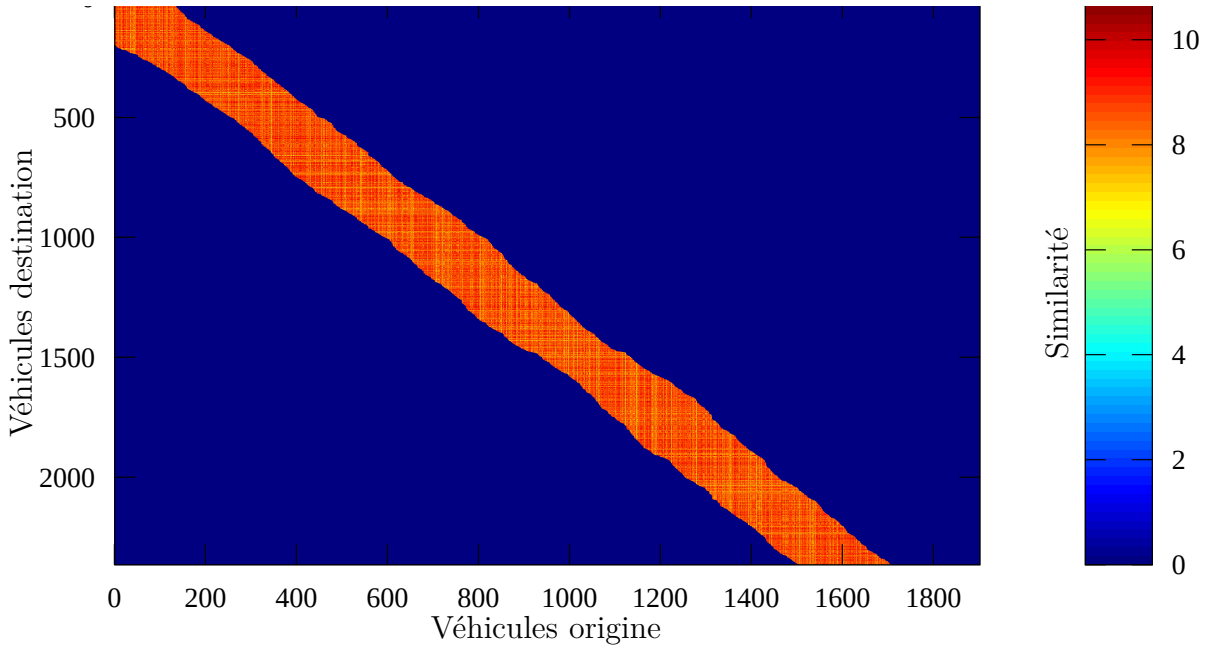


FIGURE 7.8 – Matrice des distances réalisée à partir des données de Rennes.

Par ailleurs, il est intéressant de séparer les populations lors de l'assemblage des signatures, afin de réduire la quantité de données à croiser. La figure 7.9 montre que le TBR en fonction du ratio PL – VL, mais en gardant un nombre total de véhicules constant (400). Les signatures utilisées sont issues de la base idéale d'Angers. La figure 7.9 montre que certaines méthodes sont très sensibles à la répartition VL – PL. Pour réaliser cette analyse, les valeurs d'initialisation sont calculées pour une population comportant 10 % de PL quel que soit le ratio PL – VL. En effet, les performances des différentes méthodes peuvent dépendre de l'initialisation réalisée, notamment la répartition poids

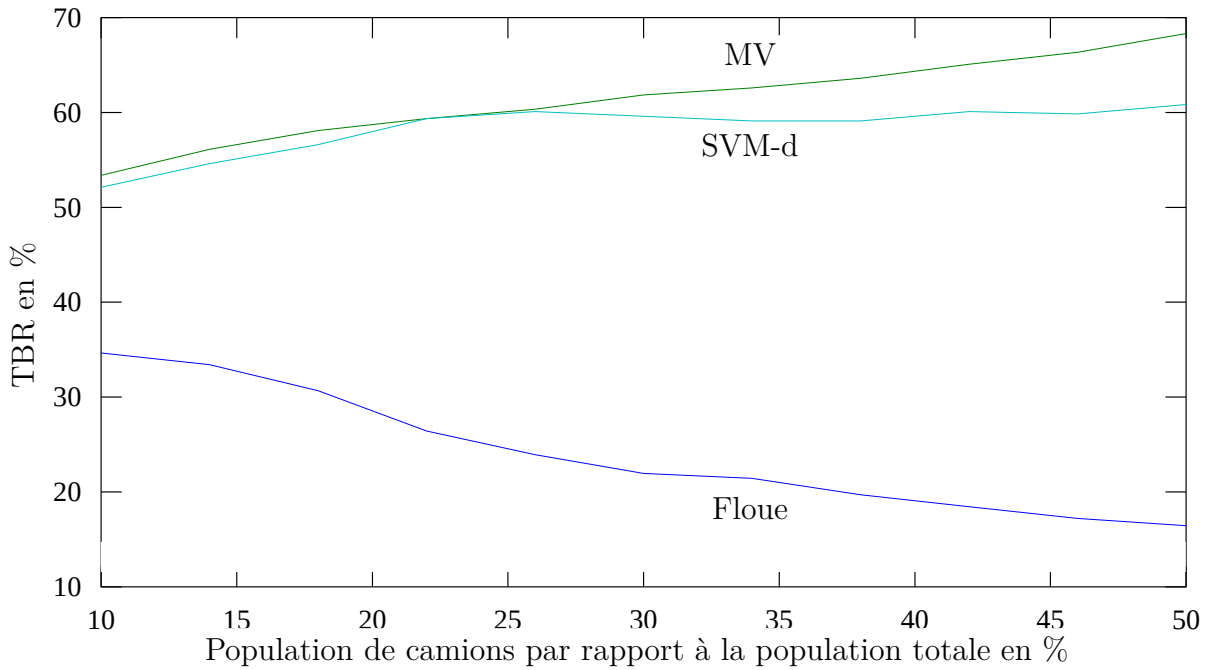


FIGURE 7.9 – TBR en fonction de la répartition poids lourds – véhicules légers pour une population constante de 400 véhicules.

lourds - véhicules légers. La méthode de logique floue est par exemple très sensible à la variation de la répartition PL – VL. C'est pourquoi l'étude sur la population mixte est abandonnée au profit des deux autres populations lors de l'étude sur la base perturbée. De plus, la classification PL – VL est déjà opérationnelle sur les stations de mesures des boucles inductives. La méthode MV2 a un meilleur TBR que la méthode MV uniquement dans le tableau 7.9 pour la population mixte, cette méthode n'est plus utilisée par suite.

Enfin, au vu des TBR présentés, l'étude sur les bases d'Angers et de Rennes s'effectue maintenant autour des trois approches suivantes : la logique floue, MV et SVM-d.

7.3.3 Étude sur une base perturbée

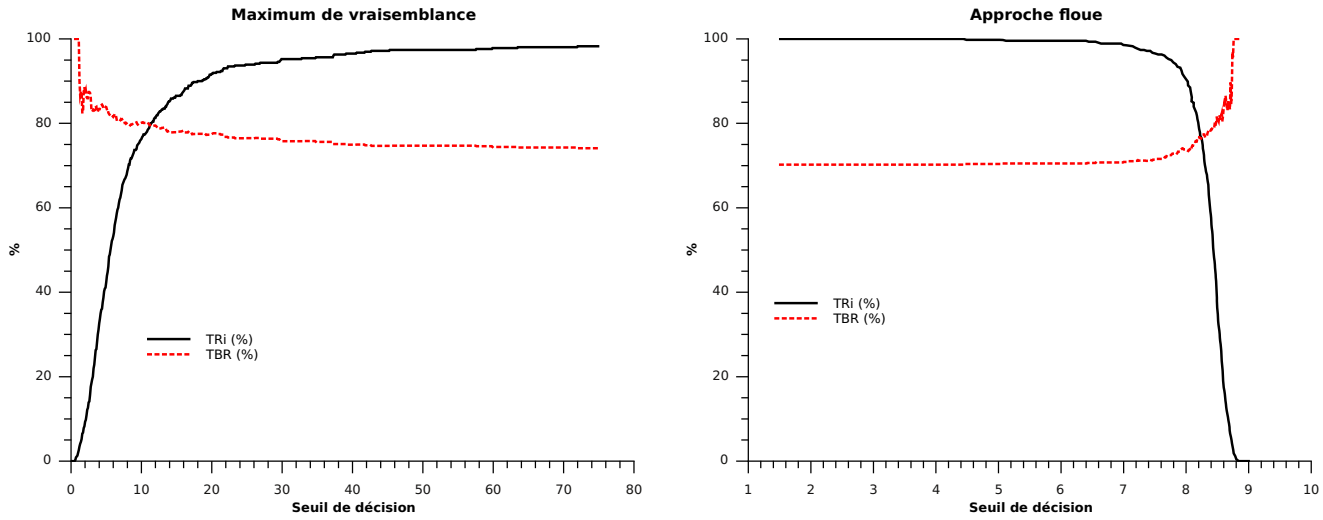
Contexte

Dans une situation réelle, l'ensemble des origines et destinations n'est pas toujours instrumenté (impossibilité technique, défaut d'un capteur, coût d'installation...) et les capteurs ne sont pas infaillibles au niveau des détections. Nous nous retrouvons donc avec des signatures origine sans signature destination et *vice versa*. Ce sont les signatures produites par les véhicules n'ayant pas circulé sur les deux zones instrumentées ou n'ayant été détectées que sur une zone. La base de données est alors constituée de paires de signatures, mais aussi de signatures perturbatrices. De plus, la déformation d'une des signatures (origine ou destination) peut être importante (accélération sur le capteur, véhicule ne passant pas dans l'axe, défaut d'acquisition...) et la réidentification devient difficile. La probabilité de commettre une erreur de réidentification augmente.

Pour tenter de réduire le taux d'erreur des algorithmes, nous mettons en place différentes propositions. Une première proposition consiste à déterminer un seuil de décision selon la similarité entre les signatures origine et destination. La seconde idée consiste à utiliser plusieurs méthodes en parallèle afin de tenir uniquement compte des couples identifiés de façon identique par chaque méthode. Les bases de données SAROT1 et Angers ont été utilisées pour valider les propositions.

Base de données SAROT1

Apprentissage et validation Nous avons remarqué lors de la phase d'apprentissage validation (7.3.2) que les algorithmes cherchent absolument à associer un véhicule destination à un véhicule origine même si la dissimilarité entre les deux est grande, ce qui conduit à des erreurs. Pour l'approche floue, cela se traduit par l'association de véhicules alors que la valeur discriminatoire est faible. Pour le maximum de vraisemblance, cela se traduit par l'association de véhicules alors que le risque est élevé. Pour remédier à ce problème, l'introduction d'un seuil de décision pour chaque méthode va supprimer les associations de véhicules présentant une dissimilarité trop importante.



(a) (b)

FIGURE 7.10 – TBR et de TRi en fonction du seuil de décision : (a) : Maximum de vraisemblance ; (b) : Approche floue.

La figure 7.10 montre la relation entre le seuil de décision et le TBR ou le TRi (cf. définition page 53). Dans le cas du maximum de vraisemblance 7.10a, l'association n'est pas réalisée lorsque la fonction est supérieure au seuil de décision. Par exemple pour un seuil de décision égal à 10, toutes les associations dont la valeur est supérieure à 10 sont rejetées. Ceci implique que des véhicules ne seront pas reconnus entre l'origine et la destination. Le TBR sera égal à 76,3 % et le TRi à 80,14 % pour un seuil fixé à 10. L'augmentation du seuil engendre un accroissement du TRi jusqu'à 100 % et une diminution du TBR. Pour un seuil de décision en-dessous de 10 le TRi connaît une diminution rapide, mais le TBR est élevé. Dans le cas de l'approche floue 7.10b, c'est l'inverse qui se produit. Un seuil de décision implique une diminution du TRi et une augmentation du TBR.

Un compromis est à trouver entre le TBR et le TRi. Un TBR élevé est intéressant mais si le TRi devient trop faible, il n'est plus possible d'estimer une matrice origine – destination. Par contre dans le cadre d'une autre application, comme les temps de parcours, un TRi moindre est moins pénalisant. Dans cette expérimentation, les seuils de décision ont été fixés à l'intersection entre la courbe du TRB et la courbe du TRi. Le seuil de détection pour le maximum de vraisemblance est de 11,2 pour un TBR et un TRi d'environ 80 %. Le seuil de détection pour l'approche floue est fixé à 8,24 pour un TBR et TRi avoisinant les 77 %. Dans les deux cas, l'introduction d'un seuil de décision améliore le TBR par rapport aux deux approches dépourvues de seuil.

Les deux méthodes sont basées sur des théories différentes et nous avons constaté que les deux approches ne commettent pas les mêmes erreurs d'associations. En regardant le tableau 7.1, en moyenne le maximum de vraisemblance donne de meilleurs résultats mais dans le détail nous constatons que l'approche floue a donné de meilleurs résultats pour les sous-ensembles 2, 7 et 8. En partant de ce constat, nous proposons d'utiliser les deux méthodes conjointement et des règles de décision additionnelles sont ajoutées. Le processus proposé consiste à dire que si les deux approches donnent un

résultat d'association identique alors le résultat d'association est validé. Au contraire, si les deux résultats ne sont pas identiques aucune association n'est réalisée. Deux cas sont considérés pour cette approche : sans seuil de décision, et avec seuil de décision.

TBR % (TRi %)	MV (Seuil)	AF (Seuil)	MVAF	MVAF2 (Seuil)
Sous-ensemble 1	57,58 (71,74)	54,84 (67,39)	68,57 (76,09)	62,96 (58,70)
Sous-ensemble 2	65,71 (76,09)	62,50 (69,57)	67,65 (73,91)	74,07 (58,70)
Sous-ensemble 3	83,78 (80,43)	74,29 (76,09)	79,49 (84,78)	83,87 (67,39)
Sous-ensemble 4	84,21 (82,61)	78,95 (82,61)	82,05 (84,78)	85,29 (73,91)
Sous-ensemble 5	86,11 (78,26)	80,56 (78,26)	84,62 (84,78)	87,50 (69,57)
Sous-ensemble 6	85,29 (73,91)	87,88 (71,74)	82,93 (89,13)	90,32 (67,39)
Sous-ensemble 7	74,36 (84,78)	71,05 (82,61)	72,97 (80,43)	75,76 (71,74)
Sous-ensemble 8	88,89 (78,26)	91,43 (76,09)	88,64 (95,65)	90,63 (69,57)
Sous-ensemble 9	80,49 (89,13)	79,49 (84,78)	83,78 (80,43)	85,29 (73,91)
Sous-ensemble 10	89,19 (80,43)	83,33 (78,26)	87,50 (86,96)	87,88 (71,74)
Moyenne	79,56 (79,57)	76,43 (76,74)	79,82 (83,70)	82,36 (68,26)

Tableau 7.10 – Performances des quatre approches de réidentification.

Le tableau 7.10 présente les résultats de TBR et de TRi pour le maximum de vraisemblance avec seuil (MV), l'approche floue avec seuil (AF), l'association du maximum de vraisemblance et de la logique floue sans seuil (MVAF) et l'association du maximum de vraisemblance et de la logique floue avec avec chacun un seuil (MVAF2). Les quatre méthodes proposées ont un TBR plus élevé par rapport au tableau 7.1. La méthode MVAF2 obtient le meilleur TBR avec un peu plus de 82 %. Ensuite MV et MVAF obtiennent un TBR de 80 %. Par contre, c'est la méthode MVAF qui obtient le meilleur résultat pour le TRi avec près de 84 % alors que MVAF2 a un TRi de 68 %. Un autre point important est l'écart-type par rapport aux différents sous-ensembles. L'écart-type pour MV et AF est respectivement de 10,53 et 11,21 alors que pour l'association des méthodes MVAF et MVAF2 cet écart-type est respectivement de 7,54 et 8,88. Un écart-type faible démontre une meilleure homogénéité dans les résultats de prédiction. Au vu des résultats obtenus, la méthode MVAF semble être le meilleur compromis entre le TBR et le TRi. De plus, c'est la méthode ayant obtenu l'écart-type le plus faible. Nous pouvons donc considérer que c'est la méthode la plus insensible à l'échantillon.

L'association de différentes méthodes est une solution avantageuse : elle permet d'éviter le calcul d'un seuil et l'utilisation des mêmes variables avec deux classifieurs différents et d'améliorer le TBR. Dans le cas de deux classifieurs, c'est un vote unanime qui est appliqué c'est-à-dire que l'ensemble des classifieurs donne une réponse identique pour valider la réidentification. Dans le cas d'un nombre de classifieurs supérieur à deux, un vote à la majorité peut être utilisé c'est-à-dire que la majorité des classifieurs donne une réponse identique pour valider la réponse.

Test Au vu des résultats obtenus, la méthode MVAF semble le meilleur compromis entre le TBR et le TRi. De plus, c'est la méthode ayant obtenu l'écart-type le plus faible. Néanmoins, les deux méthodes MVAF et MVAF2 ont été retenues pour être appliquées sur la base de test.

Pour rappel, l'apprentissage de la moyenne $\boldsymbol{\mu}$ et de la matrice de covariance $\boldsymbol{\Sigma}$, est effectué sur la base de données d'apprentissage–validation. 40,83 % et 41,15 % de TBR pour respectivement MVAF et MVAF2 sont obtenus sur la base de données de test. Ces résultats sont très inférieurs à ceux obtenus lors de la phase d'apprentissage–validation. Ces faibles résultats en réidentification sont dus à un trop grand nombre de candidats. En effet, lors de la phase d'apprentissage–validation, le nombre de candidats était de 45 et ce nombre est de 935 lors de la phase de test. Ce nombre de candidats est facilement réductible en prenant en compte la configuration des lieux et les conditions de trafic

qui vont se traduire par des contraintes temporelles [59]. En appliquant une fenêtre temporelle, le nombre de candidats est fortement réduit. La réidentification est effectuée avec une moyenne d'environ 11 candidats. Pour les deux méthodes, le TBR est d'environ 94 % mais une très grande différence apparaît au niveau du TRi en faveur de la méthode MVAF.

	TBR(%)	TRi(%)
MVAF	93,78	92,74
MVAF2 (seuil)	94,51	72,01

Tableau 7.11 – Réidentification sur la base de test avec fenêtre temporelle.

Base de données Angers

Comme présenté précédemment à la section 7.3.2, un fenêtrage temporel est utilisé. Ce fenêtrage est implanté avant la recherche des couples, afin de réduire les possibilités d'assemblage, mais aussi pour gagner du temps de calcul. La position de la fenêtre temporelle sur la liste des signatures d'origine est déterminée grâce à l'horaire de passage de la signature de destination et à différents paramètres prenant en compte les conditions de trafic. Ces paramètres correspondent à la vitesse moyenne, à l'écart-type de la vitesse moyenne ainsi qu'à la distance entre les deux zones instrumentées.

Tout d'abord, plusieurs méthodes (logique floue, MV et SVM-d) en parallèle sont utilisées afin de tenir compte uniquement des couples proposés de façon identique par chaque méthode. Nous proposons ici le vote à l'unanimité, contrairement à un vote à la majorité, celui-ci n'a pas besoin d'un nombre de méthodes impair. Cependant, ce type de discrimination est intéressant seulement si chacune des méthodes a un comportement différent. Dans la suite de ce manuscrit, la méthode utilisant le vote à l'unanimité se nommera « una ».

Dans un second temps, un deuxième post-traitement est proposé, il consiste à utiliser un seuil ou critère d'acceptabilité comme dans [85]. Chaque méthode calcule une fonction coût pour chaque couple. Dans notre cas, plus le coût est élevé, plus la probabilité de validité du couple est élevée. Il consiste à dire que q % des véhicules de la base d'apprentissage ont une fonction coût supérieure à un seuil α. Les figures 7.11 , 7.12 et 7.13 montrent le nombre de véhicules sur le nombre total de véhicules (en pourcentage) ayant une fonction coût (ou une distance à l'hyperplan) supérieure au seuil α. Ces figures sont réalisées à partir de la base d'apprentissage–validation de la base idéale. Le critère d'acceptabilité utilisé consiste à dire que l'association d'un véhicule destination avec un véhicule origine ne peut se faire que lorsque la fonction coût (ou une distance à l'hyperplan) d'une méthode est supérieure à ce seuil α. Lorsque la fonction coût (ou distance à l'hyperplan) ne dépasse pas ce seuil, l'algorithme considère que la solution calculée est incertaine. Ce seuil permet ainsi de minimiser les erreurs d'association. Cette méthode de post-traitement s'intègre juste après la désignation du candidat par la méthode. Ce post-traitement peut donc aussi s'effectuer lorsqu'une seule méthode est utilisée. Le but du post-traitement est de diminuer le taux d'erreur en minimisant les mauvaises réidentifications.

Différents seuils peuvent être utilisés mais un compromis entre mauvaises identifications et suppression de bonnes identifications est à réaliser. Les tableaux 7.12 et 7.13 montrent des seuils qui permettront de garder 90 %, 80 % ou 70 % de la population des couples corrects.

De plus, un troisième post-traitement est proposé. Il consiste à filtrer le cas où des origines ont de multiples destinations. Lors de l'association des signatures, il arrive que deux destinations pointent la même origine. Par contre, il n'est pas possible de retrouver le cas contraire car la recherche d'association s'effectue à partir des signatures de destination. Une solution pour atténuer ce problème est de rechercher dans un premier temps les origines citées à plusieurs reprises. Puis, pour chaque

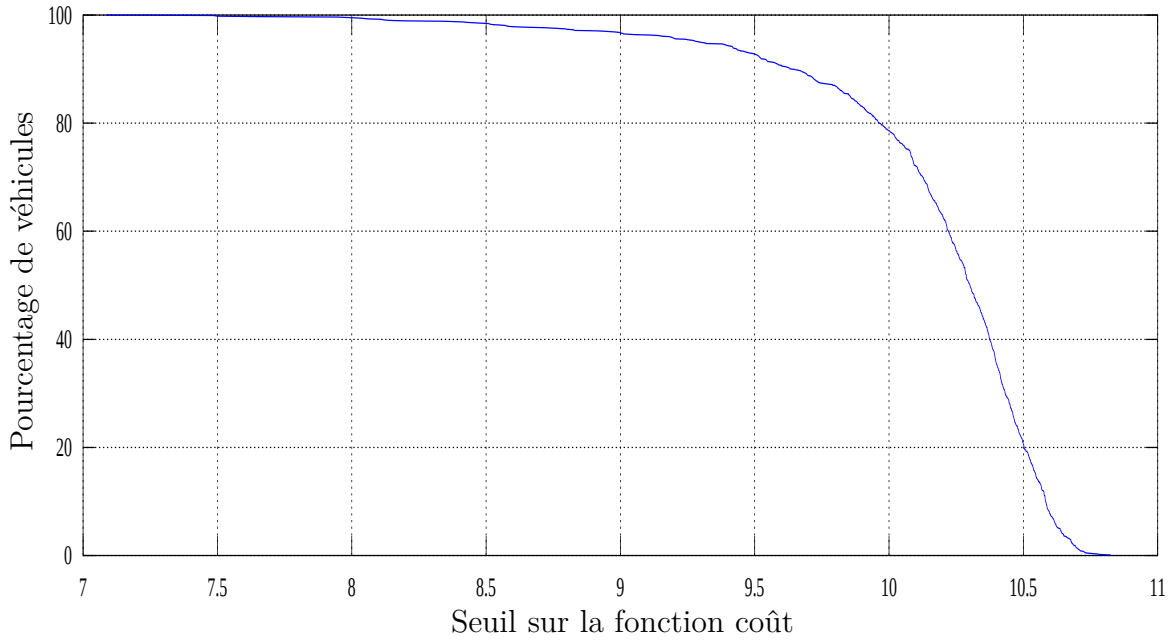


FIGURE 7.11 – Évolution du pourcentage de véhicules légers en fonction du seuil (méthode de la logique floue).

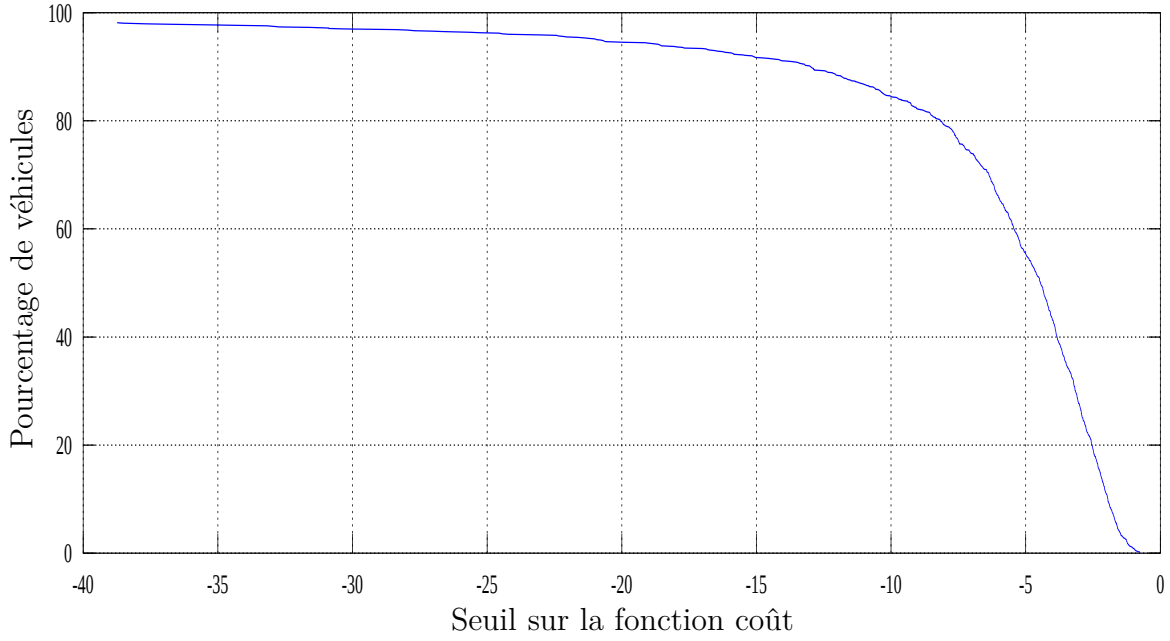


FIGURE 7.12 – Évolution du pourcentage de véhicules légers en fonction du seuil (MV).

