

RAPPORTS

Centre d'Études
Techniques de
l'Équipement de
l'Ouest

Laboratoire Régional
des Ponts et
Chaussées d'Angers

Opération **SERRES**

Estimation des temps de parcours à partir de boucles électromagnétiques pour une population mixte VL/PL

Octobre 2010

Certifié
ISO 9001



n° 9913220

Ressources, territoires, habitats et logement
Énergie et climat Développement durable
Prévention des risques Infrastructures, transports et mer

Présent
pour
l'avenir



Ministère
de l'Écologie,
de l'Énergie,
du Développement
durable
et de la Mer

Ministère de l'Écologie, de l'Énergie,
Du Développement durable et de la Mer
en charge des Technologies vertes et des Négociations sur le climat

www.cete-ouest.developpement-durable.gouv.fr

Historique des versions du document

Version	Auteur	Commentaires
1	<i>C. Le Bastard, A. Delepouille, Marc Brénugat</i>	<i>Relecture P. Landais et R. Dronneau.</i>

Affaire suivie par

Cédric Le Bastard – U46
Tél. : 02 41 79 13 05
Courriel : cedric.lebastard@developpement-durable.gouv.fr
Adresse postale : CETE de l'Ouest/Laboratoire Régional des Ponts et Chaussées d'Angers 23 avenue de l'Amiral Chauvin BP 20069 49136 Les Ponts de Cé Cédex

Référence Intranet

http://

Table des matières

Introduction	7
1 Expérimentations et base de données	9
1.1 Caractéristiques du site d'Angers	9
1.2 Caractéristiques du site de Rennes	11
1.3 Informations recueillies & Signatures	13
2 Méthode de réidentification	15
2.1 Données	15
2.1.1 Population de classe 1	16
2.1.2 Population de classe 2 à 10	17
2.1.3 Population de classe 1 à 10	17
2.1.4 Conclusion	17
2.2 Maximum de vraisemblance	18
2.3 Méthode de logique floue	19
2.4 Séparateur à vaste marge	20
3 Analyse des résultats	23
3.1 Introduction	23
3.2 Etude sur une base idéale	24
3.2.1 Performances en validation croisée	24
3.2.2 Performances sur l'ensemble de test (T)	25
3.3 Etude sur base perturbée	29
3.3.1 Méthodes proposées	29
3.3.2 Performances	32
3.4 Estimation des temps de parcours	35
Conclusion	37

Table des figures

1.1	Angers - Site Quai-Berge.	9
1.2	Angers - Répartition des véhicules recensés en fonction de la vitesse.	10
1.3	Rennes - Périphérique.	11
1.4	Rennes - Répartition des véhicules recensés en fonction de la vitesse.	12
1.5	Exemple de signature d'un véhicule de classe 10.	14
1.6	Exemple de signature d'un véhicule de classe 1.	14
2.1	Représentation des anciennes variables sur les deux axes principaux de la nouvelle base pour les PL.	16
3.1	Principe de la validation croisée pour N segments.	23
3.2	Matrice de coût réalisée à partir des données d'Angers.	27
3.3	Matrice de coût réalisée à partir des données de Rennes.	28
3.4	Taux d'identification en fonction de la répartition poids lourds - véhicules légers.	28
3.5	Evolution du pourcentage de véhicules légers en fonction du seuil (Flou)	30
3.6	Evolution du pourcentage de véhicules légers en fonction du seuil (MV1)	30
3.7	Evolution du pourcentage de véhicules légers en fonction du seuil (SVM-d)	31
3.8	Estimation des temps de parcours pour des VL	35
3.9	Estimation des temps de parcours pour des PL	36

Liste des tableaux

1.1	Tableau récapitulatif des caractéristiques des sites de récupération de données	12
3.1	Performances en validation croisée sur la base d'Angers.	24
3.2	Performances en validation croisée sur la base de Rennes.	25
3.3	Paramètres établis pour les méthodes utilisant les séparateurs à vaste marge.	25
3.4	Angers - Performances sans fenêtrage dans le cadre d'une mesure sur base totale.	25
3.5	Rennes - Performances sans fenêtrage dans le cadre d'une mesure sur base totale.	25
3.6	Performances maximales potentielles avec le fenêtrage utilisé.	26
3.7	Angers - Performances avec fenêtrage dans le cadre d'une mesure sur base totale.	26
3.8	Rennes - Performances avec fenêtrage dans le cadre d'une mesure sur base totale	26
3.9	Valeur de seuil pour un pourcentage donné de véhicules légers (Site d'Angers)	31
3.10	Valeur de seuil pour un pourcentage donné de poids lourds (Site d'Angers)	31
3.11	Angers - Performances témoins.	32
3.12	Angers - filtrage "Origine à multiples destinations".	33
3.13	Angers - Post-traitement seuil à 90 %.	33
3.14	Angers - Post-traitement seuil à 80 %.	33
3.15	Angers - Post-traitement seuil à 70 %.	34
3.16	Angers - Post-traitement seuil à 70 % et origines à multiples destinations	34

Introduction

Les déplacements sont, depuis quelques décennies, au centre des priorités nationales par les problèmes qui leurs sont liés comme la résorption des congestions, l'insécurité routière ou, plus récemment, la pollution engendrée. Aujourd'hui, il est acquis que les solutions respectueuses de l'environnement passent aussi par une bonne gestion du trafic sur le réseau existant.

Pour gérer efficacement le trafic routier, si possible en temps réel, le recueil et l'analyse des informations sont indispensables.

Pour cette raison, la recherche en trafic s'intéresse à l'utilisation des différents systèmes de détection toujours plus performants : vidéo, radar, infrarouge, GPS, téléphone portable ...

Malgré les avancées réalisées sur ces systèmes, la boucle électromagnétique reste encore le capteur le plus utilisé sur les routes en raison de sa facilité de mise en oeuvre, de sa capacité de fonctionnement par tous temps, de sa robustesse, de la faible maintenance associée mais aussi de sa bonne acceptabilité par les usagers du fait de la préservation de l'anonymat de ces derniers.

Depuis quelques années, avec l'aide de systèmes électroniques plus performants, les signaux caractéristiques des véhicules appelés aussi "signatures électromagnétiques" permettent un accès à une information plus complète et sont disponibles en temps réels.

De ce fait, grâce au nombre important de boucles électromagnétiques présentes sur le réseau routier Français, et sous réserve de la modification de la partie électronique des stations, l'information accessible aux gestionnaires pourrait très rapidement être beaucoup plus riche que celle dont ils disposent actuellement avec les dispositifs à une seule boucle (donnée trafic uniquement).

Cette étude se situe au niveau de la reconnaissance des signatures électromagnétiques des véhicules afin de fournir des informations en temps réel directement utilisables. Les applications traitées sont l'estimation des temps de parcours individuels des véhicules d'une population mixte VL/PL.

Le principe de la reconnaissance des véhicules par boucles est fondé sur la comparaison des caractéristiques d'une paire de signatures recueillies sur deux boucles successives d'un tronçon. Ces méthodes de reconnaissance, appelées aussi « méthodes de réidentification », associées aux systèmes « boucles électromagnétiques » permettent d'identifier, de réidentifier les véhicules et donc d'effectuer de la poursuite de véhicules sur une certaine portion de route. Ainsi, une estimation des temps de parcours peut-être réalisée. Pour effectuer

cette poursuite de véhicules en temps réel, différentes méthodes de traitement sont mises en oeuvre à différentes étapes (création des caractéristiques appelées aussi identifiants, méthode de réidentification, fenêtre d'analyse statique ou dynamique,...). L'optimisation de ces méthodes et la proposition de nouvelles méthodes de traitement seraient à étudier dans différentes situations. A ce jour, seule l'estimation des temps de parcours individuels des VL a été étudiée au sein du RST à partir des boucles électromagnétiques. De nombreuses études sont encore à réaliser pour différentes situations; chacune de ces situations ayant des problématiques différentes en terme de traitement.

Dans cette étude, nous proposons d'estimer les temps de parcours à partir de boucles existantes pour une population mixte VL/PL. La capacité de reconnaissance individualisée des véhicules par analyse de signature à partir de boucles électromagnétiques a déjà permis quelques expérimentations pour mesurer les temps de parcours des VL seulement [21]. Il est proposé de généraliser cette étude en prenant en compte la population entière des véhicules (VL/PL).

Dans un premier temps, les rapports [2, 9, 17] se sont intéressés à la faisabilité d'un système permettant d'effectuer de la poursuite de véhicules. Ensuite, l'étude présentée dans [10] a permis de déterminer les 9 caractéristiques les plus pertinentes. Une première expérimentation présentée dans [10] a permis de tester les méthodes de réidentification dans un flux non congestionné et pour de faibles temps de parcours. Cette étude a montré que le taux de réidentification des méthodes était de l'ordre de 50 %. Ces méthodes de réidentifications ont ensuite été testées dans trois situations différentes [21] :

- estimation des temps de parcours dans un flux faiblement congestionné : expérimentation de Rennes ;
- estimation dynamique des matrices Origine-Destination : expérimentation sur le site de SAROT zone 2 ;
- estimation des temps de parcours de longue distance : expérimentation de Porcaro-Chavagne.

Les performances des méthodes de réidentification sur ces trois sites ont été analysées pour mesurer les temps de parcours et les matrices Origine-Destination des VL seulement. Le présent rapport s'inscrit dans la suite de cette étude avec cette fois-ci des données provenant d'une population mixte VL et PL.

Tout d'abord, les expérimentations et les bases de données seront présentées, puis dans un second temps, trois méthodes de réidentification seront présentées. Ensuite, les performances de ces méthodes seront discutées. Enfin, une conclusion et des perspectives seront données.

Une partie du travail présenté dans ce rapport a été réalisée durant le stage d'Antoine Delepouille, étudiant à l'INPG.

Chapitre 1

Expérimentations et base de données

Cette étude est effectuée à partir de deux bases de données provenant du site "Quai-Berges" d'Angers et de la rocade sud de Rennes. Ces bases de données ont déjà été utilisées dans [16, 9, 10, 19, 21] pour des véhicules légers (VL) seulement. Ainsi, ces bases de données ont été retravaillées pour toutes les catégories de la classe 1 (C1) à la classe 10 (C10), i.e. pour des VL et PL (poids lourds).

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 mesure du site de test est équipé de caméras numériques en nombre suffisant pour filmer correctement les plaques d'immatriculation de tous véhicules passant sur chaque couloir. Le champ de la caméra est réglé de manière que les plaques soient enregistrées au droit des boucles de mesure. Avant de débiter l'expérimentation, les horloges des caméras et des stations à boucles électromagnétiques sont synchronisées avec celle du PC d'acquisition. A noter que chaque caméra nécessite la présence d'un opérateur. Les informations enregistrées sur les différentes zones de mesure sont les signatures électromagnétiques et la vidéo. Ensuite un traitement fastidieux est réalisée pour associer les signatures électromagnétiques aux informations vidéo.

1.1 Caractéristiques du site d'Angers

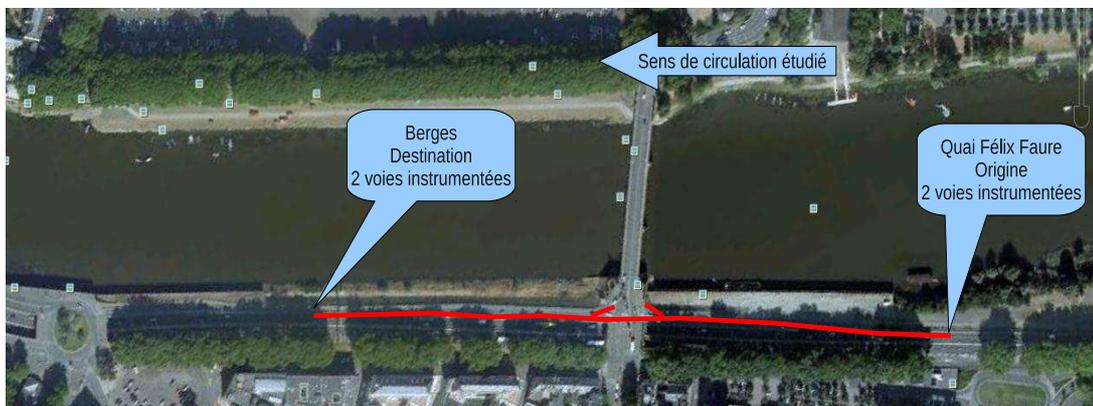


FIGURE 1.1 – Angers - Site Quai-Berge.

