



UNIVERSITÉ DE LARBI TÉBESSI, TÉBESSA



FACULTÉ DES SCIENCES EXACTES ET SCIENCES DE LA NATURE ET DE LA VIE

DÉPARTEMENT DES MATHÉMATIQUES ET INFORMATIQUE

LABORATOIRE DES MATHÉMATIQUES, INFORMATIQUE ET SYSTÈMES (LAMIS)

# Thèse

Spécialité : Informatique

Présentée et soutenue publiquement par

**ABDELATIF SAHRAOUI**

EN VUE DE L'OBTENTION DU DIPLÔME DE

DOCTEUR (CYCLE LMD)

INTITULÉE :

---

## Une infrastructure hybride de Cloud Computing pour l'optimisation des VANET

---

Dr. Makhlouf Derdour  
Dr. Philippe ROOSE

Université de Larbi Tébessi  
Université de LIUPPA/UPPA

Encadreur  
Co-Encadreur

devant le jury ci-dessous :

Pr. Mohammed Rida Laouar  
Pr. Nacira Ghoualmi  
Dr. Abdelkrim Bouramoul  
Dr. Hakim Bendjenna

Université de Larbi Tébessi  
Université de Badji Mokhtar  
Université de Constantine 2  
Université de Larbi Tébessi

Président  
Examinatrice  
Examineur  
Examineur

21 MARS 2018

UNE INFRASTRUCTURE HYBRIDE DE CLOUD COMPUTING POUR  
L'OPTIMISATION DES VANET

SAHRAOUI ABDELATIF

UNIVERSITÉ DE LARBI TÉBESSI



LABORATOIRE LAMIS



## ملخص

تطور تكنولوجيا الأجهزة الذكية والاتجاهات و تطور قدرات الإنترنت لديها القدرة على تقارب الشبكة الخاصة بالمركبات (VANET) إلى سحابة الحوسبة بالعربات (CCV). في ضوء هذا التقارب، العديد من التطبيقات المبتكرة أخذت في الظهور لتحسين توجيه السيارة، والتخفيف من الازدحام المروري من خلال حساب كفاءة البيانات الثابتة. ولتحقيق المستوى المطلوب من الموثوقية، تعتمد هذه التطبيقات على الاستخدام الفوري لمراكز البيانات البعيدة وموارد المركبات غير القابلة للاستعمال لتوفير خدمات التخزين والحوسبة والاتصالات يمكن أن تكون موجودة في المركبات غير مستعملة في مواقف السيارات والمطارات والطرق السريعة.

خلال فترات الذروة، قد يتجاوز عدد المركبات قدرة الطريق، وذلك راجع الى عدد طلبات لاستعمال الطريق أو جزء منه، مما يبطئ سرعة التدفق. في ظل هذه الظروف التي يوجد فيها ازدحام حركة مرورية، يبرز مخطط القاعدة الأساسي (MFD) لنماذج تدفق الحركة علاقة غير خطية بين تدفق البيانات وتدفق الحركة. وهذه العلاقة تساهم بشكل فعال في إنخفاض السرعة الملحوظة مع تمديد وقت وصول المقرر للسائق، الأمر الذي سيجعل حركة المرور أكثر تعقيدا. واستنادا إلى هذا الاتجاه، فإن الحل الأكثر موثوقية لتخطيط الطرق على المدى الطويل سيكون أفضل استخدام لسعة الطرق بدلا من توسيع القدرات الحالية للطرق. هذا الحل سوف يحسن لتدفق حركة المرور المتأثر بالحالة الراهنة من الطرق (أي توقف والممرات) والحد من تكلفة الازدحام للحد من ساعات الذروة.

في هذه الرسالة، نسعى إلى تحسين نماذج تدفق السيارات في الحالات المكتظة. أولاً، استنادا إلى البيانات حركة المرور التجريبية، ونموذج التفاعل بين عوامل الحركة باستخدام نموذج التصنيف الثنائي (الانحدار اللوجستي). ويتوقع النموذج إمكانية وقوع حوادث الطرق في حالة حدوث تباطؤ مفاجئ في السرعة. لتبرير أهمية التصنيف، يخضع نموذجنا لعدة اختبارات إحصائية. نتائج الاختبار تبين أن هناك العديد من التفاعلات بين عوامل حركة المرور لشرح بدقة ظاهرة الازدحام.

بعد ذلك، سوف نركز على استخراج وظيفة التعلم غامضة جديدة للتنبؤ موثوق بها وكفاءة المعلومات حركة المرور في الوقت الحقيقي. يتم جمع بيانات التعلم من مصادر حركة المرور متعددة مثل أجهزة كشف حركة المرور، تبادل البيانات من مركبة إلى مركبة، وبيانات حركة المرور المسجلة سابقا. ويسمح النموذج المقترح بنمذجة التفاعلات بين متغيرات الحركة المخزنة في هذه المصادر. وأخيرا، سوف نقوم بدمج وظيفة التنبؤ المقترحة في طبقة خدمات سحابية ، والتي بدورها تتضمن نوعين من خدمات المرور: الخدمات المحتملة (الخدمات حسب الطلب) والخدمات التلقائية. ومن الواضح أن الهدف هو تصميم نموذج تسليم مرنة لدعم العديد من الميزات من حيث التنبؤ والإخطار.

**كلمات البحث:** الحوسبة السحابية، الشبكة الخاصة بالمركبات ، والاتصالات V2V ، والاتصالات V2C ، والسلامة على الطرق، البيانات غامضة، نموذج التعلم، التنبؤ، تدفق حركة المرور، الازدحام.

# RÉSUMÉ

---

L'évolution de la technologie des dispositifs intelligents et l'évolution des capacités d'Internet font que les réseaux de véhicules ad hoc (VANET) convergent vers le paradigme Cloud Computing Véhiculaire (CCV). A la lumière de cette convergence, de nombreuses applications innovantes émergent pour améliorer le routage des véhicules, atténuer la congestion du trafic grâce à un calcul efficace des données statiques. Pour atteindre le niveau de fiabilité souhaité, ces applications reposent sur l'utilisation instantanée de centres de données distantes, et des ressources de véhicules inutilisées afin d'assurer des prestations de stockage, de calcul et de communication. Les ressources véhiculaires inutilisées peuvent être localisées dans les parkings, les aéroports et les autoroutes.

En période de pointe, la capacité de la route peut être dépassée, car le nombre de demandes pour traverser des sections ou tronçon de route est trop important, ce qui ralentit la vitesse de circulation. Dans de telles conditions où le trafic présente un encombrement, le Diagramme Macroscopique Fondamental (MFD) des modèles de flux de trafic présente une corrélation non linéaire entre les données de flux de trafic et la capacité routière. La faible corrélation prolongera l'heure prévue d'arrivée du conducteur vers sa destination, rendant ainsi la prévision du processus de flux de trafic plus complexe. Sur la base de cette tendance, la solution la plus fiable pour la planification routière à long terme serait une meilleure utilisation de la capacité routière plutôt que l'extension de la capacité des routes existantes. Cette solution améliorera la fluidité de la circulation qui est affectée par l'état actuel des routes (i.e., les panneaux d'arrêt et les passages pour piétons) et réduira le coût de la congestion afin de minimiser la perte de temps aux heures de pointe.

Dans cette thèse, nous nous intéressons à l'amélioration des modèles de circulation automobile dans des situations d'encombrement. Tout d'abord, sur la base de données de trafic empirique, nous adaptons un modèle d'interaction entre les paramètres de trafic en utilisant un modèle de classification binaire (régression logistique). Le modèle prédit la probabilité d'incidents entrants sur la route en cas de ralentissement soudain de la vitesse. Pour justifier la pertinence et la qualité de la classification, notre modèle est soumis à plusieurs tests statistiques. Les résultats des tests montrent qu'il existe plusieurs interactions entre les paramètres de trafic permettant d'interpréter précisément le phénomène de congestion.

Ensuite, nous nous concentrerons sur la dérivation d'une nouvelle fonction d'apprentissage à base d'intégrale floue pour prédire des informations de trafic fiables et efficaces en temps réel. Les données d'apprentissage sont collectées à partir de sources de trafic multiples telles que les détecteurs de trafic, les données échangées entre les véhicules en temps réel et les données historiques du trafic. Le modèle proposé permet de modéliser les interactions entre les variables de trafic sauvegardées dans ces sources. Enfin, nous intégrerons la fonction de prédiction proposée dans notre couche VANET-Cloud, qui comprend à son tour deux types de services de trafic : les services potentiels (services à la demande) et les services automatiques. L'objectif est de concevoir un modèle de livraison élastique pour prendre en charge de nombreuses fonctionnalités en termes de prédiction et de notification.

**MOTS CLÉS :** Cloud Computing, VANET, communication V2V, communication V2C, sécurité routière, données floue, Modèle d'apprentissage, prédiction, flux du trafic, congestion.

# ABSTRACT

---

The evolution of smart device technology and the advancement of Internet capabilities are converging ad hoc vehicle networks (VANETs) into the Cloud Computing Vehicle (CCV) paradigm. In the light of this convergence, many innovative applications are emerging to improve vehicle routing, mitigate traffic congestion through efficient calculation on static data. To achieve a desired level of reliability, these applications rely on instantaneous use of remote data centers, and unusable vehicular resources provide storage, compute, and communication services. Inoperable vehicular resources can be located in car parks, airports and highways.

In peak periods, the road capacity can be exceeded critical capacity, since the number of road demand to pass the sections of the road is quite enough to slow the traffic flow speed. At congested traffic conditions, the Macroscopic Fundamental Diagram (MFD) for traffic flow models presents a non-linear correlation between traffic flow data and road capacity. The lower correlation will extend the predicted time of arriving at a driver's time to its destination, thus making the traffic flow forecast without road capacity a more complex process. Based on this trend, the best use of road capacity is a reliable solution for long-term road planning rather than extending existing road capacity. This solution will improve the fluidity of traffic that is affected by the current state of the roads (eg, stop signs and pedestrian crossing), and reduce the cost of congestion to minimize wasted time at peak periods.

The goal of current research aims to achieve intelligent mobility based on how to identify alternative routes for each driver according to several parameters such as, optimal travel time, minimum journey, normal speed, ..., etc.