	90 %	80 %	70 %
Floue	9,64	9,96	10,12
MV	-12,99	-8,16	-6,36
SVM-d	0,31	0,88	1,08

Tableau 7.12 – Valeur de seuil pour un pourcentage donné de véhicules légers (Site d'Angers).

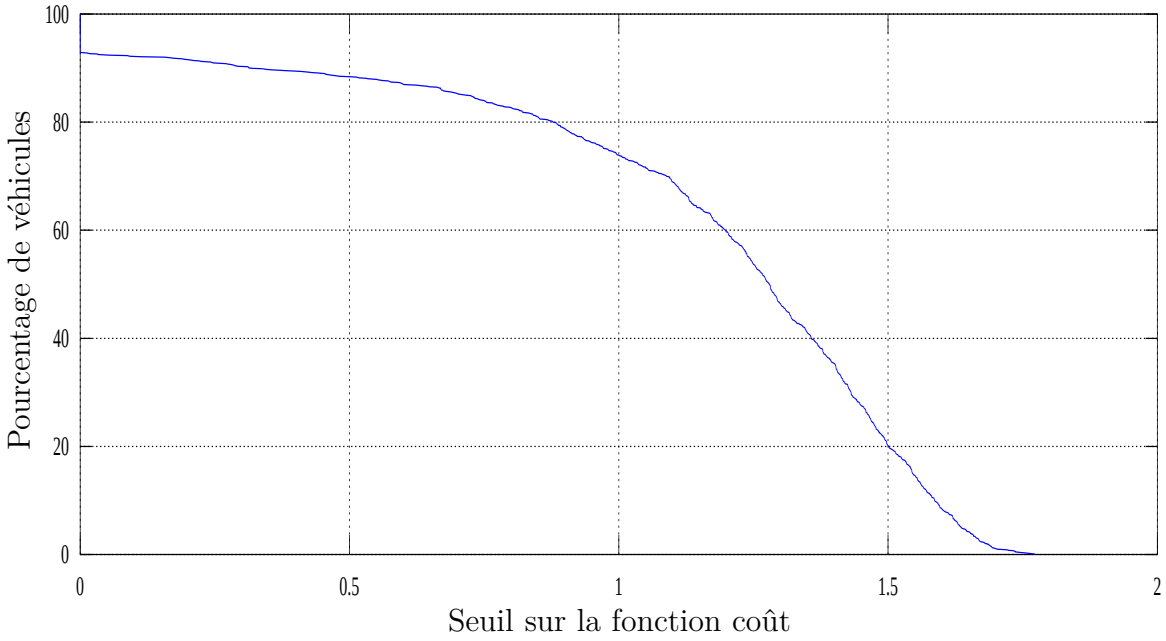


FIGURE 7.13 – Évolution du pourcentage de véhicules légers en fonction du seuil (SVM-d).

	90 %	80 %	70 %
Floue	8,14	8,3	8,37
MV	-13,7	-10	-8,1
SVM-d	0,80	1,06	1,23

Tableau 7.13 – Valeur de seuil pour un pourcentage donné de poids lourds (Site d'Angers).

origine, seul le couple dont la fonction coût (ou la distance à l'hyperplan) est maximum est conservé. Tous les autres couples citant la même origine sont éliminés.

Performances Les différents tableaux présentés dans cette sous-partie correspondent chacun à une configuration particulière des méthodes. Ces configurations sont les différents post-traitements proposés :

- filtrage par vote à l'unanimité (una),
- filtrage par seuil,
- filtrage des cas où des origines ont de multiples destinations.

Chaque tableau présente deux taux, le premier est le TBR, le deuxième est le TR. TBR donne la qualité des prédictions et TR informe de la quantité du trafic pouvant être analysée.

Performances témoins Cette configuration est simplement composée d'un traitement du type fenêtrage temporel avec les méthodes de réidentification. Par conséquent, les performances de cette situation représentent un élément de comparaison pour évaluer les bénéfices des traitements proposés.

Le tableau 7.14 montre que les TBR pour les PL sont très élevés surtout pour les deux cas SVM-d et vote à l'unanimité. De plus, les TR montrent que la quasi totalité des couples recensés sont prédits pour les poids lourds. Dans le cas des VL, il existe des TR légèrement inférieurs aux PL. Les TBR sont corrects dans le cas des méthodes seules. SVM-d permet d'obtenir pour le cas d'une méthode seule les meilleures performances. Les moins bons TBR des VL peuvent s'expliquer par le grand

		Floue	MV	SVM-d	una
PL	TBR (%)	90,6	91,3	98,5	98,5
	TR (%)	97,8	98,6	95,7	94,9
VL	TBR (%)	46,8	47,6	66,0	73,6
	TR (%)	90,7	92,3	87,3	84,3

Tableau 7.14 – Angers - Performances témoins.

nombre de perturbations. En effet, les méthodes du type MV ou Floue donnent une association avec une signature d'origine quelle que soit la signature destination, même pour une signature destination ne possédant pas d'origine. La méthode SVM-d utilise par définition un seuil hyperplan. Ainsi, dans certaines situations, la méthode permet déjà de filtrer les véhicules perturbateurs. A contrario, les bons taux des PL sont en partie dus aux faibles perturbations (peu de véhicules) de ce type de population.

Performances origine à multiples destinations Ce cas reprend la configuration « témoin ». Un filtre permet de supprimer les paires estimées ayant une origine à multiples destinations.

		Floue	MV	SVM-d	una
PL	TBR (%)	97,8	97,8	99,2	99,2
	TR (%)	97,8	98,4	95,7	94,9
VL	TBR (%)	79,9	82,6	84,1	90,0
	TR (%)	87,0	89,2	83,2	78,3

Tableau 7.15 – Angers - filtrage « Origine à multiples destinations ».

Le tableau 7.15 montre une très nette augmentation des TBR pour la population des VL comparée au tableau 7.14 mais en contrepartie une légère baisse des TR. Pour les PL, peu de différences de performance entre les deux configurations sont observées. La seule observation notable est l'augmentation des TBR pour les méthodes seules. Le filtrage permet d'éliminer un grand nombre de couples contenant des perturbations. Cependant, quelques couples corrects disparaissent aussi, comme le montrent les TR des VL sur le tableau 7.15 comparativement au TR des VL du tableau 7.14.

Performances avec seuil Dans cette section, les mauvaises associations sont filtrées par un seuil (ou critère d'acceptabilité) sur les fonctions coûts ou distance couple–hyperplan. Les trois tableaux suivants correspondent à trois ensembles de seuils différents présentés dans les tableaux 7.12 et 7.13 à la page 121.

		Floue	MV	SVM-d	una
PL	TBR (%)	99,2	98,4	99,1	99,1
	TR (%)	89,9	89,1	84,1	80,4
VL	TBR (%)	66,3	69,0	70,5	78,2
	TR (%)	85,6	85,9	85,8	80,1

Tableau 7.16 – Angers - Post-traitement seuil à 90 %.

Dans le tableau 7.16, les TBR des PL ont augmenté en comparaison au tableau 7.14. Cependant les TR ont légèrement diminué. La même observation peut être faite pour les VL. Les TBR des PL ne peuvent plus être augmentés de façon significative. Le tableau 7.17 montre les performances des quatre méthodes pour un seuil à 80 %. La population des PL n'est pas représentée dans ce tableau. Un gain des TBR en comparaison au tableau 7.16 est constaté. La baisse des TR est justifiée car les

		Floue	MV	SVM-d	una
VL	TBR (%)	77,3	78,1	78,9	84,4
	TR (%)	77,4	76,5	73,5	66,7

Tableau 7.17 – Angers - Post-traitement seuil à 80 %.

		Floue	MV	SVM-d	una
PL	TBR (%)	99,0	98,9	100	100
	TR (%)	68,8	68,1	44,9	38,4
VL	TBR (%)	81,1	82,2	81,0	86,6
	TR (%)	68,2	67,3	60,7	52,2

Tableau 7.18 – Angers - Post-traitement seuil à 70 %.

seuils sont obtenus tels que les TR soient ici limités à 80 %. Comme pour le tableau 7.17, dans le tableau 7.18 une faible augmentation des TBR est observée, et une baisse des TR.

Les trois tableaux 7.16, 7.17 et 7.18 mettent en valeur le compromis du seuil, qui nécessite de réduire le TR afin d'augmenter les performances (TBR). De plus, le taux de représentation est diminué par le vote à l'unanimité. Cependant, ce compromis est à relativiser au vu de la valeur du TBR avant compromis car ce dernier est déjà très haut. Par comparaison du tableau 7.18 avec le tableau 7.14, dans le cas du vote à l'unanimité, pour augmenter le TBR de la population PL de 1,5 %, il faut perdre plus de 50 % de TR.

Performances avec seuil et origines à multiples destinations La configuration « seuil et origines à multiples destinations » reprend la configuration « témoin » avec deux traitements post méthode, qui sont : seuil et origines à multiples destinations. Dans cette section, le cas des PL n'est pas présenté. En effet au vu des traitements antérieurs, ce cas testé dans cette section ne présente plus d'intérêt.

		Floue	MV	SVM-d	una
VL	TBR (%)	86,7	88,8	86,3	91,7
	TR (%)	66,5	65,9	58,8	49,5

Tableau 7.19 – Angers - Post-traitement seuil à 70 % et origines à multiples destinations.

En comparant le tableau 7.19 au tableau 7.18, une réduction des erreurs de prédiction est observée avec l'assemblage des deux méthodes. En contrepartie, le taux de représentation subit un léger recul.

Synthèse Suite aux résultats des trois méthodes utilisées (MV, floue, SVM-d) et afin de profiter des avantages de chacune d'elles, le vote à l'unanimité apparaît comme un choix intéressant, notamment pour éliminer une partie des perturbations. Néanmoins, cela demande de multiplier les calculs liés à la réidentification par le nombre de méthodes.

Le seuil est une solution logique au vu des informations à notre disposition telle la valeur de la fonction « coût ». Il permet d'éliminer des mauvais couples. Cependant, le compromis entre l'augmentation du taux d'identification et la limitation du taux de représentation demeure difficile à résoudre (en particulier le choix des valeurs de seuil, qui peuvent dépendre par exemple du nombre de variables utilisées). La solution « origines à multiples destination » possède de bonnes performances sans avoir de paramètre à définir. De plus, l'élimination des perturbations est ciblée.

7.3.4 Conclusion

Dans des situations réelles, nous avons constaté que toutes les signatures origine et destination ne sont pas associées entre elles. Nous sommes donc obligés de tenir compte de cette contrainte et de mettre en place des critères empêchant les réidentifications de ces signatures orphelines. L'introduction d'un seuil sur la mesure de similarité évite des erreurs de réidentification. Une autre solution consiste à utiliser le vote à l'unanimité pour éliminer en partie des erreurs de réidentification. Sur les bases de données testées, les deux méthodes ont permis d'améliorer le TBR. Nous avons remarqué que le vote à l'unanimité a à chaque fois donné un meilleur TBR que l'utilisation d'un seuil de décision. De plus, le vote à l'unanimité ne dépend pas de la détermination d'un seuil.

Nous avons aussi démontré qu'une sélection des variables en fonction du type de population obtient de meilleurs résultats en matière de TBR. L'utilisation des mêmes variables pour tous les types de population est donc préjudiciable à la réidentification.

Il reste cependant un problème. Le TBR dépend fortement du nombre de candidats. Nous avons eu recours à une fenêtre temporelle pour limiter le nombre de candidats des bases de données de test. Les résultats sont fortement dépendants de la détermination de la fenêtre temporelle fixe. Celle-ci revient à réaliser une modélisation du trafic simple. Un algorithme performant de modélisation du trafic va réduire le nombre de candidats en dynamique et améliorera par conséquent le TBR. Afin d'évaluer les méthodes de réidentification indépendamment d'un modèle de trafic, nous proposons dans la section suivante une nouvelle méthode d'évaluation.

7.4 Nouvelle méthode d'évaluation

7.4.1 Contexte

Pour s'affranchir de l'impact du nombre de candidats sur les résultats de l'évaluation, nous avons évalué la classification des couples origine – destination. La classification a pour objectif de déterminer si une signature origine comparée à une signature destination appartient à la classe identique ou non. À partir de la classification, les courbes de caractéristique de performance sont tracées et une analyse comparative des différents classifieurs est réalisée. Cependant, nous pensons que ces courbes ne reflètent pas entièrement les spécificités de la réidentification. Nous avons donc établi un nouvel indicateur pour l'évaluation de la réidentification en fonction du nombre de candidats.

Dans la suite, nous nous intéressons à la classification et à la réidentification de signaux issus de magnétomètres et de boucles inductives. Pour les boucles inductives, nous comparons le cas de signaux bruts avec celui de signaux déconvolués par la méthode décrite à la section 6.3. Les signaux employés sont ceux de la base de données SAROT2 présentée dans la section expérimentation. Pour le magnétomètre, ce sont les signaux de chaque axe qui seront utilisés séparément ou conjointement. Les signaux des magnétomètres proviennent de la base données du projet MOCoPo décrite dans la section expérimentation.

7.4.2 Classification et comparaisons

Boucle inductive

L'objectif est de comparer la signature sans déconvolution, à la signature avec déconvolution que la déconvolution soit effectuée à partir de la moyenne ou non de la fonction de transfert $\tilde{\mathbf{h}}$, dans le cadre de la classification.

La figure 7.14 montre que la signature estimée comporte plus d'informations sur les caractéristiques de la forme du signal que la signature sans déconvolution. En outre dans cette section, aucune méthode de sélection ou réduction de l'information n'est utilisée. Les mesures de similarité utilisées

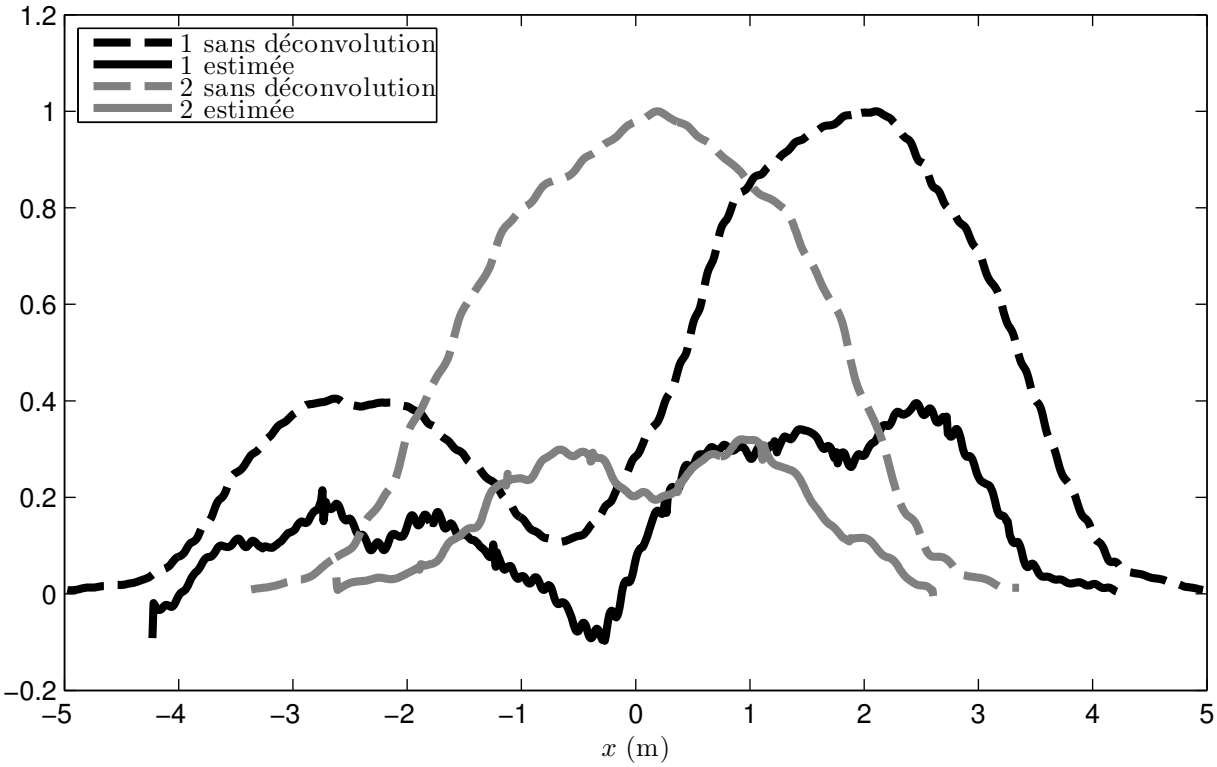


FIGURE 7.14 – Signatures sans et avec déconvolution pour les voitures de tourismes « 1 » et « 2 ».

pour la classification sont le coefficient de corrélation r (cf. équation (7.2)) entre la signature destination et les signatures origine ainsi que les distances de Canberra (cf. équation (7.25)), euclidienne (cf. équation (7.5)) et Manhattan (cf. équation (7.1)). Chaque algorithme donne une mesure de dissimilarité. Les conditions de trafic ne sont pas prises en compte de manière à n'évaluer que la qualité de l'information de la signature. De manière à illustrer la performance, les courbes de caractéristique de performance ou plus fréquemment appelées les courbes ROC (de l'anglais Receiver Operating Characteristic) sont tracées sur la figure 7.15 pour trois cas :

- à partir des signatures sans déconvolution,
- à partir des signatures estimées à l'aide de la fonction de transfert estimée à chaque véhicule,
- à partir des signatures estimées à l'aide de la moyenne des fonctions de transfert estimées d'une base d'apprentissage.

Les figures 7.15 montrent que les courbes ROC tracées à partir des signatures estimées échantillonnées à 64 points sont très proches l'une de l'autre alors que la courbe ROC issue des signatures sans déconvolution échantillonnées à 64 points est clairement en-dessous. Ceci est vérifié pour les quatre approches.

Le tableau 7.20 exprime l'aire en-dessous de la courbe (AUC : Area Under Curve) en fonction de la mesure de dissimilarité utilisée, de la fréquence d'échantillonnage et du signal. L'aire en-dessous de la courbe pour les signatures sans déconvolution varie entre 0,9122 et 0,9375 suivant la mesure de dissimilarité et la fréquence d'échantillonnage utilisées. Sans déconvolution, la classification avec la distance de Canberra obtient les meilleurs résultats quelle que soit la fréquence d'échantillonnage utilisée. En appliquant la déconvolution, les performances sont améliorées de manière significative lors de cette expérimentation, pour toutes les mesures de dissimilarité et quelle que soit la fréquence d'échantillonnage. La meilleure AUC de 0,9633 est obtenue pour le signal estimé avec la distance euclidienne à partir de 512 points. Les AUC sont sensiblement supérieures dans le cas des fonctions de

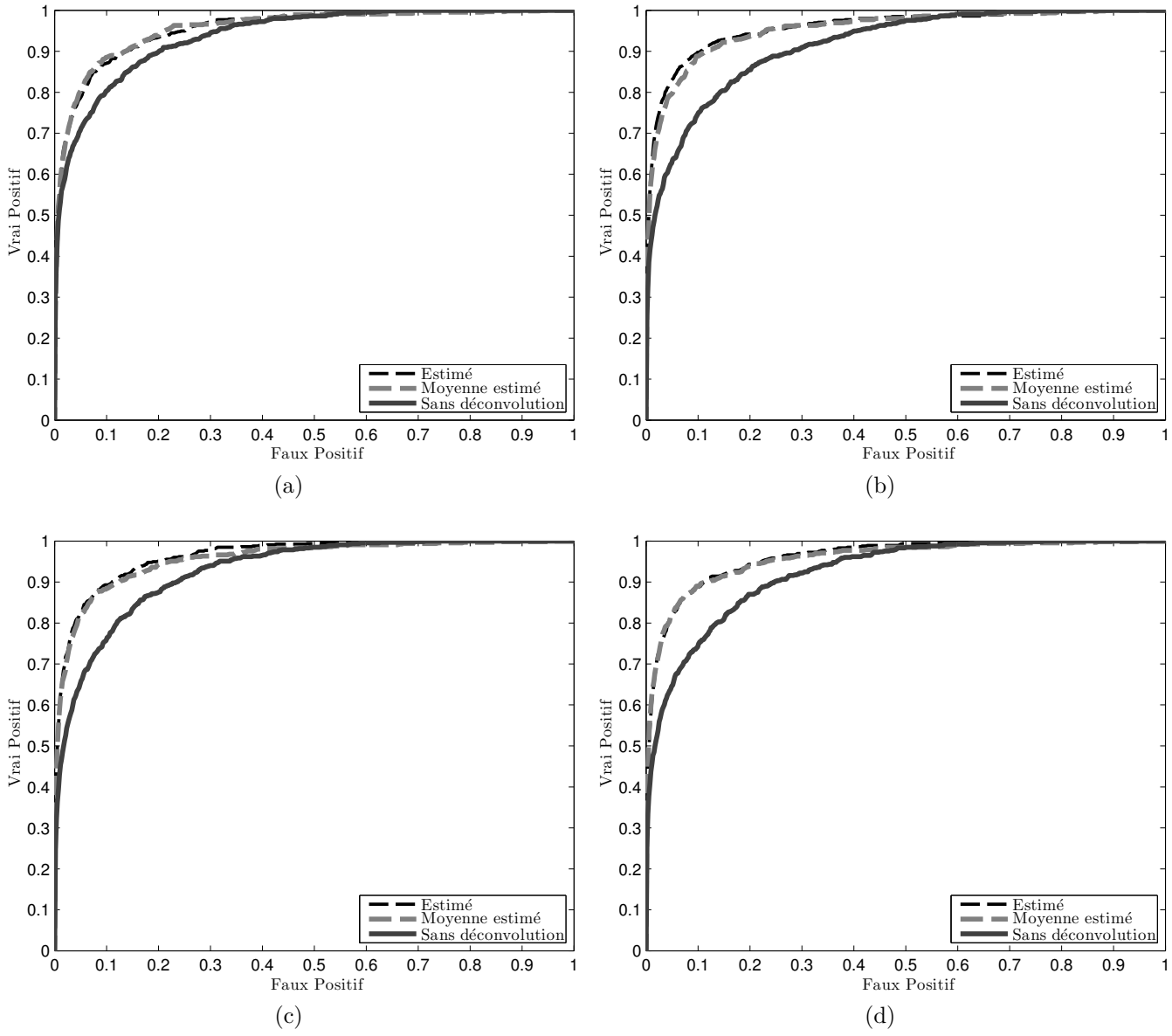


FIGURE 7.15 – Courbes ROC pour les signatures en 64 points : sans déconvolution, estimées par $\tilde{\mathbf{h}}$ et estimées à partir de la moyenne de $\tilde{\mathbf{h}}$: (a) Distance de Canberra ; (b) coefficient de corrélation ; (c) Distance Euclidienne ; (d) Distance de Manhattan.

		Nombre de points						
Distance	Signal	16	32	64	128	256	512	1024
Canberra	Sans déconvolution	0,9338	0,9365	0,9376	0,9375	0,9375	0,9375	0,9375
	Estimé	0,9471	0,9523	0,9546	0,9554	0,9553	0,9554	0,9555
	Estimé par la moyenne	0,9397	0,9494	0,9551	0,9564	0,9568	0,9569	0,9568
Corrélation	Sans déconvolution	0,9122	0,9134	0,9145	0,9242	0,9142	0,9142	0,9142
	Estimé	0,9513	0,9561	0,9572	0,9580	0,9579	0,9578	0,9578
	Estimé par la moyenne	0,9363	0,9480	0,9531	0,9551	0,9550	0,9551	0,9548
Euclidienne	Sans déconvolution	0,9250	0,9256	0,9268	0,9267	0,9267	0,9267	0,9267
	Estimé	0,9559	0,9603	0,9625	0,9631	0,9631	0,9633	0,9633
	Estimé par la moyenne	0,9330	0,9485	0,9553	0,9578	0,9583	0,9585	0,9584
Manhattan	Sans déconvolution	0,9181	0,9201	0,9212	0,9210	0,9210	0,9210	0,9210
	Estimé	0,9541	0,9574	0,9591	0,9598	0,9596	0,9598	0,9598
	Estimé par la moyenne	0,9406	0,9521	0,9555	0,9569	0,9571	0,9573	0,9572

Tableau 7.20 – AUC des courbes en fonction de la fréquence d'échantillonnage et de la distance utilisée.

transfert estimées à chaque passage de véhicule comparé au cas d'une fonction de transfert utilisée pour l'ensemble des véhicules. Cette constatation est inversée pour les AUC des signaux dont le nombre de points est supérieur ou égal à 64 points avec la distance de Canberra. Afin de limiter les calculs, de réaliser des applications temps réel et à la vue des résultats semblables pour les signaux déconvolués, l'utilisation d'une fonction de transfert pour l'ensemble des véhicules est à privilégier.

Magnétomètre

Nous avons utilisé la première demi-heure de la première session d'expérimentation du projet MOCoPo. La base de données, pour cette vérification, est constituée des signatures détectées par l'algorithme $Detect_{Modif}$ présenté à la section 5.3 et évalué à la section 5.4. Seules les signatures non tronquées sont conservées dans la base de données. Les signatures tronquées sont identifiées lors de la détection. Elles sont considérées comme tronquées lorsque nous ne sommes pas certains de détenir le début ou la fin de la signature. Nous éliminons aussi de la base de données toutes les signatures destination ne possédant pas de signature origine associée et réciproquement. Au final, la base de données est composée de 927 signatures destination et de 927 signatures origine, chaque signature destination étant associée avec une signature origine.