L'expérimentation d'Angers a été réalisée au niveau de la voie "Quai-Berge" sur 2 périodes (matin et après-midi). La vitesse est réglementée à 70 km/h. Le site d'Angers, présenté sur la figure 1.1, 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 mètres. A 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, seules 1538 VL et 206 PL sont passés sur les deux points de mesure. 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 1.2 représente le pourcentage de véhicules en fonction de leur vitesse en km/h. 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. Nous pouvons observer que les poids lourds ont une vitesse moyenne inférieure de 10 km/h par rapport aux véhicules légers. De plus, nous remarquons que la répartition des vitesses des poids lourds est moins étalée. Cette répartition des vitesses permettra de préciser les paramètres de fenêtrage qui seront utilisés avec les méthodes de réidentification. Dans cette situation, nous utiliserons une fenêtre rectangulaire centrée sur la valeur moyenne **60 km/h**, de demi-largeur **20 km/h**.

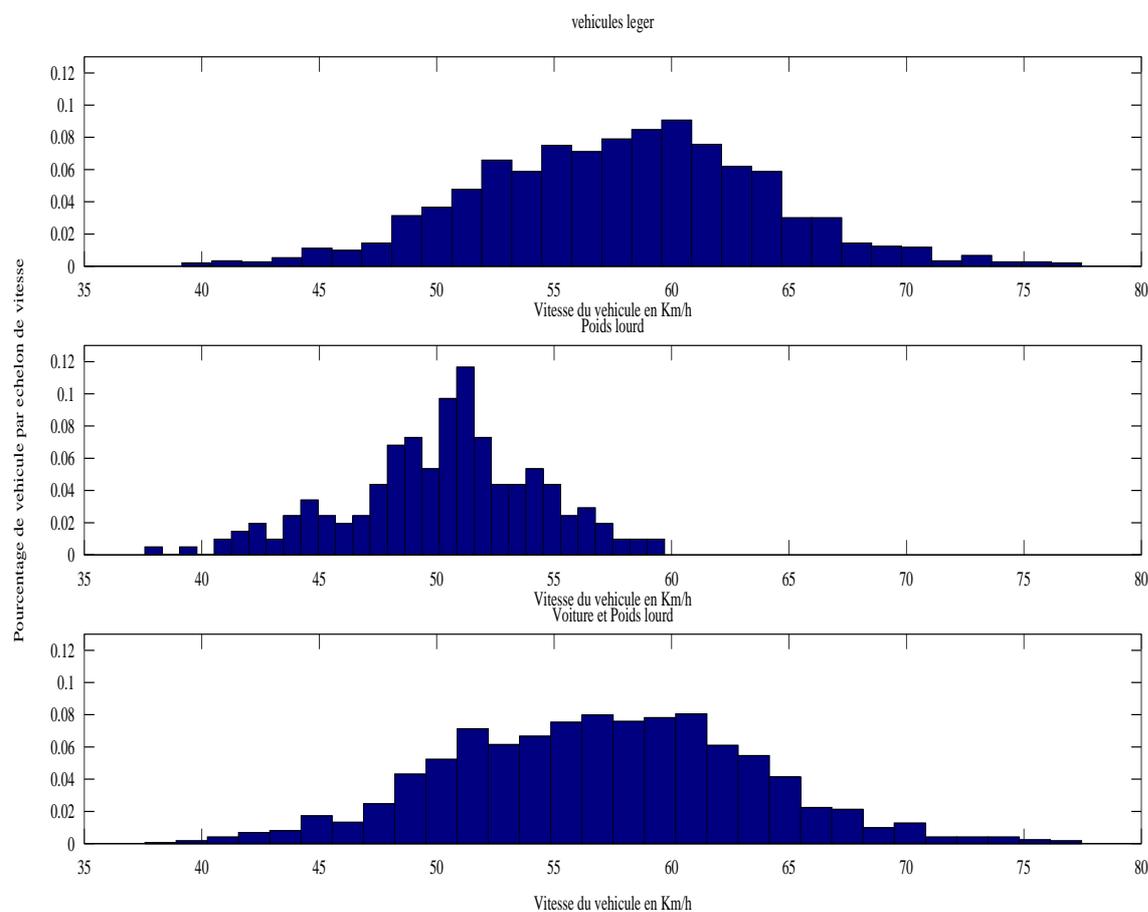


FIGURE 1.2 – Angers - Répartition des véhicules recensés en fonction de la vitesse.

1.2 Caractéristiques du site de Rennes

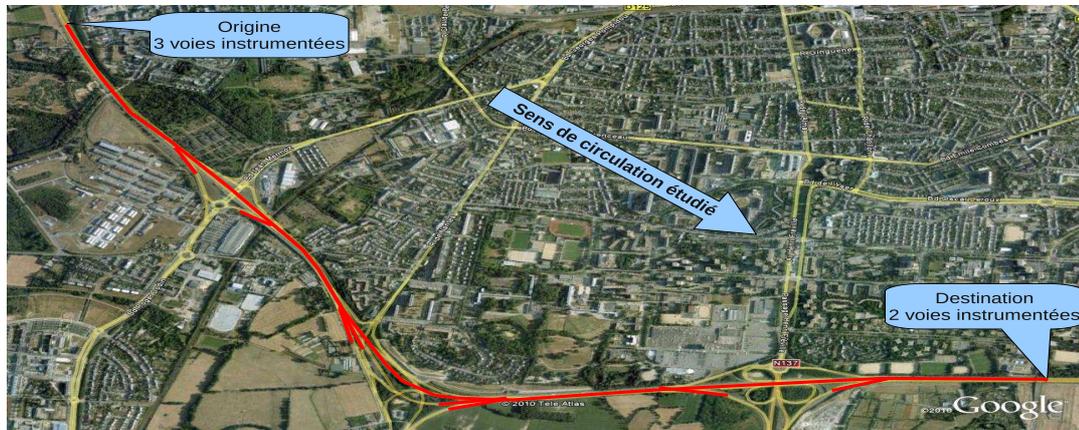


FIGURE 1.3 – Rennes - Périphérique.

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 [21] 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 mètres. La vitesse règlementée est de 90 km/h mais la vitesse moyenne enregistrée est de 50,5 km/h. Le début du parcours (origine) est composé de 3 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 : 5 entrées et 4 sorties.

Chaque point de mesure du site de test est équipé de caméras numériques en nombre suffisant pour filmer correctement les plaques d'immatriculation de tous véhicules passant sur chaque couloir. Le champ de la caméra est réglé de manière que les plaques soient enregistrées au droit des boucles de mesure. Avant de débiter l'expérimentation, les horloges des caméras et des stations à boucles électromagnétiques sont synchronisées avec celle du PC d'acquisition.

2490 VL, 217 PL et 3550 VL, 189 PL ont été respectivement relevées 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 mesurés. 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 1.4 représente le pourcentage de véhicules en fonction de leur vitesse en km/h. 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, nous appliquerons une fenêtre centrée sur la valeur moyenne **50 km/h**, de demi-largeur **15 km/h**. Cependant, la grande distance entre l'origine et la destination est néfaste au fenêtrage malgré une fenêtre plus serrée que dans le cas précédent. Les figures 1.2 et 1.4 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. Le tableau 1.1 récapitule les caractéristiques du site de Rennes ainsi que celui d'Angers.

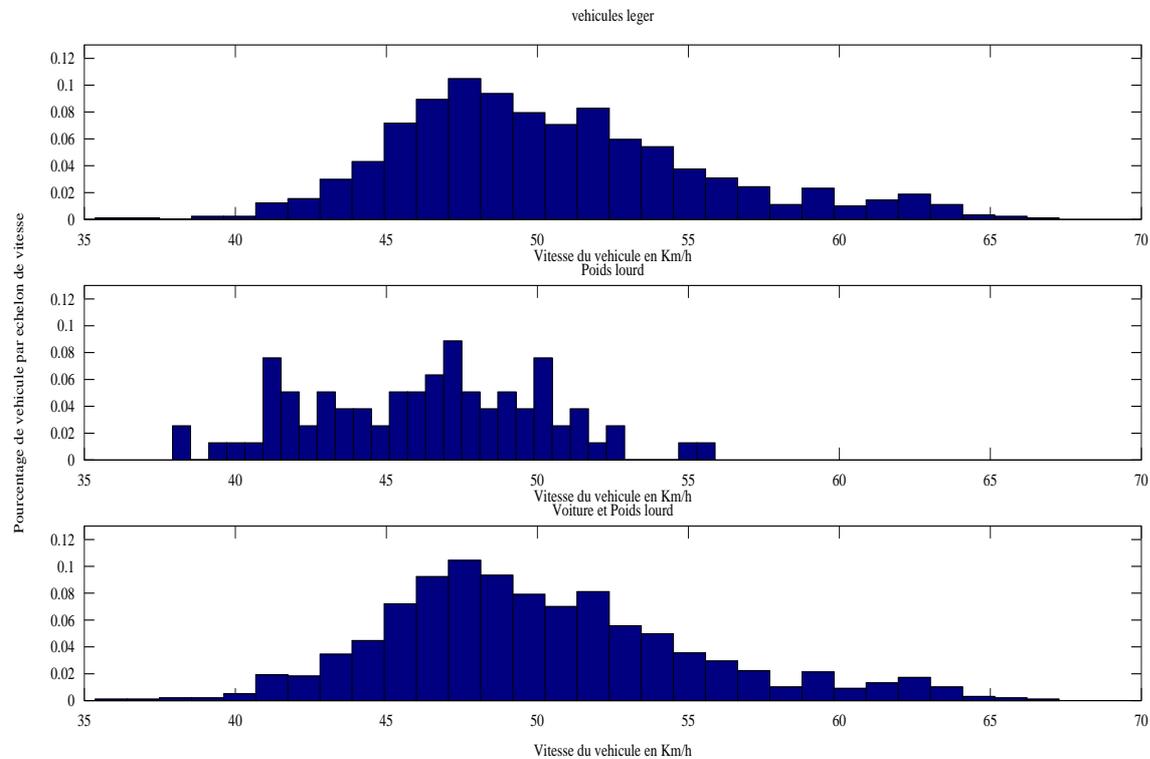


FIGURE 1.4 – Rennes - Répartition des véhicules recensés en fonction de la vitesse.

	Site angevin		Site rennais	
Distance entre les 2 zones	560 m		5150 m	
Entrées	1		5	
Sorties	1		4	
	Origine	Destination	Origine	Destination
Nombre de voies	2	2	3	2
Nombre de VL	1844	2881	2490	3550
Nombre de PL	210	237	217	189
	Véhicules passés sur les 2 zones			
Nombre de VL	1538		930	
Nombre de PL	206		80	
Observations	-Faible distance -Non congestionnée -Peu d'E/S non instrumentées		-Longue distance -Faiblement congestionnée -Nombre important d'E/S non instrumentées	

TABLE 1.1 – Tableau récapitulatif des caractéristiques des sites de récupération de données

1.3 Informations recueillies & Signatures

Lors de ces expérimentations, la lecture des plaques est nécessaire et se réalise manuellement. Un opérateur visionne les vidéo et recense tous les véhicules étant passés sur les zones instrumentées. Différents champs sont renseignés notamment la plaque d'immatriculation¹, le type de véhicule, l'horaire de passage et le nom du fichier contenant la signature liée à ce véhicule. Après analyse des vidéo, nous obtenons une base de données comprenant tous les véhicules passés sur un ou plusieurs des points de mesure ainsi que les véhicules passés sur les deux points de passage.