In this thesis, we are interested in improving vehicular traffic models in congested conditions. First, based on empirical traffic data, we adapt a model of interaction between traffic parameters using a binary classification model (logistic regression). The model predicts the probability of incoming incidents on the road in the event of a sudden slowdown in speed. To justify the relevance and quality of the classification, our model is put under several statistical tests. The results of the tests show that there are several interactions between the traffic parameters that make it possible to accurately interpret the phenomenon of congestion.

Second, we will focus on the derivation of a new fuzzy integral learning function to predict low and efficient traffic information in real time. Learning data is collected from multiple traffic sources such as traffic detectors, data exchanged between real-time vehicles and historical traffic data. The proposed model allows to model the interactions between the traffic variables saved in these sources. Finally, we will integrate the prediction function proposed in our VANET-Cloud layer, which in turn includes two types of traffic service : potential services (on-demand services) and automatic services. Clearly, the goal is to design an elastic delivery model to support many features for drivers in terms of prediction and notification.

**KEYWORDS :** Cloud Computing, Réseau Véhiculaires, Fuzzy data fusion, Learning model, Traffic flow prediction, Prediction error, Connected loop detectors, Multiple Linear Regression, Choquet integral, Congested conditions.

# DÉDICACE

---

*A ma mère et mon père,  
A ma grand-mère Aïcha,  
A mes sœurs,  
A mon encadreur Dr. Derdour Makhlouf*

# REMERCIEMENT

---

*Je fais partie des personnes qui croient mordicus qu'il n'y a de force ni de puissance que par Dieu. Mes remerciements vont tout premièrement à dieu tout puissant pour la volonté, la santé et la patience qui m'a donné durant tous ces années d'études.*

*Cette thèse a été réalisée avec le soutien et l'assistance d'un grand nombre de personnes. Je profite cette occasion pour exprimer ma gratitude à ces personnes pour leurs contributions comme mes collègues, ma famille et mes amis.*

*Je suis très reconnaissant aux membres du jury qui ont accepté de réviser et d'évaluer ma thèse et de fournir des critiques pertinentes. Je vous assure de ma profonde gratitude pour l'attention que vous avez portée à ce manuscrit et le temps que vous avez consacré à son évaluation.*

*Je voudrai exprimer tout d'abord ma gratitude à ma belle-famille et surtout, à ma mère, mon père, ma grand-mère Aïcha, mes sœurs. Ils sont toujours intéressés à mes études avec le succès et m'ont aidé beaucoup plus qu'elles ne peuvent le croire.*

*Je voudrai exprimer mes sincères remerciements à mon encadreur, le Dr. Derdour Makhoulouf, pour avoir acceptée de diriger ma thèse. Il m'a fait connaître un nouveau domaine de recherche réseau, de l'informatique de nuage véhiculaire. Au cours de ma thèse, Dr. Derdour Makhoulouf m'a orienté en proposant de nombreuses idées et perspectives. Grâce à son pensé logique dans ce domaine, mes idées et mes propositions initiales ont été améliorés et développés.*

*Je tiens également ma gratitude à mon co-encadreur le Dr. Philippe Roose pour avoir acceptée de diriger cette thèse et pour ses conseils et son suivi durant mon stage en France, n'oublie pas non plus toute sa famille.*

*Je voudrais m'adresser aussi sincèrement que possible aux membres du jury : Prof. Mohammed Rida Laouar, Pr. Nacira Ghoualmi, Dr Hakim Bendjenna et Dr Abdelkrim Bouramoul pour avoir accepté de participer au jury de cette thèse. Je voudrais également exprimer mes sincères remerciements à tous les membres de notre laboratoire de recherche (LAMIS), qui cherchent à développer le travail de recherche de ce laboratoire dans des domaines particuliers de l'informatique, et à valoriser le travail de recherche en organisant et en présentant des journées doctorales et des conférences internationales*

*Je voudrais également remercier tous mes collègues pour leur patience et leurs encouragements durant mon doctorat. Donc, heureux d'avoir des amis aimants comme vous. Vraiment merci.*

*Merci à vous tous !*



# TABLE DES MATIÈRES

<b>Résumé</b>	1
<b>Abstract</b>	2
<b>Dédicace</b>	3
<b>Remerciement</b>	4
<b>Liste des Tables</b>	9
<b>Introduction Générale</b>	13

## CHAPITRE 1

### INTRODUCTION AU CLOUD COMPUTING VÉHICULAIRE \_\_\_\_\_ PAGE 17

Introduction	18
1 Les réseaux véhiculaires	19
2 Les standards de la communication dans les VANET(s)	19
3 La convergence des applications VANET vers les applications Cloud Computing Véhiculaire	22
4 Le Cloud Computing véhiculaire	23
5 Les plateformes générales du Cloud Computing véhiculaire	24
5.1 Les plateformes Cloud Véhiculaire (CV)	24
5.2 Les plateformes VANET-Cloud (VuC)	25
5.3 Les plateformes Cloud Véhiculaires Hybrides (HVC)	25
6 Les applications de Cloud Computing véhiculaire	26
6.1 Centre de données pour l'aéroport	26
6.2 La gestion des feux de signalisation	27
6.3 La gestion dynamique du flux du trafic	27
7 Les outils d'évaluation	28
Conclusion	29

## CHAPITRE 2

### LA DYNAMIQUE NON LINÉAIRE DU TRAFIC VÉHICULAIRE \_\_\_\_\_ PAGE 30

Introduction	32
--------------	----

1	Les théories et les variables du flux de trafic	33
1.1	Les caractéristiques fondamentales du flux de trafic	33
1.2	La modélisation du flux de trafic	35
1.3	Le diagramme fondamental du trafic	36
2	Les techniques d'agrégation des données du trafic	38
2.1	L'agrégation complémentaire	38
2.2	L'agrégation redondante	39
2.3	L'agrégation coopérative	40
3	Les fonctions d'agrégations	40
3.1	Les fonctions linéaires	40
3.2	Les fonctions non-linéaires	42
4	L'apprentissages automatique supervisé	42
4.1	Définition formel	43
4.2	Le problème d'apprentissage supervisé	43
4.3	Les fonctions de perte	43
5	Les modèles d'apprentissages pour la prédiction du flux de trafic	44
6	Discussion	45
	Conclusion	47

### CHAPITRE 3

## UNE ÉVALUATION PRÉCISE DES PARAMÈTRES DU TRAFIC EN UTILISANT LA RÉGRESSION LOGISTIQUE PAGE 48

	Introduction	49
1	Scénario d'étude	50
2	Estimation du temps de retard des véhicules	50
3	Un modèle de régression logistique pour la détection des incidents routiers inattendus	53
3.1	La vérification de la pertinence du modèle (Bonne adaptation)	54
3.2	Évaluation des variables expliquées	56
	Conclusion	59

### CHAPITRE 4

## UN SYSTÈME DE FUSION FLOUE POUR UNE PRÉDICTION PRÉCISE DU TRAFIC PAGE 60

	Introduction	61
1	Un système de fusion floue de données du trafic basé sur les données d'apprentissage	62
1.1	La source de données du trafic	62
1.2	Les niveaux de fusion des données	63

1.3	La phase d'évaluation	66
2	Modèle d'apprentissage flou	66
2.1	La régression linéaire multiple (MLR)	67
2.2	La mesure floue et l'intégral floue de Choquet	67
2.3	Le modèle de prédiction Choquet-MLR	69
3	Les prédicteurs du trafic	70
3.1	Le temps de passage du segment routier	71
3.2	Le prédicteur du temps de déplacement	73
4	Expérimentation et évaluation	74
4.1	Les performances d'agrégation	75
4.2	Les critères d'évaluation de la prévision	75
4.3	L'impact du volume de données sur le processus de prédiction	77
	Conclusion	80

## CHAPITRE 5

## UNE PLATEFORME VANET-CLOUD POUR LA GESTION DES INFORMATIONS DU TRAFIC PAGE 81

	Introduction	83
1	Une plateforme VANET-CLOUD pour la gestion du trafic	84
1.1	Une couche de collecte des données du trafic	84
1.2	Infrastructure en tant que couche de service (IaaS)	85
1.3	Une plate-forme en tant que couche de service	85
1.4	Une couche de service VANET-Cloud	85
2	Modèle d'apprentissage non linéaire monotone en utilisant la régression de Choquet	86
3	Mécanisme de diffusion de données avec la communication V2C	87
3.1	Le mécanisme d'échange de données	89
3.2	Les algorithmes de traitement des messages	90
3.3	L'organigramme de flux de données	90
4	Simulation et évaluation	92
4.1	Une forme étendue de la plateforme iCanCloud pour évaluer les performances des architectures VANET-Cloud	95
4.2	Taux de livraison de données	96
4.3	La haute performance de communication	97
4.4	La latence du service Cloud	97
	Conclusion	99
	<b>Conclusion générale</b>	100
	<b>Perspectives</b>	102

<b>Liste Des Publications de Nos Travaux</b>	103
<b>Bibliographie</b>	104

# LISTE DES FIGURES

FIGURE 1.1:	Le réseau véhiculaire. . . . .	19
FIGURE 1.2:	Architecture du WAVE (Jiang et Luca, 2008). . . . .	21
FIGURE 1.3:	Architecture du CALM (P1609.1, 2006). . . . .	21
FIGURE 1.4:	Architecture du protocole C2CC (Bossom et al, 2009). . . . .	22
FIGURE 1.5:	La convergence des différentes technologies qui conduisent à l'émergence de l'Internet intelligent du véhicule. . . . .	23
FIGURE 1.6:	Le Cloud Véhiculaire. . . . .	25
FIGURE 1.7:	Les VANET(s)-Cloud. . . . .	25
FIGURE 1.8:	Les Cloud véhiculaires hybrides. . . . .	26
FIGURE 2.1:	Vitesse moyenne vs vitesse instantané. . . . .	34
FIGURE 2.2:	la densité ( $\rho$ ). . . . .	34
FIGURE 2.3:	Diagramme fondamental (modèle macroscopique). Cas d'étude dans une autoroute japonaise. . . . .	37
FIGURE 2.4:	Agrégation complémentaire. . . . .	39
FIGURE 2.5:	L'agrégation redondante. . . . .	39
FIGURE 2.6:	L'agrégation coopérative. . . . .	40
FIGURE 3.1:	L'autoroute Est-Ouest de Constantine. . . . .	50
FIGURE 3.2:	Modéliser la vitesse du trafic entre un couple de boucles inductives. . .	51
FIGURE 3.3:	La corrélation entre la vitesse observée et la vitesse prévue des véhicules. . . . .	52
FIGURE 3.4:	Le niveau de corrélation entre le temps de déplacement observé et le temps de déplacement prévu après l'incident. . . . .	52
FIGURE 3.5:	La courbe Roc pour la précision de la classification et la sensibilité du modèle de régression. . . . .	57
FIGURE 4.1:	L'architecture de base du système proposé pour la fusion floue de données à partir de plusieurs sources de données du trafic. . . . .	62