Les signatures retenues sont ensuite normalisées en amplitude par la formule suivante :

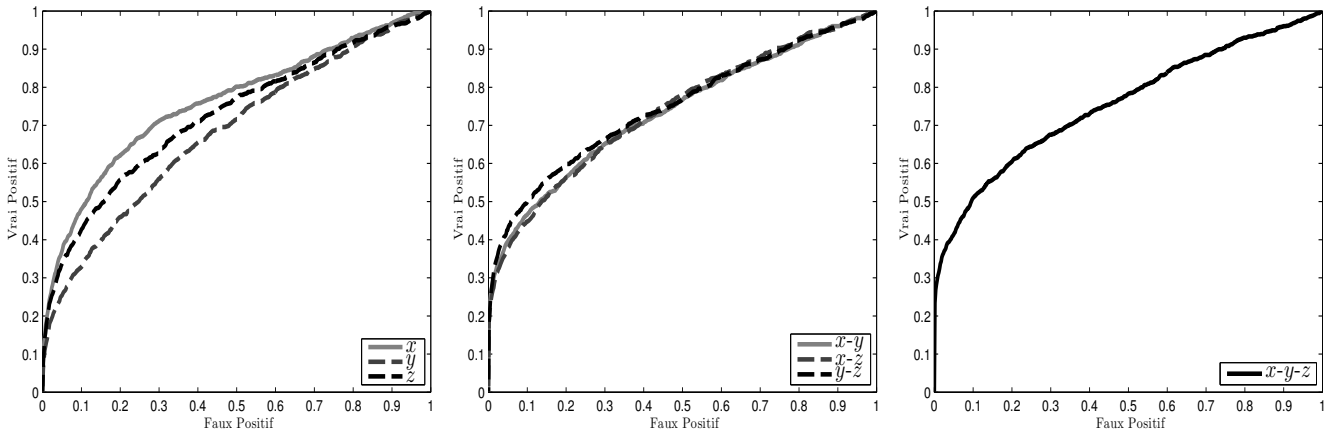
$$\mathbf{x} = \frac{\mathbf{x_c}}{\max(\mathbf{x_c}) - \min(\mathbf{x_c})} \tag{7.42}$$

avec $\mathbf{x_c}$ la signature selon l'axe x centrée sur la valeur zéro. La même relation est adoptée pour les axes y et z.

La première approche exploite la déformation temporelle dynamique (DTW) exposée à la page 102 avec la distance euclidienne pour réaliser la classification. Sept possibilités sont comparées entre-elles pour le calcul de la distance euclidienne : l'utilisation d'un seul axe x, y ou z ; l'utilisation de deux axes x-y, x-z ou y-z ; l'utilisation des trois axes x-y-z. Les courbes ROC sont représentées sur la figure 7.16 et les aires en-dessous des courbes (AUC) sont données dans le tableau 7.21.

	x	y	z	x-y	x-z	y-z	x-y-z
AUC	75,83	67,92	72,49	73,62	74,80	73,76	75,55

Tableau 7.21 – Aire en-dessous de la courbe pour la classification par DTW (distance euclidienne).



(a) (b) (c)

FIGURE 7.16 – Classification avec les magnétomètres par DTW (distance euclidienne) : (a) : selon un axe x, y ou z ; (b) : selon deux axes x-y, x-z ou y-z ; (c) : selon trois axes x-y-z.

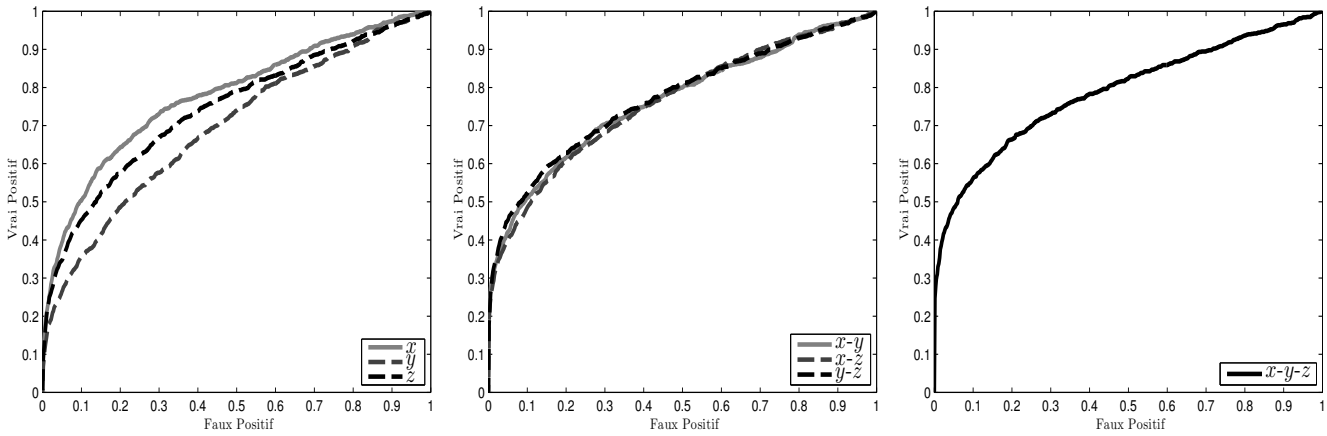
La courbe ROC de l'axe x est au-dessus des autres courbes pour la classification à partir d'un seul axe (figure 7.16a). Cette constatation se traduit dans le tableau 7.21 par une AUC supérieure à celle des axes y et z. La figure 7.16b montre des courbes très proches les unes des autres. En deux dimensions, la courbe x - z obtient la meilleure courbe ROC, mais l'AUC reste inférieure à celle de l'axe x. Cependant, le début de la courbe pour l'axe x révèle qu'environ 20 % des couples origine – destination peuvent être classés correctement en commettant peu d'erreur. Avec la courbe y - z, plus de 20 % des couples origine – destination peuvent être classés correctement en commettant peu d'erreur. La figure 7.16c indique qu'environ 30 % des couples origine – destination peuvent être classés correctement en commettant peu d'erreur. Au final, même si l'axe x donne *a priori* le plus d'information et obtient l'AUC la plus importante, l'association des trois axes semblent produire un meilleur résultat pour réaliser le suivi de véhicules par la suite.

La figure 7.17 et le tableau 7.22 montrent les résultats dans le cadre de la DTW avec la distance de Manhattan. Une grande similarité des courbes obtenues entre les figures 7.16a et 7.17a (7.16b et 7.17b, 7.16c et 7.17c) est observable. La différence entre les deux distances se remarque lors de la comparaison des tableaux 7.21 et 7.22. La distance de Manhattan donne une AUC supérieure pour chaque situation référencée. L'axe x obtient la meilleure AUC avec la distance euclidienne. Tandis qu'en utilisant la distance de Manhattan, c'est l'ensemble des trois axes x - y - z qui acquière la meilleure AUC.

	x	y	z	x-y	x-z	y-z	x-y-z
AUC	77,71	69,24	74,27	76,49	77,10	76,09	78,86

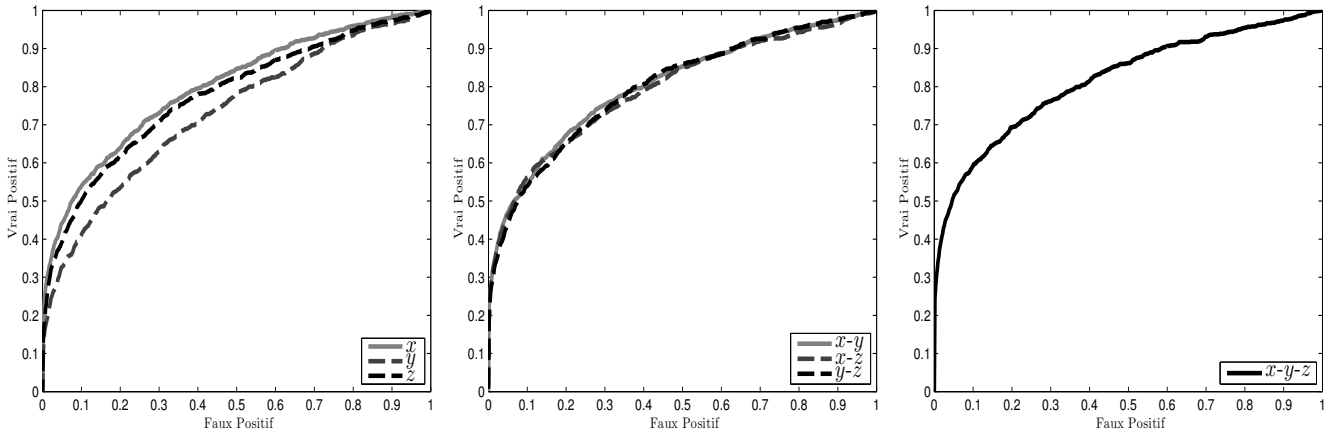
Tableau 7.22 – Aire en-dessous de la courbe pour la classification par DTW (distance de Manhattan).

Une autre solution est proposée pour s'affranchir de la déformation temporelle des signatures : la normalisation en nombre de points par une interpolation spline. Le nombre de points a été fixé arbitrairement à 64 points. La signature est dans un premier temps normalisée en 64 points, puis dans un second temps normalisée en amplitude suivant l'équation (7.42). La classification est élaborée avec la distance de Manhattan. Lorsque plusieurs axes sont utilisés, la somme des distances de Manhattan des différents axes est calculée. Cette somme représente la mesure de similarité employée pour la classification. La forme des courbes ROC de la figure 7.18 reste similaire à celle des figures précédentes 7.16 et 7.17. Les résultats sont en légère amélioration et sont confirmés par le tableau 7.23.



(a) (b) (c)

FIGURE 7.17 – Classification avec les magnétomètres par DTW (distance de Manhattan) : (a) : selon un axe x, y ou z ; (b) : selon deux axes x-y, x-z ou y-z ; (c) : selon trois axes x-y-z.



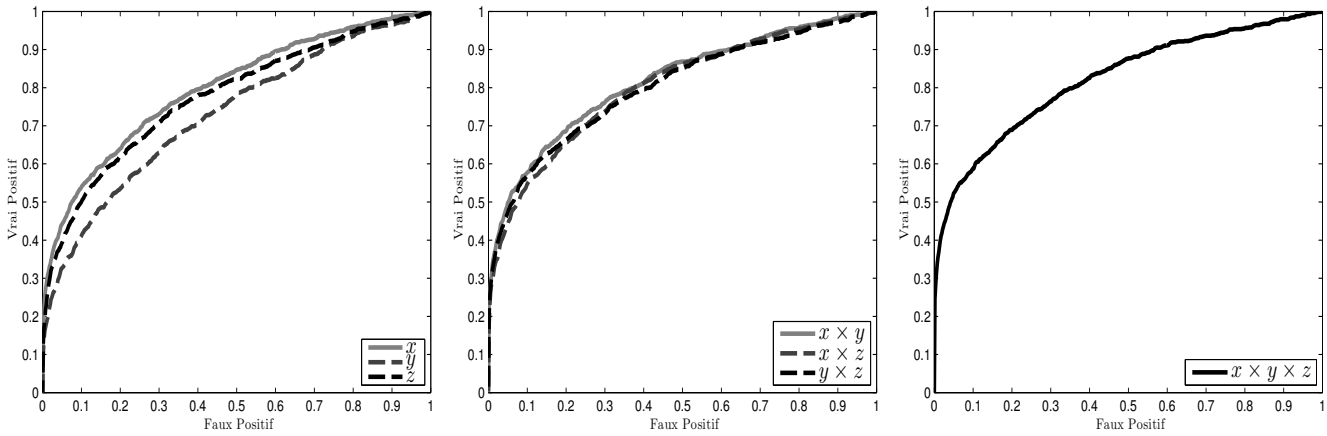
(a) (b) (c)

FIGURE 7.18 – Classification avec les magnétomètres par l'addition des distances de Manhattan après normalisation des signatures en 64 points : (a) : selon un axe x, y ou z ; (b) : selon deux axes x-y, x-z ou y-z ; (c) : selon trois axes x-y-z.

	x	y	z	x-y	x-z	y-z	x-y-z
AUC	80,05	73,29	77,81	80,86	80,25	79,99	82,14

Tableau 7.23 – Aire en-dessous de la courbe pour la classification par l'addition des distances de Manhattan après normalisation des signatures en 64 points.

Au lieu d'additionner les distances de Manhattan des différents axes, il est proposé de procéder aux produits des distances de Manhattan. Le produit des distances donne un poids plus fort pour les distances proches de zéro. La forme des courbes ROC obtenue sur la figure 7.19 est semblable à celles des autres figures 7.16, 7.17 et 7.18. Une amélioration minime est perceptible en analysant le tableau 7.23 et 7.24. La meilleure solution de classification, du point de vue de l'AUC, est obtenue avec la normalisation en nombre de points et en amplitude, et le produit des distances de Manhattan des trois axes. Cependant, la classification entre la classe « identique » et « différent » reste difficile. La meilleure AUC de 82,60 obtenue avec les magnétomètres pour la réidentification est largement inférieure aux AUC obtenues avec les boucles inductives. L'apparence de la courbe ROC de la figure 7.19c montre que les deux classes sont difficiles à séparer avec une augmentation rapide du taux de faux positifs vis-à-vis d'un taux de vrais positifs supérieur à 0,4.



(a) (b) (c)

FIGURE 7.19 – Classification avec les magnétomètres par la multiplication des distances de Manhattan après normalisation des signatures en 64 points : (a) : selon un axe x, y ou z ; (b) : selon deux axes $x \times y$, $x \times z$ ou $y \times z$; (c) : selon trois axes $x \times y \times z$.

	x	y	z	$x \times y$	$x \times z$	$y \times z$	$x \times y \times z$
AUC	80,05	73,29	77,81	82,04	80,46	80,55	82,60

Tableau 7.24 – Aire en-dessous de la courbe pour la classification par la multiplication des distances de Manhattan après normalisation des signatures en 64 points.

7.4.3 Réidentification et comparaisons

La classification permet de déterminer si un couple origine – destination appartient à la classe « identique » ou « différent ». Pour un même véhicule destination, la classification peut prédire que plusieurs véhicules origine appartiennent à la classe « identique » sans déterminer l'association origine – destination la plus vraisemblable. L'objectif de la réidentification est de déterminer le couple origine – destination dont la distance de dissimilarité est minimale parmi un ensemble de couple origine – destination possible. La réidentification se différencie de la classification en ne conservant au maximum qu'un seul couple origine – destination. Le couple origine – destination considéré comme « identique » est celui dont la mesure de dissimilarité est minimale parmi l'ensemble des couples.

Présentation du cas simple avec la boucle inductive

Dans le cas présent, nous avons une base de données composée de 654 signatures origine et 654 signatures destination. Les mesures de dissimilarité sont calculées pour l'ensemble des couples possibles afin de créer une matrice de dimension 654×654. Cette matrice est donc composée de 654 couples « identique » sur la diagonale et de $654 \times 653 = 427062$ couples « différent ». La matrice $\mathbf{M_c}$ représente l'ensemble des mesures de dissimilarité en utilisant la distance de Canberra entre les couples origine – destination. Les véhicules destinations sont sur les lignes de cette matrice et les véhicules origine sont sur les colonnes de cette matrice. La distance de Canberra $d_c(i,j)$ représente la distance entre la i^{e} signature destination et la j^{e} signature origine. Lorsque $i = j$, les signatures origine et destination appartiennent au même véhicule.

$$\mathbf{M_C} = \begin{pmatrix} d_C(1,1) & d_C(1,2) & \dots & d_C(1,654) \\ d_C(2,1) & d_C(2,2) & \dots & d_C(2,654) \\ \vdots & \vdots & & \vdots \\ d_C(654,1) & d_C(654,2) & \dots & d_C(654,654) \end{pmatrix} \tag{7.43}$$

Les matrices $\mathbf{M_E}$, $\mathbf{M_M}$ et $\mathbf{M_r}$ sont construites de façon identique avec comme mesure de dissimilarité la distance euclidienne, la distance de Manhattan et le coefficient de corrélation, respectivement. Nous prendrons l'exemple de la matrice $\mathbf{M_C}$, mais le raisonnement s'applique aussi aux autres matrices. Nous proposons pour évaluer la réidentification de comparer le couple origine – destination réel « identique » à un ou plusieurs autres couples origine – destination appelés candidats supplémentaires. Les candidats supplémentaires possèdent la même signature destination que le couple origine – destination réel « identique » mais par contre, la signature origine est différente. Pour évaluer le pourcentage de bonne réidentification, une première approche est de tirer les candidats supplémentaires au sort. Parmi les candidats supplémentaires, nous vérifions si la réidentification est correcte ou non. Pour avoir un résultat significatif et non biaisé, le tirage au sort doit être effectué plusieurs fois. Par exemple, nous allons décrire les modalités d'estimation du TBR pour un candidat supplémentaire. La distance de référence pour le premier véhicule est la distance $d_c(1,1)$ qui représente la bonne association. Le candidat supplémentaire est tiré au sort sur la même ligne de $\mathbf{M_C}$ et a une distance $d_c(1,j)$ avec $j = 2 \dots 654$. Si $d_c(1,1) < d_c(1,j)$ alors la réidentification est correcte. Dans le cas inverse, nous commettons une erreur de réidentification. Le tirage au sort est effectué n fois et nous comptabilisons le nombre de bonnes réidentifications pour estimer le TBR du premier véhicule. Ce processus est réalisé pour l'ensemble des lignes de $\mathbf{M_C}$, ainsi le TBR est évalué pour chaque ligne. Le TBR retenu pour un candidat supplémentaire est la moyenne des TBR de chaque ligne. Cette démarche est appliquée pour plusieurs candidats supplémentaires. La réidentification est considérée incorrecte si au moins l'un des candidats supplémentaires a une distance inférieure à la distance $d_c(i,i)$ de la bonne réidentification.

Une deuxième approche consiste à dénombrer le nombre d'éléments suivant les différents cas possibles. Cette approche équivaut à réaliser un nombre de tirages au sort infini pour la première approche. Cette approche est possible car les différents cas sont simples à identifier et à dénombrer. Pour estimer le taux de bonne réidentification par la deuxième approche, chaque ligne de la matrice est divisée en trois sous-ensembles : le sous-ensemble E_{id} composé d'un unique élément $\mathbf{M_C}(i,i)$; le sous-ensemble E_{diff} composé de l'ensemble des éléments de la ligne i dont $d_C(i,j) > d_C(i,i)$ avec $j = 1 \dots 654$ et $j \neq i$; et le sous-ensemble E_{err} composé de l'ensemble des éléments de la ligne i dont $d_C(i,j) \leqslant d_C(i,i)$ avec $j = 1 \dots 654$ et $j \neq i$. E_{err} représente l'ensemble des véhicules qui provoque une erreur de réidentification et est composé de n_{err} éléments. E_{diff} représente l'ensemble des véhicules qui ne pose pas de difficulté à la bonne réidentification et est composé de n_{diff} éléments. La bonne réidentification de la i^{e} signature dépend des différents sous-ensembles et du nombre de candidats supplémentaires n_{cand}. Pour comparer l'ensemble E_{id} à n_{cand}, le nombre possible de combinaisons

c_{total} est de :

$$c_{total} = \frac{(n_{diff} + n_{err})!}{n_{cand}!(n_{diff} + n_{err} - n_{cand})!} \tag{7.44}$$

Le nombre de combinaisons c_{cand} appartenant à l'ensemble E_{diff} en fonction du nombre de candidats supplémentaires est de :

$$c_{cand} = \frac{n_{diff}!}{n_{cand}!(n_{diff} - n_{cand})!} \tag{7.45}$$

La probabilité qu'une bonne réidentification se réalise en fonction du nombre de candidats supplémentaires est de :

$$p_{c,cand} = \frac{c_{cand}}{c_{total}} \tag{7.46}$$

Le TBR pour chaque ligne et un nombre de candidats donné est équivalent à $p_{c,cand}$. Le TBR retenu pour un nombre de candidats supplémentaires est la moyenne des $p_{c,cand}$.

La figure 7.20 montre les taux de bonne réidentification en fonction du nombre de candidats et de la mesure de dissimilarité utilisée pour les signaux avec ou sans déconvolution. Comme le montre la figure 7.20a, la distance de Canberra donne les meilleurs résultats pour des signaux sans déconvolution. Dans le cadre de signaux sans déconvolution, les autres mesures de dissimilarité ont des courbes pratiquement similaires et en-dessous de celle de la distance de Canberra. La distance de Canberra est particulièrement sensible pour des valeurs de la signature proches de zéro, ces valeurs se trouvent en début et en fin de signature. Le début et la fin de la signature sont les parties les moins affectées par convolution entre la fonction de transfert de la boucle inductive et le signal d'entrée. En effet, lors de son passage au-dessus de la boucle inductive, le véhicule ne recouvre pas la totalité de la largeur de la boucle inductive lors de son entrée et de sa sortie.

La figure 7.20b montre que le coefficient de corrélation est la meilleure mesure de dissimilarité parmi celles testées pour la réidentification à partir du signal estimé par $\tilde{\mathbf{h}}$. Les autres mesures de dissimilarités sont toutes inférieures.

La figure 7.20c montre des différences moindres entre les quatre mesures de dissimilarité pour les signaux estimés à partir de la moyenne de $\tilde{\mathbf{h}}$. La distance de Canberra obtient des performances moindres que les autres mesures. La distance de Manhattan a des performances légèrement supérieures au coefficient de corrélation et à la distance euclidienne.

Lors de la comparaison des performances de réidentification sur la figure 7.20d pour les trois types de signaux différents, les performances de réidentification pour les signaux sans déconvolution sont inférieures aux signaux déconvolués. Pour un candidat supplémentaire, la différence de performance entre les signaux avec et sans déconvolution est de l'ordre de 3,35. Cet écart s'accentue lorsque le nombre de candidats supplémentaires augmente. Une différence de l'ordre de 12,3 points est atteinte pour 50 candidats supplémentaires. Les signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$ ont des performances légèrement supérieures aux signaux déconvolués en calculant $\tilde{\mathbf{h}}$ à chaque fois. Cette différence est inférieure de 0,5 point. Au final, ce sont les signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$ avec une distance de Manhattan qui présentent le meilleur pourcentage de bonne réidentification.