Lors de ces expérimentations, divers caractéristiques sont aussi mesurées à partir de la station tels que le numéro de capteur, la signature, le pas temporel entre chaque point (qui n'est pas constant), le nombre de points, la catégorie du véhicule, le maximum du signal, le coefficient de calibration, la date, l'horodatage, le temps de présence, la vitesse et un statut de fonctionnement.

La station mesure la signature électromagnétique. Elle est la mesure de la variation de la déformation du champ magnétique lors du passage d'un véhicule sur une boucle électromagnétique. La signature représente le reflet de la répartition de la masse métallique du véhicule. Les figures 1.5 et 1.6 montrent les signatures électromagnétiques d'un véhicules léger (signature avec un maximum) et d'un poids lourds (signature avec plusieurs maximums).

Dans les études présentées dans [2, 3, 4, 5], afin de pouvoir comparer au mieux les deux signaux d'un même véhicule au passage sur deux boucles différentes, divers prétraitements ont été réalisés :

- pour s'affranchir des problèmes liés aux temps d'acquisition aberrants, 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, mais aussi de la position du véhicule sur la boucle. 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 seront donc différentes et par conséquent leur amplitude aussi. Pour remédier à ce problème, chaque signal a été normalisé en amplitude [2]. A noter qu'un véhicule qui ne passe pas au même niveau sur les deux boucles, ne possède pas exactement la même signature, même après normalisation [20].
- 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 [2] plus de détails sur cette normalisation.

Les figures 1.5 et 1.6 mettent en lumière une partie du pré-traitement. Les croix rouges correspondent au relevé de la boucle normalisée à 5000. La courbe bleue correspond à la modélisation du relevé extrapolé à partir des croix rouges. Enfin, les ronds verts correspondent au signal ré-échantillonné à pas constant. A l'issue de ces divers traitements, un signal normalisé en 96 points est créé. Dans un second temps, différentes variables globales sont calculées. L'analyse de ces variables sera présentée dans la section 2.1.

1. La plaque d'immatriculation est cryptée afin de ne pas pouvoir identifier le véhicule réel.

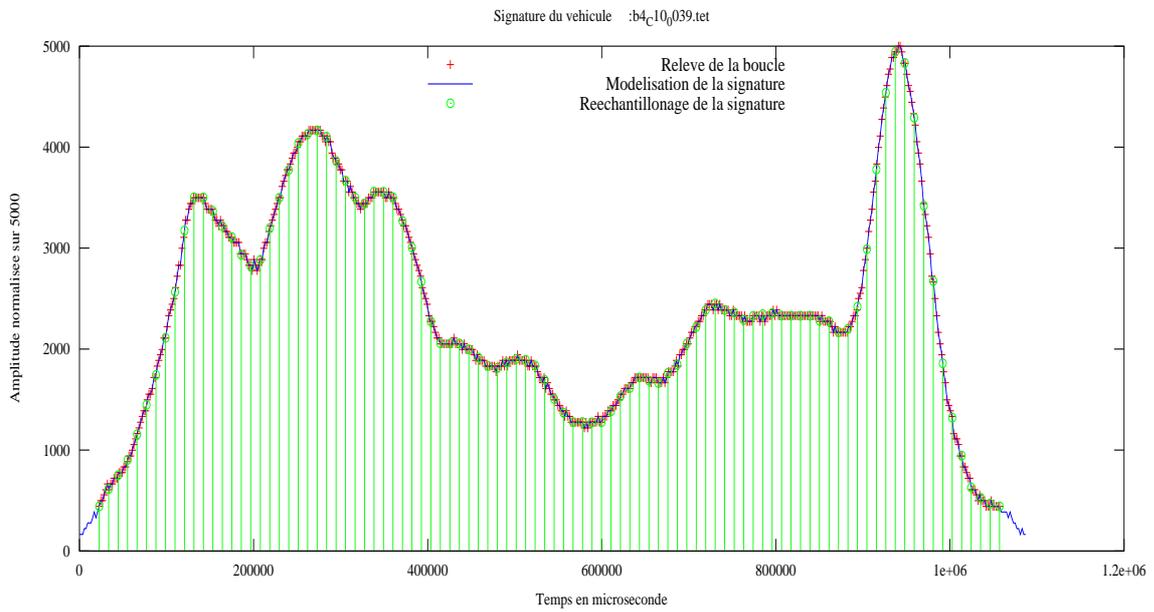


FIGURE 1.5 – Exemple de signature d'un véhicule de classe 10.

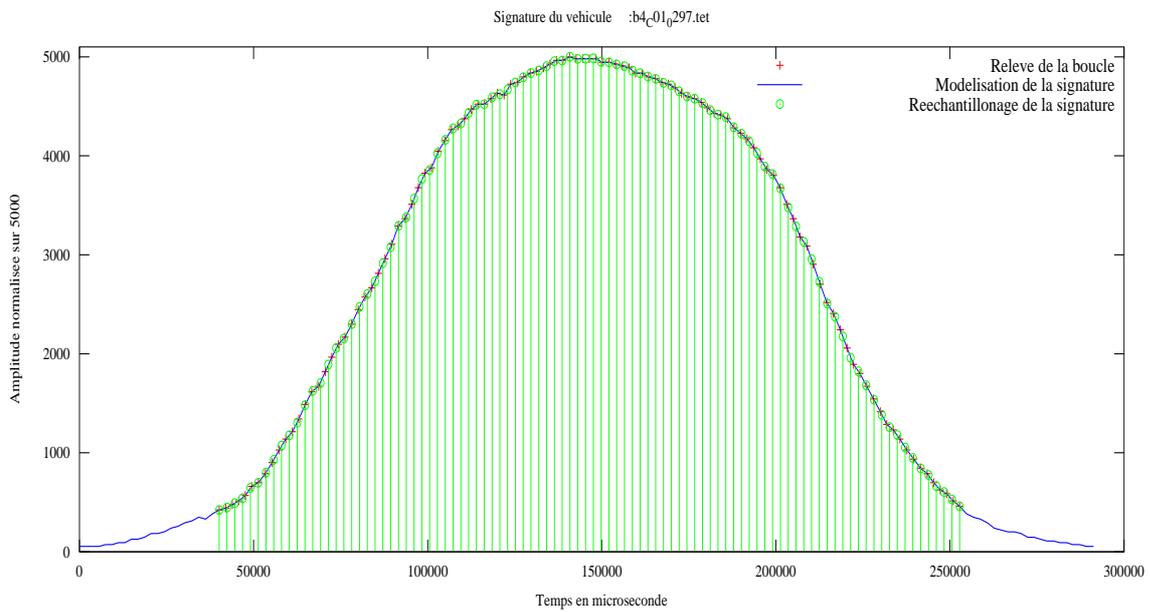


FIGURE 1.6 – Exemple de signature d'un véhicule de classe 1.

Chapitre 2

Méthode de réidentification

2.1 Données

Dans cette étude, nous travaillons à partir des signatures électromagnétiques. Chaque signature est ré-échantillonnée sur 96 points, quel que soit le type de véhicule (VL ou PL). Cette valeur a été choisie afin de ne pas perdre trop d'informations (notamment pour les poids lourds), mais aussi pour conserver une certaine rapidité de calcul en ne surchargeant pas le calculateur de données. D'après la littérature [2], 204 variables sont calculées à partir de la signature : 96 valeurs liées à la signature traitée, 47 variables fréquentielles obtenues à partir de la transformée de Fourier et 63 variables dites globales qui sont obtenues par calcul. Ainsi, chaque signature est caractérisée par 204 variables, trop de donnée et il y a certainement de la redondance. Cela conduit aussi à des temps de calculs trop longs et des traitements inutiles. Nous allons donc réduire le nombre de variables tout en conservant au mieux l'information contenue dans ces données. Pour ce faire, comme dans [16], nous avons réalisé une analyse en composantes principales qui permettra de déterminer quelles sont les variables à conserver et celles à rejeter. La procédure adoptée dans cette étude est celle présentée dans [16]. Le but est de sélectionner les variables les plus représentatives et les moins corrélées entre elles. Contrairement à [16], nous travaillons ici avec une base composée de VL et de PL. Cette analyse permet de sélectionner les variables les plus pertinentes pour chaque type de population.

Le tableau analysable est constitué de z véhicules caractérisés chacun par 204 variables ; z dépend de la population analysée. Trois types de populations sont analysées :

- VL (classe1)
- PL (classe 2 à 10)
- VL-PL (classe 1 à 10)

Chaque véhicule a la même importance dans l'analyse. Les variables étant hétérogènes, elles ont été normalisées (moyenne=0 et écart-type=1). Ainsi, après ACP, nous obtenons les valeurs propres, les taux d'inertie et les coordonnées des variables dans le plan 1-2 (1ère et 2ème composantes principales). On peut ensuite projeter l'ensemble des variables dans le plan 1-2 afin de rechercher les corrélations ou non corrélations entre elles comme il est illustré sur la figure 2.1.

L'objectif est de choisir un nombre minimum de variables 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.

aussi retenues : la bande passante¹, l'écart-type total et le coefficient de Skewness².

2.1.2 Population de classe 2 à 10

Les classes 2 à 10 correspondent au véhicule "poids lourds". Lors de cette étude, neuf caractéristiques ont été retenues. Cinq caractéristiques sont issues de la transformée de Fourier : la partie réelle de la 1^{ère} et 2^{ème} composante, la partie imaginaire de la 2^{ème} et 3^{ème} composante et le module de la 3^{ème} composante. De plus, il y a quatre caractéristiques temporelles globales : la valeur du maximum secondaire, la valeur quadratique moyenne totale (cf. équation 2.1), la valeur quadratique moyenne gauche (cf. équation 2.2) et le coefficient de Skewness.

$$\frac{\sum_{k=1}^{\frac{longueur}{2}} [signature(k)]^2}{\frac{longueur}{2}} \quad (2.1)$$

$$\frac{\sum_{k=1}^{longueur} [signature(k)]^2}{longueur} \quad (2.2)$$

2.1.3 Population de classe 1 à 10

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 du 42^{ème}, 43^{ème} et 44^{ème} points de la signature. Huit caractéristiques proviennent de la transformée de Fourier : la partie réelle de la 2^{ème}, 3^{ème} et 7^{ème} composante, la partie imaginaire de la 2^{ème}, 3^{ème} et 4^{ème} composante, le module de la 2^{ème} composante ainsi que le taux de distorsion harmonique (cf. équation 2.3). Il y a aussi une caractéristique temporelle qui est la somme de la partie droite de la signature divisée par la somme générale.

$$THD \approx \frac{\sqrt{A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2 + A_6^2 + A_7^2 + A_8^2}}{A_0} \quad (2.3)$$

avec A_i est le module de la $i^{\text{ème}}$ composante de la transformée de Fourier.

2.1.4 Conclusion

Cette section a permis de réduire le nombre de variables par Analyse en Composantes Principales (ACP) et a déterminé les p caractéristiques les plus pertinentes. Ainsi, dans cette étude, seules les p^3 variables (ou identifiants) les plus pertinentes seront utilisées. Alors, chaque véhicule est caractérisé par p variables (x_1, \dots, x_p), Dans la suite du document, nous définissons un vecteur \mathbf{x} pour représenter le vecteur caractérisant la signature d'un véhicule :

$$\mathbf{x} = (x_1, \dots, x_p)^T \quad (2.4)$$

La première boucle électromagnétique d'un tronçon fournit la signature origine des véhicules, et la boucle électromagnétique à l'autre extrémité du tronçon fournit la signature

-
1. Bande dont les valeurs sont supérieures au maximum divisé par 2
 2. Coefficient de dissymétrie, mesurant l'asymétrie de la densité de probabilité
 3. 11 pour les VL; 9 pour les PL et 12 pour les VL/PL

destination des véhicules. Nous noterons \mathbf{x}_o les caractéristiques d'une signature origine et \mathbf{x}_d celles d'une signature destination. Nous considérons que la signature d'un véhicule est unique et que les acquisitions sont noyées dans un bruit blanc gaussien centré \mathbf{n} :

$$\mathbf{x}_o = \mathbf{x} + \mathbf{n} \quad (2.5)$$

$$\mathbf{x}_d = \mathbf{x} + \mathbf{n} \quad (2.6)$$

Toutes les méthodes présentées dans cette section ont pour entrée les \mathbf{x} variables décrites précédemment.