FIGURE 4.2:	La collection de données du trafic sur le Cloud pour la prévision des flux du trafic à court terme. . . . .	63
FIGURE 4.3:	L'adaptation du modèle de données du trafic. . . . .	66
FIGURE 4.4:	L'organigramme du processus de prédiction de la réponse en utilisant Choquet-MLR. . . . .	72
FIGURE 4.5:	La description du modèle de flux du trafic. . . . .	75
FIGURE 4.6:	La prévision du flux du trafic en utilisant l'agrégation non linéaire (IC) par rapport à l'agrégation linéaire (WAO). . . . .	76
FIGURE 4.7:	La prévision à court terme du trafic. . . . .	79
FIGURE 4.8:	Prévision quotidienne du trafic. . . . .	79
FIGURE 4.9:	Les données du trafic observé versus les données prédites en 1 min et 3 min de temps. . . . .	79
FIGURE 5.1:	Une plateforme VANET-Cloud pour la gestion du trafic. . . . .	84
FIGURE 5.2:	Une couche de service VANET-Cloud pour les services à la demande et automatiques. . . . .	87
FIGURE 5.3:	Les modèles de diffusion de données avec la communication V2C. . . . .	88
FIGURE 5.4:	Organigramme de flux d'informations sur le trafic des véhicules. . . . .	93
FIGURE 5.5:	Description du scénario. . . . .	94
FIGURE 5.6:	Le modèle de flux du trafic. . . . .	94
FIGURE 5.7:	Vue conceptuelle de l'intégration. . . . .	96
FIGURE 5.8:	Débit de transmission de données par le Cloud pour chaque modèle. . . . .	97
FIGURE 5.9:	Les performances de communication des services Cloud pour les trois modèles de diffusion de données. . . . .	98
FIGURE 5.10:	La latence des services Cloud avec les normes V2C et IEEE 802.11p. . . . .	98

# LISTE DES TABLES

TABLEAU 3.1: Les paramètres de trafic sélectionnés pour la prédiction de l'incident. . . . .	54
TABLEAU 3.2: Le test Chi-squared pour les coefficients du modèle . . . . .	55
TABLEAU 3.3: Interprétation des variables impliquées dans le modèle (model summary)	55
TABLEAU 3.4: Tableau de classification à l'étape 0. . . . .	56
TABLEAU 3.5: Tableau de classification à l'étape 1. . . . .	56
TABLEAU 3.6: Le test Hosmer et Lemeshow. . . . .	56
TABLEAU 3.7: La zone sous la courbe. . . . .	57
TABLEAU 3.8: Variables dans l'équation à l'étape 0. . . . .	57
TABLEAU 3.9: Variables dans l'équation à l'étape 1. . . . .	58
TABLEAU 4.1: les termes qui décrit la description de la collecte de données. . . . .	71
TABLEAU 4.2: Coefficients pour agréger les variables du trafic. . . . .	75
TABLEAU 4.3: Comparaison entre la puissance d'agrégation du prédicteur de vitesse à l'aide d'opérateurs CI et WAO. . . . .	76
TABLEAU 4.4: Comparaison entre la méthode Choquet-MLR proposée avec les autres méthodes de prédiction. . . . .	78
TABLEAU 5.1: Paramètres de simulation. . . . .	95

---

# CHAPITRE : INTRODUCTION GÉNÉRALE

---

1. CONTEXTE DU TRAVAIL ET PROBLÉMATIQUE
2. CONTRIBUTIONS
3. STRUCTURE DE LA THÈSE



## Contexte du travail et problématique

Aujourd'hui, les individus, dans leur vie quotidienne, sont mobilisés par des trajets en voitures de plus en plus longs en particulier pour les personnes habitant et travaillant en banlieue des grandes agglomérations. Les déplacements entre leurs lieux de résidence et de travail ainsi que leurs points d'intérêt (ex., le domicile, le lieu de travail, lieu de rendez-vous, restaurant, etc.) sont devenus un problème de vie quotidien pour eux. Ces individus essaient de toujours rendre les déplacements en voiture entre deux points d'intérêt (POI) les plus confortables, les plus sûrs et les plus agréables possible. Face à ces besoins, les autorités publiques et les communautés de transport travaillent à l'amélioration de la sécurité de nos systèmes de transport tout en réduisant le temps de déplacement, le nombre d'accidents, le contrôle efficace de la fluidité du trafic, etc.

En dépit des évolutions techniques adoptées par les acteurs de transports comme l'intégration des capteurs dans le bitume, les applications de gestion de trafic, la présence de feux de signalisation « *intelligents* », le comptage des véhicules, le nombre et l'ampleur des embouteillages ne cesse d'augmenter. La gestion des embouteillages aux heures de pointe est l'un des challenges à résoudre aujourd'hui. Atténuer le coût des embouteillages dans les zones urbaines est une perspective clé pour les services de transport et des usagers de la route. Selon le Centre d'Economie et de Recherche d'Entreprise (CERE) (INRIX, 2014), le temps passé par les véhicules (et donc les automobilistes) sur les routes engorgées est extrêmement coûteux en termes économiques, écologiques et commerciaux. Les prévisions concernant les Etats-Unis, l'Allemagne, la France et le Royaume-Uni indiquent que l'augmentation du coût du trafic augmentera de 46 % en 2030 par rapport aux évaluations en 2013.

Selon les prévisions du CERE, les embouteillages ont énormément évolués avec la révolution urbaine, et contribuent continuellement à perturber les individus dans leurs activités, leurs déplacements, mais également de manière plus générale, les activités de transport et l'économie des services gouvernementaux. En fait, de nombreuses communautés de transport se concentrent sur le développement des Systèmes Décisionnels basés sur la Gestion du Trafic (SDGT) pour contrôler la sécurité routière, le réseau de trafic et assurer une mobilité durable pour tous les usagers de la route (véhicules, passagers, piétons, etc.). À partir de leurs points de vue, les Systèmes Intelligents de Gestion du Trafic (SIGT) seront utilisés dans les villes intelligentes (Smart Cities) comme une infrastructure de base pour faciliter la vie des individus en résolvant des problèmes quotidiens inhérents au transport tels que la congestion du trafic, le covoiturage, etc. En outre, les solutions offertes par les systèmes intelligents de gestion du trafic offriront un avantage non seulement pour les usagers de la route, mais aussi pour les centres de gestion du transport et les autres secteurs gouvernementaux.

D'autre part, les solutions mises en place par les centres de recherche en réseau pour réduire les embouteillages sont obtenues à partir des informations détectées par l'infrastructure routière (ex. camera, boucle inductive, etc.) et les informations échangées dans les réseaux véhiculaire ad-hoc (VANET). Les VANET sont des réseaux coopératifs se basant sur la communication sans fil où l'échange des données est effectué entre les véhicules et leurs voisins ou entre les véhicules et l'infrastructure. Ainsi, sur la base des informations de trafic échangées en temps réel et de cet horizon télématique plus large offert aux utilisateurs de la route, des phénomènes potentiels et dangereux de la route peuvent être prédits et interprétés de manière précoce permettant ainsi un temps de réaction suffisant et permettant ainsi adopter des contre-mesures appropriées.

Le réseau ad hoc de véhicules (VANET) a suscité dernièrement beaucoup d'intérêts pour être une partie intégrante dans la prochaine génération des paradigmes de réseau tels que l'Internet des objets et le Cloud Computing (CC). En fait, cette convergence est soutenue par un certain nombre de solutions pour relever le défi et les problèmes des réseaux de véhicules.

Le Cloud Computing Véhiculaire (CCV) est l'une de ces solutions. Le CCV est une nouvelle technologie hybride qui a un impact important sur la gestion du trafic et la sécurité routière en utilisant instantanément les ressources des véhicules, tels que les ressources de traitement, le stockage et l'Internet pour la prise de décision. Le but principal de cette intégration est d'offrir des plateformes unifiées indispensables à la communication entre les véhicules et le traitement de l'information. L'ensemble des applications avancées fournies par ces plateformes de communication est destiné à faciliter la conduite en douceur, améliorer la sécurité routière et à contrôler le trafic. Compte tenu des caractéristiques de ces véhicules intelligents (capacités de traitement, de stockage et de communication, accès aux services Cloud) et de leur impact potentiel sur la sécurité routière et la sensibilisation du conducteur, l'ensemble des applications promises par le CCV ne se limite pas seulement à la sécurité routière (ex. les accidents), mais s'étendent à l'optimisation du flux trafic, comme le contrôle de la congestion des flux, l'optimisation des itinéraires vers des applications et des plateformes commerciales telles que la planification du voyage, le partage de fichiers, le covoiturage et l'accès à Internet.

À la lumière de cette convergence, comprendre le comportement précis de la circulation devient une tâche cruciale dans les Systèmes de Transport Intelligent (STI), la communication entre les véhicules et l'avenir de la sécurité routière. Plus précisément, anticiper le flux du trafic en temps réel est essentiel pour libérer l'encombrement du trafic, améliorer la sécurité routière et réduire l'impact du trafic sur les incidents routiers. Par conséquent, la capacité à prévoir rapidement, de manière fiable et précise la dynamique du trafic en des périodes à court terme devient de plus en plus importante. Les modèles de prévision du trafic à court terme sont devenus un élément intégral du jeu d'instrument nécessaire au contrôle et à la gestion du trafic en temps réel. En outre, de tels instruments sont importants pour fournir aux voyageurs des informations fiables sur le temps de déplacement, l'optimisation des feux de signalisation et améliorer les pratiques du déplacement.