Présentation lors du vote à l'unanimité avec la boucle inductive

Dans cette section, le calcul de plusieurs mesures de dissimilarité est effectué simultanément sur les signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$. La réidentification est validée lorsque l'ensemble des mesures de dissimilarité donne le même résultat. Lorsque les mesures de dissimilarité donnent des résultats différents, la signature destination n'est pas associée à une signature origine. L'ensemble des signatures n'est pas réidentifié. Pour évaluer les approches avec vote à l'unanimité, le TBR et le TR sont calculés. Cette fois-ci, les sous-ensembles sont difficiles à dénombrer et l'approche par tirage au sort est utilisée pour estimer les TBR et TR. Pour chaque signature destination, les moyennes sur mille tirages au sort pour chaque séquence de un à cinquante candidats supplémentaires sont effectuées. La

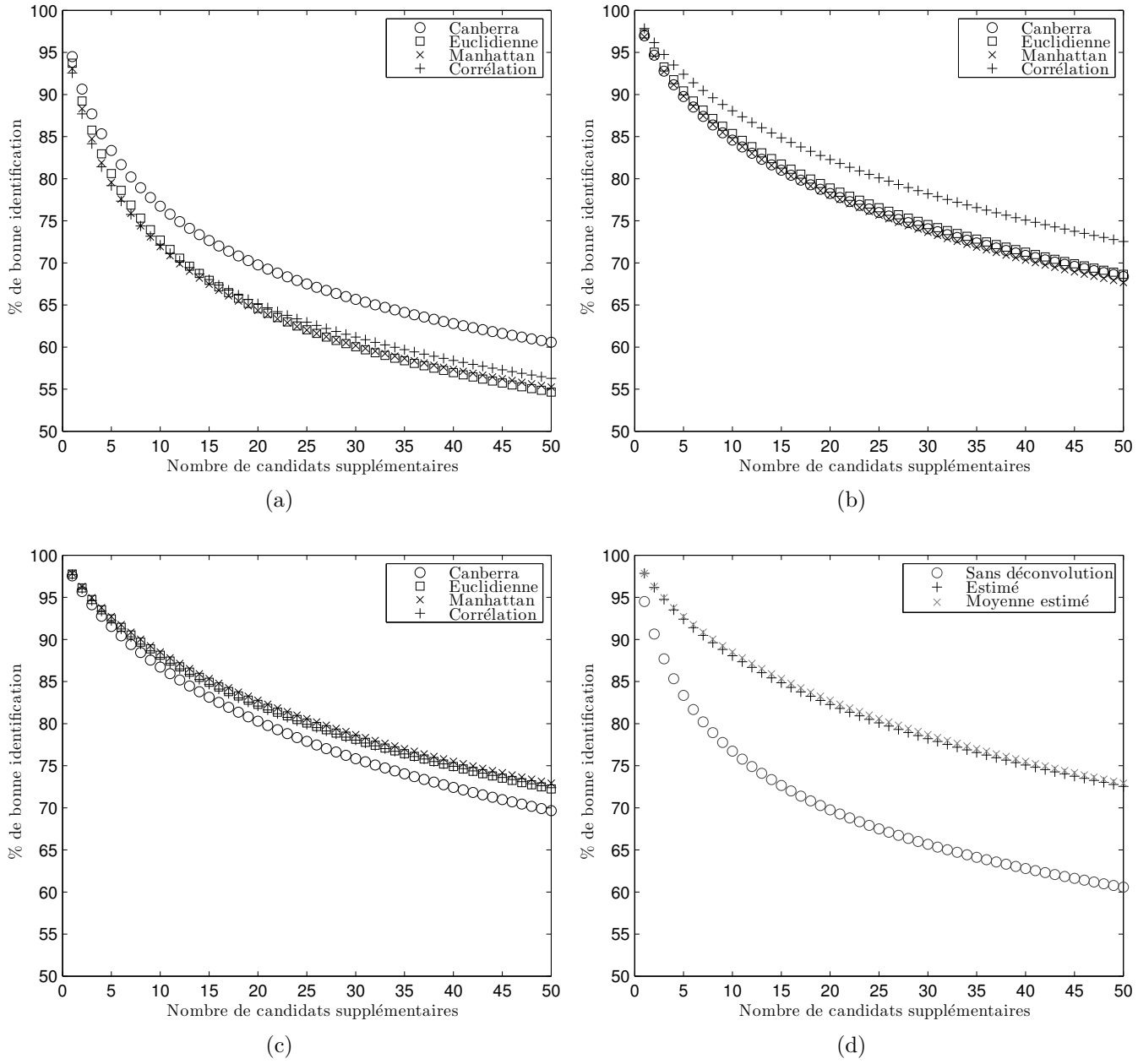
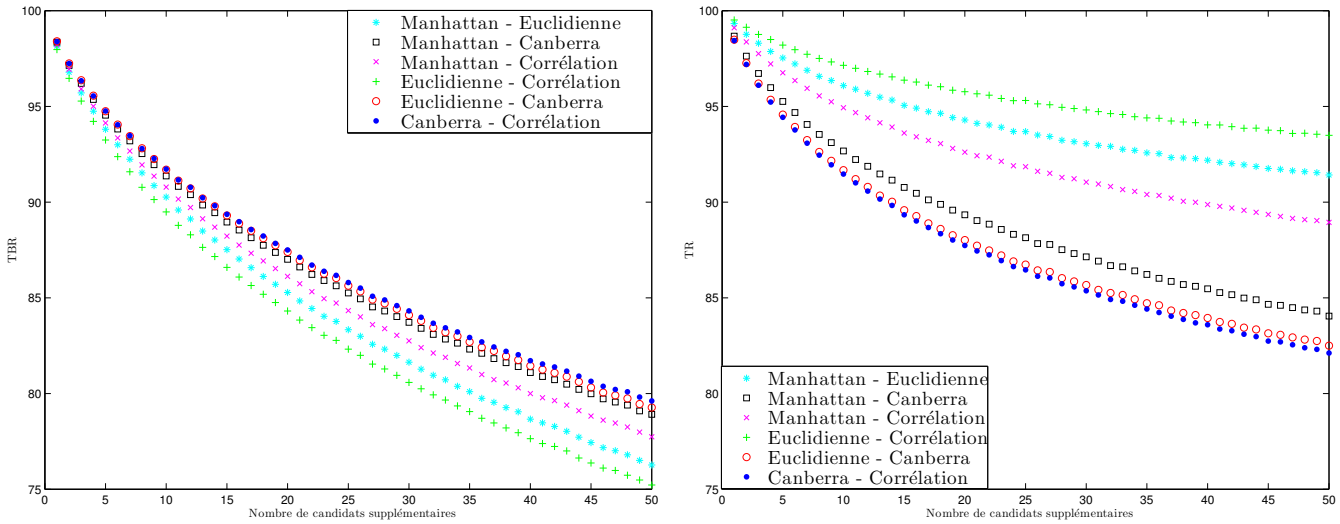


FIGURE 7.20 – Taux de bonne réidentification en fonction du nombre de candidats supplémentaires et des distances utilisées : (a) Signal sans déconvolution ; (b) Signal estimé par $\tilde{\mathbf{h}}$; (c) Signal estimé à partir de la moyenne de $\tilde{\mathbf{h}}$; (d) Distance de Canberra pour le signal sans déconvolution, coefficient de corrélation pour le signal estimé par $\tilde{\mathbf{h}}$ et distance de Manhattan pour le signal estimé à partir de la moyenne de $\tilde{\mathbf{h}}$.

figure 7.21 montre le TBR et le TR pour les votes à l'unanimité entre deux mesures de dissimilarité différentes. Quelles que soient les mesures de dissimilarité associées, la figure 7.21a montre que les TBR ont augmenté. Les TBR sont tous supérieurs à 75 % pour 50 candidats supplémentaires alors que l'utilisation d'une seule mesure de dissimilarité est inférieure à cette valeur (cf. figure 7.20). La courbe de TBR de la distance euclidienne avec le coefficient de corrélation est en-dessous des autres courbes. Ce sont les trois courbes dont l'une des mesures de dissimilarité est la distance de Canberra qui présentent les meilleurs TBR. La distance de Canberra avec le coefficient de corrélation affiche les meilleurs TBR sur la figure 7.21a. L'amélioration des TBR se fait au dépend du TR. Les courbes présentant les meilleurs TBR sont celles qui montrent les TR les plus faibles sur la figure 7.21b. Pour 50 candidats supplémentaires, le vote à l'unanimité entre la distance euclidienne et le coefficient de corrélation obtient un TBR de 75,22 % et un TR de 93,49 %. Toujours pour 50 candidats supplémentaires, le vote à l'unanimité entre la distance de Canberra et le coefficient de corrélation obtient un TBR de 79,61 % et un TR de 82,12 %. L'augmentation du TBR est moins importante que la diminution du TR.

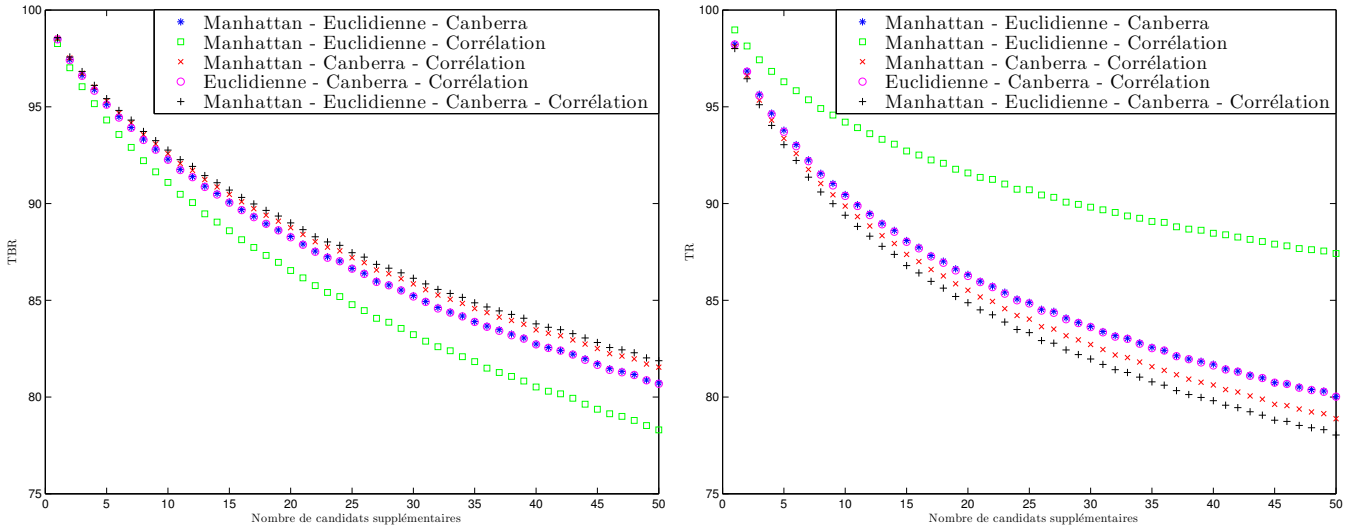


(a) (b)

FIGURE 7.21 – Utilisation de deux mesures de dissimilarité pour estimer le TBR (a) et le TR (b)) en fonction du nombre de candidats supplémentaires avec le vote à l'unanimité pour les signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$.

Comme le montre la figure 7.22, nous avons aussi effectué les estimations des TBR et TR pour un vote à l'unanimité sur trois et quatre mesures de dissimilarité différentes. À part pour l'association des distances de Manhattan, euclidienne et du coefficient de corrélation, les TBR sont améliorés par rapport aux votes à l'unanimité à partir de deux mesures de dissimilarité. Le TBR pour l'association des distances de Manhattan, euclidienne et du coefficient de corrélation est de 78,3 % pour 50 candidats supplémentaires. Il est donc légèrement inférieur à celui de l'association de la distance de Canberra et du coefficient de corrélation, néanmoins le TR pour l'association des trois mesures de dissimilarité est de 87,41 % contre 82,12 % pour les deux mesures de dissimilarité. Les courbes des TBR sur la figure 7.22a pour les associations « Manhattan - Euclidienne - Canberra » et « Euclidienne - Canberra - Corrélation » se superposent. Il en est de même pour les courbes de TR sur la figure 7.22b pour ces deux associations. Nous remarquons sur la figure 7.22a que l'association des quatre mesures de dissimilarité obtient les meilleurs TBR. Hormis pour l'association de mesures de dissimilarité « Manhattan - Euclidienne - Corrélation », les quatre autres associations (« Manhattan - Euclidienne - Canberra », « Manhattan - Canberra - Corrélation », « Euclidienne - Canberra - Corrélation » et « Manhattan - Euclidienne - Canberra- Corrélation ») ont des TBR similaires. Pour 50 candidats supplémentaires, les TBR sur la figure 7.22a se situent entre 80,69 % et 81,87 %. Ces

différences se réduisent pour un nombre de candidats supplémentaires inférieur à 50. Les différences entre les TR pour ces quatre associations sur la figure 7.22b sont plus marquées, et se situent entre 78,04 % et 80,20 %. De la même manière que pour les TBR, les différences entre les TR décroissent avec la diminution du nombre de candidats supplémentaires.



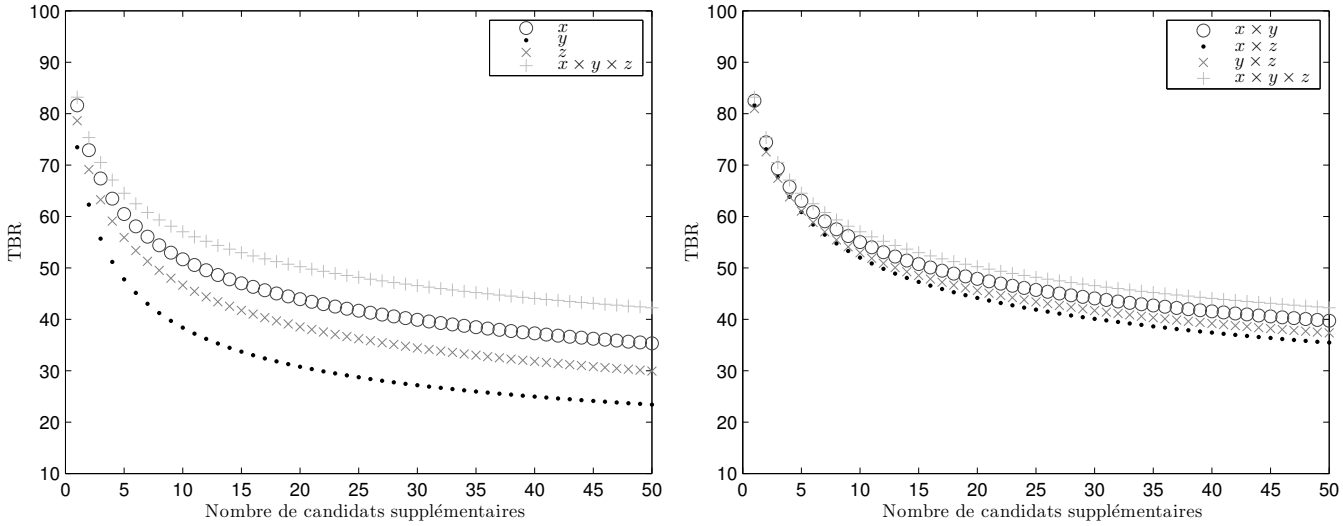
(a) (b)

FIGURE 7.22 – Utilisation de trois ou quatre mesures de dissimilarité pour estimer le TBR (a) et le TR (b)) en fonction du nombre de candidats supplémentaires avec le vote à l'unanimité pour les signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$.

Application au magnétomètre

La base de données étudiée comporte 927 véhicules représentés par une signature destination et une signature origine. Les 927 associations de ces signatures origine et destination forment la classe « identique ». La classe « différent » est conçue en combinant les signatures destination avec les signatures origine qui ne correspondent pas au même véhicule. $927 \times 926 = 858\,402$ couples de la classe « différent » sont obtenus. L'évaluation du TBR est effectuée en fonction du nombre de candidats supplémentaires. Pour cela la dissimilarité du couple de classe « identique » est comparée à la dissimilarité des autres couples possédant la même signature destination de la classe « différent ». La méthode utilisée pour effectuer le calcul est décrite dans la section 7.4.3 (page 131). La mesure de dissimilarité utilisée est la distance de Manhattan, et le produit des distances est calculé pour l'utilisation de plusieurs axes. La figure 7.23 présente les TBR en fonction du nombre de candidats supplémentaires. La combinaison de plusieurs axes améliore le TBR, les meilleurs résultats sont obtenus avec les trois axes ensembles. Le TBR diminue rapidement avec le nombre de candidats supplémentaires. Avec seulement un candidat supplémentaire et en disposant des mesures sur les trois axes, le TBR est seulement de 83,22 % et au-delà de 20 candidats supplémentaires, il est inférieur à 50 %.

Pour améliorer le TBR, nous proposons de calculer la distance de Manhattan pour chacun des trois axes. Nous effectuons ensuite un vote à l'unanimité entre deux axes ou les trois axes. Le couple origine – destination est uniquement sélectionné si l'ensemble des axes du couple présente la distance minimale parmi l'ensemble des candidats. Si aucun couple origine – destination ne satisfait cette condition, aucune association entre l'origine et la destination n'est effectuée. L'évaluation s'effectue selon le processus décrit à la section 7.4.3 (page 133). Pour chaque signature destination, les moyennes sur mille tirages au sort pour chaque séquence de un à cinquante candidats supplémentaires sont calculées. La figure 7.24 présente les résultats obtenus pour le vote à l'unanimité pour les différentes



(a) (b)

FIGURE 7.23 – Taux de bonne réidentification en fonction du nombre de candidats supplémentaires en utilisant la distance de Manhattan : (a) : pour les axes pris individuellement et le produit des distances des trois axes ; (b) : pour le produit des distances des axes pris deux à deux et le produit des distances des trois axes.

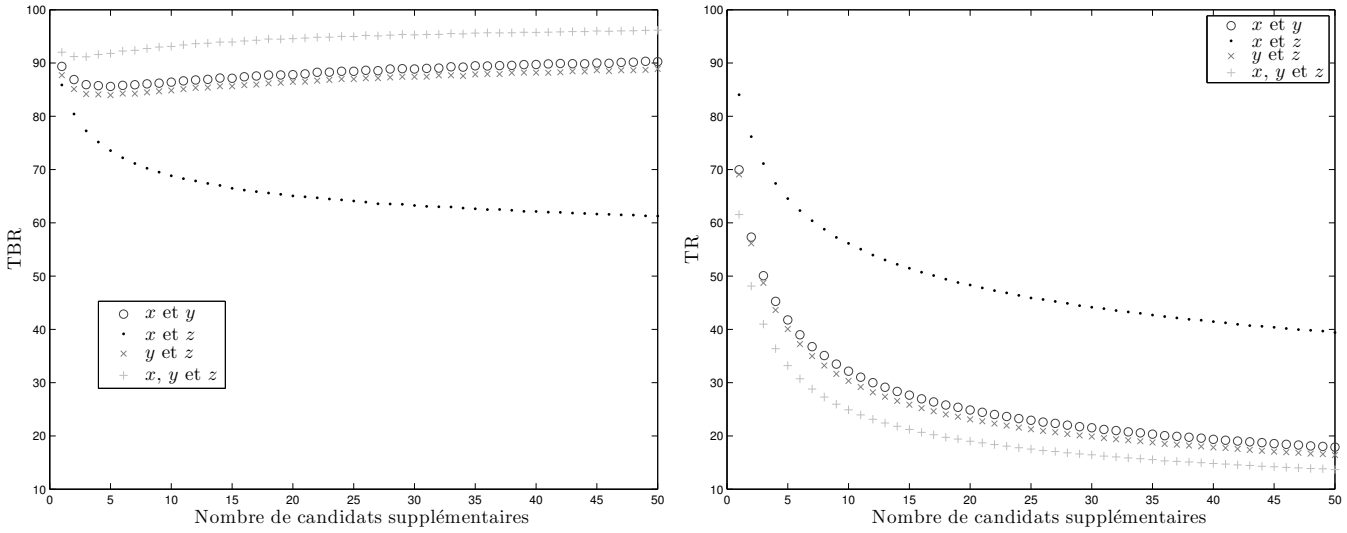
combinaisons de deux axes et la combinaison des trois axes. Quel que soit le nombre de candidats supplémentaires, le TBR est supérieur à 90 % pour le vote à l'unanimité en se servant des trois axes. D'ailleurs, l'augmentation du nombre de candidats favorise le TBR puisqu'en ajoutant des candidats, il est plus « difficile » d'obtenir un vote à l'unanimité pour un couple origine – destination faux. Seul le vote à l'unanimité pour les axes x et z ne s'améliore pas avec l'augmentation du nombre de candidats, cela traduit une forte similitude entre ces deux axes. L'amélioration du TBR s'accompagne d'une diminution significative du TR. Celui-ci est de 61,53 % pour un seul candidat supplémentaire avec le vote à l'unanimité pour les trois axes, il se réduit à 30,76 % pour six candidats supplémentaires et est inférieur à 20 % au-delà de 17 candidats supplémentaires.

7.5 Conclusion

La réidentification s'effectue en estimant la similarité entre la signature origine et la signature destination. Des méthodes sans apprentissage comme la distance euclidienne, la distance de Manhattan...ou avec apprentissage comme la logique floue, les approches bayésiennes sont possibles. Outre les méthodes usuelles, nous avons proposé d'employer la distance de Canberra et d'adapter les SVM pour réaliser la réidentification.

Quelle que soit la méthode utilisée, nous avons remarqué que l'introduction d'un seuil de décision améliore les performances de la réidentification. Nous avons proposé d'adopter le vote à l'unanimité. Dans le cadre des expérimentations menées, le vote à l'unanimité a obtenu de meilleures performances que l'introduction d'un seuil. De plus, nous évitons ainsi la détermination du seuil de décision.

Nous nous sommes rendus compte que les performances des algorithmes de classification diminuent en fonction de la taille des échantillons. Pour restreindre le nombre de candidats, une solution est de mettre en place des fenêtres temporelles pour sélectionner les candidats potentiels à la réidentification. Les fenêtres temporelles reviennent à une modélisation du trafic simpliste. Des algorithmes de modélisation du trafic plus complexes et plus performants seraient en mesure d'améliorer la réidentification. Pour s'affranchir des conditions de trafic lors du suivi de véhicules, nous avons proposé d'évaluer la réidentification des véhicules en fonction uniquement du nombre de candidats.



(a) (b)

FIGURE 7.24 – TBR (a) et TR (b) en fonction du nombre de candidats supplémentaires en utilisant la distance de Manhattan pour le vote à l'unanimité sur une combinaison de deux axes ou les trois axes.

L'ensemble des résultats obtenus nous a permis de démontrer l'intérêt du vote à l'unanimité et l'avantage de déconvoluer le signal de la boucle inductive.

8

Expérimentations

Sommaire

8.1 Temps de parcours **140**
8.1.1 Contexte 140
8.1.2 Présentation des sites 140
8.1.3 Processus 142
8.1.4 Temps de parcours individuels 143
8.1.5 Conclusion 143
8.2 Matrice Origine – Destination **145**
8.2.1 Contexte 145
8.2.2 SAROT1 145
8.2.3 SAROT2 147
8.2.4 Conclusion 150
8.3 Projet MOCoPo **150**
8.3.1 Contexte 150
8.3.2 Présentation du site et des capteurs 151
8.3.3 Acquisition 153
8.3.4 Base de données 154
8.3.5 Matrice origine – destination 155
8.3.6 Conclusion 156
8.4 Conclusion **157**

Dans ce chapitre, nous nous intéressons aux applications de temps de parcours et matrice origine – destination une fois le suivi de véhicules réalisé. Nous avons fait le choix de ne présenter qu'une seule application par expérimentation, cependant il est à chaque fois possible d'utiliser les données de suivi de véhicules pour réaliser également l'autre application.

8.1 Temps de parcours

8.1.1 Contexte

Le temps de parcours est une donnée essentielle aux gestionnaires de réseaux. Nous allons montrer au travers d'expérimentations que la réidentification donne des informations sur les temps de parcours. En effet, la réidentification permet de connaître les temps de parcours individuels de chaque véhicule.

Cette étude est effectuée à partir de deux bases de données provenant du site « Quai–Berge » d'Angers et de la rocade sud de Rennes. Ces bases de données ont déjà été utilisées dans [59, 60, 71] pour des véhicules légers (VL) seulement. Ainsi, ces bases de données ont été réutilisées pour toutes les catégories de la classe 1 (C1) à la classe 10 (C10), c'est-à-dire pour des VL et PL (Poids Lourds). Ces différentes catégories sont définies par la norme NF P 99-300.

Le début du parcours (origine) pour les deux sites est composé de x voies et la fin de ce parcours de y voies. Chaque point de mesures du site de test est équipé de caméras numériques en nombre suffisant pour filmer correctement les plaques d'immatriculation de tous les véhicules passant sur chaque couloir. Le champ de la caméra est réglé de manière que les plaques soient enregistrées au droit des boucles inductives. Les informations enregistrées sur les différentes zones de mesures sont les signatures inductives et la vidéo. Ensuite un traitement fastidieux est réalisé pour associer les signatures inductives aux informations vidéo.

8.1.2 Présentation des sites

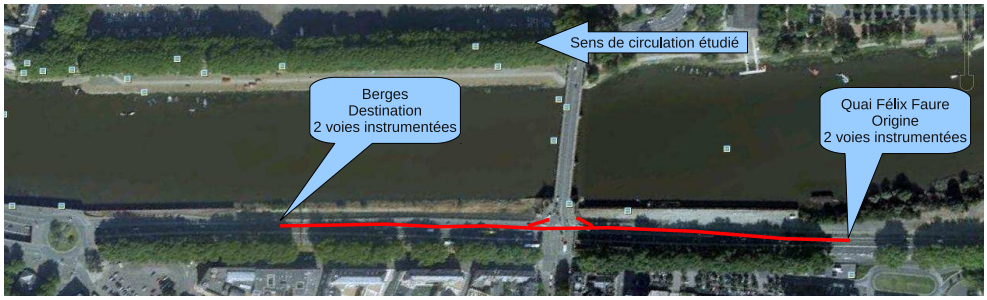


FIGURE 8.1 – Angers - Site Quai–Berge.

Site d'Angers L'expérimentation d'Angers a été réalisée au niveau de la voie « Quai–Berge » sur deux périodes (matin et après-midi). La vitesse est réglementée à 70 km h^{-1}. Le site d'Angers, présenté sur la figure 8.1, est une 2×2 voies qui traverse la ville le long de la Maine (x=2, y=2). La distance entre l'origine et la destination est de 560 m. À partir de cette expérimentation, il a été recensé sur la zone Quai (origine) 1844 VL ainsi que 210 PL, et sur la zone Berge (destination) 2881 VL et 237 PL. Cependant, parmi tous ces véhicules, seuls 1538 VL et 206 PL sont passés sur les deux points de mesures. L'explication de cet écart de chiffres est l'existence d'une sortie et d'une entrée sur la voie Quai–berge entre les deux zones de capteurs.

La figure 8.2 représente le pourcentage de véhicules en fonction de leur vitesse en km h^{-1}. Trois populations y sont représentées, soit de haut en bas : les véhicules légers, les poids lourds et une population mixte comprenant les véhicules légers et les poids lourds. Pour les poids lourds, une vitesse moyenne inférieure de 10 km h^{-1} par rapport aux véhicules légers est observée. De plus, la répartition des vitesses des poids lourds est moins étalée. Cette répartition des vitesses permettra de préciser les paramètres de la fenêtre temporelle pour réduire le nombre de candidats qui seront utilisés avec les méthodes de réidentification. Dans cette situation, une fenêtre rectangulaire centrée sur la valeur moyenne **60 km h^{-1}**, de demi-largeur **20 km h^{-1}** sera utilisée.

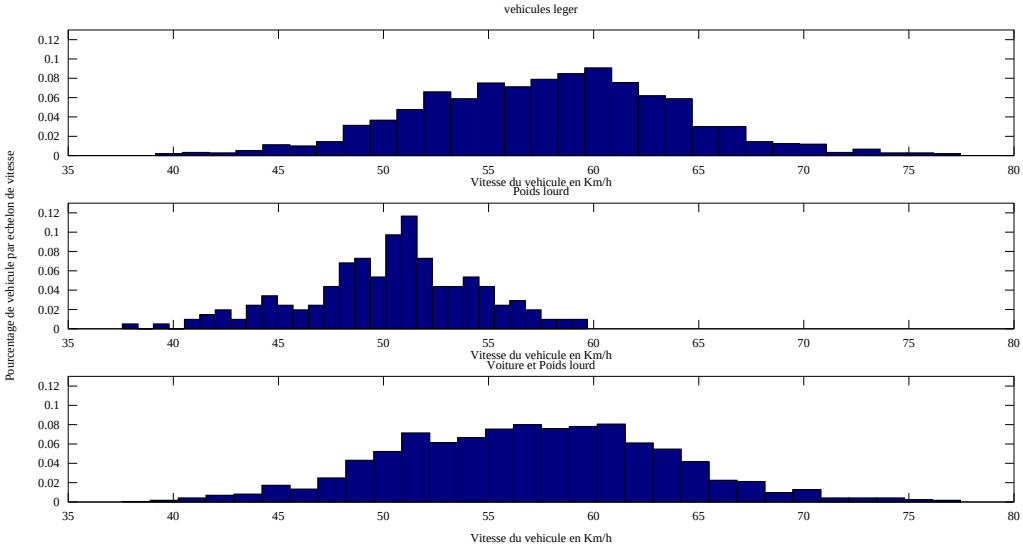


FIGURE 8.2 – Angers - Répartition des véhicules recensés en fonction de la vitesse.