2.2 Maximum de vraisemblance

Le but est de déterminer la meilleure signature origine possible parmi n signatures origines connaissant la signature de destination [17]. Nous cherchons donc à maximiser la probabilité de \mathbf{x}_{oi} sachant \mathbf{x}_d . La règle de décision appliquée est donc la suivante :

$$\max_{i=0\dots n} P(\mathbf{x}_{oi}|\mathbf{x}_d) \quad (2.7)$$

Or d'après le théorème de Bayes, nous avons :

$$P(\mathbf{x}_{oi}|\mathbf{x}_d) = \frac{P(\mathbf{x}_d|\mathbf{x}_{oi})P(\mathbf{x}_{oi})}{P(\mathbf{x}_d)} \quad (2.8)$$

Nous en déduisons donc qu'il faut maximiser $P(\mathbf{x}_d|\mathbf{x}_{oi})P(\mathbf{x}_{oi})$ car $P(\mathbf{x}_d)$ est une constante de normalisation et n'intervient pas dans la décision. Nous appliquons le principe du maximum de vraisemblance qui donne un ensemble de n fonctions discriminantes g_i :

$$g_i(\mathbf{x}_d) = \ln(P(\mathbf{x}_d|\mathbf{x}_{oi})) + \ln(P(\mathbf{x}_{oi})) \quad (2.9)$$

Nous supposons que $\mathbf{x}_d|\mathbf{x}_{oi}$ suit une loi gaussienne de moyenne $\boldsymbol{\mu}_i$ et d'écart-type σ_i estimée par un apprentissage supervisé. En supposant que les composantes des \mathbf{x}_{oi} sont des variables indépendantes deux à deux, la matrice de covariance $\boldsymbol{\Sigma}$ estimée par l'apprentissage est diagonale. Ainsi, nous obtenons :

$$P(\mathbf{x}_d|\mathbf{x}_{oi}) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|^{1/2}}} \exp\left(-\frac{1}{2}(\mathbf{x}_d - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_d - \boldsymbol{\mu}_i)\right) \quad (2.10)$$

Alors, la fonction discriminante s'écrit sous la forme :

$$g_i(\mathbf{x}_d) = -\frac{1}{2}(\mathbf{x}_d - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_d - \boldsymbol{\mu}_i) + \ln(P(\mathbf{x}_{oi})) - \frac{1}{2}(\ln(|\boldsymbol{\Sigma}_i|) + \ln(2\pi)) \quad (2.11)$$

Or $-\frac{1}{2}(\ln(|\boldsymbol{\Sigma}_i|) + \ln(2\pi))$ est identique pour tout \mathbf{x}_d , donc l'objectif est de rechercher pour chaque signature destination la signature origine parmi les n signatures origines qui maximise :

$$g_i(\mathbf{x}_d) = -\frac{1}{2}(\mathbf{x}_d - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_d - \boldsymbol{\mu}_i) + \ln(P(\mathbf{x}_{oi})) \quad (2.12)$$

Nous considérons que les véhicules origines candidats sont équiprobables, ainsi l'équation (2.12) devient :

$$g_i(\mathbf{x}_d) = -\frac{1}{2} (\mathbf{x}_d - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_d - \boldsymbol{\mu}_i) \quad (2.13)$$

Ainsi, nous cherchons le maximum de la fonction coût (2.13).

2.3 Méthode de logique floue

Le passage des véhicules sur les boucles électromagnétiques engendre des imprécisions dans le recueil des données. En effet, les véhicules ne passent pas tous au même endroit sur les boucles, ce qui a une incidence sur l'amplitude des signatures. De plus, les capteurs mesurent la vitesse avec 5 % d'erreur, ce qui modifie la longueur des signatures. Pour tenir compte de ces incertitudes et imprécisions des données recueillies, nous utilisons dans cette section une méthode de logique floue. De cette façon, nous n'avons plus besoin de faire des hypothèses sur la nature de la distribution des variables aléatoires. Nous devons cependant définir des sous-ensembles flous pour ce problème.

Soient \mathbf{x}_{di} l'ensemble des caractéristiques des signatures destinations pour $i = (0..n)$ et \mathbf{x}_{oj} l'ensemble des caractéristiques signatures origines pour $j = (0..m)$. Nous définissons un sous-ensemble $E_{i,j}$ flou ayant une fonction d'appartenance trapèze $f_{E_{i,j}}$:

$$f_{E_{i,j}}(x_k) = \begin{cases} 0 & \text{si } x_k \notin [a_k - \alpha_k, b_k + \alpha_k], \\ 1 & \text{si } x_k \in [a_k, b_k], \\ 1 + \frac{x_k - a_k}{\alpha_k} & \text{si } x_k \in]a_k - \alpha_k, a_k[, \\ 1 + \frac{b_k - x_k}{\alpha_k} & \text{si } x_k \in]b_k, b_k + \alpha_k[. \end{cases} \quad (2.14)$$

avec :

$$\begin{aligned} x_k &= x_{d,k} - x_{u,i,k} \quad k \in [0, p] \\ a_k &= m_k - p_1 \times \sigma_k \\ b_k &= m_k + p_1 \times \sigma_k \\ \alpha_k &= p_2 \times \sigma_k \end{aligned} \quad (2.15)$$

Nous avons choisi une fonction d'appartenance trapèze car elle est fréquemment utilisée dans la littérature [16]. Les paramètres m_k and σ_k sont obtenus à l'aide d'un apprentissage supervisé, les paramètres p_1 et p_2 sont calculés empiriquement dans [16] et sont respectivement égaux à 0 et 15. Dans notre cas, et toujours selon [16], $a_k = b_k$ donc nous avons une fonction d'appartenance triangulaire centrée sur la moyenne. La fonction discriminante est définie par :

$$F_i(X_d) = \sum_{k=0}^p f_{E_{i,j}}(x_k) \quad (2.16)$$

Lorsque la signature origine est identique à celle de destination, la fonction discriminante tend vers le nombre de variables d'entrée, par exemple $p=11$ pour les VL. Une valeur qui tend vers p , le nombre de variables d'entrée, traduit une grande similitude entre la signature origine et destination.

2.4 Séparateur à vaste marge

Les Séparateurs à Vaste Marge (SVM) est une méthode de classification qui fut introduite par V.Vapnik en 1995 dans son ouvrage "The nature of statistical learning theory" [22]. Cette méthode repose sur l'existence d'un classifieur linéaire dans un espace approprié et utilise la notion de marge maximale.

La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. Une des originalités de la méthode est de produire une fonction de décision qui n'utilise qu'un sous-ensemble de la base d'apprentissage, les vecteurs de support (VS). Dans le domaine des SVM, un hyperplan peut-être considéré comme optimal dans le sens où il maximise la marge de séparation. Le problème est de trouver cette frontière séparatrice optimale, à partir d'un ensemble d'apprentissage. Ceci est réalisée en formulant le problème comme un problème d'optimisation quadratique, pour lequel il existe des algorithmes connus.

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 de représentation des données d'entrées en un espace de plus grande dimension, dans lequel ces données sont linéairement séparables. Cette transformation est réalisée par l'intermédiaire des fonctions "noyau". Le lecteur est invité à lire les références [22, 23] pour plus d'information sur cette méthode.

Un grand nombre de recherche ont déjà été effectué avec les SVM (bio-informatique, finance, médicale, télécommunication, micro-électronique,...). Dans cette section, nous proposons d'utiliser cette méthode pour ré-identifier les véhicules. Ainsi, notre problème de classification est binaire : ré-identification du véhicule ou pas. Pour le formalisme des SVM, posons $\mathbf{v} = \mathbf{x}_d - \mathbf{x}_{on}$. Soit un ensemble d'apprentissage $A = (\mathbf{v}_1, u_1), \dots, (\mathbf{v}_k, u_k)$ composé de k couples avec $\mathbf{v}_i \in \mathbb{R}^m$. Dans cette situation, l'étiquette u_i correspond à l'identification ou non-identification du véhicule. L'algorithme des SVMs projette les vecteurs \mathbf{v}_i dans un espace de travail \mathbf{H} à partir d'une fonction non linéaire $\phi : \mathbb{R} \rightarrow \mathbf{H}$. L'hyperplan optimal de séparation des deux classes dans l'espace est ensuite recherché. Cet hyperplan (\mathbf{w}, b) matérialise la frontière de séparation entre les deux classes (identification et non-identification dans notre application). La classe u d'un nouvel exemple \mathbf{v} est définie par $u = \text{sign}(\mathbf{w} \cdot \phi + b)$. L'hyperplan est optimal s'il maximise la distance qui le sépare des exemples dont il est le plus proche. Cette distance est usuellement appelée marge du classifieur. En choisissant une fonction noyau, $K(\mathbf{v}_i, \mathbf{v}_j) = \langle \phi(\mathbf{v}_i), \phi(\mathbf{v}_j) \rangle$, il est montré que la résolution du problème dual correspondant produit une fonction de décision de la forme $f(\mathbf{v}) = \sum_i \alpha_i u_i K(\mathbf{v}_i, \mathbf{v}) + b$ où α_i et b correspondent à la solution optimale du problème dual. Une propriété importante de cette méthode est que la solution ne dépend pas de la dimension des données, mais du nombre d'échantillons et en fait du nombre de vecteurs support.

Dans cette étude, nous avons utilisé *C-SVM* de la librairie LIBSVM accessible en ligne à l'adresse : www.csie.ntu.edu.tw/~cjlin/libsvmtools. Dans la suite de notre étude, seule le noyau "RBF" (Radial Basis Function) sera présenté. En effet, des tests ont été réalisés avec le cas linéaire, le noyau RBF et le noyau polynomial. De ces tests, les meilleurs résultats ont été obtenus par le noyau RBF. Ainsi, plusieurs hyper-paramètres seront à estimer (C

et σ) pour *C-SVM* à noyau "RBF". La forme générique du noyau à base radiale est :

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (2.17)$$

Afin d'utiliser cette méthode SVM dans un problème de ré-identification, nous avons rajouté une contrainte supplémentaire : seuls les couples dont l'origine n'est pas citée à de multiples reprises ne sont pas filtré. Cette version se nommera dans la suite du rapport SVM. Une deuxième contrainte supplémentaire est rajoutée. Elle consiste à prendre 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.

Chapitre 3

Analyse des résultats

3.1 Introduction

Nous proposons de tester les différents algorithmes d'identification présentés précédemment. Les méthodes étudiées sont le maximum de vraisemblance avec un modèle de temps de parcours gaussien présenté dans [10] (MV2), le maximum de vraisemblance présenté au paragraphe 2.2 (MV1), la méthode de logique floue avec une fonction d'appartenance en trapèze, la méthode SVM et la méthode SVM prenant en compte la distance échantillon-hyperplan. Les méthodes seront comparées par la méthode de validation croisée [14] afin de déterminer celle fournissant les meilleurs résultats.

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 ($A-V$) comprendra $z/3$ véhicules et sera subdivisé en N sous-ensembles disjoints. Chaque sous-ensemble sera une et une seule fois un sous ensemble de validation et les $N - 1$ autres sous ensembles seront présents pour l'apprentissage comme le montre la figure 3.1. Puis, nous calculons la moyenne des performances. Dans un second temps, la base $A-V$ sera utilisée pour l'apprentissage et les performances seront calculées sur la base de test (T).