Dans cette thèse, nous nous sommes intéressés au traitement de l'information du trafic en temps réel et de son impact potentiel sur les applications CCV en se basant sur différents scénarios expérimentaux. Dans un premier temps, nos contributions à la gestion du trafic s'étendent de la collecte, à l'agrégation des données du trafic jusqu'à la prédiction des données. Ensuite, nous adaptons notre mécanisme de prédiction à des contextes variés qui méritent une optimisation dynamique comme la communication véhiculaire et les pratiques de déplacements quotidiens en tenant en compte des conditions d'embouteillage. Chacune de ces contributions se concentre sur un domaine de recherche particulier du monde Cloud Computing Véhiculaire (VCC) et tente de relever les défis pertinents des études VANET de nos jours.

## Contributions

Les principales contributions de cette thèse sont résumées comme suit :

- **L'évaluation précise du modèle de flux de trafic à l'aide de la régression logistique :** dans cette partie de la thèse, une étude est présentée dans le cadre du développement d'un modèle prédictif basé sur un modèle de régression binaire pour interpréter le phénomène de congestion d'une manière précise. L'objectif du modèle développé est double : (1) modéliser les interactions entre les variables de trafic impliquées dans le modèle, (2) anticiper avec précision la probabilité d'occurrence d'un incident imprévu sur la route. En outre, la vitesse du véhicule observée, le temps de déplacement prévu, le temps de déplacement observé et le temps de retard sont quatre variables de trafic impliquées dans le modèle. De plus, la précision de la prédiction est prouvée en vérifiant la pertinence du modèle en fonction de la qualité de l'ajustement et de la signification

statistique de chaque variable expliquée.

- **Une approche basée sur les intégrales floues pour préciser la prédiction du flux du trafic à court terme :** cette partie de la thèse se concentre sur la conception d'une nouvelle technique de collection, d'agrégation et de prédiction du flux de trafic en temps réel. L'objectif principal est d'optimiser le taux d'erreur de la prédiction dans les conditions d'embouteillage dont les paramètres de trafic ont des valeurs instables. Pour cela, nous proposons une méthode de prédiction de flux de trafic basé sur une intégrale de floue et l'analyse de régression. La méthode intègre des paramètres de trafic ajustés via l'intégrale floue de Choquet (CI) et l'analyse de Régression Linéaire Multiple (MLR) pour augmenter la précision de prédiction du flux de trafic. L'intégrale choisie vise à modéliser les interactions entre les données de trafic observées à partir de plusieurs sources de données (par exemple, des détecteurs de boucle connectés) pour ajuster la stabilité des paramètres du trafic dans le processus de prédiction. Notre étude expérimentale repose sur un modèle de flux de trafic où des détecteurs de boucle sont installés sur un segment de l'autoroute algérienne Est-Ouest (A1). Les résultats montrent une forte corrélation entre les données prédites et l'état réel du flux de trafic. De plus, l'erreur de prédiction de l'analyse de régression est considérablement réduite. En outre, les performances de prédiction en utilisant l'intégrale floue de Choquet ont surpassé les intégrales non linéaires d'agrégation et les autres méthodes de prédiction à l'aide de métriques Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) et EC Index (EC).
- **Une couche VANET-Cloud pour prédire et diffuser les informations du trafic :** dans cette partie de la thèse, nous proposons une couche VANET-Cloud pour réduire la congestion routière en se basant sur la prédiction et la diffusion d'informations de trafic. La couche proposée offre des services Cloud potentiels et des services automatisés aidant à contrôler la mobilité des véhicules. Les services Cloud potentiels utilisent des détecteurs de boucle connectés comme une base pour prédire les informations de trafic (comme le temps de déplacement). L'information de trafic prédite sera diffusée ensuite selon trois modèles basées sur la communication Vehicule-to-Cloud (V2C) : modèles réactifs, proactifs et hybrides. Notre évaluation et notre expérimentation ont pour but d'examiner les performances réseau de ces services pour chaque modèle de diffusion. Les résultats de la simulation indiquent que la couche proposée offre de meilleures performances par rapport aux réseaux conventionnels VANET.

## Structure de la thèse

La thèse est structurée en cinq chapitres qui organisent l'ensemble des travaux effectués pendant notre recherche, les cinq chapitres sont détaillés de la manière suivante. Tout d'abord, un aperçu sur les concepts de base du réseau véhiculaire et sa convergence vers les applications Cloud Computing Véhiculaires (CCV) sont détaillés dans le chapitre 1.

Le chapitre 2 présente un état de l'art sur les théories et les variables du flux de trafic et les solutions pour libérer la congestion routière. En particulier, l'état de l'art est centré autour les techniques d'agrégation et de prédiction des données du trafic, les modèles d'apprentissage supervisées et leurs défis jusqu'à présent. L'évaluation critique de ces techniques pour la prédiction dans le domaine de transport est également brièvement décrite dans ce chapitre.

Les chapitres 3, 4 et 5 réunissent nos contributions qui s'appuient sur la technologie CCV pour optimiser à la fois la congestion routière et les performances de la communication véhiculaire pendant les conditions d'embouteillage. Pour cela, dans ces chapitres, nous

présenterons, dans un premier temps, une étude empirique pour évaluer les paramètres de trafic dans la phase de la congestion. Ensuite, nous présenterons, dans le chapitre 4, notre technique de collection et d'agrégation à base intégrale du floue de Choquet pour préciser à court terme la prédiction du flux du trafic. Ultérieurement, nous proposerons, dans le chapitre 5, une couche de services pour le trafic véhiculaire, dite VANET-Cloud, reposant sur la technique de la précision de la prédiction afin d'augmenter l'efficacité de la gestion du trafic. Enfin, les résultats de simulation sont fournis pour montrer la précision de la prédiction et son impact sur les performances de la communication véhiculaire.

# Introduction au Cloud Computing Véhiculaire

## SOMMAIRE: CHAPITRE 1

Introduction	18
1 Les réseaux véhiculaires	19
2 Les standards de la communication dans les VANET(s)	19
3 La convergence des applications VANET vers les applications Cloud Computing Véhiculaire	22
4 Le Cloud Computing véhiculaire	23
5 Les plateformes générales du Cloud Computing véhiculaire	24
5.1 Les plateformes Cloud Véhiculaire (CV)	24
5.2 Les plateformes VANET-Cloud (VuC)	25
5.3 Les plateformes Cloud Véhiculaires Hybrides (HVC)	25
6 Les applications de Cloud Computing véhiculaire	26
6.1 Centre de données pour l'aéroport	26
6.2 La gestion des feux de signalisation	27
6.3 La gestion dynamique du flux du trafic	27
7 Les outils d'évaluation	28
Conclusion	29

## INTRODUCTION

Les réseaux VANET appartiennent à un domaine de recherche d'actualité à savoir les communications V2V, V2C et leurs apports à la sécurité routière. De nos jours, les véhicules ordinaires sont dotés d'équipements de haute technologie permettant de fournir aux automobilistes plus de confort pour leur conduite. En effet, l'intégration de ce type de réseaux avec la technologie Cloud Computing (CC) est devenue un sujet d'actualité. L'objectif de cette intégration est de construire des plateformes globales de communication et de traitement d'information accessibles en ligne et fondant une partie cruciale du Système de Transport Intelligent (STI). L'ensemble des applications avancées fourni par ces plateformes vise à augmenter la souplesse de la conduite, améliorer la sécurité routière et contrôler la fluidité de la circulation, etc.

Nous allons mettre en évidence, dans ce chapitre, les concepts fondamentaux de la technologie Cloud Computing Véhiculaire (VCC). D'abord, nous commencerons par un aperçu sur les réseaux véhiculaires, les architectures générales et les standards de communications. Ensuite, nous exposerons la convergence de ce genre de réseaux avec les applications de type Cloud Computing véhiculaire en présentant ses architectures et ses applications innovantes.

## 1 Les réseaux véhiculaires

Les réseaux véhiculaires ad hoc (VANET) émergent grâce aux progrès technologiques dans le domaine de l'automobile. Ce type de réseau diffère des autres types de réseaux mobiles en raison de son architecture hybride et de sa grande mobilité liée au déplacement des véhicules. Les nœuds de ce type de réseau se distinguent par des caractéristiques uniques telles qu'une courte portée de communication, une faible bande passante, une topologie hautement dynamique, une connectivité spontanée et une mobilité dirigée. Un réseau de véhicules peut être formé par des véhicules connectés qui pourraient avoir des dispositifs de communication sans fil, un système de localisation et des interfaces radio. Dans le monde d'aujourd'hui, les véhicules modernes sont équipés de divers dispositifs embarqués tels que l'OBU (On Board Unit), des caméras, un gestionnaire d'événement (Event Data Recorder, EDR), un radar, etc. Ces appareils de haute technologie constituent une partie indispensable du réseau de capteurs du véhicule.

Le réseau de capteurs intégré dans les véhicules leur permet d'effectuer de nombreuses fonctionnalités complexes telles que l'acquisition, le stockage et le partage d'informations avec d'autres véhicules, et avec de technologies telles que les services Cloud, Internet des objets, etc. L'objectif de la plupart des travaux de recherches dans ce domaine est d'accroître la sécurité routière et de réduire le taux de collisions routières. Comme l'illustre la figure 1.1, ces réseaux présentent une classe émergente des réseaux mobile. Ils peuvent être considérés comme une particularité des réseaux de type MANET (*Mobile Ad hoc NETWORK*) où les nœuds mobiles sont des véhicules intelligents. On parle de notion de « *véhicule intelligent* » quand un véhicule est capable d'effectuer des tâches liées à l'intelligence afin de prendre des décisions autonomes.

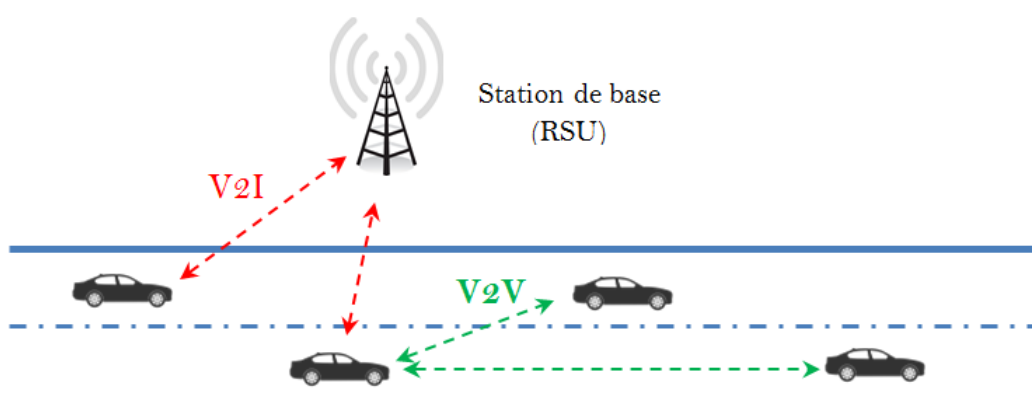


FIGURE 1.1 – Le réseau véhiculaire.