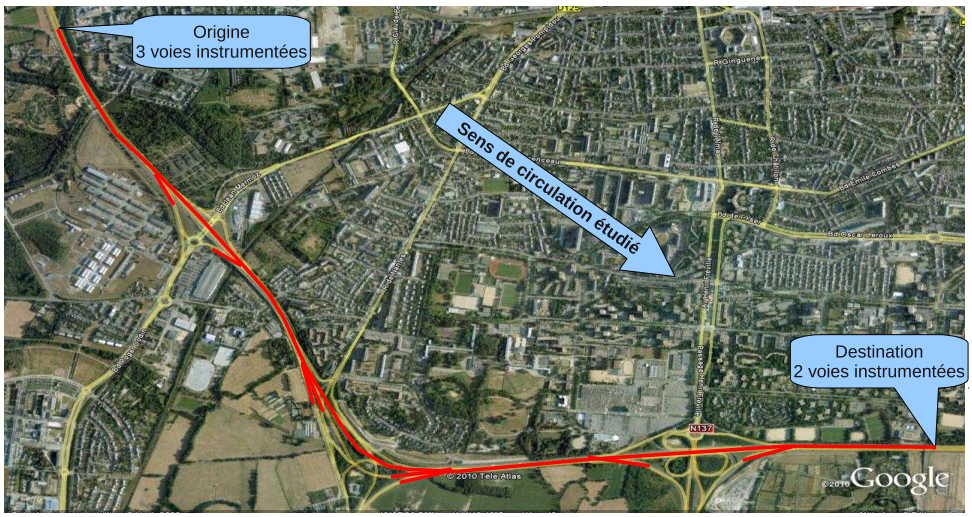


FIGURE 8.3 – Rennes - Périphérique.

Site de Rennes Une base de données a été réalisée à partir des signatures recueillies sur un tronçon du périphérique de Rennes. Cette étude a été réalisée dans [85] pour analyser le comportement des méthodes de réidentification dans le cas d'un flux de VL faiblement congestionné. La distance entre l'origine et la destination est de 5146 m. La vitesse réglementée est de 90 km h^{-1} mais la vitesse moyenne enregistrée est de 50,5 km h^{-1}. Le début du parcours (origine) est composé de trois voies et la fin de ce parcours de deux voies (x=3, y=2). Entre l'origine et la destination, il y a plusieurs entrées–sorties non instrumentées : cinq entrées et quatre sorties. Chaque point de mesures du site de test est équipé de caméras numériques en nombre suffisant pour filmer correctement les plaques d'immatriculation de tous les véhicules passant sur chaque couloir. Le champ de la caméra est réglé de manière que les plaques soient enregistrées au droit des boucles électromagnétiques de mesure.

2490 VL, 217 PL et 3550 VL, 189 PL ont été respectivement relevés à l'origine et à la destination. Parmi tous ces véhicules, uniquement 930 VL et 80 PL sont passés à la fois par l'origine et par la destination. Au total 6446 véhicules différents ont été comptabilisés. Le passage de trois voies à l'origine à deux voies a tendance à provoquer des ralentissements. Les nombreuses entrées et sorties non instrumentées sur l'itinéraire rendent aussi le site très complexe pour l'application envisagée.

La figure 8.4 représente le pourcentage de véhicules en fonction de leur vitesse en km h^{-1}. Comme pour la base d'Angers, trois populations y sont représentées, soit de haut en bas : les véhicules légers, les poids lourds et une population mixte comprenant les véhicules légers et les poids lourds. La vitesse des poids lourds est plus faible. Pour cette base de données, une fenêtre centrée sur la valeur moyenne 50 km h^{-1}, de demi-largeur 15 km h^{-1} est appliquée. Cependant, la grande distance entre l'origine et la destination rend l'utilisation d'un fenêtrage complexe. Les figures 8.2 et 8.4 montrent que la répartition de la vitesse de la population mixte est similaire au cas de la population de véhicules légers. Ceci est lié à la répartition poids lourds – véhicules légers dans la population mixte, où les VL ont une place majoritaire.

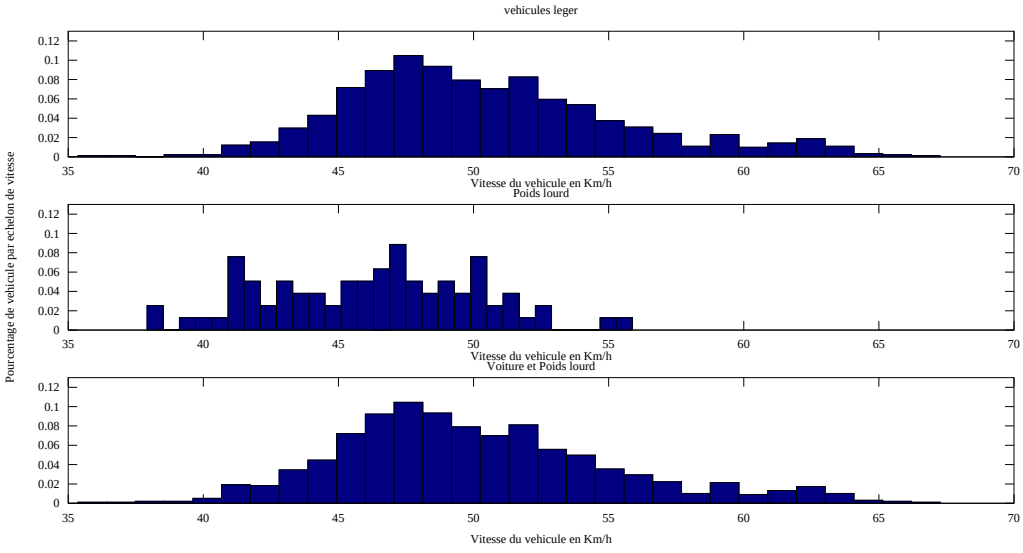


FIGURE 8.4 – Rennes - Répartition des véhicules recensés en fonction de la vitesse.

Le tableau 8.1 récapitule les caractéristiques du site de Rennes ainsi que celui d'Angers.

8.1.3 Processus

Le signal recueilli au passage d'un véhicule sur une boucle inductive implantée dans la chaussée est représentatif du véhicule. La séparation des véhicules en deux populations distinctes VL et PL a permis de mettre en évidence des caractéristiques différentes pour chacune des populations. Une réduction du nombre de caractéristiques pour chacune des populations est réalisée par l'Analyse en Composantes Principales. Les caractéristiques employées sont celles décrites dans la section 6.2.2.

	Site angevin		Site rennais	
Distance	560 m		5150 m	
Entrées (E) Sorties (S)	1 1		5 4	
	Origine	Destination	Origine	Destination
Nombre de voies	2	2	3	2
Nombre de VL Nombre de PL	1844 210	2881 237	2490 217	3550 189
	Véhicules passés sur les 2 zones			
Nombre de VL Nombre de PL	1538 206		930 80	
Observations	-Faible distance -Non congestionnée -Peu d'E/S non instrumentées		-Longue distance -Faiblement congestionnée -Nombre important d'E/S non instrumentées	

Tableau 8.1 – Tableau récapitulatif des caractéristiques des sites de récupération de données

Cette réduction permet de limiter le nombre d'informations à transmettre et de diminuer par la suite le temps de calcul pour le suivi de véhicules. Après avoir testé individuellement la méthode de logique floue, une approche bayésienne et les séparateurs à vaste marge, nous avons retenu un algorithme de comparaison fondé sur l'utilisation conjointe des trois méthodes présentées dans la section 7.3. Le nombre de candidats est réduit par l'application d'une fenêtre temporelle 7.3.2.

8.1.4 Temps de parcours individuels

Cette section présente les temps de parcours individuels des VL et PL à partir de la base de données d'Angers. La méthode utilisée est la méthode « una » (vote à l'unanimité) avec un filtrage « origine à multiples destinations » dont les performances sont présentées dans la section 7.3.3. Les figures 8.5 et 8.6 montrent les temps de parcours observés et estimés des VL et des PL sur cette base de données en fonction du temps écoulé. Sur les deux figures, les ronds rouges représentent les temps de parcours des véhicules réidentifiés. Les croix bleues représentent les temps de parcours réellement réalisés par les véhicules (ils sont établis par la réidentification manuelle des véhicules par la vidéo). Lorsque le véhicule est correctement réidentifié, le rond rouge se superpose avec la croix bleue. Une mauvaise réidentification se discerne par un rond rouge seul et un véhicule non réidentifié par une croix bleue seule. Sur la figure 8.5, des mauvaises réidentifications ainsi que des non réidentifications sont visibles pour les VL. Leurs répartitions ne semblent pas perturber l'estimation du temps de parcours. Sur la figure 8.6, une seule mauvaise réidentification est constatée. L'estimation des temps de parcours pour les PL est donc plus fiable. Malheureusement sur la durée de l'expérimentation, aucune période transitoire entre un trafic fluide et congestionné n'est constatée. L'estimation du temps de parcours semble réaliste dans une situation stable mais nécessiterait une évaluation lors de régime transitoire.

8.1.5 Conclusion

Afin de répondre à la demande des gestionnaires qui doivent faire face à l'augmentation progressive du trafic sur nos routes, il est de plus en plus important de réaliser un système permettant de suivre les véhicules. En effet, avec un tel système, il devient possible de connaître les temps de parcours entre deux points. Ces informations sont importantes pour maîtriser et optimiser la circulation.

Les résultats obtenus pour l'estimation des temps de parcours sont corrects dans le cadre d'un

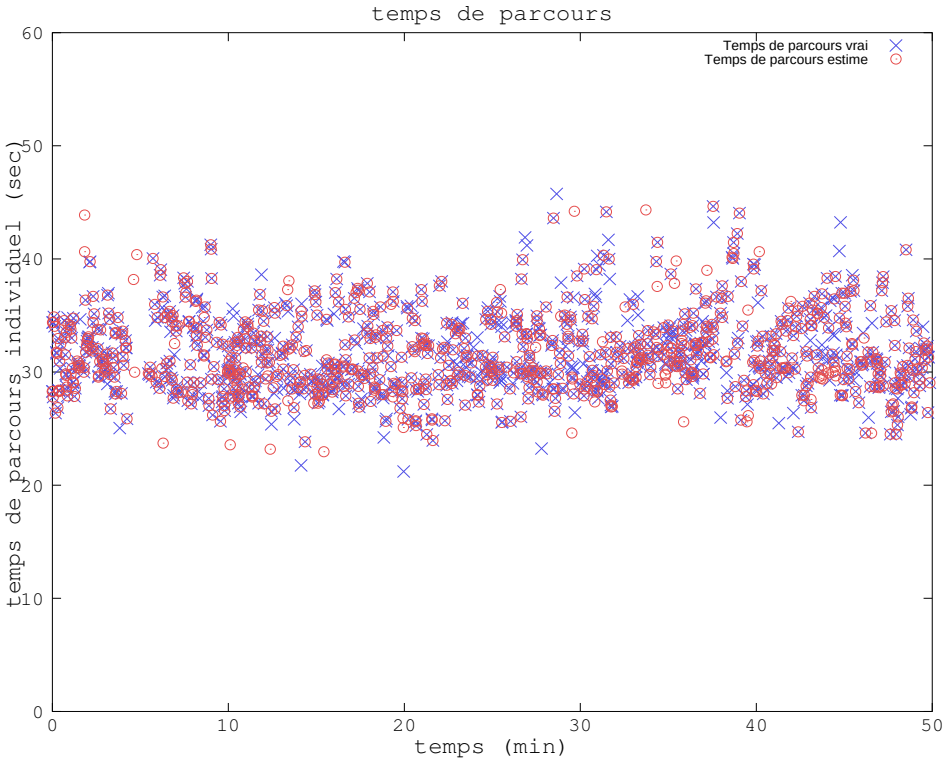


FIGURE 8.5 – Estimation des temps de parcours pour des VL.

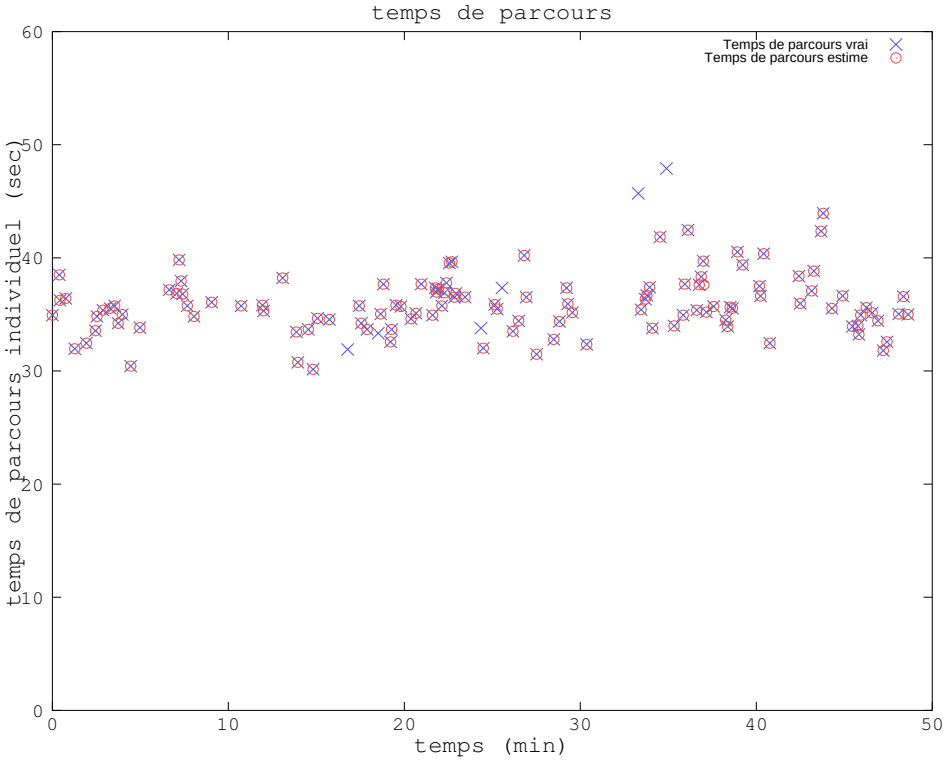


FIGURE 8.6 – Estimation des temps de parcours pour des PL.

trafic stable (sans changement d'états : fluide, congestion), spécialement pour le cas des Poids Lourds. En perspective, une expérimentation sur les états transitoires de trafic serait intéressante à réaliser. La mise en place d'algorithmes de modélisation du trafic pour limiter le nombre de candidats à la réidentification apporterait certainement une amélioration du TBR (évoqué à la section 7.3.2).

8.2 Matrice Origine – Destination

8.2.1 Contexte

La plate-forme SAROT (Site Angevin de Référence pour l'Observation du Trafic) a été développée pour permettre d'effectuer des mesures à des fins d'expérimentations. SAROT est situé sur la rocade Est de la ville d'Angers. Deux voies et une rampe d'accès, en direction de Paris, sont instrumentées. La section instrumentée a une longueur d'environ 1050 m et est limitée a 90 km h^{-1}. Le trafic journalier est d'environ 22 000 véhicules.

Dans notre cas, la plate-forme est utilisée pour réaliser le suivi de véhicule et pour calculer une matrice origine – destination. Deux expérimentations ont été menées et ont permis de constituer les base de données SAROT1 et SAROT2.

8.2.2 SAROT1

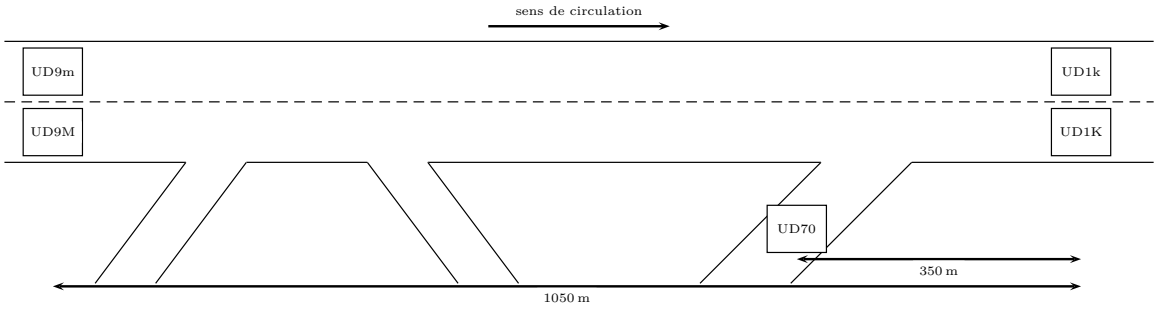


FIGURE 8.7 – Expérimentation sur le site SAROT pour la base de donnée SAROT1.

Présentation de l'expérimentation Comme le montre la figure 8.7, les boucles inductives UD9M, UD9m et UD70 sont considérées comme les origines et les boucles inductives UD1K et UD1k comme les destinations. Les informations enregistrées sur les différentes zones de mesure sont les signatures inductives et la vidéo. Ensuite un traitement fastidieux est réalisé pour associer les signatures inductives aux informations vidéo. La vidéo est considérée comme la référence. L'existence d'une entrée et d'une sortie non instrumentées crée des véhicules sans destination ou sans origine. Tous les véhicules provenant de l'entrée non instrumentée ou allant à la destination non instrumentée sont supprimés de la base de données et seules les signatures des voitures passant à l'origine et à la destination sont conservées. L'étude ne conserve que les voitures étant donné que les poids lourds sont plus simples à réidentifier. Au final la base de données contient 1396 véhicules. La base de données est divisée en deux. 460 véhicules sont utilisés pour la phase d'apprentissage par k validation croisée, avec $k = 10$, et 936 pour la phase de test. En ce basant sur [59], seulement neuf caractéristiques par signature sont retenues. Les caractéristiques sont les suivantes : la longueur sur l'indice du maximum global de la signature ; le premier et le second coefficients des parties réelles de la transformée de Fourier ; le premier, le second et le troisième coefficients des parties imaginaires de la transformée de Fourier ; le module des quatrième, cinquième et sixième coefficients de la transformée de Fourier.

La phase d'apprentissage a été effectuée lors de la section 7.3. Pour les origines UD9m et UD9M allant aux destinations UD1k et UD1K, la vitesse individuelle des véhicules est considérée comme

étant comprise entre 60 km h^{-1} et 115 km h^{-1}. Pour ceux provenant de la voie d'insertion UD70 et allant aux destinations UD1k et UD1K, la vitesse individuelle des véhicules est considérée comme étant comprise entre 50 km h^{-1} et 100 km h^{-1}. La fenêtre temporelle, décrite à la section 7.3.2 pour réduire le nombre de candidats est paramétrée avec les informations ci-dessus.

La méthode de réidentification utilisée est présentée à la section 7.3.3.

Estimation de matrice origine – destination La matrice origine – destination de référence R pour la base de donnée de test, représentée par le tableau 8.2, a été produite à partir des données de la vidéo. La matrice donne le nombre de déplacements effectués d'une origine vers une destination. Le nombre entre parenthèses représente le pourcentage des déplacements effectués entre une origine et une destination par rapport à l'ensemble des déplacements. 50 % des déplacements sont réalisés entre l'origine UD9M et la destination UD1K. Le flux de véhicules entre l'origine UD70 et la destination UD1k est faible avec environ 2 %. La répartition entre les quatre autres origine – destination possibles est plus homogène.

		Destination	
		UD1K	UD1k
Origine	UD9M	465 (49,95 %)	116 (12,46 %)
	UD9m	80 (8,59 %)	132 (14,18 %)
	UD70	118 (12,67 %)	20 (2,15 %)

Tableau 8.2 – Matrice origine – destination de référence R de la base de test.

À partir des réidentifications effectuées sur la base de test, une matrice origine – destination T est constituée (tableau 8.3). Comme le TRi est de 92,74 %, le nombre total de déplacements effectués pour cette matrice va être inférieur à celui de la matrice de référence. En effet, la matrice T comporte 868 déplacements alors que la matrice R dénombre 936 déplacements. Pour pouvoir comparer les deux matrices entre elles, les flux de déplacements sont exprimés en pourcentage.

		Destination	
		UD1K	UD1k
Origine	UD9M	434 (50 %)	105 (12,10 %)
	UD9m	81 (9,33 %)	116 (13,36 %)
	UD70	113 (13,02 %)	19 (2,19 %)

Tableau 8.3 – La matrice origine – destination T de la base de test.

		Destination	
		UD1K	UD1k
Origine	UD9M	0	0,03
	UD9m	-0,09	0,06
	UD70	-0,03	-0,02

Tableau 8.4 – La matrice d'erreur E pour la base de test.

Une matrice E représentant l'erreur relative entre les pourcentages des flux de la matrice R et de la matrice T est calculée. L'erreur relative entre les deux matrices est définie pour chaque élément comme suit :

$$E(i,j) = \frac{R(i,j) - T(i,j)}{R(i,j)} \tag{8.1}$$

avec $i \in 1,2,3$ représentant les lignes de la matrice et $j \in 1,2$ représentant les colonnes.

L'erreur relative globale est définie comme suit :

$$RE = \sqrt{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R(i,j) - T(i,j)}{R(i,j)}\right)^2} \tag{8.2}$$

avec $n = 3$ et $m = 2$.

L'erreur relative globale RE est de 0,08. Le tableau 8.4 montre que l'erreur relative, quel que soit le déplacement effectué, est inférieure à 0,1. La matrice T est proche de la réalité même si le TRi est de l'ordre de 93 %.

8.2.3 SAROT2

Présentation de l'expérimentation Une autre expérimentation a été menée sur le site de SAROT. Une portion de la zone comme le montre la figure 8.8 a été utilisée.

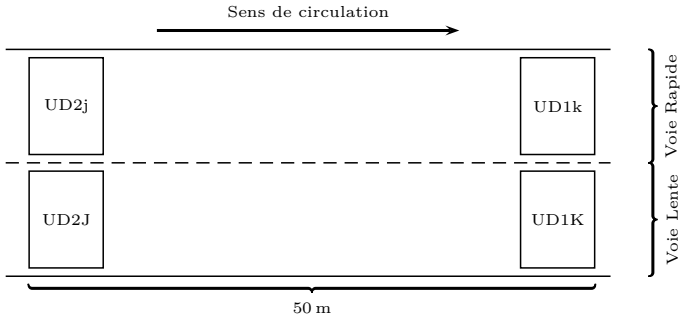


FIGURE 8.8 – Schéma simplifié pour l'expérimentation avec la déconvolution.

La figure 8.8 représente un schéma simplifié des quatre boucles inductives utilisées. Chacune des boucles inductives mesure la signature inductive, la vitesse du véhicule avec une tolérance de 10 % et la longueur du véhicule. Les capteurs UD2J et UD2j sont considérés comme les origines et les capteurs UD1K et UD1k comme les destinations. La distance entre les origines et les destinations est courte, 50 m, mais le but est de comparer la réidentification en s'affranchissant des conditions de trafic comme le décrit la section 7.4. Pour permettre la vérification des résultats, deux caméscopes enregistrent le passage des véhicules à l'origine et à la destination durant l'acquisition des signaux. Les vidéos servent à identifier les véhicules à l'origine et à la destination mais aussi à connaître le modèle du véhicule. En connaissant le modèle du véhicule, il est possible en reprenant les informations techniques, de connaître la longueur réelle du véhicule. Une base de données de 654 paires de signatures de véhicules étant passés à l'origine et à la destination a été constituée.

La réidentification est effectuée à partir des signaux non déconvolués et déconvolués selon la méthode décrite à la section 6.3. La méthode de réidentification choisie a été déterminée à la section 7.4.

Estimation de matrice origine – destination La matrice origine – destination de référence R pour la base de données de 654 véhicules, représentée par le tableau 8.5, est produite à partir des données de la vidéo. La matrice donne le nombre de déplacements effectués d'une origine vers une destination. Le nombre entre parenthèses représente le pourcentage des déplacement effectués entre

une origine et une destination par rapport à l'ensemble des déplacements. En observant le tableau 8.5, nous apercevons que les véhicules provenant de l'origine UD2J (ou UD2j) vont quasiment tous vers la destination UD1K (ou UD1k, respectivement).

		Destination	
		UD1K	UD1k
Origine	UD2J	509 (77,83 %)	2 (0,31 %)
	UD2j	1 (0,15 %)	142 (21,71 %)

Tableau 8.5 – Matrice origine – destination de référence R.

Néanmoins, nous allons réaliser l'estimation des matrices origine – destination à partir des signaux déconvolués à partir de la moyenne de $\tilde{\mathbf{h}}$. La comparaison est réalisée sur l'utilisation de la distance de Manhattan et le vote à l'unanimité sur l'ensemble des mesures de dissimilarité (noté par la suite una).

		Destination	
		UD1K	UD1k
Origine	UD2J	499 (76,30 %)	9 (1,37 %)
	UD2j	16 (2,44 %)	130 (19,89 %)

(a)

		Destination	
		UD1K	UD1k
Origine	UD2J	459 (77,14 %)	11 (1,85 %)
	UD2j	9 (1,51 %)	116 (19,50 %)

(b)

Tableau 8.6 – Matrice origine – destination estimée avec 10 candidats supplémentaires : (a) Distance de Manhattan ; (b) una.