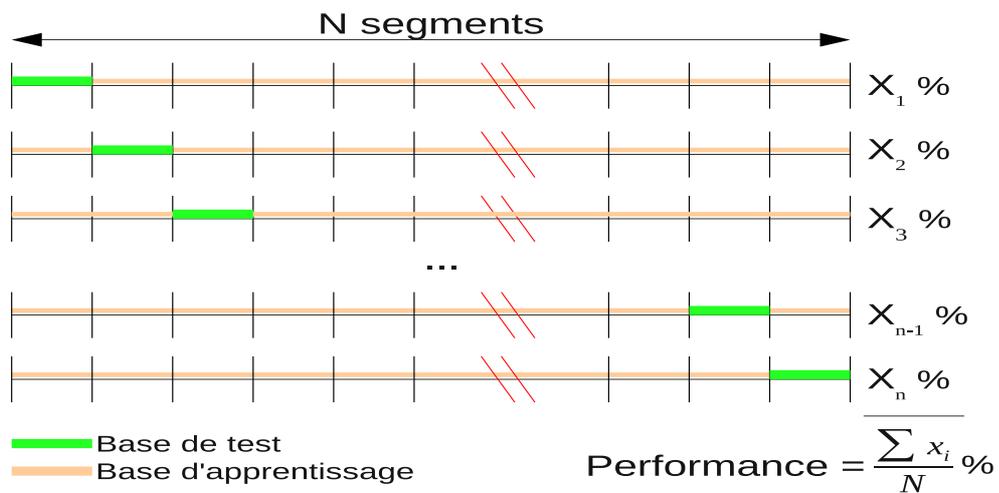


FIGURE 3.1 – Principe de la validation croisée pour N segments.

Les performances utilisées dans cette étude seront le taux d'identification et le taux de

représentation définis comme suit :

$$Taux\ d'identification = \frac{Nombre\ de\ couples\ corrects\ proposés}{Nombre\ total\ de\ couples\ proposés} \quad (3.1)$$

$$Taux\ de\ représentation = \frac{Nombre\ de\ couples\ corrects\ proposés}{Nombre\ total\ de\ couples\ recensés} \quad (3.2)$$

Le taux de représentation indique le pourcentage de couples corrects trouvés par rapport au nombre total de couples recensés grâce à la vidéo. Il donne le pourcentage du trafic pouvant être analysé.

3.2 Etude 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 élaborées, afin de sélectionner les approches les plus efficaces. Pour cela la performance choisie est le **taux de représentation**, qui correspond dans le cas d'une base idéale, au **taux d'identification**.

3.2.1 Performances en validation croisée

Dans ce cas, seul le premier ensemble ($A-V$) comprenant les $z/3$ véhicules est utilisé. De plus, cet ensemble est subdivisé en $N = 5$ sous-ensembles disjoints. Le nombre de PL étant très faible, nous avons utilisé la validation croisée sur la totalité de la base idéale divisée en $N = 5$ segments. Pour utiliser la méthode du maximum de vraisemblance avec un modèle de temps de parcours gaussien, nous avons pris les paramètres présentés dans la section 1.1 (vitesse moyenne : 60km/h, écart-type : 20 km/h) pour la base d'Angers et les paramètres présentés dans la section 1.2 (vitesse moyenne : 50km/h, écart-type : 15 km/h) pour la base de Rennes. Pour le classifieur C-SVM à noyau RBF, les deux hyperparamètres à régler sont C et σ . Ces paramètres ont été estimés par validation croisée pour les trois types de population (VL, PL et mixte) et pour les deux sites (Angers et Rennes). Seuls les paramètres du site d'Angers sont présentés dans le tableau 3.3.

Méthode	Floue Triangulaire	MAP Uniforme	MAP Gaussien	SVM (Noyau RBF)	SVM (Noyau RBF) Distance Hyperplan
Abréviation	Floue	MV1	MV2	SVM	SVM-d
PL	81,0 %	89,3 %	92,7 %	66,3 %	87,3 %
VL	68,0 %	70,0 %	82,9 %	40,4 %	64,1 %
PL & VL	51,6 %	69,6 %	87,1 %	34,3 %	62,2 %

TABLE 3.1 – Performances en validation croisée sur la base d'Angers.

Le tableau 3.1 permet de remarquer que la méthode SVM est moins efficace que les quatre autres méthodes. De plus, il est à noter que quel que soit la méthode, il existe une différence de performances selon le type de la population. En effet, le taux d'identification est plus grand pour les PL, suivi du taux pour les VL. Le taux le plus faible est obtenu par la population mixte.

Les tableaux 3.1 et 3.2 montrent que la méthode SVM présentée a des performances réduites

Méthode	Floue	MV1	MV2	SVM	SVM-d
PL	90,0 %	92,5 %	93,7 %	70,0 %	86,2 %
VL	51,2 %	55,3 %	62,4 %	13,9 %	52,0 %
PL & VL	34,9 %	52,8 %	61,5 %	7,8 %	39,1 %

TABLE 3.2 – Performances en validation croisée sur la base de Rennes.

comparativement aux autres méthodes. Cela peut s'expliquer par le filtre mis en place dans cette approche, qui est assez restrictif. Les différences de performance selon la population s'expliquent quant à elles par la taille de celle étudiée, mais aussi par les variables choisies pour chaque type de population.

	SVM		SVM-d	
	C	Γ	C	Γ
PL	1,414	64	2	11,314
VL	16	128	16	1
PL & VL	0,5	1024	2	64

TABLE 3.3 – Paramètres établis pour les méthodes utilisant les séparateurs à vaste marge.

3.2.2 Performances sur l'ensemble de test (T)

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.

Méthode	Floue	MV1	MV2	SVM	SVM-d
PL	69,6 %	71,7 %	84,1 %	49,3 %	79,7 %
VL	36,7 %	38,8 %	67,1 %	14,7 %	27,9 %
PL & VL	17,0 %	41,4 %	82,1 %	8,6 %	33,6 %

TABLE 3.4 – Angers - Performances sans fenêtrage dans le cadre d'une mesure sur base totale.

Le tableau 3.4 montre que les performances des méthodes sans fenêtrage dans le cadre d'une mesure sur base totale avec MV2 sont plus élevées. On constate aussi de meilleurs résultats pour la population de PL ainsi que de moins bons résultats pour la population mixte. On constate dans le tableau 3.5 qu'il y a aussi de meilleurs résultats pour la méthode MV2, ainsi que pour les poids lourds.

Méthode	Floue	MV1	MV2	SVM	SVM-d
PL	79,6 %	81,5 %	87,0 %	57,4 %	75,9 %
VL	37,7 %	41,0 %	57,4 %	8,2 %	31,8 %
PL & VL	18,5 %	41,2 %	56,4 %	4,3 %	27,2 %

TABLE 3.5 – Rennes - Performances sans fenêtrage dans le cadre d'une mesure sur base totale.

Les tableaux 3.4 et 3.5 montrent que la méthode MV2 obtient de meilleurs résultats. Ceci s'explique par la pondération effectuée avec cette méthode. En effet, cette méthode fait appel à une information a priori supplémentaire par rapport à MV1. Les meilleures performances des PL s'expliquent aussi par la population réduite qu'ils constituent comparée à celle des VL.

Dans un second temps, nous effectuons un fenêtrage temporel lors de la phase d'évaluation. Cela consiste à chercher une signature origine qui appartient à la fenêtre temporelle. Le positionnement de la fenêtre est déterminé avec l'horaire de passage sur le point de destination du véhicule que l'on cherche à lier. Ces paramètres sont un facteur limitant, les performances maximales possibles sont explicitées dans le tableau 3.6 .

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

TABLE 3.6 – Performances maximales potentielles avec le fenêtrage utilisé.

Méthode	Floue	MV1	MV2	SVM	SVM-d
PL	97,8 %	98,6 %	98,6 %	76,8 %	95,7 %
VL	92,1 %	93,5 %	92,5 %	62,4 %	88,2 %
PL & VL	81,2 %	82,4 %	86,4 %	58,7 %	83,0 %

TABLE 3.7 – Angers - Performances avec fenêtrage dans le cadre d'une mesure sur base totale.

Méthode	Floue	MV1	MV2	SVM	SVM-d
PL	92,6 %	90,7 %	90,7 %	59,3 %	79,6 %
VL	65,0 %	69,4 %	62,9 %	31,1 %	59,8 %
PL & VL	47,9 %	65,1 %	55,8 %	11,3 %	51,3 %

TABLE 3.8 – Rennes - Performances avec fenêtrage dans le cadre d'une mesure sur base totale

Nous constatons dans l'ensemble une nette augmentation des performances sur le tableau 3.7 (3.8) en comparaison avec le tableau 3.4 (3.5). 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 l'on n'utilise pas de fenêtrage temporel. Les tableaux 3.7 et 3.8 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.

Nous observons que la mise en oeuvre du fenêtrage permet une nette amélioration des taux de détection. 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 en comparaison à celles obtenue sur le site d'Angers. Cette différence peut s'expliquer par le

nombre de candidat possible qui est plus important sur le site de Rennes. Les figures 3.2 et 3.3 montrent une matrice origine-destination avec la valeur de la fonction coût obtenue par la méthode de logique floue avec fenêtrage sur les sites de Rennes et d'Angers. Ces matrices sont obtenues en calculant la fonction coût de tous les couples possibles. Cependant, il est nécessaire de restreindre ces couples avec le fenêtrage précédemment utilisé. Les véhicules étant triés par ordre de passage, nous observons donc simplement une diagonale. Ensuite, selon la quantité de véhicules comparés la diagonale est plus ou moins large. Dans le cas de Rennes (3.3), la diagonale observée est très large par rapport au cas d'Angers (figure 3.2). 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 lors de l'étude sur base perturbée.

Le fenêtrage temporel apporte un gain de temps important grâce à la réduction de données effectuée. Par conséquent, la méthode MV2 n'est pas retenue pour l'étude sur la base perturbée. En effet, cette méthode perd de l'intérêt au profit de la méthode MV1 lorsque l'on utilise un fenêtrage temporel.

Par ailleurs, nous avons vu aussi qu'il est intéressant de séparer les populations lors de l'assemblage des signatures, afin de réduire la quantité de données à croiser. La figure 3.4 montre le taux d'identification en faisant varier le ratio PL - VL, mais en gardant un nombre total de véhicules constant (400). Les signatures utilisées sont issues de la base idéale d'Angers.

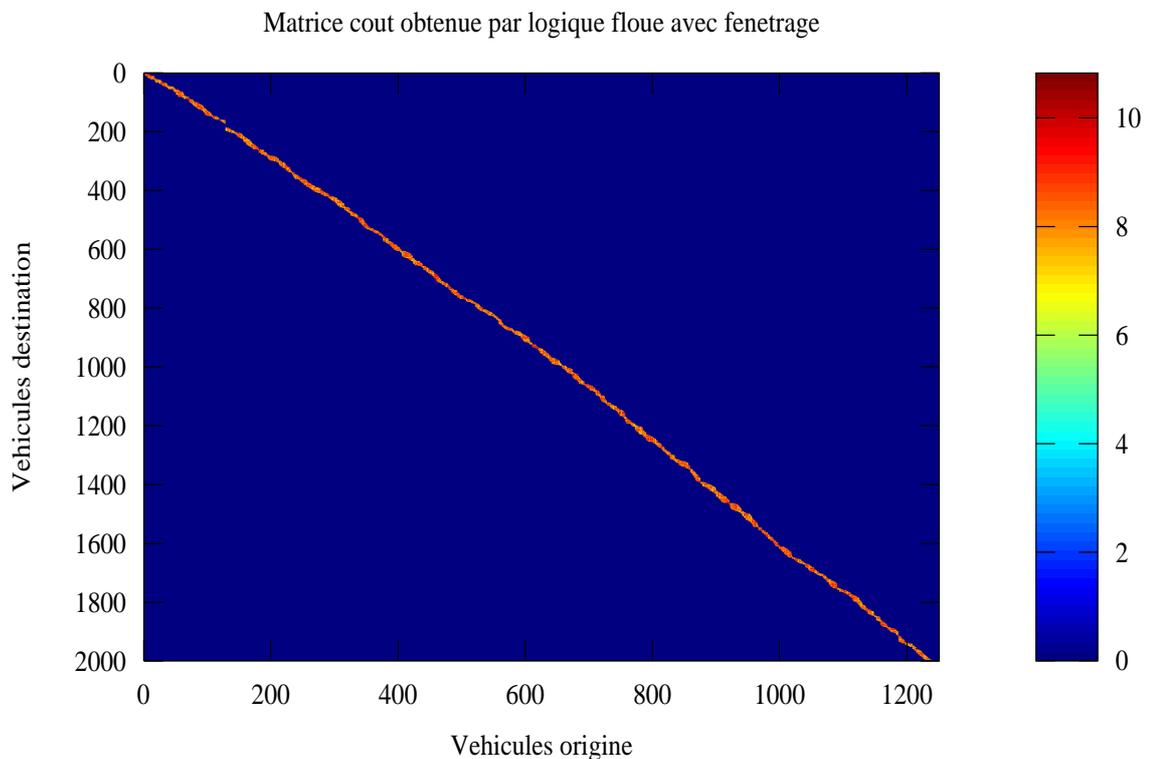


FIGURE 3.2 – Matrice de coût réalisée à partir des données d'Angers.