## 2 Les standards de la communication dans les VANET(s)

La Commission fédérale des communications des États-Unis (FCC, 1999) développe DSRC (en anglais Dedicated Short Range Communication) pour « *augmenter la sécurité du voyageur, réduire la consommation de carburant et la pollution, et continuer à faire avancer l'économie du pays* » (Kenney, 2011). Le DSRC est un standard de diffusion d'informations par le biais des ondes radar. Il ne représente que la partie « air » de la communication, c'est à dire la transmission radio établie entre des équipements fixes jalonnant l'infrastructure routière RSU (Road Side Unit) et des équipements embarqués à bord des véhicules connus sous le nom OBU (On board Unit). DSRC œuvre dans la bande de fréquence de 5.9 GHz (Europe et États-Unis) ou de 5.8 GHz (Japon) et sur une largeur de bande de 75 MHz (Brickley et al, 2010). Elle est



généralement segmentée en 7 canaux de 10 MHz chacun. L'ensemble des canaux se répartissant fonctionnellement en un canal de contrôle et six canaux de service. Le canal de contrôle est réservé à la transmission des messages de gestion du réseau et des messages de très haute priorité à l'instar de certains messages critiques liés à la sécurité routière. Quant aux 6 autres canaux, ils sont dédiés à la transmission des données des différents services annoncés sur le canal de contrôle.

Actuellement, la communication entre les véhicules attire toujours beaucoup de recherches dans le monde. Elle mène à la maturité des projets intéressants comme par exemple : la communication à courte portée DSRC/WAVE (Li, 2010), l'accès aux communications des services mobiles terrestres (CALM) (D31, 2009), et Car-to-Car Consortium (C2CC) (Eichler et al, 2006). Ces projets présentent les architectures VANET les plus connues ayant comme objectif principal l'initialisation et l'augmentation de la sécurité sur la route.

Le modèle de communication à courte portée DSRC/WAVE (Li, 2010), en anglais Wireless Access in Vehicular Environments (Accès sans fil dans les environnements véhiculaires), est un standard qui diffère du Wi-Fi et des autres réseaux cellulaires sans fils en raison de sa spécification et des familles collectives des standards comme : IEEE 802.11p, IEEE 1609 et SAE, Society of Automotive Engineers (Li, 2010); (SAE, 2010). Il comprend deux types de dispositifs (Morgan et Yasser, 2010; Hartenstein et Hannes, 2009) : des stations de base (RSU) et une unité de bord (OBU).

Les stations de base se réfèrent aux équipements fixés le long de la route pour fournir la connectivité internet et agir comme des passerelles pour les véhicules. L'unité de bord (On Board Unit, OBU) se réfère à des appareils qui équipent chaque véhicule dans lequel elle peut être utilisées la communication *véhicule-à-véhicule* (V2V) et celle de *véhicule-à-infrastructure* (V2I), respectivement comme communication directe et indirecte. Ainsi, il est conforme à l'architecture OSI mais avec une importance accordée aux trois couches suffisantes pour les besoins de la communication (Jerbi et Moez, 2008). La couche 1 dite « *Physique* » est basée sur le lien micro-ondes à 5.8 GHz. La couche 2 dite « *Liaison* », elle permet de gérer les données et la communication véhicule-à-infrastructure. La couche 7 dite « *Application* », est composée d'un ensemble de commandes qui traitent les applications envisagées, i.e., les applications embarquées sur le dispositif. Pareillement, le reste de la pile protocolaire de WAVE se situe entre la couche liaison de données et la couche application. La figure 1.2 représente l'architecture de WAVE que le groupe IEEE1609 a standardisé sous quatre catégories, l'architecture comporte les couches suivantes :

- **IEEE 1609.1** - *WAVE Resource Manager* : permet aux applications distantes une bonne gestion des ressources des OBU telles que la mémoire et l'interface utilisateur en garantissant l'interopérabilité (P1609.1, 2006).
- **IEEE 1609.2** - *WAVE Security Services for Applications and Management Messages* : décrit le format des paquets et les fonctions de sécurité dans un système WAVE pour les messages de sécurité, données et gestion (P1609.2, 2006).
- **IEEE 1609.3** - *WAVE Networking Services* : définit les services de niveau réseau et transport incluant l'adressage, le routage et la transmission. L'envoi est basé sur le protocole WSMP (WAVE Short Messages Protocol) qui permet un échange efficace des messages WSM (WAVE Short Messages) (P1609.3, 2006).
- **IEEE 1609.4** - *WAVE Multi-Channel Operation* : définit les mécanismes d'accès en priorité, la coordination et la gestion des 7 canaux DSRC lors du routage et de la transmission de données (P1609.3., 2006).



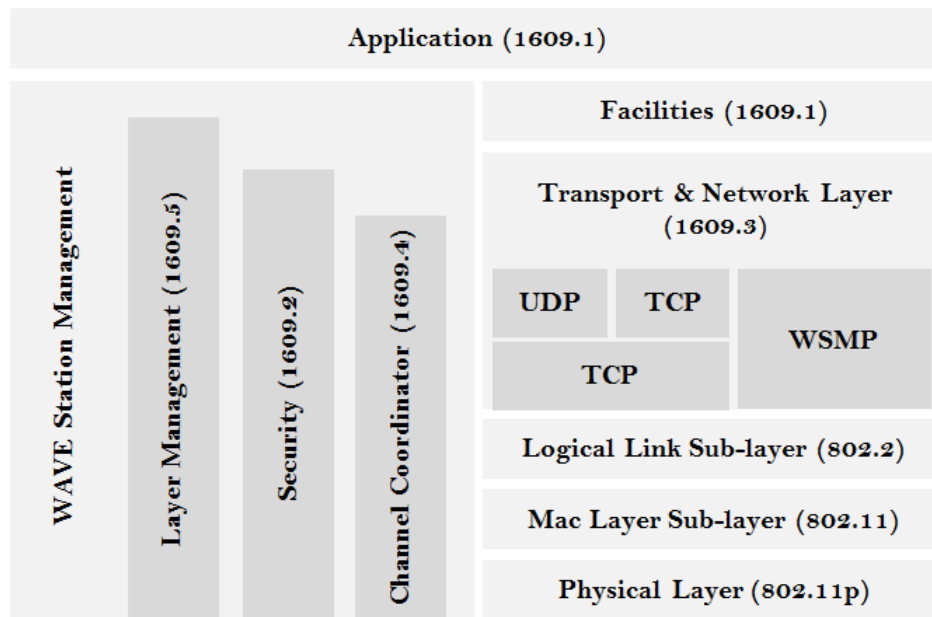


FIGURE 1.2 – Architecture du WAVE (Jiang et Luca, 2008).

L'Organisation Internationale de Standardisation (ISO) du groupe de travail 1 ISO TC204 WG16 (D31, 2009) initie le CALM (Communications Access for Land Mobiles) pour fournir plusieurs protocoles de communication, comme le Communication Air Interface (CAI), de longue et de moyenne portée. ISO WG a conçu le CALM pour que son architecture de communication aborde le principal challenge de la continuité de la communication entre les véhicules sur la route. L'architecture de CALM comprend toutes les couches du model ISO, de la couche physique à la couche application. La couche réseau est destinée à charger IPv6 pour les applications orientées utilisateur et non-IP (FAST) pour des applications de sécurité. Ainsi, le WG couvre tous les types de communication (multi-Saute, directe) comme véhicule-à-véhicule (V2V), véhicule-à-infrastructure (V2I), et de véhicule à l'Internet. La classification des protocoles est basée unicast, multicast et géo-cast. L'architecture prend en charge tout support physique de communications telles que la communication de moyenne portée ou la communication de courte portée du réseau cellulaire. Elle couvre aussi toutes sortes de ses applications (applications de sécurité et de non-sécurité, les applications Internet classiques). La figure 1.3 illustre l'architecture du CALM.

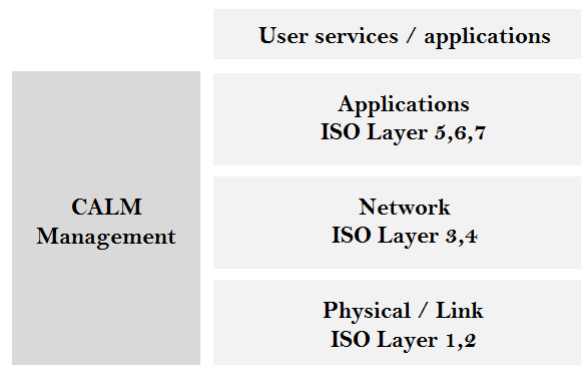


FIGURE 1.3 – Architecture du CALM (P1609.1, 2006).

**Car-to-Car Consortium (C2CC)** a été lancé par six constructeurs automobiles européens (BMW, Audi, Daimler Benz, Fiat, Honda, Opel, Renault, Volkswagen). L'objectif de ce projet

est double : (1) rendre les véhicules comme des « *objets intelligents* » sur la route, (2) standardiser les interfaces et les protocoles de communication sans fil entre les véhicules et leur environnement. Ainsi, il permet de développer des applications de sécurité active basées sur les spécifications de (Bossom et al., 2009). Il couvre aussi une longue gamme d'applications telles que la sécurité et d'autres applications de confort, la figure 1.4 illustre l'architecture du protocole C2CC.