		Destination	
		UD1K	UD1k
Origine	UD2J	484 (74,01 %)	24 (3,67 %)
	UD2j	22 (3,36 %)	124 (18,96 %)

(a)

		Destination	
		UD1K	UD1k
Origine	UD2J	411 (74,60 %)	17 (3,08 %)
	UD2j	17 (3,08 %)	106 (19,24 %)

(b)

Tableau 8.7 – Matrice origine – destination estimée avec 20 candidats supplémentaires : (a) Distance de Manhattan ; (b) una.

Que ce soit pour la distance de Manhattan ou la méthode una, les tableaux 8.6, 8.7, 8.8, 8.9 et 8.10 montrent une sur-estimation des déplacements de l'origine UD2j à la destination UD1K et de l'origine UD2J à la destination UD1k. Nous remarquons que sur ces tableaux, cette surestimation a tendance à augmenter avec le nombre de candidats supplémentaires. La surestimation est moins importante en général avec la méthode una qu'avec la distance de Manhattan. Pour chacune des matrices origine – destination estimées, l'erreur relative globale est calculée suivant l'équation (8.2). Cette fois-ci n et m valent 2. Le tableau 8.11 confirme l'augmentation de l'erreur relative avec l'accroissement du nombre de candidats supplémentaires. Un cas particulier apparaît pour la méthode una pour 40 candidats supplémentaires, une baisse de l'erreur relative est constatée. Pour l'ensemble des erreurs relatives, la méthode una obtient de meilleurs résultats que la méthode avec la distance de Manhattan.

		Destination	
		UD1K	UD1k
Origine	UD2J	477 (72,93 %)	28 (4,28 %)
	UD2j	30 (4,59 %)	119 (18,20 %)

(a)

		Destination	
		UD1K	UD1k
Origine	UD2J	395 (74,95 %)	12 (2,28 %)
	UD2j	20 (3,79 %)	100 (18,98 %)

(b)

Tableau 8.8 – Matrice origine – destination estimée avec 30 candidats supplémentaires : (a) Distance de Manhattan ; (b) una.

		Destination	
		UD1K	UD1k
Origine	UD2J	474 (72,48 %)	34 (5,20 %)
	UD2j	33 (5,04 %)	113 (17,28 %)

(a)

		Destination	
		UD1K	UD1k
Origine	UD2J	384 (75,44 %)	14 (2,75 %)
	UD2j	16 (3,14 %)	95 (18,67 %)

(b)

Tableau 8.9 – Matrice origine – destination estimée avec 40 candidats supplémentaires : (a) Distance de Manhattan ; (b) una.

		Destination	
		UD1K	UD1k
Origine	UD2J	469 (71,71 %)	37 (5,66 %)
	UD2j	40 (6,12 %)	108 (16,51 %)

(a)

		Destination	
		UD1K	UD1k
Origine	UD2J	373 (74,30 %)	17 (3,39 %)
	UD2j	23 (4,58 %)	89 (17,73 %)

(b)

Tableau 8.10 – Matrice origine – destination estimée avec 50 candidats supplémentaires : (a) Distance de Manhattan ; (b) una.

Nombre de candidats supplémentaires	10	20	30	40	50
Distance de Manhattan	7,51	16,76	22,47	25,30	30,23
una	7,23	15,01	17,45	14,94	21,68

Tableau 8.11 – Erreur relative globale en fonction du nombre de candidats supplémentaires et de la méthode utilisée : distance de Manhattan ou una.

8.2.4 Conclusion

Au cours des deux expérimentations, une estimation de la répartition des flux entre les différentes origines et destinations est obtenue par le suivi de véhicules. Les principaux flux de déplacements sont facilement identifiables à partir des matrices origine – destination estimées par le suivi de véhicules. Les estimations sont délivrées directement à partir du suivi de véhicules, cependant l'adjonction de méthodes statistiques pour l'estimation de matrice origine – destination améliorerait certainement les estimations. L'estimation des erreurs relatives en fonction du nombre de candidats montre que la prise en compte d'un indicateur supplémentaire de confiance pourrait être pris en compte lors de l'évaluation des matrices origine – destination.

8.3 Projet MOCoPo

8.3.1 Contexte

Le projet MOCoPo s'intéresse à la Mesure et Modélisation de la COngestion et de la POllution. Le site de la rocade intérieure sud de Grenoble a été choisi pour réaliser ce projet. C'est une autoroute périurbaine dont la congestion est importante et récurrente. Pour l'instant, les relations entre les conditions de trafic et les émissions de polluants sont appréhendées par des modèles horaires. Les deux principaux objectifs vont être de mieux comprendre la formation et la disparition d'une congestion, ainsi que les relations entre les conditions de trafic et la pollution.

Afin de réaliser ce projet pluridisciplinaire, différents organismes se sont réunis : Ascoparg, CEREA, Cerema, DIR-CE, IFSTTAR, INRIA-NeCS. Le projet est divisé en huit tâches (figure 8.9) : une tâche d'organisation (T0), trois tâches de recueil de données (T1, T2 et T3) et quatre tâches de modélisation (T4, T5, T6 et T7).

La tâche 1 consiste à collecter des données de trajectoires à partir de films vidéo réalisés par un hélicoptère en vol stationnaire au-dessus de différents sites représentatifs. La tâche 2 concerne les mesures de pollution, de la météo et la détermination du parc en circulation. La tâche 3 s'intéresse à la réalisation d'un fichier de données individuelles de trafic. La tâche 4 utilise principalement les données de la tâche 3 pour effectuer la réidentification de véhicules à partir des signatures des magnétomètres. Parmi les tâches de modélisation, la tâche 5 se focalise sur la modélisation des changements de voie. La tâche 6 se concentre sur la modélisation statistique des trajectoires. Enfin, la tâche 7 concerne le lien entre la modélisation des émissions de polluants et la pollution mesurée.

Les tâches 3 et 4 seront présentées plus en détail dans ce mémoire. Les données individuelles sont recueillies à l'aide d'un réseau de capteurs installés sur différentes zones de la rocade sud de Grenoble. L'équipement consiste pour chaque point de mesures en un magnétomètre permettant d'obtenir les caractéristiques magnétiques tridimensionnelles de chaque véhicule et de deux boucles inductives ou de deux magnétomètres industriels Sensys pour mesurer la vitesse. Ces réseaux de capteurs mesurent la date de passage, la longueur et la vitesse des véhicules, ainsi qu'une information très complète sur la répartition tridimensionnelle des masses métalliques du véhicule.

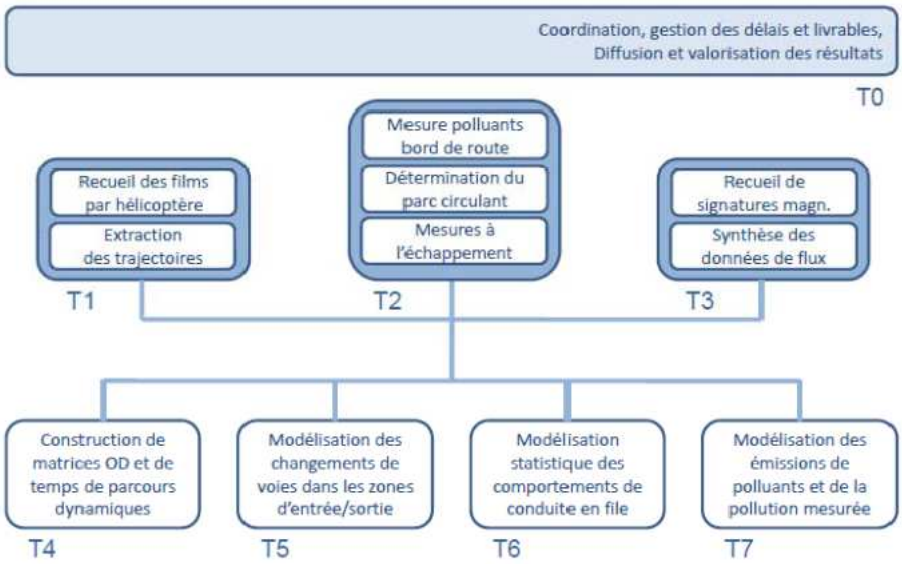


FIGURE 8.9 – Organisation du projet MOCoPo.

8.3.2 Présentation du site et des capteurs

Lors du montage du projet, il était prévu d'utiliser les magnétomètres industriels de Sensys Network pour mesurer la signature magnétique en trois dimensions. Malheureusement, nous n'avons pas eu accès à la signature magnétique du véhicule. Ainsi, l'équipe NeCS de l'INRIA a développé un magnétomètre sans fil mesurant la signature magnétique des véhicules pour les besoins de ce projet. Pour des raisons de confidentialité et de propriétés intellectuelles, nous n'avons pas eu accès aux méthodes développées par l'INRIA, ainsi qu'aux caractéristiques des magnétomètres. Les signatures des véhicules détectés sont mesurées et nous sont transmises seulement à l'échéance des expérimentations.

Les capteurs industriels ont été utilisés pour relever la vitesse des véhicules. Pour chaque point de mesure, un réseau de capteurs a été mis en place. Il est constitué de deux capteurs industriels encadrant un magnétomètre « signature ». Pour vérifier la détection des véhicules et leur identification, des caméscopes sont utilisés. Les caméscopes sont positionnés de manière à filmer les plaques d'immatriculation des véhicules au niveau des réseaux de capteurs.

FIGURE 8.10 – Zones d'implantation des réseaux de capteurs sur la rocade sud de Grenoble.

Les réseaux de capteurs ont été implantés sur trois zones de la rocade sud de Grenoble comme le montre la figure 8.10. Ce choix permet d'avoir une section courte et une section longue. Les deux sections sont représentées schématiquement sur la figure 8.11. Il est à noter que comme le montrent les figures 8.10 et 8.11, les réseaux de capteurs à la sortie de l'échangeur du Rondeau sont utilisés en tant que destination pour la section courte et la section longue.

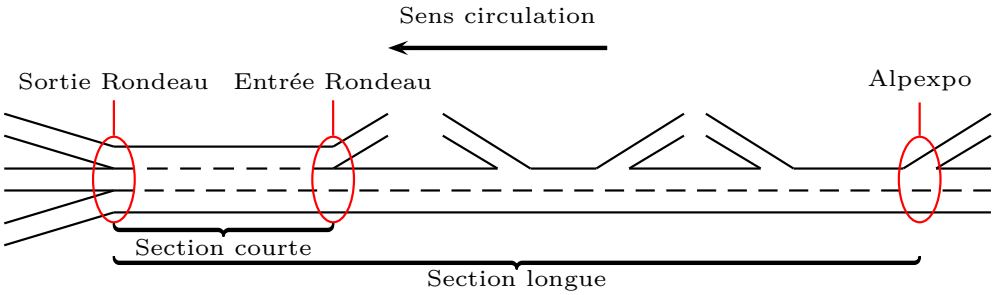
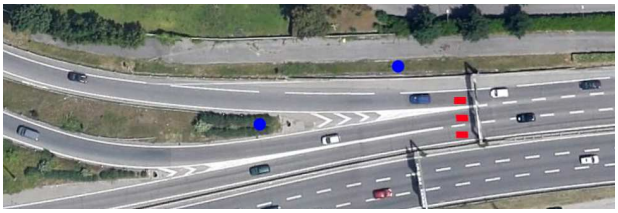


FIGURE 8.11 – Schéma de la rocade sud de Grenoble.

La figure 8.12 représente la section courte. Elle fait environ 300 m et se situe au niveau de l'échangeur du Rondeau. L'ensemble des véhicules passe sur un réseau de capteurs à l'entrée comme à la sortie de cette section. Cette section est une zone d'entrecroisement avec deux entrées et trois sorties : c'est un site privilégié pour l'observation des changements de voies.

FIGURE 8.12 – Entrées – sorties de l'échangeur du Rondeau. En rouge l'emplacement des réseaux de capteurs et en bleu l'emplacement des caméscopes.

La section longue fait environ 2,5 km. Elle est comprise entre l'entrée Alpexpo (figure 8.13b) et la sortie de l'échangeur du Rondeau (figure 8.13a). Sur cette section, une partie des informations ne peut être complète car deux sorties et une entrée entre l'origine et la destination ne sont pas instrumentées. Il est à noter que l'entrée du Rondeau n'est pas observée lors de l'expérimentation sur la section longue et peut donc être considérée comme une entrée non instrumentée.



(a) Sortie de section : Rondeau.

(b) Entrée de section : Alpexpo.

FIGURE 8.13 – Réseaux de capteurs de Alpexpo à l'entrée de l'échangeur du Rondeau. En rouge l'emplacement des réseaux de capteurs et en bleu l'emplacement des caméscopes.

Sur l'ensemble des figures 8.10, 8.13a, 8.13b, les points bleus représentent la position des caméscopes et les traits rouges les réseaux de capteurs.

Alpexpo La figure 8.13b représente l'entrée de la section longue qui se situe au niveau de l'échangeur Alpexpo. La figure 8.13b montre que les réseaux de capteurs sont implantés au centre de : la voie d'insertion (VI), la voie lente (VL) et la voie rapide (VR). Les réseaux de capteurs sont constitués (figure 8.14) : de deux magnétomètres Sensys Network dit « industriels » représentés en noir et d'un magnétomètre dit « signature » représenté en rouge pouvant acquérir les caractéristiques magnétiques tridimensionnelles du véhicule. Le magnétomètre « signature » est entouré de deux magnétomètres industriels. Ces derniers sont utilisés pour mesurer la vitesse du véhicule.

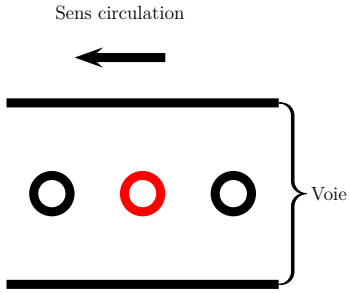


FIGURE 8.14 – Réseaux de capteurs magnétomètres.

Pour la zone Alpexpo, deux caméscopes sont utilisés : l'un pour la voie d'insertion et la voie lente ; l'autre pour la voie lente et la voie rapide. Au vu de la configuration de la zone Alpexpo et l'emplacement des caméscopes, des masquages entre les véhicules sont susceptibles de se produire. Une voiture passant sur la voie rapide en même temps qu'un poids lourd passant sur la voie lente ou d'insertion risque de ne pas être visible sur la vidéo.

Entrée de l'échangeur du Rondeau La figure 8.13a représente l'entrée de l'échangeur du Rondeau. Comme pour la zone Alpexpo, les réseaux de capteurs sont implantés au centre de VI, VL et VR. Ce sont les mêmes réseaux de capteurs que pour Alpexpo (figure 8.14). Les VI, VL et VR sont considérées comme les origines de la section courte.

Deux caméscopes (figure 8.13a) sont placés à l'entrée de l'échangeur du Rondeau : le caméscope au niveau du pont enregistre la VL et la VR ; le deuxième caméscope situé à droite de la VI enregistre la VI et la VL.

Sortie de l'échangeur du Rondeau La sortie de l'échangeur du Rondeau comporte 3 voies nommées par la suite : droite (D) dans le prolongement de la VI ; milieu (M) dans le prolongement de la VL ; gauche (G) dans le prolongement de la VR. Chacune des voies représente les destinations de la section courte et de la section longue. Chaque voie est équipée d'un réseau de capteurs. Le réseau de capteurs est différent de ceux utilisés pour l'entrée du Rondeau et Alpexpo. Les boucles inductives déjà implantées et fonctionnelles sur chaque voie avant la réalisation du projet MOCoPo ont été réutilisées. Ce réseau de capteurs est constitué (figure 8.15) : de deux boucles inductives représentées en noir ; d'un magnétomètre « signature » représenté en rouge. Le magnétomètre est entouré de deux boucles inductives. Les boucles inductives mesurent la vitesse des véhicules.

Deux caméscopes (figure 8.12) sont placés à la sortie de l'échangeur du Rondeau :

- le caméscope à droite de la voie D enregistre les voies de D et M ;
- le second caméscope sur l'îlot enregistre les voies M et G.

8.3.3 Acquisition

Des expérimentations ont été réalisées les 26 et 27 juillet 2012. Suite à un problème matériel sur la zone Alpexpo, aucune acquisition n'a pu être effectuée sur cette zone à cette date. En revanche deux

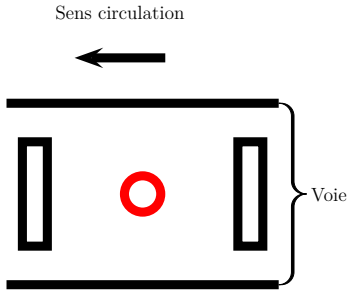


FIGURE 8.15 – Réseaux de capteurs boucles inductives - magnétomètre.

sessions d'acquisitions ont eu lieu sur la section courte de l'échangeur du Rondeau : de 19 h à 20 h le 26 juillet 2012 ; de 8 h30 à 9 h30 le 27 juillet 2012. Des fichiers obtenus à partir des magnétomètres industriels, des boucles inductives et des magnétomètres « signature » ont été générés.

Deux autres expérimentations ont été réalisées le 30 mai 2013 sur la section longue : de 14 h à 15 h environ et de 18 h à 19 h environ.

8.3.4 Base de données

De manière à pouvoir analyser l'ensemble des données fournies, nous avons effectué un travail d'agrégation des données. Les images des caméscopes sont prises comme référence. Pour chaque vidéo, le passage des véhicules a été relevé avec les informations suivantes :

- le numéro de l'image dans la vidéo ;
- le temps écoulé ;
- la plaque d'immatriculation (qui est ensuite cryptée) ;
- le modèle du véhicule ;
- la voie de circulation.

Ensuite, les données issues des différents capteurs sont agrégées avec les données de référence des vidéos pour l'entrée et la sortie de chaque section. La dernière étape consiste à agréger les données de l'entrée avec celles de la sortie de la section, l'entrée du Rondeau avec la sortie du Rondeau pour la section courte, et l'entrée Alpexpo avec la sortie du Rondeau pour la section longue.

Les données représentent quatre heures d'enregistrement. Même si les deux sessions sur la section courte ont eu lieu à des périodes différentes de la journée, 19 h pour l'une et 8 h pour l'autre, les conditions météorologiques et de trafic étaient quasiment les mêmes (météo clémente, circulation relativement fluide). Par contre pour les deux sessions sur la section longue, une faible congestion est présente lors des acquisitions. Les conditions météorologiques étaient aussi clémentes.

Pour chaque session, la détection des véhicules par les différents capteurs mis en place a été vérifiée. Le taux de fausses détections a aussi été déterminé. Ces fausses détections sont souvent provoquées par :

- le chevauchement du véhicule sur deux voies, impliquant une double détection ;
- un poids lourd sur la voie adjacente ;
- un véhicule comportant une remorque considéré comme deux véhicules...

La session d'acquisition qui a eu lieu le 26 juillet 2012 est appelée la première session et celle qui a eu lieu le 27 juillet 2012 la deuxième session. La session sur la section longue à 15 h du 30 mai 2013 est considérée comme la troisième session et celle de 18 h comme la quatrième session.

Seule la première session est exploitée dans ce manuscrit. Durant cette première session, un dysfonctionnement des magnétomètres « signature » à la sortie de l'échangeur du Rondeau n'a pas permis de relever les signatures sur l'ensemble de l'expérimentation. Les signatures ont été acquises pendant la première demi-heure.

8.3.5 Matrice origine – destination

La matrice origine – destination de référence pour la première session d'expérimentation sur la section courte est donnée par le tableau 8.12. Cette matrice est établie à partir des vidéos. Les 1753 véhicules sont répartis selon leur origine et leur destination et la représentation en pourcentage d'un déplacement est donnée entre parenthèses.

		Destination		
		D	M	G
Origine	VI	72 (4,11 %)	136 (7,76 %)	66 (3,76 %)
	VL	592 (33,77 %)	315 (17,97 %)	13 (0,74 %)
	VR	81 (4,62 %)	228 (13,01 %)	250 (14,26 %)

Tableau 8.12 – Matrice origine – destination de référence R pour la première session.

La répartition des temps de parcours entre l'origine et la destination est présentée sur la figure 8.16. Les temps de parcours sont compris entre 9 s et 44 s mais la majorité des temps de parcours se situe entre 10 s et 25 s.

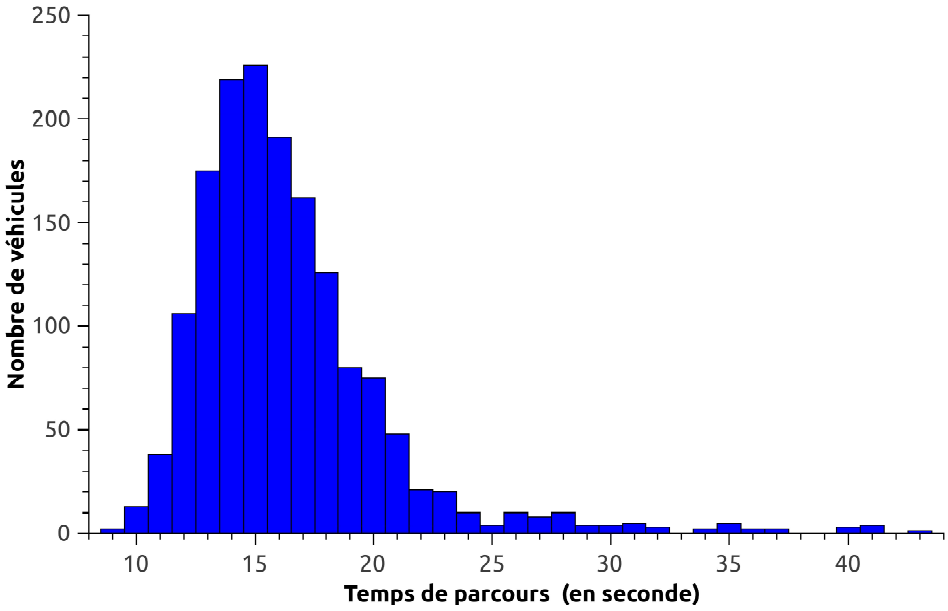


FIGURE 8.16 – Temps de parcours des véhicules entre l'origine et la destination.

La réidentification pour estimer la matrice origine – destination est effectuée en utilisant la distance de Manhattan avec le vote à l'unanimité des trois axes. Toutes les détections sont prises en

compte ce qui implique que des signatures à l'origine ne sont pas associées à des signatures à la destination et inversement. En comptabilisant toutes les détections à la destination, nous obtenons 1225 signatures destination. Pour limiter le nombre de candidats, une fenêtre temporelle fixe est appliquée. La signature origine doit appartenir à l'intervalle de temps compris entre l'horaire de la signature de destination moins 44 s et l'horaire de la signature de destination moins 9 s. Avec cette fenêtre temporelle, le nombre de candidats varie entre 12 et 44 candidats, avec une moyenne de 25,46 candidats et un écart-type de 5,35. Sur les 1225 signatures destination, nous avons obtenu les résultats suivants : 154 associations correctes, 13 associations fausses et 1058 signatures destination sans associations. Le TBR représente 92,22 % et le TR 13,63 %. Le TR est faible, cependant nous avons estimé la matrice origine – destination T du tableau 8.13.

		Destination		
		D	M	G
Origine	VI	4 (2,90 %)	9 (6,52 %)	3 (2,17 %)
	VL	36 (26,09 %)	29 (21,01 %)	1 (0,72 %)
	VR	27 (19,57 %)	27 (19,57 %)	2 (1,45 %)

Tableau 8.13 – Matrice origine – destination estimée T pour la première session.

La matrice E de l'erreur relative (tableau 8.14) entre les pourcentages des flux de la matrice origine – destination R et de la matrice origine – destination T est calculée. Les éléments de la matrice E sont obtenus par l'équation suivante :

$$E(i,j) = \frac{R(i,j) - T(i,j)}{R(i,j)} \tag{8.3}$$

avec $i \in 1,2,3$ représentant les lignes de la matrice et $j \in 1,2,3$ représentant les colonnes.

		Destination		
		D	M	G
Origine	VI	0,29	0,16	0,42
	VL	0,23	-0,17	0,03
	VR	-3,24	-0,50	0,90

Tableau 8.14 – Matrice d'erreur E pour la première session.

Les différences vis-à-vis de la matrice de référence R sont trop importantes. Les déplacements entre l'origine VR et les destinations D et M sont surestimés et le déplacement entre l'origine VR et la destination G est très largement sous-estimé. Malgré un TBR correct, le TR est trop faible pour que la matrice origine – destination soit représentative.

8.3.6 Conclusion

L'implantation des magnétomètres directement sur le site de MOCoPo ne nous a pas permis de réaliser des expérimentations sous des conditions maîtrisées. Cependant nous avons proposé des algorithmes de détection des véhicules pour les capteurs magnétomètre. Une comparaison vis-à-vis

de la boucle inductive au niveau de la détection a été effectuée. Les résultats tendent à démontrer que le magnétomètre peut avoir des performances identiques à la boucle inductive pour la détection. Le contexte difficile du trafic sur l'échangeur du Rondeau explique la difficulté de séparer les classes « identique » et « différente » à partir des magnétomètres. Dans le cadre de notre expérimentation, l'utilisation de la DTW en trois dimensions avec la distance euclidienne ne permet pas d'atteindre 78 % de réidentification sans commettre d'erreur comme Charbonnier et coauteurs [80]. Pour cette expérimentation, l'utilisation de la distance de Manhattan à la place de la distance euclidienne améliore l'AUC obtenue par la DTW. La normalisation en nombre de points associée à la distance de Manhattan accroît l'AUC. Le vote à l'unanimité en fonction des trois axes se révèle plus pertinent pour augmenter significativement le TBR que l'utilisation conjointe des axes dans le calcul de la distance. Cependant, le TR diminue nettement et la réidentification n'est plus représentative pour l'estimation de matrice origine – destination. Pour améliorer les résultats, il faudrait effectuer une étude plus approfondie en milieu contrôlé de certaines situations (véhicule excentré...) pour les analyser. Une autre solution consisterait à l'ajout de magnétomètres (réalisation d'un réseau de capteurs) pour obtenir plus d'informations.