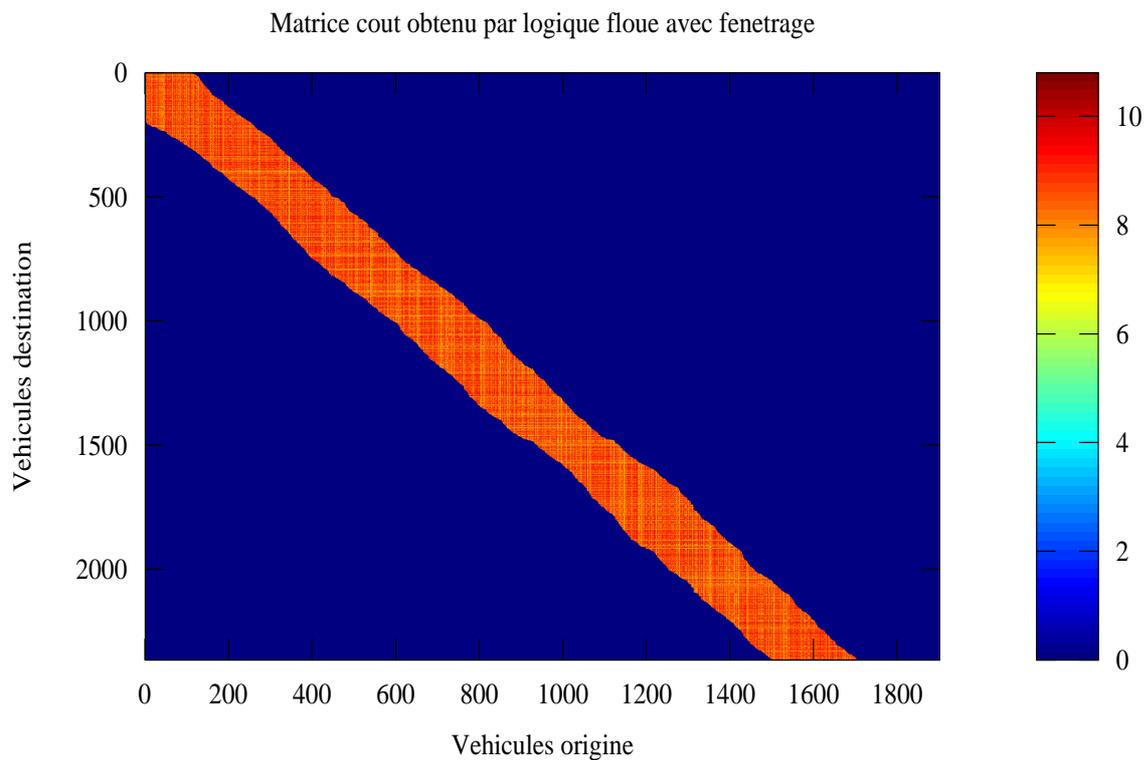


FIGURE 3.3 – Matrice de coût réalisée à partir des données de Rennes.

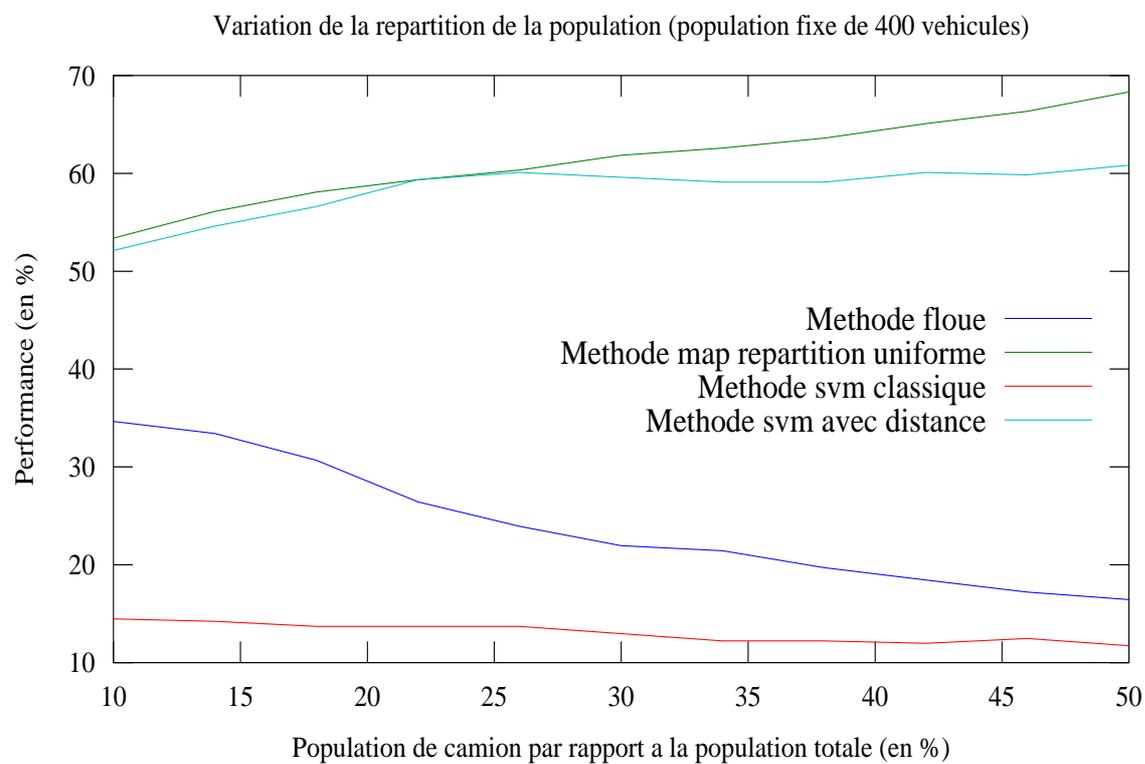


FIGURE 3.4 – Taux d'identification en fonction de la répartition poids lourds - véhicules légers.

La figure 3.4 montre que certaines méthodes sont très sensibles à la répartition VL-PL. En effet, les performances des différentes méthodes peuvent dépendre de l'initialisation réalisée, notamment la répartition poids lourds - véhicules légers. La méthode de logique floue est par exemple très sensible à la variation de la répartition PL - VL. Quel que soit le ratio PL - VL, les valeurs d'initialisation restent les mêmes, c'est à dire celles pour une population ayant 10 % de PL. C'est pourquoi l'étude sur la population mixte est abandonnée au profit des deux autres populations lors de l'étude sur base perturbée. De plus, la classification PL - VL est déjà opérationnelle sur les stations de mesures. Enfin, au vu des performances de la méthode SVM présentée, l'étude sur la base perturbée se déroule autour des trois approches suivantes : la logique floue, MV1 et SVM-d.

3.3 Etude sur base perturbée

Cette base est constituée de paires de signatures, mais aussi de signatures perturbatrices. Ce sont les signatures produites par les véhicules n'ayant pas circulé sur les deux zones instrumentées.

3.3.1 Méthodes proposées

Pour tenter de réduire le taux d'erreur des algorithmes, différentes propositions ont été mises en place. Il s'agit tout d'abord de contraindre l'algorithme à ne pas citer une origine pour de multiples destinations qui sont principalement dues aux perturbations. Ensuite l'idée consiste à utiliser plusieurs méthodes en parallèle afin de tenir uniquement compte des couples proposés de façon identique par chaque méthode. Enfin, la dernière proposition consiste à utiliser un seuil sur la fonction coût de chaque méthode afin de discriminer un maximum de couples faux.

Comme nous l'avons vu dans la section 3.2.2, nous utilisons pour cette base un fenêtrage temporel. 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 que l'on veut assembler 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, nous proposons d'utiliser plusieurs méthodes en parallèle afin de tenir compte uniquement des couples proposés de façon identique par chaque méthode. Le vote à l'unanimité, contrairement à un vote à la majorité, n'a pas besoin d'un nombre de méthode impair. Cependant, ce type de discrimination est intéressant seulement si chacune des méthodes a un comportement différent. Dans la suite de ce document, la méthode utilisant le vote à l'unanimité se nommera "una".

Dans un second temps, un deuxième post-traitement est proposée, il consiste à utiliser un seuil ou critère d'acceptabilité comme dans [21]. Chaque méthode calcule une fonction coût pour chaque couple. Plus le coût est élevé, plus la probabilité de validité du couple est élevée. Ce critère d'acceptabilité a été déterminé à partir de la base d'apprentissage. Il consiste à dire que q % des véhicules de la base d'apprentissage ont une fonction coût supérieure à un seuil α . L'association d'un véhicule destination avec un véhicule origine

ne peut se faire que lorsque le résultat du calcul de la distance par la méthode est supérieur à ce seuil. Ce seuil permet de minimiser les erreurs d'association. Cette méthode de post-traitement s'intègre juste après la désignation du candidat par la méthode. Il peut donc aussi s'effectuer lorsqu'on utilise une seule méthode. Le but du post-traitement est de diminuer le taux d'erreur en minimisant la perte de couples corrects. Pour déterminer les seuils, nous utilisons la base idéale.

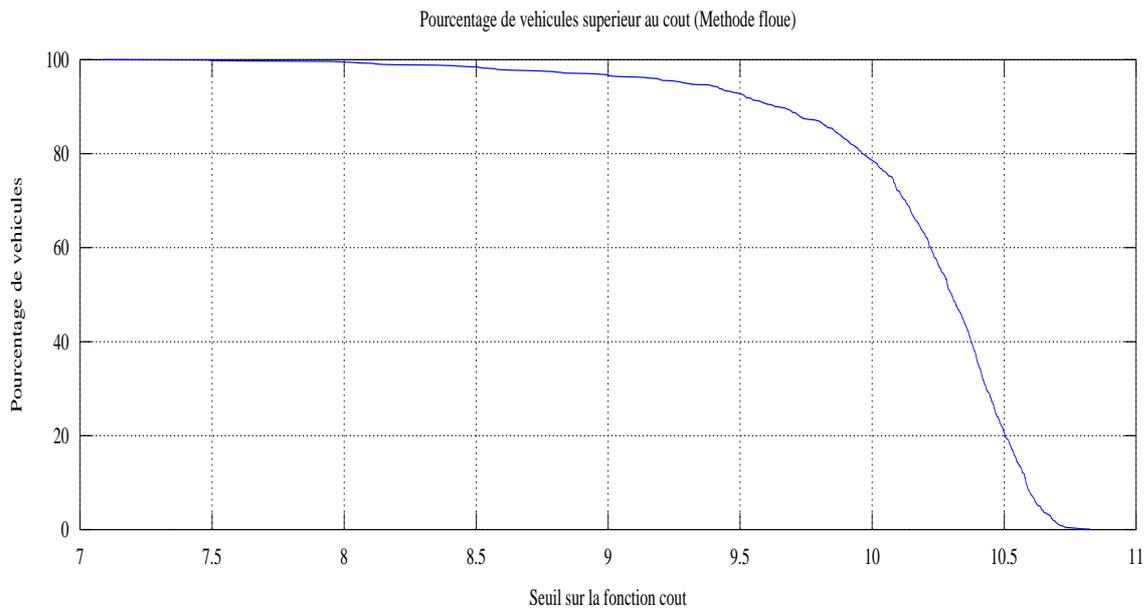


FIGURE 3.5 – Evolution du pourcentage de véhicules légers en fonction du seuil (Flou)