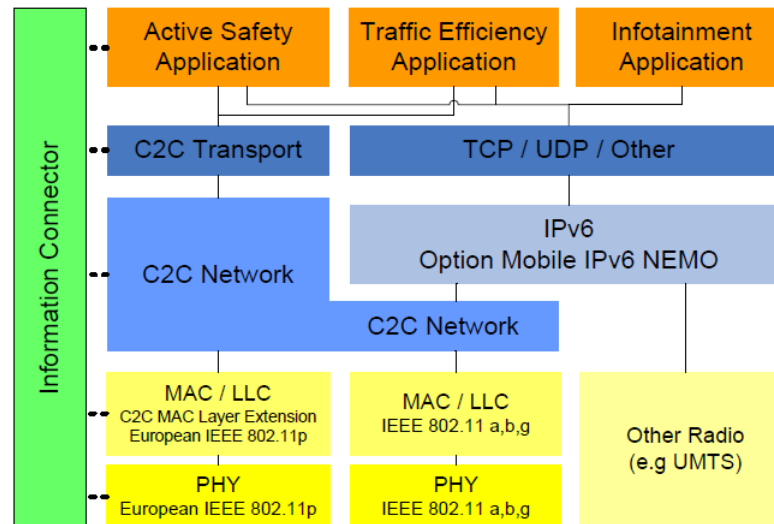


FIGURE 1.4 – Architecture du protocole C2CC (Bossom et al, 2009).

### 3 La convergence des applications VANET vers les applications Cloud Computing Véhiculaire

Le Cloud Computing (CC) est un nouveau paradigme basé sur la technologie Internet pour fournir un accès à la demande à un ensemble de services informatiques tels que les services de gestion, informatique utilitaire, ressources informatiques infinies en termes d'élasticité, de capacité de stockage et de calcul (Mell et Grance, 2011). Récemment, la technologie CC a été impliquée dans les environnements véhiculaires pour ajouter une valeur aux services de sécurité routière. De plus, les véhicules ordinaires deviennent de plus en plus sophistiqués en termes de capacité de stockage, de traitement et de communication. Cette convergence est appelée Cloud Computing Véhiculaire (CCV) (Gerla, 2012). Ce paradigme présente un nouveau changement technologique, qui profite à la technologie CC pour fournir aux utilisateurs des services utilitaires, des services à faible coût, des gestionnaires du trafic, des plateformes de planification et de voyage, etc.

Les Réseaux de Capteur Sans Fil (RCSF) et le CCV sont combinés entre eux via l'Internet des objets (IoTh) pour l'intérêt croissant des réseaux VANET conventionnels et leur convergence avec la nouvelle génération de réseaux sans fil. Au lieu d'équiper chaque véhicule d'une variété de dispositifs et de capteurs, les conducteurs peuvent déployer les ressources disponibles en accédant aux plateformes et aux infrastructures d'allocation. D'un autre côté, les véhicules peuvent fournir leurs dispositifs qui pourraient être alloués par d'autres véhicules. Ces plateformes sont fournies par le Cloud en tant que services à la demande, c'est le Cloud Computing Véhiculaire. L'avantage est de bénéficier des aspects économiques de l'installation qui s'expriment en termes de temps et de coût.

Toutes les applications fournies par ces plateformes seront disponibles via Internet pour fournir des informations aux conducteurs sur l'état du trafic, les conditions routières et d'autres informations. En outre, les véhicules peuvent former des nuages temporels tels que des mini-centres de données mobiles pour le stockage et le traitement. Ceci permettra aux véhicules possédant cette technologie, de fournir des services aux véhicules proches ou à d'autres véhicules connectés à distance. En effet, le conducteur peut décider de louer les ressources de sa voiture à la demande par jour, par semaine ou par mois. Cet avantage attire les propriétaires de grandes entreprises de la TI qui cherchent à augmenter leur capacité informatique en ajoutant des ressources de véhicules comme des ressources excédentaires pour des raisons pécuniaires (Yan et al, 2012).

En particulier, nous pouvons suivre les racines de cette convergence, en observant plusieurs technologies émergentes où chaque sous-convergence conduit à l'apparition d'une nouvelle génération d'applications. La convergence de ces domaines technologiques est illustrée dans la figure 1.5. Comme nous l'avons mentionné, les réseaux VANET(s) présentent une sous classe de Mobile Ad hoc Network (MANET). Ils représentent la partie importante dans les Système de Transport Intelligent (STI) qui forment des réseaux véhiculaires intelligents (IVN) (Karthikeyan et Bhuvaneshwari, 2012; Zhang et al, 2012; Serebinski et al, 2013; Boskovich and Barth, 2013). En outre, l'attribution d'un système robuste de gestion du trafic impose toujours un problème qui persiste. Mais avec l'avènement de la technologie CC, la tâche de gestion sera n'importe où, en tout temps, et elle fonctionnera selon trois modèles : Cloud privé, Cloud public et Cloud hybride. Ce passage s'appelle le CCV où les véhicules possèdent un accès public et/ou privé aux services Cloud via une connexion Internet (Talib et al, 2017; He et al, 2014; Whaiduzzaman et al, 2014; Lee et al, 2014; Gerla, 2012; Hussain et al, 2012). Par conséquent, l'Internet-de-Véhicule (IoV) présente une classe particulière de l'internet des objets (IoTh) pour différencier le concept de véhicule dans les réseaux VANET/IoV en termes de logiciel et matériel. Les solutions prometteuses par l'intégration de ces technologies dans les STI constituent la base essentielle d'une transition vers une sécurité routière plus sûre, une prise de décision optimale et un temps de retard négligeable. Il existe donc un besoin absolu de proposer différents types de réseaux de nuage de véhicules pour supporter une telle solution (Leng et Zhao., 2011).

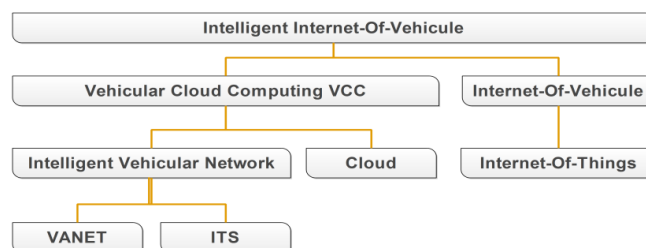


FIGURE 1.5 – La convergence des différentes technologies qui conduisent à l'émergence de l'Internet intelligent du véhicule.

## 4 Le Cloud Computing véhiculaire

Les véhicules sont des témoins d'incidents associés à la congestion et ne pouvant pas participer à l'atténuation de leurs effets (Whaiduzzaman et al, 2014). Dans les conditions complexes du trafic, les véhicules ont besoin de coopérer entre eux et avec les infrastructures routières pour échapper à une telle situation. Néanmoins, les véhicules peuvent prendre beaucoup du temps dans le processus de coopération et même leurs temps de déplacement prévu au départ peuvent être prolongés. Ainsi, les véhicules risquent de ne pas pouvoir

échapper aux embouteillages en raison du manque des ressources informatiques suffisantes, ces ressources rendraient la solution fournie performante ou bien ils sont connectés à des réseaux de nuage conventionnel limité, en termes de puissance de calcul et de stockage.

Le CCV est apparue comme un nouveau paradigme pour augmenter l'efficacité et la robustesse de ces systèmes. L'avantage essentiel de ce paradigme est de fournir des solutions à la demande aux événements qui se sont produits mais qui ne peuvent être satisfaits raisonnablement de manière proactive. Cet avantage n'existe pas dans les VANET conventionnels et s'avère être une caractéristique capitale pour les services innovants. Ce nouvel éventail de services a contribué également à solutionnées certains défis techniques des systèmes de transport de plus en plus complexes et leur incertitude émergente. Les tendances des recherches actuelles visent à utiliser les ressources informatiques inexploitées des immenses flottes de véhicules coincés dans les embouteillages pour fournir des solutions aux problèmes liés à leurs communications et à leurs mobilités en temps réel. Le grand nombre de véhicules sur les routes sera traité comme des ressources informatiques abondantes et sous-utilisées, qui peuvent servir à fournir des services.

L'idée de Cloud Véhiculaire est apparu récemment, les auteurs Olariu et Weigle ont mis accent sur les applications potentiels et les aspects importants qui est en relation avec les défis de la recherche (Olariu et Weigle, 2009). Ensuite, le paradigme CCV est défini par (Eltoweissy et al 2010) comme suit : « *un groupe étendu de véhicules qui peuvent agir soit d'une façon autonome soit en coopération pour la réalisation des différentes tâches comme : le calcul, la détection, la communication et des ressources physiques qui peuvent être coordonnées et allouées dynamiquement aux utilisateurs autorisés* ». Une extension de Cloud Computing traditionnel avec plusieurs nouvelles dimensions sera discutée ultérieurement dans les travaux (Olariu et al, 2011-2013). L'objectif principal de la majorité de ces travaux est d'aider les chercheurs à mieux comprendre les mécanismes fondamentaux des Cloud Computing Véhiculaires (CCV) et les différentes applications potentielles existantes pour améliorer la sécurité routière.

## 5 Les plateformes générales du Cloud Computing véhiculaire

Les plateformes de CCV sont désireuses de résoudre les problèmes inhérents accumulés dans les réseaux de véhicules, tels que la latence accrue des messages, l'efficacité des messages et le débit insuffisant. Hussain et al., (2012) ont divisé les Cloud VANET en trois plateformes architecturales potentielles : les Cloud Véhiculaires (VC), les VANET-Cloud (VuC) et les Cloud Véhiculaires Hybrides (HVC).

### 5.1 Les plateformes Cloud Véhiculaire (CV)

Dans les Cloud Véhiculaires (CV), les véhicules servant de prestataires de services et la formation de Cloud dépendent principalement de la mobilité des véhicules (par exemple, des Cloud statiques et des Cloud dynamiques), comme mentionné dans la figure 1.6.

Les véhicules servent de fournisseurs de services et des clients en même temps dans ce paradigme. Pour former le Cloud Véhiculaire, les véhicules lancent un protocole pour sélectionner entre eux des fournisseurs de services (Brokers). Ainsi, parmi ces entités candidats, les véhicules sélectionnent des entités (appelées des Entité Autorisées, EA) parmi les fournisseurs de services qui ont une forte confidentialité pour fournir des autorisations de formation du Cloud. Un fois la permission est lancée, les véhicules participants forment un environnement virtuel qui regroupe leurs ressources. Plus spécifiquement, le regroupement

des ressources est destiné essentiellement pour atteindre l'objectif de la création. Une fois la fonctionnalité du CV est satisfaite, le CV est détruit et les EA quittent le groupe.