8.4 Conclusion

Les algorithmes développés pour le suivi de véhicules s'appliquent indépendamment des sites d'expérimentations. Pour les boucles inductives, le suivi de véhicules donne des résultats très intéressants aussi bien pour les temps de parcours que pour l'estimation des matrices origine – destination. Cependant les résultats obtenus dépendent du taux de bonne réidentification mais aussi du nombre de candidats. Afin d'obtenir de meilleures estimations possibles, il faut envisager une modélisation du trafic (un exemple simple de modélisation du trafic est la fenêtre temporelle).

Pour l'expérimentation avec les magnétomètres, les résultats obtenus sont insuffisants pour l'estimation de matrice origine – destination. Il est à noter que la situation complexe du site et les conditions de trafic font que malgré une fenêtre temporelle le nombre de candidats reste élevé. Il faut optimiser les algorithmes de suivi pour les magnétomètres pour améliorer le taux de bonne réidentification avant de passer à des applications de temps de parcours et de matrice origine – destination.

9

Conclusion

Sommaire

9.1 Bilan 159
9.2 Perspectives 161

9.1 Bilan

Cette thèse a pour objectif l'amélioration du suivi anonyme de véhicules pour le développement des algorithmes « Intelligents » de gestion du trafic. Pour réaliser cet objectif, nous nous sommes imposés deux contraintes : une première contrainte vis-à-vis du gestionnaire de réseau et une seconde vis-à-vis de l'usager. La première consiste à délivrer au gestionnaire l'ensemble des principales données de gestion de trafic en plus du suivi de véhicules. La seconde considère la perception de l'usager à l'égard de l'anonymat. En pratique, quelles que soient les données relevées par n'importe quel type de capteur, celles-ci sont rendues anonymes en les cryptant, par contre l'usager n'aura pas forcément l'impression que son anonymat est conservé. Ceci guide notre choix vers des capteurs ne relevant pas d'informations personnelles et étant discrets. Ces exigences ont orienté en partie notre choix vers deux capteurs magnétiques :

- la boucle inductive ;
- et le magnétomètre.

La boucle inductive a aussi été sélectionnée pour les raisons suivantes : c'est l'un des capteurs les plus répandus en France ; son implantation et développement depuis plusieurs décennies en ont fait un capteur fiable et robuste ; ce capteur est insensible aux conditions météorologiques. Le magnétomètre a été adopté quant à lui pour différentes considérations. Depuis 2004, cette technologie commence à être déployée dans le domaine du trafic routier aux États-Unis ; son coût et son temps d'installation sont concurrentiels ; l'information est sensée être plus abondante ; les différentes technologies de capteurs évoluent sérieusement ces dernières années en matière de sensibilité, résolutions et miniaturisation. Cependant, le magnétomètre AMR choisi, présente l'inconvénient d'être sensible aux variations de la température.

Dans le cadre des capteurs magnétiques, le suivi de véhicules est la capacité à identifier le véhicule lors de son passage au niveau d'un capteur origine et de le réidentifier lors de son passage au

niveau d'un capteur destination (Chapitre 4). Le processus se décompose en trois grandes étapes : la détection, le prétraitement des informations et la réidentification. Une fois le suivi de véhicules réalisé, les temps de parcours individuels et l'estimation des matrices origine – destination sont des applications possibles pour la gestion « intelligente » du trafic. La problématique du suivi de véhicules reste complexe car comme dans tout système de mesures, l'information acquise par les capteurs est bruitée.

Les boucles inductives obtiennent de très bon taux de détection et présentent peu de fausses détections. Par contre, la détection pour le magnétomètre est plus difficile. La détection et la segmentation des signaux issus des magnétomètres n'étaient pas optimales pour la réidentification des signatures. Nous avons utilisé l'algorithme développé par Chinrungrueng et coauteurs pour la détection. De manière à prendre en compte la situation complexe de notre site d'expérimentation, l'algorithme a été modifié. Les modifications apportées ont permis d'améliorer sensiblement le taux de bonne détection et ont surtout diminué fortement les fausses détections, par rapport à l'algorithme de Chinrungrueng et coauteurs. Au niveau de l'entrée de l'échangeur du Rondeau, l'algorithme développé a un taux de détection supérieur (93,62 %) à celui des capteurs industriels Sensys (81,90 %). Mais le taux de fausse détection (0,90 %) est légèrement supérieur (0,45 %) à celui des magnétomètres Sensys. Pour la sortie de l'échangeur du Rondeau, le taux de détection (94,87 %) se rapproche de celui des boucles inductives (97,26 %). En ne commettant aucune fausse détection, le magnétomètre obtient un meilleur taux de fausse détection que la boucle inductive (0,62 %).

Au niveau des prétraitements, nous avons mis en place deux techniques pour améliorer les résultats du suivi de véhicules. Ces techniques ont été testées sur des signaux issus de la boucle inductive. La première technique consiste à retenir les variables les plus pertinentes au sens de l'Analyse en Composantes Principales (ACP) en fonction de la population (Section 6.2). Les trois populations considérées sont les Véhicules Légers (VL), les Poids Lourds (PL) et l'ensemble des deux populations (VL – PL). Le taux de bonne réidentification est en général pour les populations de VL et de PL supérieur à celui de la population mixte VL – PL. La deuxième technique repose sur la déconvolution aveugle (Section 6.3). Les algorithmes de déconvolution proposés par Kwon obligent à recalculer pour chaque véhicule la fonction de transfert. Nous avons défini un cadre mathématique rigoureux pour l'algorithme de déconvolution aveugle basé sur la régularisation de Tikhonov. Nous avons démontré la convergence de l'algorithme proposé vers une solution unique pour la fonction de transfert. La fonction de transfert est maintenant calculée une seule fois lors d'une phase d'apprentissage, et cette fonction permet d'estimer le signal déconvolué pour le passage de tous les véhicules. De plus, nous avons prouvé que le signal déconvolué améliore la réidentification par rapport au signal brut. Les deux techniques sont proposées séparément mais elle pourrait être utilisée conjointement.

Le dernier point abordé concerne la réidentification. La distance de Canberra a été introduite pour le suivi de véhicules à partir des signaux déconvolués. Celle-ci a obtenu de meilleures performances pour la réidentification que les distances de Manhattan, Euclidienne et de corrélation. Une adaptation des Séparateurs à Vastes Marges a été utilisée. La distance entre le nouvel élément et la marge est prise en compte pour la réidentification. Cette méthode a donné des taux de bonne réidentification similaire à ceux du maximum de vraisemblance avec un taux de représentation souvent inférieur. Dans la littérature, l'application d'un seuil est courant pour éviter les erreurs d'associations entre les signatures origine et les signatures destinations. Nous avons dans un premier temps proposé une méthode pour estimer la valeur du seuil de décision en fonction du taux de bonne réidentification et du taux de réidentification. Le problème de la variation du seuil en fonction des conditions de trafic a été relevé. Pour pallier cette difficulté, nous avons développé la méthode de vote à l'unanimité. Le vote à l'unanimité peut être exécuté avec une même méthode d'évaluation de dissimilarité et des variables différentes ou plusieurs méthodes différentes avec les mêmes variables (Section 7.3.3). L'association du couple origine – destination n'est effectif que si l'ensemble des solutions obtenues est identique. Les comparaisons entre le vote à l'unanimité et l'utilisation d'un seuil de décision, au cours de nos expérimentations, ont toujours donné de meilleurs taux de bonne réidentification pour

le vote à l'unanimité. Cette solution évite la détermination du seuil de décision. Lors des évaluations, nous avons introduit une fenêtre temporelle pour réduire le nombre de candidats (Section 7.3.2). Cette réduction du nombre de candidats permet d'augmenter le taux de bonne réidentification, cependant l'évaluation de la réidentification devient dépendante de la modélisation du trafic. Nous avons appliqué une solution simple pour modéliser le trafic avec une fenêtre temporelle fixe, d'autres algorithmes de modélisation du trafic plus performants augmenteraient le taux de bonne réidentification. Pour évaluer uniquement la réidentification, nous avons créé un indicateur en fonction du nombre de candidats.

Nous avons mené plusieurs expérimentations pour évaluer les algorithmes que nous avons développés. Sur ces expérimentations, deux types d'applications ont été analysées : les temps de parcours individuels et les matrices origine – destination. Les résultats obtenus au niveau du suivi de véhicules avec les signaux issus de la boucle inductive montrent que les performances atteintes permettent la réalisation d'application telles que les temps de parcours individuels et les matrices origine – destination. Par contre, l'expérimentation menée avec les magnétomètres présente un contexte complexe pour le suivi de véhicules. Sur cette expérimentation, nous avons amélioré la détection à un niveau de performance presque équivalent à celui des boucles inductives (Section 5.4). Cependant, malgré les améliorations au niveau du suivi de véhicules que nous avons apportées, celles-ci restent insuffisantes pour l'estimation des matrices origine – destination (Section 8.3). Le magnétomètre n'offre certes pas les mêmes performances que la boucle inductive dans la configuration testée, néanmoins l'ensemble des techniques utilisées pour les boucles inductives n'a pas encore été expérimenté avec les magnétomètres et le site expérimental s'est révélé très complexe pour le suivi de véhicules.

9.2 Perspectives

Une des perspectives qui apparaît à l'issue de cette thèse concerne l'industrialisation de notre travail. En effet, au niveau de la boucle inductive, la séparation en différentes populations existe déjà sur certains modèles de détecteur. Pour pouvoir être implémenté au niveau des détecteurs, le calcul des caractéristiques individuelles retenues en fonction de la population nécessite juste de tenir compte des contraintes de mémoire et de temps d'exécution. Cette adaptation ne représente pas de difficulté particulière. Une fois les caractéristiques calculées, les données individuelles sont ensuite transmises vers un serveur et stockées. Pour une application en temps réel, il faut s'assurer que la transmission des données vers le serveur est dimensionnée correctement. Lors de la réception des données individuelles, un module développé sur le serveur effectuera la réidentification. Nous recommandons d'utiliser un vote à l'unanimité pour la réidentification. À partir du suivi de véhicules par boucle inductive, les applications de temps de parcours et d'estimation de matrices origine – destination sont possibles. Une première version a été réalisée en partenariat avec l'industriel Thalès sur un site expérimental. Les données individuelles de quatre stations de mesures étaient transmises en temps réel vers une base de données. Ainsi, le gestionnaire disposait des informations de temps de parcours et de matrice origine – destination en temps réel. Cette version a démontré la faisabilité de l'industrialisation, mais elle ne prenait pas en compte la séparation des populations ni le vote à l'unanimité. Le calcul du signal déconvolué est aussi une fonction qu'il est plausible d'implémenter rapidement sur les détecteurs. La solution que nous avons proposée limite fortement les calculs au niveau du détecteur. La fonction de transfert de la boucle inductive étant maintenant connue, le calcul se limite à une opération entre le signal observé et la fonction de transfert. Ainsi, les principaux apports de cette thèse sont susceptibles d'être transférés dans le domaine industriel.

D'autres perspectives sont envisageables au niveau de la recherche pour les boucles inductives. Nous avons expérimenté la déconvolution aveugle des boucles inductives pour le standard de boucle le plus répandu en France. Nous souhaiterions appliquer cette méthode à d'autres formats de boucles inductives. Nous pensons qu'un véhicule passant sur une boucle standard et une boucle d'un autre format doit avoir le même signal déconvolué. Cette hypothèse reste à vérifier. Elle permettrait de

réaliser le suivi de véhicules avec un ensemble de boucles inductives hétérogènes, situation qui se rencontre sur certains réseaux (en région parisienne par exemple). L'algorithme de déconvolution développé pourrait être aussi étendu à d'autres capteurs. La séparation en deux populations distinctes est basée sur la norme NF P 99-300, cette séparation n'est peut être pas optimale. Nous proposons d'utiliser une méthode de classification non supervisée. Cette classification a pour objectif de regrouper les individus dans des classes. Chaque classe doit être le plus homogène possible. Entre elles, les classes sont le plus distinctes possibles. Les classes déterminées seront certainement différentes de celles basées sur la norme et elles ne seront pas forcément au nombre de deux. Une fois les classes désignées, une sélection de variables est effectuée pour chaque classe. Nous pensons que la réidentification des véhicules pourrait ainsi être améliorée.

Au regard des résultats obtenus avec les magnétomètres, nous pensons que des améliorations doivent être encore apportées avant d'entrer dans une phase d'industrialisation. Pour l'instant, une seule des expérimentations parmi les quatre réalisées a été exploitée. L'analyse de ces expérimentations pourrait nous donner des informations complémentaires pour améliorer le suivi de véhicules par les magnétomètres. Pour la détection, une comparaison entre la méthode que nous avons proposée et celle développée en prenant la ligne de base devrait être effectuée. Nous avons analysé les signaux pour un capteur indépendamment de ceux des voies adjacentes, or une partie des fausses détections est due à des chevauchements entre les voies. Le véhicule se trouve être détecté deux fois, une comparaison entre les deux détections est susceptible de supprimer le signal présentant la variation d'amplitude la plus faible. La détection a été considérée en plaçant les capteurs au milieu des voies, il serait judicieux d'examiner d'autres configurations dans le positionnement des capteurs. Localiser les capteurs sur le bord de la voie est une des suggestions à étudier. Les capteurs deviennent indépendants de la voie, ce qui implique que les véhicules en chevauchement sur deux voies n'ont plus d'impact sur la détection. Mais cela nécessite de réaliser un réseau de capteurs en bord de voie et de développer des algorithmes de séparation de sources pour la détection des véhicules. De nombreuses questions se posent sur le nombre de capteurs, la topologie du réseau de capteurs... En effet, pour effectuer la séparation de sources, il faut au moins un capteur de plus que le nombre de sources à détecter. L'implantation en bord de voie semble plus complexe à réaliser mais présente l'avantage de ne pas créer d'interruption de la circulation lors de la pose et la maintenance des équipements. En ce qui concerne la réidentification, une recherche plus approfondie doit être menée pour sélectionner des variables plus caractéristiques des signatures. Nous évoquons aussi la possibilité que la topologie d'un seul capteur soit insuffisante pour des situations aussi complexes que celles rencontrées lors de notre expérimentation. Il faudrait tester un réseau de capteurs par voie et répondre aux questions de la topologie et du nombre de capteurs à utiliser.

La démocratisation de l'implantation de magnétomètre comme capteur de mesures de trafic laisse suggérer qu'une étude de suivi entre des capteurs de natures différentes serait envisageable. La question de prévoir le remplacement progressif d'un capteur par un autre nous amène à poser la question de réaliser le suivi de véhicules entre un magnétomètre et une boucle inductive.

Publications

Revues internationales avec comité de lecture

- D. Guilbert, S. Ieng, C. Le Bastard et Y. Wang : « Robust blind deconvolution process for vehicle re-identfication by an Inductive Loop Detector ». Sensors Journal, IEEE, vol. 14, n° 12, pages 4315 – 4322, Dec. 2014

  DOI : 10.1109/JSEN.2014.2345755.

Revues internationales professionnelles

- F. Bosc, D. Guilbert, C. Le Bastard et J. Gaudin : « Mesure du trafic des deux-roues motorisés par magnétomètre ». Revue générale des routes et de l'aménagement, octobre – novembre 2014, n° 923.

Conférences internationales avec actes

- D. Guilbert, C. Le Bastard, S. Ieng et Y. Wang : « Re-identification by Inductive Loop Detector : Experimentation on target origin - destination matrix ». Intelligent Vehicle Symposium (IV), 2013 IEEE, 23-26 June 2013, Gold Coast, QLD, pages 1421 – 1427

  DOI : 10.1109/IVS.2013.6629666.

- C. Le Bastard, D. Guilbert, A. Delepoulle, A. Boubezoul, S. Ieng, Y. Wang : « Vehicule identification from inductive loops application : Travel time estimation for a mixed population of cars and trucks ». Intelligent Transportation Systems (ITSC), 2011 IEEE, 5 – 7 October 2011, Washington DC, pages 507 – 512

  DOI : 10.1109/ITSC.2011.6082802.

- D. Guilbert, F. Bosc, C. Le Bastard et G. Louah : « SAROT : an experimental platform for traffic analysis ». 8^{th} ITS European Congress, 6 – 9 June 2011, Lyon, France.

- D. Guilbert, C. Le Bastard et A. Bacelar : « Origin - Destination matrix by Using Inductive Loop Detector », Transport Research Arena, June 2010, Brussel.

Conférences internationales

- D. Guilbert, C. Le Bastard, S. Ieng et Y. Wang : « Association of two Inductive Loop Based Algorithms for Vehicle Re-identification ». International Conference on Computational Intelligence and Software Engeneering, 9 – 11 December 2011, Wuhan, Chine.

- D. Guilbert, F. Bosc, C. Le Bastard et V. Boucher : « Presentation of an experimental platform and applications : SAROT ». Models and Measurements in Traffic (MMT10), 17 – 19 November 2010, Paris-Sud Université d'Orsay.

Conférences et séminaires

- D. Guilbert, C. Lopez, C. Le Bastard, S. Ieng et Y. Wang : « Détection et ré-identification par magnétomètre ». Séminaire de clôture du projet Solutions pour une Exploitation de la Route Respectueuse de l'Environnement et de la Sécurité (SERRE), 20 mai 2014, Paris.
- D. Guilbert et C. Le Bastard : « Temps de parcours individuel ». Journée technique Méthodologie de la validation des Temps de Parcours, 24 mai 2013, CERTU, Lyon.
- D. Guilbert et C. Le Bastard : « Analyse des données recueillies lors du projet MOCoPo ». Séminaire du projet Solutions pour une Exploitation de la Route Respectueuse de l'Environnement et de la Sécurité (SERRE), 21 – 22 mars 2013, Nantes.
- D. Guilbert, C. Le Bastard, S. Ieng, A. Ihamouten et Y. Wang : « Gestion du patrimoine routier par la connaissance du trafic : suivi de véhicules ». Journée Scientifiques Structural Health Monitoring, 22 – 23 mars, 2012, Nantes.
- D. Guilbert : « Connaissance du trafic routier », Journée de l'école doctorale, 2012, Nantes.
- D. Guilbert, F. Bosc et C. Le Bastard : « Aller plus loin avec les capteurs au sol : l'analyse de signatures magnétiques ». Séminaire du projet Solutions pour une Exploitation de la Route Respectueuse de l'Environnement et de la Sécurité (SERRE), 22 – 23 septembre 2011, Paris.

Rapports de recherche

- D. Guilbert et C. Le Bastard : « Mesure et mOdélisation de la COngestion et de la POllution Tâche 4 ». Programme PREDIT - Groupe Opérationnel 2 - Gestion du trafic Contrat n° 2010 MT CVS 121, mars 2014.
- D. Guilbert et C. Le Bastard : « Mesure et mOdélisation de la COngestion et de la POllution Tâche 3 ». Programme PREDIT - Groupe Opérationnel 2 - Gestion du trafic Contrat n° 2010 MT CVS 121, mars 2014.
- D. Guilbert : « MOCoPo Bibliographie sur les magnétomètres - Les données trafic et leurs applications ». Rapport d'étude numéro 46.1099.105, avril 2011.

Bibliographie

[1] Denis Baupin et Fabienne Keller : *Les nouvelles mobilités sereines et durables : concevoir et utiliser des véhicules écologiques*. Rapport technique N° 1713 ASSEMBLÉE NATIONALE - N° 293 SÉNAT. L'office parlementaire d'évaluation des choix scientifiques et technologiques, jan. 2014 (cf. pages 16, 17).
URL : www.ladocumentationfrancaise.fr/rapports-publics/144000098/
(visité le 20/01/2015).

[2] Christian de Boissieu : *Rapport du groupe de travail « Division par quatre des émissions de gaz à effet de serre de la France à l'horizon 2050 »*. Paris : La documentation Française, oct. 2006. ISBN : 2-11-006280-0 (cf. page 16).
URL : www.ladocumentationfrancaise.fr/rapports-publics/064000757/
(visité le 20/01/2015).

[3] Bernard Suard et Laurent Mignaux : *Chiffres clés du transport - Édition 2014*. Chromatiques Éditions, 2014 (cf. page 16).
URL : http://www.statistiques.developpement-durable.gouv.fr/publications/p/2113/873/chiffres-cles-transport-edition-2014.html
(visité le 20/01/2015).

[4] Michal Krzyzanowski, Brigit Kuna-Dibbert et Jürgen Schneider : *Health effects of transport-related air pollution*. World Health Organization Europe Copenhagen, jan. 2005. ISBN : 9289013737 (cf. page 16).

[5] CERTU : *Les temps de parcours - Estimation, diffusion et approche multimodale*. Éditions du Certu. Collection Dossiers. Certu, 2008. ISBN : 978-2-11-097163-0 (cf. pages 17, 18).
URL : http://www.certu-catalogue.fr/les-temps-de-parcours.html
(visité le 20/01/2015).

[6] Shuang Wang, Lijuan Cui, Dianchao Liu, Robert Huck, Pramode Verma, James J. Sluss et Samuel Cheng : « Vehicle Identification Via Sparse Representation ». In : *IEEE Transactions on Intelligent Transportation Systems* 13.2 (juin 2012), pages 955–962. ISSN : 1524-9050 (cf. page 17).
DOI : 10.1109/TITS.2011.2171034.

[7] Christine Buisson et Jean-Baptiste Lesort : *Comprendre le trafic routier - Méthodes et calculs*. Éditions du Certu. Collection Références 96. Certu, 2010. ISBN : 978-2-11-098892-8 (cf. page 18).

[8] Ré-Mi Hage : « Estimation du temps de parcours d'un réseau urbain par fusion de données de boucles magnétiques et de véhicules traceurs : Une approche stochastique avec mise en oeuvre d'un filtre de Kalman sans parfum ». Thèse de doctorat. Nantes : Université de Nantes, sept. 2012 (cf. page 18).
URL : http://tel.archives-ouvertes.fr/tel-00919589
(visité le 20/01/2015).

[9] Torgil Abrahamsson : *Estimation of origin-destination matrices using traffic count–a literature survey*. Rapport technique Interim Report IR98-021. Laxenburg : International institute for Applied System Analysis, 1998 (cf. page 18).
URL : http://www.iiasa.ac.at/publication/more_IR-98-021.php
(visité le 20/01/2015).

[10] Martin L. Hazelton : « Some comments on Origin-Destination Matrix Estimation ». In : *Transportation Research Part A : Policy and Practice* 37.10 (déc. 2003), pages 811–822. ISSN : 0965-8564 (cf. page 18).
DOI : 10.1016/S0965-8564(03)00044-2.

[11] Pei-Wei Lin et Gang-Len Chang : « A generalized model and solution algorithm for estimation of the dynamic freeway origin–destination matrix ». In : *Transportation Research Part B : Methodological* 41.5 (juin 2007), pages 554–572. ISSN : 01912615 (cf. page 18).
DOI : 10.1016/j.trb.2006.09.004.

[12] JunWei Li, BoLiang Lin, ZhiHui Sun et XueFei Geng : « An estimation model of time-varying origin -destination flows in expressway corridors based on unscented Kalman filter ». In : *Science in China Series E : Technological Sciences* 52.7 (juin 2009), pages 2069–2078. ISSN : 1006-9321 (cf. page 18).
DOI : 10.1007/s11431-009-0150-0.

[13] Dominique Guichon, Florient Piel, Frédéric Bosc, Christine Buisson et Yannick Pech : *Panorama des systèmes de recueil de données de trafic routier*. Rapport technique ISRN : EQ-SETRA–12-ED30-FR. SETRA, nov. 2012 (cf. pages 22–25, 27–29, 33).

[14] Sing Yiu Cheung et Pravin Varaiya : *Traffic Surveillance by Wireless Sensor Networks : Final Report*. California PATH Research Report UCB-ITS-PRR-2007-4. Berkeley : University of California, jan. 2007 (cf. pages 22, 25–28, 31, 38, 41, 42, 49, 56–58, 63, 64, 74–76, 101, 102).
URL : www.its.berkeley.edu/publications/UCB/2007/PRR/UCB-ITS-PRR-2007-4.pdf
(visité le 20/01/2015).

[15] Sébastien Aubin : « Capteurs de position innovants : application aux Systèmes de Transport Intelligents dans le cadre d'un observatoire de trajectoires de véhicules ». Thèse de doctorat. Institut National Polytechnique de Toulouse, 2009 (cf. pages 25, 26, 30).
URL : https://tel.archives-ouvertes.fr/tel-00484751/
(visité le 20/01/2015).

[16] Lawrence A. Klein, Milton K. Mills et David R. P. Gibson : *Traffic Detector Handbook : Third Edition-Volume I*. Rapport technique FHWA-HRT-06-108. Federal Highway Administration, oct. 2006 (cf. pages 26, 33–37, 73).
URL : www.fhwa.dot.gov/publications/research/operations/its/06108/
(visité le 20/01/2015).

[17] Martin Isaksson : « Vehicle Detection using Anisotropic Magnetoresistors ». Master's thesis. Göteborg : Chalmers University of Technology, mai 2008 (cf. pages 29, 38, 76).
URL : http://studentarbeten.chalmers.se/publication/70865-vehicle-detection-using-anisotropic-magnetoresistors
(visité le 20/01/2015).