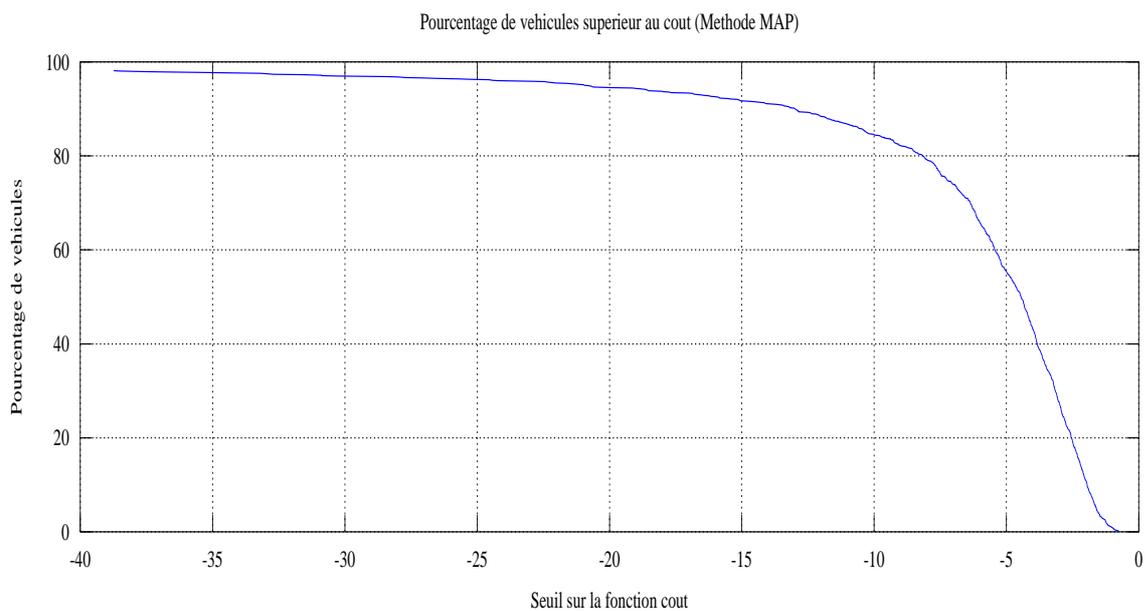


FIGURE 3.6 – Evolution du pourcentage de véhicules légers en fonction du seuil (MV1)

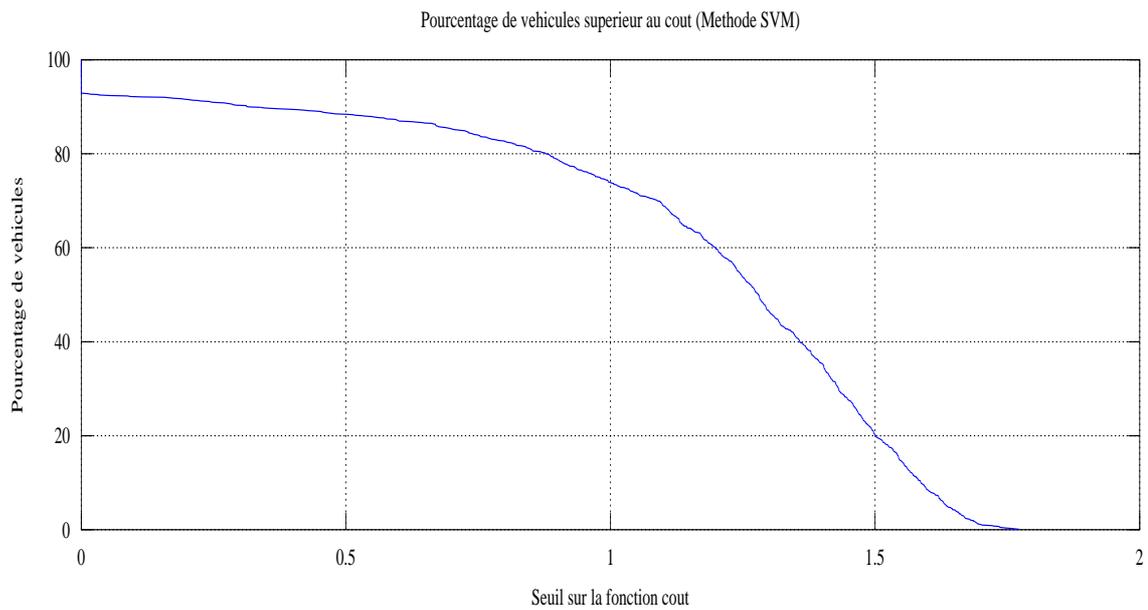


FIGURE 3.7 – Evolution du pourcentage de véhicules légers en fonction du seuil (SVM-d)

Les figures 3.5 , 3.6 et 3.7 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. 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. Différents seuils peuvent être utilisés mais un compromis entre mauvaises identifications et suppression de bonne identification est à réaliser. Ces trois figures concernent seulement les VL. Des figures similaires sont obtenues pour les PL, mais avec une valeur de seuil différente. Les tableaux 3.9 et 3.10 montrent des seuils qui permettront de garder 90 %, 80 % ou 70 % de la population des couples corrects.

Méthode	90 %	80 %	70 %
Floue	9,64	9,96	10,12
MV1	-12,99	-8,16	-6,36
SVM-d	0,31	0,88	1,08

TABLE 3.9 – Valeur de seuil pour un pourcentage donné de véhicules légers (Site d'Angers)

Méthode	90 %	80 %	70 %
Floue	8,14	8,3	8,37
MV1	-13,7	-10	-8,1
SVM-d	0,80	1,06	1,23

TABLE 3.10 – Valeur de seuil pour un pourcentage donné de poids lourds (Site d'Angers)

De plus, un troisième post-traitement est proposée. Il consiste à filtrer le cas où nous avons des origines ayant 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 signature de destination par signature de destination. Une solution pour atténuer ce problème est de rechercher dans un premier temps les origines citées à plusieurs reprises. Puis, pour chaque 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.

3.3.2 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 nous avons des origines ayant de multiples destinations

Chaque tableau présente deux taux, le premier est le taux d'identification¹, le deuxième est le taux de représentation². TI donne la qualité des prédictions et TR informe de la quantité du trafic pouvant être analysé.

3.3.2.1 Performances témoins

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

		Floue	MV1	SVM-d	una
PL	TI	90,6 %	91,3 %	98,5 %	98,5 %
	TR	97,8 %	98,6 %	95,7 %	94,9 %
VL	TI	46,8 %	47,6 %	66,0 %	73,6 %
	TR	90,7 %	92,3 %	87,3 %	84,3 %

TABLE 3.11 – Angers - Performances témoins.

Le tableau 3.11 montre que les taux d'identification pour les PL sont très élevés surtout pour les deux cas SVM-d et vote à l'unanimité. De plus, les taux de représentation 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 TI 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 TI des VL peuvent s'expliquer par le grand nombre de

1. TI : Taux d'Identification, voir la définition sur l'équation 3.1 à la page 24

2. TR : Taux de Représentation, voir la définition sur l'équation 3.2 à la page 24

perturbations. En effet, les méthodes du type MV1 ou Floue donnent une association avec une signature d'origine quel que soit la signature destination. La méthode SVM-d utilise par définition un seuil hyperplan. Ainsi, dans certaine situation, la méthode permet déjà de filtrer les véhicules perturbateurs. A contrario, les bons taux des PL sont en partie dus aux faibles perturbations de ce type de population.

3.3.2.2 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	MV1	SVM-d	una
PL	TI	97,8 %	97,8 %	99,2 %	99,2 %
	TR	97,8 %	98,4 %	95,7 %	94,9 %
VL	TI	79,9 %	82,6 %	84,1 %	90,0 %
	TR	87,0 %	89,2 %	83,2 %	78,3 %

TABLE 3.12 – Angers - filtrage "Origine à multiples destinations".

Le tableau 3.12 montre une très nette augmentation des TI pour la population des VL comparé au tableau 3.11 mais en contrepartie une légère baisse des TR. Pour les PL, on observe peu de différences de performance entre les deux configurations. La seule observation est l'augmentation des TI 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 3.12 comparativement au tableau 3.11.

3.3.2.3 Performances avec seuil

Dans cette section, nous filtrons les mauvaises associations par un seuil (ou critère d'acceptabilité) sur les fonctions coûts ou distance couple-hyperplan. Les 3 tableaux suivants correspondent à trois ensembles de seuils différents présentés dans les tableaux 3.9 et 3.10 à la page 31.

		Floue	MV1	SVM-d	una
PL	TI	99,2 %	98,4 %	99,1 %	99,1 %
	TR	89,9 %	89,1 %	84,1 %	80,4 %
VL	TI	66,3 %	69,0 %	70,5 %	78,2 %
	TR	85,6 %	85,9 %	85,8 %	80,1 %

TABLE 3.13 – Angers - Post-traitement seuil à 90 %.

		Floue	MV1	SVM-d	una
VL	TI	77,3 %	78,1 %	78,9 %	84,4 %
	TR	77,4 %	76,5 %	73,5 %	66,7 %

TABLE 3.14 – Angers - Post-traitement seuil à 80 %.

		Floue	MV1	SVM-d	una
PL	TI	99,0 %	98,9 %	100 %	100 %
	TR	68,8 %	68,1 %	44,9 %	38,4 %
VL	TI	81,1 %	82,2 %	81,0 %	86,6 %
	TR	68,2 %	67,3 %	60,7 %	52,2 %

TABLE 3.15 – Angers - Post-traitement seuil à 70 %.

Nous constatons dans le tableau 3.13 que les TI des PL ont augmenté en comparaison au tableau 3.11. Cependant les TR ont légèrement diminué. La même observation peut-être faite pour les VL. Les TI des PL ne peuvent plus être augmentés de façon significative. Le tableau 3.14 montre les performances des 4 méthodes pour un seuil à 80%. La population des PL n'est pas représentée dans ce tableau. On observe un gain des TI en comparaison au tableau 3.13. La baisse des TR est justifiée car les seuils sont obtenus tels que les TR soient ici limités à 80%. Comme pour le tableau 3.14, nous observons dans le tableau 3.15 une augmentation des TI, faible dans ce cas, et une baisse des TR.

Les trois tableaux 3.13, 3.14 et 3.15 mettent en valeur le compromis du seuil, qui nécessite de réduire le TR afin d'augmenter les performances (TI). De plus, nous pouvons voir aussi que le taux de représentation est diminué par le vote à l'unanimité. Cependant, ce compromis est à relativiser au vu de la valeur du TI avant compromis car ce dernier est déjà très haut. Par comparaison du tableau 3.15 avec le tableau 3.11, dans le cas du vote à l'unanimité, nous pouvons voir que pour augmenter le TI de la population PL de 1,5 %, il faut perdre plus de 50 % de TR.

3.3.2.4 Performances avec seuil et origines à multiples destinations

La configuration "seuil et origines à multiples destinations" reprend la configuration "témoin" avec un traitement post-méthode, qui est composé de deux traitements présentés dans les cas précédents : 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'intérêts.

		Floue	MV1	SVM-d	una
VL	TI	86,7 %	88,8 %	86,3 %	91,7 %
	TR	66,5 %	65,9 %	58,8 %	49,5 %

TABLE 3.16 – Angers - Post-traitement seuil à 70 % et origines à multiples destinations

Lorsque nous comparons le tableau 3.16 au tableau 3.15, nous remarquons que l'assemblage des deux méthodes permet une réduction des erreurs de prédiction. En contrepartie, le taux de représentation subit un léger recul.