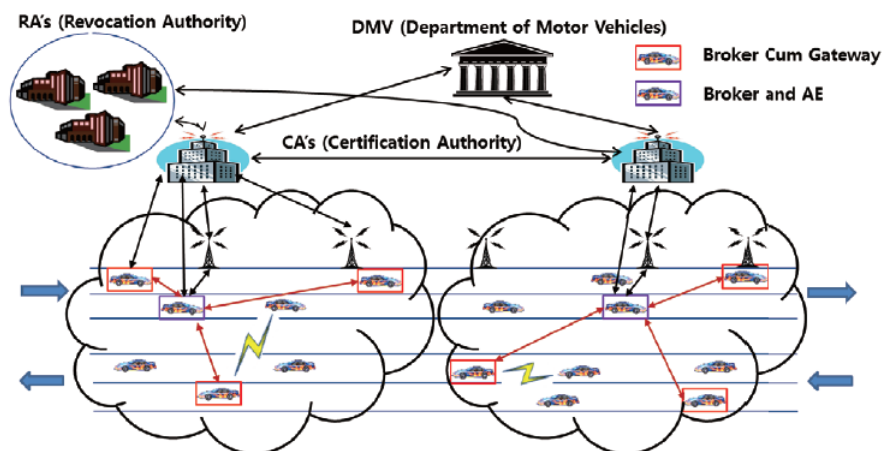


FIGURE 1.6 – Le Cloud Véhiculaire (Hussain et al. 2012).

## 5.2 Les plateformes VANET-Cloud (VuC)

Dans les véhicules utilisant des Cloud (VuC), voir la figure 1.7, nommés aussi modèle VANET-Cloud, les véhicules et les RSU(s) agissent comme des passerelles dans le réseau VANET pour échanger des services et assurer la communication avec les Cloud conventionnels. La communication entre les RSU et les services Cloud peut être établie par un câble de haute vitesse de transmission.

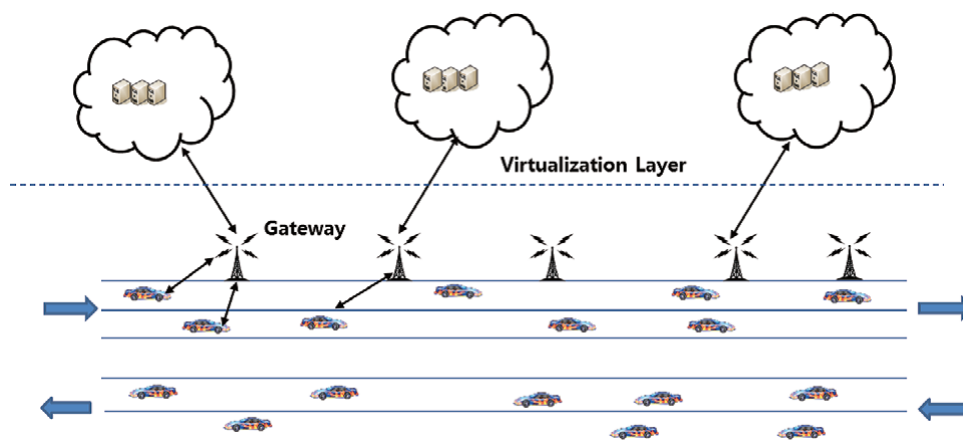


FIGURE 1.7 – Les VANET(s)-Cloud (Hussain et al. 2012).

## 5.3 Les plateformes Cloud Véhiculaires Hybrides (HVC)

Les plateformes Cloud Véhiculaires Hybrides se réfèrent à la combinaison entre les Cloud Véhiculaires (VC) et les véhicules utilisant les Cloud (HVC) où chaque véhicule agit en tant que fournisseur de services et client en même temps, (voir la figure 1.8). Les véhicules qui se déplacent sur la route peuvent offrir à la location leurs ressources et utiliser les services Cloud en même temps. Le réseau en tant que service (Network as a Service (NaaS)) et les réseaux Pair-à-Pair (P2P) sont les deux exemples les plus appropriés pour de tels scénarios. Néanmoins,

la forte mobilité des véhicules augmente la probabilité de réseaux de véhicules intermittents où un groupe de véhicules peut perdre les liens de connexion avec ses groupes voisins.

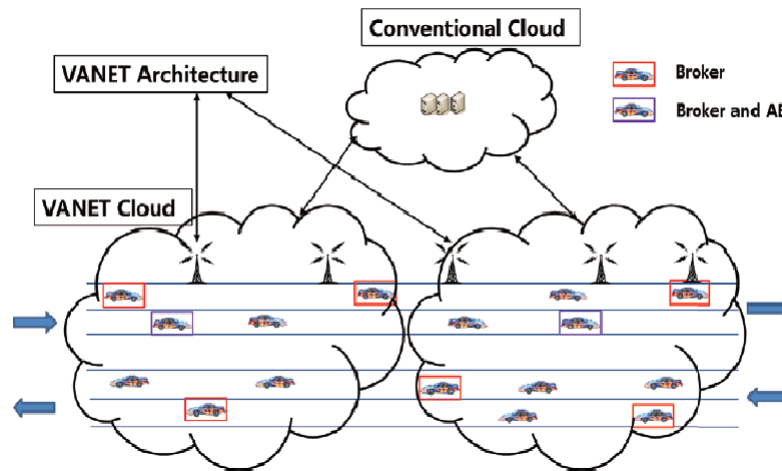


FIGURE 1.8 – Les Cloud véhiculaires hybrides (Hussain et al. 2012).

## 6 Les applications de Cloud Computing véhiculaire

Les plateformes CCV peuvent fournir plusieurs applications et informations de sécurité à d'autres conducteurs de véhicules concernant l'état réel du trafic, la notification, la prévention, etc. Ces applications de sécurité pourraient fournir des informations sur un environnement véhiculaire plus large par rapport aux informations fournies par un réseau de véhicules traditionnels ou par le Cloud traditionnel (Bitam et al., 2015). Les plateformes CCV peuvent prendre en charge plusieurs services englobant les applications suivantes :

### 6.1 Centre de données pour l'aéroport

Arif et al., (2012) proposent un scénario VC caractérisé par une disponibilité relativement stable et à long terme des ressources où, durant les jours ouvrables, les voitures des employés sont stationnées pendant des heures dans le parking. Ils ont été les premiers à mettre en œuvre l'aspect le plus simple d'un VC, à savoir l'agrégation des capacités informatiques des véhicules garés dans un environnement plutôt stable, le stationnement à long terme dans un aéroport international. Ainsi, ils ont traité ce problème en proposant le temps d'occupation dans le parking sur la base d'une architecture de CC open source nommé « Eucalyptus » (Nurmi et al., 2009). L'architecture proposée s'appuie sur une API coté client qui communique avec le contrôleur Cloud, le contrôleur des nœuds, plusieurs contrôleurs de cluster. L'extension de l'architecture est conçue pour assurer une intégration transparente et une gestion distribuée des ressources informatiques excédentaires des véhicules participants. En particulier, l'architecture est conçue pour adapter dynamiquement les ressources de véhicules dans le processus de gestion, allouer dynamiquement les ressources à une application en fonction des besoins des demandeurs définis à l'avance. Ainsi, les conditions de fonctionnement de l'architecture sont modifiables et sensibles au contexte dans lequel elles se trouvent. L'architecture de base du système proposé englobe les composants suivants :

- Des interfaces communicantes (des API coté client).



- Contrôleurs de centre de données qui joue le rôle d'un gestionnaire de Cloud avec une interface communicante avec le monde extérieur.
- Des contrôleurs de ressources dont le rôle de chacun est de gérer un groupe de ressources des véhicules. Ces contrôleurs ont des interfaces communicantes avec les groupes de véhicules et peuvent contenir des contrôleurs des tâches (ex., allocation, agrégation,..., etc.).

Une fois que les ressources excédentaires des véhicules sont prêtes à être utilisées, le centre de données de l'aéroport peut étendre ses ressources en ajoutant dynamiquement une extension des ressources du véhicule (c.-à-d., le parking).

## 6.2 La gestion des feux de signalisation

La planification des feux de circulation dynamiques est l'un des exemples les plus appropriés pour les Cloud dynamiques. Supposons un événement sportif auquel des milliers de personnes vont assister. L'évènement sera organisé dans un stade situé au centre de la ville. A la fin de l'évènement, les spectateurs essaient de partir le plutôt possible, leur déplacement à ce moment provoque un gros embouteillage. Le fonctionnement de feux de signalisation ne sera pas efficace et l'effet de changer un feu de circulation affecterait beaucoup d'autres. Après un certain temps. Le fonctionnement normal des feux aggravera encore la situation et la rendra difficile à résoudre, ce qui exigerait la réorganisation des feux à grande échelle. Pour éviter cette paralysie, les entités autorisées d'un environnement véhiculaire envoient un signal de reprogrammation des feux de signalisation au centre de gestion de trafic, par conséquent, les problèmes de circulation peuvent être résolus en temps opportun.

## 6.3 La gestion dynamique du flux du trafic

Comprendre le comportement du trafic en temps réel est un élément nécessaire pour optimiser le fonctionnement du réseau routier tout en réduisant ses effets négatifs. Pour éviter une telle paralysie sur le réseau routier, la mobilité des véhicules sur les routes principales et les routes secondaires doit être surveillée en temps réel. À l'heure actuelle, la surveillance du trafic repose essentiellement sur un large éventail de détecteurs fiables comme le détecteur inductif de boucle, les caméras, le système de suivi acoustique et les capteurs de radar à micro-ondes, etc. Parmi ces détecteurs, les boucles inductives sont les plus utilisées pour mesurer les données de trafic. Les détecteurs ILD sont intégrés dans l'infrastructure routière et peuvent être utilisés pour enregistrer le signal chaque fois qu'une voiture passe la boucle. Mais la mise en œuvre d'un système de surveillance basé sur les boucles inductives est trop coûteuse car chaque détecteur nécessite du matériel et des contrôleurs. Olariu et Weigle., (2009) estiment que chaque détecteur coûte environ 8 \$, relié à un câble de fibre optique qui coûte environ 300 \$ pour le kilomètre. De plus, ces statistiques montrent que plus de 50 % des détecteurs sont imprécis. Pour augmenter la précision des détecteurs, (Al Mamun et al., 2012) proposent une architecture de Cloud Véhiculaire pour surveiller le trafic du manière efficace et moins coûteuse. L'architecture proposée combine le concept de réseau de véhicules ad hoc (VANET) et le CC pour assurer un meilleur contrôle de la circulation, augmenter la sécurité sur la route et concevoir un système de gestion de trafic évolutive. Avec ce modèle de CV, la prolifération du réseau de capteurs dans la ville est utilisée comme une infrastructure de base pour surveiller le réseau du trafic

---

## 7 Les outils d'évaluation

La congestion routière est devenue un problème majeur dans les zones urbaines, où ses effets négatifs sur les heures de pointe peuvent avoir un impact sur les services de transport et services logistiques. L'avènement de la technologie de Cloud Computing véhiculaire promet d'atténuer la congestion du trafic en fournissant des solutions flexibles pour les usagers de la route, pour lesquels ces solutions sont accessibles via des applications reposant sur la connectivité internet.