[18] Milton K. Mills : « Self inductance formulas for quadrupole loops used with vehicle detectors ». In : *Vehicular Technology Conference, 1985. 35th IEEE*. Tome 35. Mai 1985, pages 81–87 (cf. page 35).
DOI : 10.1109/VTC.1985.1623335.

[19] Janusz Gajda, Ryszard Sroka, Marek Stencel, Andrzej Wajda et Tadeusz Zeglen : « A vehicle classification based on inductive loop detectors ». In : *Proc. 18th IEEE Instrumentation and Measurement Technology Conf. IMTC 2001*. Tome 1. 2001, pages 460–464 (cf. page 37).
DOI : 10.1109/IMTC.2001.928860.

[20] Taek Mu Kwon : *Blind Deconvolution of Vehicle Inductance Signatures for travel-Time Estimation*. Rapport technique MN/RC-2006-06. Mimnesota Departement of Transportation, fév. 2006 (cf. pages 37, 56, 74, 76–82, 86).
URL : http://www.lrrb.org/PDF/200606.pdf
(visité le 20/01/2015).

[21] Michael J. Caruso et Lucky S. Withanawasam : « Vehicle Detection and Compass Applications using AMR Magnetic Sensors ». In : *Sensors Expo Proceedings*. Tome 477. Published : Honeywell Inc. 1999 (cf. page 39).

[22] Slawomir Tumanski : « Modern magnetic field sensors – a review ». In : *Przegląd Elektrotechniczny* 10 (oct. 2013), pages 1–12 (cf. pages 38–41).
URL : pe.org.pl/articles/2013/10/1.pdf
(visité le 20/01/2015).

[23] Paul Mesa et Sasank Reddy : *Magnetometer Detection and Temperature Noise Control with SOS*. 2007 (cf. page 43).
URL : nesl.ee.ucla.edu/courses/ee202a/2007f/sample/projects/Magnetometer%20Faults%20-%20Report.pdf
(visité le 20/01/2015).

[24] Benjamin Coifman : « A New Algorithm for Vehicle Reidentification and Travel Time Measurement on Freeways ». In : *Applications of Advanced Technologies in Transportation, American Society of Civil Engineers*. 1998, pages 167–174 (cf. pages 47, 48).
URL : http://http.cs.berkeley.edu/~zephyr/resume/AATT.pdf
(visité le 20/01/2015).

[25] Benjamin Coifman : « Vehicle Reidentification and Travel Time Measurement Using Loop Detector Speed Traps ». Thèse de doctorat. Berkeley : University of California, jan. 1998 (cf. page 47).
URL : http://www.escholarship.org/uc/item/5d69n86x
(visité le 20/01/2015).

[26] Benjamin Coifman : « Vehicle Re-Identification and Travel Time Measurement in Real-Time on Freeways Using Existing Loop Detector Infrastructure ». In : *Transportation Research Record* 1643.1 (jan. 1998), pages 181–191. ISSN : 0361-1981 (cf. page 47).
DOI : 10.3141/1643-22.

[27] Benjamin Coifman : *Vehicle Reidentification and Travel Time Measurement on Congested Freeways*. Rapport technique UCB-ITS-PWP-99-18. Ohio State University, 1999 (cf. page 47).
URL : www.its.berkeley.edu/publications/UCB/99/PWP/UCB-ITS-PWP-99-18.pdf
(visité le 20/01/2015).

[28] Benjamin Coifman et Michael Cassidy : « Vehicle Reidentification and Travel Time Measurement on Congested Freeways ». In : *Transportation Research Part A : Policy and Practice* 36.10 (2002), pages 899–917. ISSN : 0965-8564 (cf. page 49).
DOI : 10.1016/S0965-8564(01)00046-5.

[29] Benjamin Coifman et Pravin Varaiya : *Deployment and Evaluation of Real-Time Vehicle Reidentification from an Operations Perspective*. California PATH Research Report UCB-ITS-PRR-2002-37. Institute of Transportation Studies, UC Berkeley, 2002 (cf. page 49).
URL : htpps://escholarship.org/uc/item/6tp5w2gt
(visité le 20/01/2015).

[30] Benjamin Coifman et Edgar Ergueta : « Improved Vehicle Reidentification and Travel Time Measurement on Congested Freeways ». In : *Journal of Transportation Engineering* 129.5 (sept. 2003), pages 475–483. ISSN : 0733-947X (cf. page 49).
DOI : 10.1061/(ASCE)0733-947X(2003)129:5(475).

[31] Benjamin Coifman et Sivaraman Krishnamurthy : « Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure ». In : *Transportation Research Part C : Emerging Technologies* 15.3 (2007), pages 135–153. ISSN : 0968-090X (cf. page 49).
DOI : 10.1016/j.trc.2007.03.001.

[32] Gen Li : « Vehicle Reidentification and Travel Time Estimation On Congested Highway ». Thèse de doctorat. Kongens Lyngby : Informatics et Mathematical Modelling, Technical University of Denmark, DTU, 2007 (cf. page 49).
URL : http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=5445
(visité le 20/01/2015).

[33] Stephen G. Ritchie et Sun Carlos : *Section Related Measures of Traffic System Performance : Final Report*. California PATH Research Report UCB-ITS-PRR-98-33. Irvine : California PATH Research Report, nov. 1998 (cf. pages 49, 74, 76, 100, 101).
URL : http://escholarship.org/uc/item/4sc0t3bv
(visité le 20/01/2015).

[34] Seyed Mohammad Tabib : « Vehicle re-identification based on inductance signature matching ». Thèse de doctorat. University of Toronto, 2001 (cf. pages 49, 74, 100, 101).
URL : http://hdl.handle.net/1807/16354
(visité le 20/01/2015).

[35] Gérard Dreyfus, Jean-Marc Martinez, Manuel Samuelides, Mirta B. Gordon, Fouad Badran et Sylvie Thiria : *Apprentissage statistique : Réseaux de neurones - Cartes topologiques - Machines à vecteurs supports*. Eyrolles, juil. 2011. ISBN : 9782212042986 (cf. pages 52, 111).

[36] Mandoye Ndoye, Virgil Totten, James V. Krogmeier et Darcy M. Bullock : « A signal processing framework for vehicle re-identification and travel time estimation ». In : *Proc. 12th Int. IEEE Conf. Intelligent Transportation Systems ITSC '09*. St. Louis, MO, oct. 2009, pages 1–6. ISBN : 978-1-4244-5519-5 (cf. pages 55, 74).
DOI : 10.1109/ITSC.2009.5309763.

[37] Sing Yiu Cheung, Sinem Coleri, Baris Dundar, Sumitra Ganesh, Tan Chin-Woo et Pravin Varaiya : *Traffic Measurement and Vehicle Classification with a Single Magnetic Sensor*. Institute of Transportation Studies, Research Reports, Working Papers, Proceedings. Institute of Transportation Studies, UC Berkeley, 2004, page 26 (cf. page 56).
URL : http://escholarship.org/uc/item/2gv111tv
(visité le 20/01/2015).

[38] Sing Yiu Cheung, Sinem Coleri Ergen et Pravin Varaiya : « Traffic surveillance with wireless magnetic sensors ». In : *Proceedings of the 12th ITS world congress*. Tome 1917. 2005, pages 173–181 (cf. page 56).

[39] Pisit KANATHANTIP, Wuttipong KUMWILAISAK et Jatuporn CHINRUNGRUENG : « Robust vehicle detection algorithm with magnetic sensor ». In : *Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON) Conf.* Chaing Mai : IEEE, 2010, pages 1060–1064. ISBN : 978-1-4244-5606-2 (cf. pages 58, 59, 63).

[40] Sangwon LEE, Dukhee YOON et Bhaskar KRISHNAMACHARI : *Car Counting Implementation Using Magnetometer and Ultrasonic Sensor in Wireless Sensor Network*. Rapport technique. Jan. 2008 (cf. pages 58, 63).

[41] Sangwon LEE, Dukhee YOON et Amitabha GHOSH : « Intelligent parking lot application using wireless sensor networks ». In : *Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on*. Irvine, CA, mai 2008, pages 48–57 (cf. pages 58, 63).
DOI : 10.1109/CTS.2008.4543911.

[42] Jatuporn CHINRUNGRUENG et Saowaluck KAEWKAMNERD : « Wireless magnetic sensor network for collecting vehicle data ». In : *Proc. IEEE Sensors*. Christchurch, oct. 2009, pages 1792–1795 (cf. pages 60, 67).
DOI : 10.1109/ICSENS.2009.5398447.

[43] Carlos SUN, Stephen G. RITCHIE et Kevin TSAI : « Algorithm Development for Derivation of Section-Related Measures of Traffic System Performance Using Inductive Loop Detectors ». In : *Transportation Research Record : Journal of the Transportation Research Board* 1643 (jan. 1998), pages 171–180 (cf. pages 74, 100).
DOI : 10.3141/1643-21.

[44] Carlos SUN, Stephen G. RITCHIE, Kevin TSAI et R. JAYAKRISHNAN : « Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Re-Identification on Freeways ». In : *Transportation Research Part C : Emerging Technologies* 7.4 (1999), pages 167–185. ISSN : 0968-090X (cf. pages 74, 100).
DOI : 10.1016/S0968-090X(99)00018-2.

[45] Stephen G. RITCHIE, Carlos SUN, Seri OH et Cheol OH : *Section-Related Measures of Traffic System Performance : Prototype Field Implementation*. California PATH Research Report UCB-ITS-PRR-2001-32. Irvine : Institute of Transportation Studies, UC Berkeley, oct. 2001 (cf. page 74).
URL : http://www.its.berkeley.edu/publications/UCB/2001/PRR/UCB-ITS-PRR-2001-32.pdf
(visité le 20/01/2015).

[46] Cheol OH et Stephen G. RITCHIE : « Real-Time Inductive-Signature-Based Level of Service for Signalized Intersections ». In : *Transportation Research Record : Journal of the Transportation Research Board* 1802 (jan. 2002), pages 97–104 (cf. page 74).
DOI : 10.3141/1802-12.

[47] Cheol OH et Stephen G. RITCHIE : « Anonymous Vehicle Tracking for Real-Time Traffic Surveillance and Performance on Signalized Arterials ». In : *Transportation Research Record* 1826.1 (jan. 2003), pages 37–44. ISSN : 0361-1981 (cf. pages 74, 81).
DOI : 10.3141/1826-06.

[48] Cheol OH, Andre TOK et Stephen G. RITCHIE : « Real-Time Freeway Level of Service Using Inductive-Signature-Based Vehicle Reidentification System ». In : *IEEE Transactions on Intelligent Transportation Systems* 6.2 (juin 2005), pages 138–146. ISSN : 1524-9050 (cf. page 74).
DOI : 10.1109/TITS.2005.848360.

[49] Stephen G. Ritchie, Seri Park, Cheol Oh, Shin-Ting Jeng et Andre Tok : *Field Investigation of Advanced Vehicle Reidentification Techniques and Detector Technologies - Phase 2*. California PATH Research Report UCB-ITS-PRR-2005-8. Irvine : Institute of Transportation Studies, UC Berkeley, 2005 (cf. page 74).
URL : https://escholarship.org/uc/item/3jq8609j
(visité le 20/01/2015).

[50] Stephen G. Ritchie, Seri Park, Cheol Oh, Shin-Ting Jeng et Andre Tok : *Anonymous Vehicle Tracking for Real-Time Freeway and Arterial Street Performance Measurement*. California PATH Research Report UCB-ITS-PRR-2005-9. Irvine : University of California, Berkeley, mar. 2005 (cf. page 74).
URL : https://escholarship.org/uc/item/50c6z6zh
(visité le 20/01/2015).

[51] Baher Abdulhai et Seyed Mohammad Tabib : « Towards anonymous vehicle tracking via inductance-pattern recognition ». In : *Transportation Research Board*. Washington, D.C., 2002, page 20 (cf. page 74).
URL : http://trid.trb.org/view.aspx?id=720858
(visité le 20/01/2015).

[52] Ahmed Tawfik, Aidong Peng, Seyed Mohammad Tabib et Baher Abdulhai : « Learning Spatio-temporal context for vehicle reidentification ». In : *2nd IEEE International Symposium on Signal Processing and Information Technology*. IEEE, 2002 (cf. page 74).

[53] Shin-Ting Jeng et Stephen G. Ritchie : « A new inductive signature method for vehicle reidentification ». In : *Transportation Research Board 84th annual meeting*. Washington, D.C., jan. 2005, pages 9–13 (cf. pages 74, 101).

[54] Shin-Ting Jeng : « Real-time Vehicle Reidentification System for Freeway Performance Measurements ». Thèse de doctorat. Irvine : University of California, 2007 (cf. pages 74, 101).
URL : www.utc.net/research/diss140.pdf
(visité le 20/01/2015).

[55] Stephen G. Ritchie, Shin-Ting Jeng, Yeow Chern Tok et Seri Park : *Corridor Deployment and Investigation of Anonymous Vehicle Tracking for Real-Time Traffic Performance Measurement*. California PATH Research Report UCB-ITS-PRR-2008-23. Institute of Transportation Studies, UC Berkeley, oct. 2008 (cf. pages 74, 87).

[56] Andre Tok, Shin-Ting Jeng, Hang Liu et Stephen G. Ritchie : « Design and Initial Implementation of an Inductive Signature-Based Real-Time Traffic Performance Measurement System ». In : *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*. Oct. 2008, pages 216–221 (cf. page 74).
DOI : 10.1109/ITSC.2008.4732703.

[57] Shin-Ting Jeng, Andre Tok et Stephen G. Ritchie : « Freeway Corridor Performance Measurement Based on Vehicle Reidentification ». In : *Intelligent Transportation Systems, IEEE Transactions on* 11.3 (sept. 2010), pages 639–646. ISSN : 1524-9050 (cf. pages 74, 101).
DOI : 10.1109/TITS.2010.2049105.

[58] Shin-Ting Jeng et Lianyu Chu : « Vehicle Reidentification with the Inductive Loop Signature Technology ». In : *Journal of the Eastern Asia Society for Transportation Studies* 10 (2013), pages 1896–1915 (cf. pages 74, 101).
DOI : 10.11175/easts.10.1896.

[59] Sio-Song Ieng, Christophe Grellier, Julien Rivault, Jean Bertrand et Michel Pithon : « Inductive-Loop-Based Vehicle Signature Features Analysis and the Anonymous Vehicle Reidentification for Travel Time Estimation ». In : *Transportation Research Board 86th Annual Meeting*. Washington, D.C., jan. 2007, page 16 (cf. pages 74, 76, 81, 103–105, 111, 120, 140, 145).
URL : trid.trb.org/view.aspx?id=801332
(visité le 20/01/2015).

[60] Sio-Song Ieng, Jean Bertrand, Alexis Bacelar et Jacques Nouvier : « Travel Time by Using Widespread Inductive Loops Network ». In : *Transport Research Arena*. Ljubljana, Slovenia, avr. 2008 (cf. pages 74, 76, 103–105, 140).

[61] David Guilbert, Cédric Le Bastard et Alexis Bacelar : « Origin-Destination matrix by Using Inductive Loop Detector ». In : *Transport Research Arena*. Brussels, Belgique, juin 2010 (cf. page 74).

[62] Joseph M. Ernst, James V. Krogmeier et Darcy M. Bullock : « Non-linear compensation of vehicle signatures captured from electromagnetic sensors with application to vehicle re-identification ». In : *Proc. 13th Int Intelligent Transportation Systems (ITSC) IEEE Conf.* 2010, pages 923–928 (cf. pages 74, 75).
DOI : 10.1109/ITSC.2010.5625132.

[63] Karric Kwong, Robert Kavaler, Ram Rajagopal et Pravin Varaiya : « Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors ». In : *Transportation Research Part C : Emerging Technologies* 17.6 (déc. 2009), pages 586–606. ISSN : 0968090X (cf. pages 76, 102, 103).
DOI : 10.1016/j.trc.2009.04.003.

[64] Karric Kwong, Robert Kavaler, Ram Rajagopal et Pravin Varaiya : « Real-Time Measurement of Link Vehicle Count and Travel Time in a Road Network ». In : *Intelligent Transportation Systems, IEEE Transactions on* 11.4 (déc. 2010), pages 814–825. ISSN : 1524-9050 (cf. pages 76, 102, 103).
DOI : 10.1109/TITS.2010.2050881.

[65] Taek Mu Kwon : « Route tracking of border crossing vehicles using inductance signatures of loop detectors ». In : *Measurement Systems for Homeland Security, Contraband Detection and Personal Safety Workshop, 2005. (IMS 2005) Proceedings of the 2005 IEEE International Workshop on*. Mar. 2005, pages 103–109 (cf. pages 76, 77, 79, 80, 86).
DOI : 10.1109/MSHS.2005.1502566.

[66] Robert M Gray : « Toeplitz and Circulant Matrices : A Review ». In : *Foundations and Trends® in Communications and Information Theory* 2.3 (2006), pages 155–239. ISSN : 1567-2190 (cf. page 78).
DOI : 10.1561/0100000006.

[67] D. Godard : « Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems ». In : *IEEE Transactions on Communications* 28.11 (nov. 1980), pages 1867–1875. ISSN : 0090-6778 (cf. pages 79, 80).
DOI : 10.1109/TCOM.1980.1094608.

[68] Michel Pithon : *Détection sélective des véhicules par analyse de leur signature électromagnétique, validation sur site*. Rapport technique. DLRCA, mar. 2006 (cf. pages 81, 83).

[69] Carlos Sun : *An Investigation in the Use of Inductive Loop Signatures for Vehicle Classification*. California PATH Research Report UCB-ITS-PRR-2000-4. Institute of Transportation Studies, UC Berkeley, mar. 2000 (cf. page 81).
URL : escholarship.org/uc/item/93j2v5d8
(visité le 20/01/2015).

[70] Cédric Le Bastard, Patrice Briand, Peggy Subirats, Eric Violette, David Doucet et Ronan Queguiner : *Utilisation des boucles électromagnétiques dans les observatoires locaux : Estimation du positionnement latéral & de la vitesse des véhicules*. Rapport technique. 2009 (cf. page 83).

[71] Christophe Grellier : *Reconnaissance et suivi de véhicules par analyse de leur signature électromagnétique*. Rapport technique. LRPCA - IMA, 2005 (cf. pages 83, 103, 140).

[72] Benjamin Coifman et Seoungbum Kim : « Speed estimation and length based vehicle classification from freeway single-loop detectors ». In : *Transportation Research Part C : Emerging Technologies* 17.4 (2009), pages 349–364. ISSN : 0968-090X (cf. page 87).
DOI : 10.1016/j.trc.2009.01.004.

[73] Charles W. Groetsch : *The theory of Tikhonov regularization for Fredholm equations of the first kind*. Boston : Pitman, 1984. ISBN : 0273086421 9780273086420 (cf. page 88).

[74] Per Christian Hansen : « The discrete Picard condition for discrete ill-posed problems ». In : *BIT Numerical Mathematics* 30.4 (déc. 1990), pages 658–672. ISSN : 0006-3835, 1572-9125 (cf. pages 88, 89, 91).
DOI : 10.1007/BF01933214.

[75] Per Christian Hansen : « The Truncated SVD As a Method for Regularization ». In : *BIT* 27.4 (oct. 1987), pages 534–553. ISSN : 0006-3835 (cf. page 91).
DOI : 10.1007/BF01937276.

[76] Per Christian Hansen : « The L-Curve and its Use in the Numerical Treatment of Inverse Problems ». In : *Computational Inverse Problems in Electrocardiology*. Advances in Computational Bioengineering 5. Southampton : WIT Press, 2001, pages 119–142. ISBN : 978-1-85312-614-7 (cf. page 94).

[77] José Pelegrí Sebastiá, Jorge Alberola Lluch et J. Rafael Lajara Vizcaíno : « Signal conditioning for GMR magnetic sensors : Applied to traffic speed monitoring GMR sensors ». In : *Sensors and Actuators A : Physical* 137.2 (2007), pages 230–235. ISSN : 0924-4247 (cf. page 94).
DOI : 10.1016/j.sna.2007.03.003.

[78] Shin-Ting Jeng et Stephen G. Ritchie : « New Inductive Signature Data Compression and Transformation Method for Online Vehicle Reidentification ». In : Washington, D.C., 2006, page 26 (cf. page 101).
URL : http://trid.trb.org/view.aspx?id=777825
(visité le 20/01/2015).

[79] Rene O. Sanchez : « Wireless Magnetic Sensor Applications in Transportation Infrastructure ». Thèse de doctorat. Berkeley : University of California, Berkeley, 2012 (cf. pages 102, 103).
URL : http://connected-corridors.berkeley.edu/sites/default/files/Rene%20Sanchez%20Dissertation.pdf
(visité le 20/01/2015).

[80] Sylvie Charbonnier, A. Pitton et A. Vassilev : « Vehicle re-identification with a single magnetic sensor ». In : *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*. 2012, pages 380–385 (cf. pages 102, 103, 105, 157).
DOI : 10.1109/I2MTC.2012.6229117.

[81] Rene O. Sanchez, Christopher Flores, Roberto Horowitz, Ram Rajagopal et Pravin Varaiya : « Arterial travel time estimation based on vehicle re-identification using magnetic sensors : Performance analysis ». In : *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. Oct. 2011, pages 997–1002 (cf. pages 102, 103).
doi : 10.1109/ITSC.2011.6083003.

[82] Rene O. Sanchez, Christopher Flores, Roberto Horowitz, Ram Rajagopal et Pravin Varaiya : « Vehicle re-identification using wireless magnetic sensors : Algorithm revision, modifications and performance analysis ». In : *2011 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. Juil. 2011, pages 226–231 (cf. page 103).
doi : 10.1109/ICVES.2011.5983819.

[83] Vladimir N. Vapnik : *The Nature of Statistical Learning Theory*. New York, NY, USA : Springer-Verlag New York, Inc., 1995. isbn : 0-387-94559-8 (cf. page 105).

[84] Abderrahmane Boubezoul : « Système d'aide au diagnostic par apprentissage : application aux systèmes microélectroniques ». Thèse de doctorat. Université Paul Cézanne Aix-Marseille III, mar. 2008 (cf. pages 105, 110).

[85] David Guilbert, Cédric Le Bastard, Marc Brénugat et Patrice Briand : *Utilisation des boucles électromagnétiques dans les observatoires globaux : Estimation des temps de parcours et des matrices Origine-Destination*. Opération métrologie, trajectoire et trafic (MTT) 46.09.99.101. Angers : CETE de l'Ouest, déc. 2009 (cf. pages 120, 142).

l'université nantes angers le mans
COMMUNAUTÉ D'UNIVERSITÉS ET ÉTABLISSEMENTS

UNIVERSITÉ DE NANTES

Thèse de Doctorat

David GUILBERT

Analyse et classification des signatures des véhicules provenant de capteurs magnétiques pour le développement des algorithmes « Intelligents » de gestion du trafic

Analysis and classification of vehicle signatures from magnetic sensors for the development of “smart” traffic management algorithms

Résumé

La circulation routière est au cœur des préoccupations de la société au travers des problématiques d'aménagement du territoire, de mobilité, de lutte contre l'insécurité routière, ou plus récemment de lutte contre la pollution. La connaissance des déplacements des véhicules permet de répondre en partie à ces préoccupations. Le développement de la mesure des déplacements individuels des véhicules peut être réalisé par le suivi des véhicules. Pour réaliser le suivi anonyme des véhicules, le choix des capteurs magnétiques est appréhendé au regard des principaux capteurs de trafic. Après une étude sur les propriétés physiques de la boucle inductive et du magnétomètre, les trois étapes (détection, prétraitement et réidentification) du processus de suivi sont développées.

Tout d'abord, un automate d'état est proposé pour améliorer la détection de véhicules par magnétomètre.

Ensuite, des prétraitements sont proposés. Le premier concerne la proposition d'une méthode de déconvolution aveugle pour le capteur « boucle inductive ». Le deuxième se situe sur la sélection des variables saillantes par analyse en composantes principales.

Par la suite, la méthode SVM est adaptée à la réidentification de véhicules. Un processus de vote à l'unanimité des méthodes logique floue, approche bayésienne et mesures de similarités est proposé et comparé par rapport à l'utilisation d'un seuil de décision. Un nouvel indicateur indépendant de la modélisation du trafic est proposé pour évaluer la réidentification.

Enfin, l'ensemble des propositions est évalué lors de différentes expérimentations avec pour objectif de mesurer les temps de parcours individuels ou d'estimer les matrices origine – destination.

Mots clés

Transports, temps de parcours, matrice origine – destination, boucle inductive, oscillateur, magnétomètre, méthodes de réidentification, reconnaissance de forme.

Abstract

Road traffic is at the heart of concerns for society due to issues of spatial development, mobility, the fight for better road safety or, more recently, environmentally friendly considerations. Observation and knowledge of travel patterns can partly help to answer these concerns. The development of a way to measure individual journeys can be achieved using vehicle tracking. To be able to anonymously track vehicles, magnetic sensors are chosen rather than the main traffic sensors. After a preliminary study of the physical properties of both the inductive loop and magnetometer, three steps in the monitoring process (detection, pre-processing and re-identification) are developed.

Firstly, a state machine is provided to improve vehicle detection using a magnetometer.

Then, two new pre-processing steps are available. The first concerns the use of a novel method of blind deconvolution for the "inductive loop" sensor. The second concerns the selection of characterizing variables by principal component analysis.

Subsequently, the SVM method is adapted for the re-identification of vehicles. A unanimous voting process on either fuzzy logic, a Bayesian approach or similarity measurement is offered and compared in relation to the use of a decision threshold. A new independent predictor of traffic modelling is available to evaluate this re-identification.

Finally, all the suggestions are evaluated during different experiments with the goal of obtaining individual travel time measurements or estimates of the origin – destination matrix.

Key Words

Transport, travel time, origin – destination matrix, inductive loop, oscillator, magnetometer, re-identification methods, patterns recognition.