3.3.2.5 Synthèse

Au vu de l'imperfection des trois méthodes utilisées (MV1, 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 l'identification par le nombre de méthodes. De plus, si les méthodes utilisées sont similaires, le vote risque de ne pas apporter de valeur ajoutée. 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 de choisir les valeurs de seuil, qui peuvent dépendre par exemple du nombre de variables utilisées). La solution "origines à multiples destinations" possède de bonnes performances sans avoir de paramètre à définir. De plus, l'élimination des perturbations est ciblées.

3.4 Estimation des temps de parcours

Cette section présente une *estimation de temps de parcours* 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 3.3.2.2. Les figures 3.8 et 3.9 montrent les temps de parcours observés et estimés des VL et des PL sur cette base de données en fonction du temps écoulé.

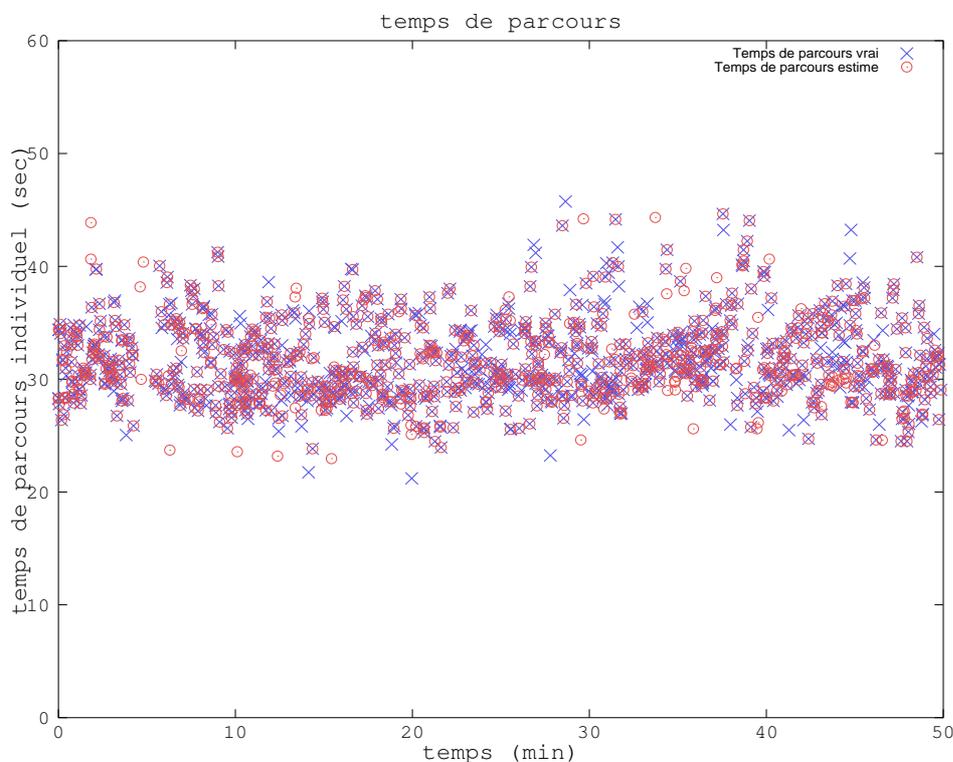


FIGURE 3.8 – Estimation des temps de parcours pour des VL

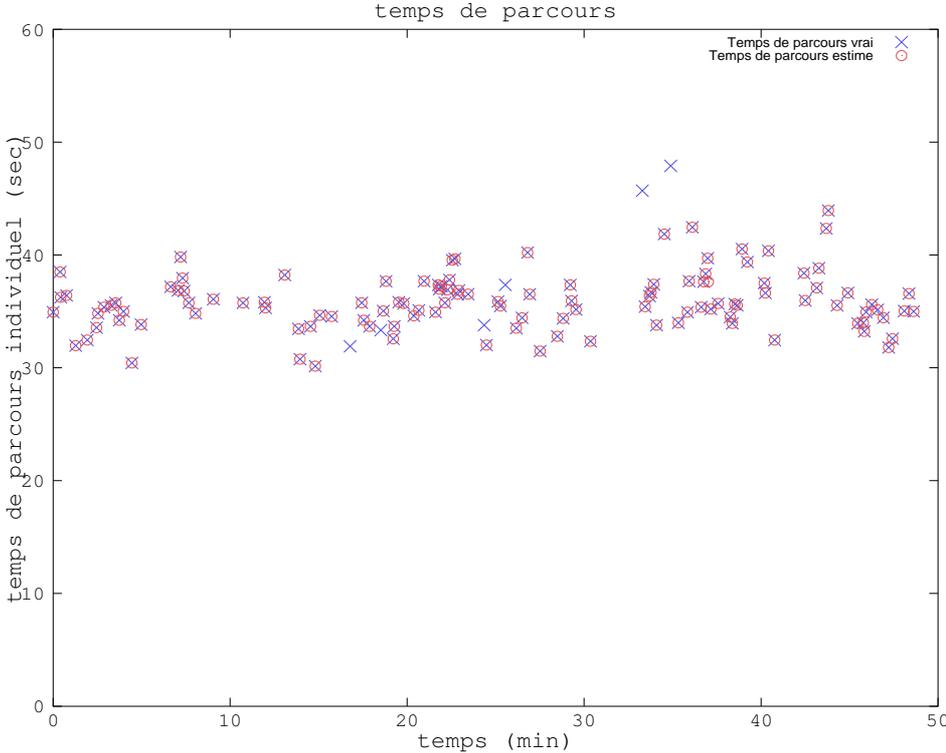


FIGURE 3.9 – Estimation des temps de parcours pour des PL

Conclusion

Ce rapport présente une nouvelle partie de notre travail sur la mesure des temps de parcours à partir des signatures électromagnétiques des véhicules. Il vise à valider expérimentalement trois algorithmes de réidentification sur deux bases de données pour une population mixte VL et PL. Les méthodes présentées permettent d'obtenir pour les VL des résultats pouvant atteindre pour la base d'Angers *90% de bonnes identifications* pour 78% de trafic analysé. De même, pour les PL, les performances de ces méthodes atteignent *99% de bonnes identifications* pour 95% de trafic analysé. Ces taux obtenus sont très prometteurs. Les résultats de ces études sont très encourageants et des améliorations peuvent encore être apportées.

Ainsi, ces méthodes ou l'association de ces méthodes devront dans un second temps être analysée sur un site "plus complexe" que celui présenté dans ce rapport afin d'une part d'analyser la robustesse de ces méthodes en fonction du site et d'autre part de prendre en compte un plus grand nombre de PL. Dans cette application, nous avons utilisé un fenêtrage temporel statique. En perspective, nous souhaitons utiliser un fenêtrage temporel dynamique à partir d'un modèle de trafic existant. Ces études sur ce type de fenêtrages feront l'objet d'études plus approfondies dans le projet MOCOPO avec *l'INRIA*. De plus, nous nous sommes focalisé sur des identifiants provenant de la signature. Sachant que les débits de transmission et les calculateurs temps réels sont de plus en plus performants et puissants, nous proposons aussi de transformer les identifiants (parfois corrélés entre eux) en de nouvelles variables indépendantes les unes des autres afin d'améliorer certaines méthodes.

Bibliographie

- [1] J. Bertrand, "Mesure du positionnement latéral des véhicules sur la chaussée avec des boucles électromagnétiques", *Rapport LRPCA*, réf. 46.07.99.107, Février, 2008.
- [2] M. Pithon, " Détection sélective des véhicules par analyse de leur signature électromagnétique, validation sur site ", Mars 2006.
- [3] S. T. Jeng and S. G. Ritchie. "A New Inductive Signature Data Compression and Transformation Method for On-Line Vehicle Reidentification", The 85th Annual TRB Meeting, Jan 22-26, 2006.
- [4] C. Sun. "An Investigation in the use of Inductive Loop Signature for Vehicle Classification", California PATH Research Report UCB-ITS-PRR-2000-4.
- [5] C. Oh and S. G. Ritchie, "Anonymous Vehicle Tracking for Real-Time Traffic Surveillance", UCI-ITS-TS-WP-02-15 August 2002.
- [6] S. G. Ritchie, S. Park, C. Oh and C. Sun. "Field Investigation of Advanced Vehicle Reidentification Techniques and Detector Technologies -Phase 1. California PATH Research Report UCB-ITS-PRR-2002-15 University of California", Berkeley U.S, March 2002.
- [7] S. G. Ritchie, S. Park, C. Oh and C. Sun. "Field Investigation of Advanced Vehicle Reidentification Techniques and Detector Technologies -Phase 2. California PATH Research Report UCB-ITS-PRR-2005-8 University of California", Berkeley U.S, March 2005.
- [8] B. Coifman, "Vehicle Re-Identification and Travel Time Measurement in Real-Time on Freeways Using Existing Loop Detector Infrastructure", Transportation Research Record 1643, 1998.
- [9] S.S. Ieng et C. Grellier, "Reconnaissance et suivi de véhicules par analyse de leur signature électromagnétique", Mars 2006, Rapport d'étude LRPC Angers.
- [10] S.S. Ieng, C. Grellier, J. Rivault, J. Bertrand et M. Pithon, "On the Inductive Loop Based Vehicle Signature Features Analysis and the Anonymous Vehicle Re-Identification for Travel Times Estimation", TRB 86th Annual Meeting, Compendium of papers CD-Rom. 21-25 janvier, 2007 Washington.
- [11] T. M. Kwon, A. Parsekar, "Deconvolution of vehicle inductance signature for vehicle reidentification" , TRB 84th Annual Meeting, Compendium of papers CD-Rom. janvier, 2005 Washington.
- [12] T. M. Kwon, A. Parsekar, "blind Deconvolution processing of loop inductance signals for vehicle reidentification, TRB 85th Annual Meeting", Compendium of papers CD-Rom. janvier, 2006 Washington.
- [13] V. F. Totten, "Application of vehicle detector Waveforms in Vehicle Re-identification and Evaluating detector Installation Performance", Electrical and Computer Engineering Masters Thesis, Purdue University, 2008

- [14] G. Dreyfus, J.-M. Martinez, M. Samuelides, M.B. Gordon, F. Badran, S. Thiria, " Apprentissage statistique ", Edition Eyrolles, Septembre 2008.
- [15] Gen Li, " Vehicle Reidentification and Travel Time Estimation On congested Highway ", Technical University of Denmark, 2007.
- [16] Grellier Christophe, " Reconnaissance et suivi de véhicules par analyse de leur signature électromagnétique ", LRPCA - IMA, 2005.
- [17] Sio-Song Ieng, Michel Pithon, Jean Bertrand, " Suivi des véhicules par analyse de leur signature électromagnétique pour le calcul de temps de parcours ", LRPCA, Journée des sciences de l'ingénieur - Marne-La-Vallée 2006.
- [18] " Les temps de parcours", Edition du CERTU, avril 2008
- [19] Ieng, Grellier, Rivault, Bertrand, Pithon, " On the inductive loop based vehicle signature features analysis and the anonymous vehicle re-identification for travel times estimation ", Transport Research Board 2007
- [20] C. Le Bastard, P. Briand, P. Subirat, E. Violette, D. Doucet, R. Quequiner, "Utilisation des boucles électromagnétiques dans les observatoires locaux : Estimation du positionnement latéral & de la vitesse des véhicules", Décembre 2009.
- [21] D. Guilbert , C. Le Bastard, M. Brénugat, P. Briand, "Opération MTT : Utilisation des boucles électromagnétiques dans les observatoires globaux : Estimation des temps de parcours et des matrices Origine-Destination", Décembre 2009.
- [22] V. N. Vapnik, "The nature of statistical learning theory" (1995).
- [23] A. Boubezoul, "Système d'aide au diagnostic par apprentissage : application aux systèmes microélectroniques", thèse de doctorat, Université Paul Cézanne Aix-Marseille III, Mars 2008.