Pour évaluer une telle solution, l'expérimentation d'un scénario réel dans les réseaux VANET est coûteuse, la simulation est aussi un outil essentiel pour prédire la performance des systèmes véhiculaires. En particulier, les modèles de simulation pour les systèmes Cloud-VANET justifient notre besoin dans ce travail où ils n'ont pas encore été définis. Les modèles de simulation des réseaux VANET sont conçus pour prédire l'état final du réseau en fonction de différentes mesures de performance, telles que la latence, la bande passante et le temps de réponse. Par exemple, OMNET++ (Varga, 2001), NS2 (Henderson et al, 2008) et OPNET (Chang, 1999) sont les outils de simulateur de réseaux les plus populaires pour les performances du réseau, telles que la communication, les protocoles de routage et la mobilité.

Compte tenu de leur utilisation répandue parmi le milieu de la recherche et du critère d'évaluation du simulateur, il existe plusieurs similitudes et différences entre ces simulateurs de réseau. Par exemple, ces simulateurs utilisent C++ comme une bibliothèque CLI et prennent en charge les fonctionnalités de la portabilité logicielle. Le NS2 et l'Omnet++ sont open source freeware, tandis que l'OPNET est livré avec une licence commerciale. Evidemment, le simulateur OPNET ne permet pas de modifier le code source des composants et les modèles de simulation VANET sont très limités. De plus, il est difficile d'observer la différence entre les simulateurs et les modèles de simulation simples. La distinction peut être explorée dans un grand scénario de simulation, où la complexité augmente, ce qui est le cas avec le nombre croissant de composants abstraits dans le modèle de simulation. De cette manière, les simulateurs de réseau peuvent être différenciés et appréciés. L'évolutivité du simulateur NS est très coûteuse, résultant du taux d'utilisation élevé des ressources machine, le mauvais modèle de mobilité (Mussa et al., 2015). L'absence de modèles de mobilité de simulation dans les couches physiques et MAC incite les chercheurs à apporter des améliorations en développant leurs propres composants, tels que l'extension de NS2 par la plateforme Miracle (Baldo et al., 2010) pour fournir une efficacité et un moteur intégré pour la manipulation de la couche transversale. Cependant, l'évolutivité de l'Omnet++ a un coût inférieur, car les topologies de réseau sont abstraites à la simulation modulaire via l'outil graphique (IDE). Cette modularité augmente l'efficacité de la simulation en faisant abstraction des réseaux complexes en composants uniques. Contrairement au simulateur NS2, le simulateur Omnet++ fournit des modèles de simulation pour les couches physiques et MAC. Par exemple, il offre INET (Varga, 2015) comme un cadre extensible pour la modélisation et la simulation des couches physiques et MAC des réseaux VANET. iCanCloud (Núñez et al., 2012) basé sur le plateforme INET pour modéliser et simuler des architectures évolutives dans les systèmes de cloud computing. Il comprend un ensemble de modules implémentés pour les protocoles réseau tels que GPRS / EGPRS (Todinca et al, 2004), ALOHA (Abramson, 1970), les couches OSI et les modèles TCP / IP. De manière simple, les développeurs peuvent intégrer leur proposition avec d'autres modules du Framework INET sans développer leurs propres modules.



---

## CONCLUSION

L'objectif principal de ce chapitre était de présenter le Cloud Computing Véhiculaire (CCV) en tant que nouveau paradigme de réseau lié à notre travail de recherche. Dans un premier temps, les notions de base des réseaux ad hoc de véhiculaires, ses standards de communication et ses architectures les plus populaires ont été brièvement présentés. Nous avons également discuté la convergence des applications VANET vers les applications Cloud Computing en présentant les plateformes générales dédiées aux réseaux Cloud Véhiculaires, les applications innovantes et les outils d'évaluation de performances.

Malgré la tendance croissante dans les applications des systèmes de transport intelligents, il existe encore de nombreux problèmes sans solutions précises comme la prévision exacte du flux de trafic. Dans ce chapitre, nous avons discuté la souplesse des applications CCV en fonction de leur impact sur les paramètres du trafic tel que le flux du trafic, le temps de déplacement et le retard. Nous mettons un accent particulier sur la plateforme VANET-Cloud car elle a été utilisée comme une architecture de référence de nos contributions. Le plus important est d'éviter de nombreux problèmes qui se produisent fréquemment dans les réseaux de trafic tel que les congestions et le blocage du trafic dans les routes.

Actuellement, augmenter l'efficacité du réseau de trafic est devenu une activité de recherche de première importance. Par ailleurs, en se basant sur les modèles de trafic, il s'agit de rationaliser le déploiement des infrastructures routières existantes, de limiter strictement les enjeux économiques et d'améliorer les modes de déplacement. En particulier, la prédiction de la circulation des véhicules devient un outil puissant pour accroître l'efficacité du réseau, réguler le flux du trafic et prédire son évolution future.

Le prochain chapitre sera consacré à la présentation d'outils permettant d'analyser le comportement du trafic et de modéliser efficacement les différents modèles de trafic.

# La dynamique non linéaire du trafic véhiculaire

## SOMMAIRE: CHAPITRE 2

Introduction	32
1 Les théories et les variables du flux de trafic	33
1.1 Les caractéristiques fondamentales du flux de trafic	33
1.2 La modélisation du flux de trafic	35
1.3 Le diagramme fondamental du trafic	36
2 Les techniques d'agrégation des données du trafic	38
2.1 L'agrégation complémentaire	38
2.2 L'agrégation redondante	39
2.3 L'agrégation coopérative	40
3 Les fonctions d'agrégations	40
3.1 Les fonctions linéaires	40
3.2 Les fonctions non-linéaires	42
4 L'apprentissages automatique supervisé	42
4.1 Définition formel	43
4.2 Le problème d'apprentissage supervisé	43
4.3 Les fonctions de perte	43
5 Les modèles d'apprentissages pour la prédiction du flux de trafic	44
6 Discussion	45
Conclusion	47



---

## INTRODUCTION

Nous avons présenté dans le chapitre précédent la dynamique du trafic véhiculaire et leurs comportements complexes. Pour l'analyse du comportement des grands embouteillages, il est crucial de mettre en évidence les effets des incidents routiers et prévoir l'évolution du trafic en prenant en compte deux dimensions spatiotemporelles. La dynamique est décrite en fonction de plusieurs variables telles que le débit, la densité et la vitesse. À cet égard, une approche basée sur l'analyse non linéaire et les théories du trafic est utilisée pour décrire la dynamique du système qui est trop complexe avec les méthodes d'analyse conventionnelles.

La prédiction précise de l'état actuel du trafic devient une piste de recherche importante pour optimiser le problème de la congestion routière. Une telle prédiction permet de tracer les changements liés à son état dans le futur. Ceci est possible lorsque nous analysons à l'avance les données collectées en temps réel par les réseaux de capteurs, les données détectés et échangés entre les véhicules, les données historiques, etc.

L'apprentissage automatique est souvent appliqué avec succès pour résoudre le problème de la congestion quotidienne et l'avenir de la sécurité routière. L'objectif de l'apprentissage est de construire un ensemble d'algorithmes raisonnant à partir des données empiriques pour produire des hypothèses générales. Ces hypothèses accomplissent ensuite des prédictions sur les instances futures (Chen et al, 2017). Les applications innovantes de trafic utilisent des modèles prédictifs pour généraliser les données au lieu des calculs statiques basés sur les données brutes.

Dans ce chapitre, nous allons mettre en évidence les théories de trafic qui permettent d'analyser le flux de trafic, ses caractéristiques, ses niveaux d'écoulement et ses diagrammes. Ensuite, nous allons présenter les approches d'agrégation de données dans les systèmes de trafic linéaire et non-linéaire. Enfin, après avoir examiné les différents modèles de trafic existants pour une prédiction précise du flux de trafic, un type de relation monotone entre les variables de trafic est présenté.

# 1 Les théories et les variables du flux de trafic

Dans les disciplines d'ingénierie du trafic, le flux des véhicules est l'étude des interactions entre les utilisateurs de la route et les dispositifs de contrôle de la circulation. Cette étude s'inscrit dans le cadre de développement des plans de transport à long terme qui garantissent des mouvements efficaces et résolvent les problèmes de la congestion quotidienne. L'ensemble des théories du flux de trafic englobe principalement des méthodes d'analyse du comportement, la mise en œuvre des outils et des techniques de contrôle de fluidité et l'élimination des incidents de la route. Ces théories présentent une partie du transport qui concerne l'évaluation de la qualité opérationnelle du trafic en tenant compte de toutes les conditions préalables telles que la conception du réseau routier (y compris les autoroutes, le pourcentage de poids lourds, etc.) (Elefteriadou, 2014). La fluidité est associée à plusieurs paramètres de trafic pour expliquer sa nature et aider les analystes à détecter toute variation qui caractérise l'écoulement. En outre, la compréhension précise du comportement nécessite une étude approfondie de ces paramètres et de leurs relations mutuelles.

## 1.1 Les caractéristiques fondamentales du flux de trafic

### La Vitesse moyenne ( $v$ )

C'est la vitesse qui indique comment un véhicule s'est déplacé, c.-à-d., la distance parcourue par unité de temps. Dans la pratique, la vitesse moyenne est mesurée en fonction de la distribution des véhicules dans un segment routier donné dans une période de temps définie. La vitesse moyenne peut être déterminée de plusieurs façons, parmi lesquelles deux définitions de la vitesse sont présentées :

- La *vitesse moyenne* est la vitesse des véhicules qui franchissent la longueur d'un segment routier pendant une période ( $t$ ). On parle parfois de distance par rapport au temps. La moyenne pondérée de ces vitesses, nommé aussi la vitesse moyenne spot, indiquée avec l'index  $m$ , en référence à la moyenne, est calculée comme suit :

$$v_m = \frac{1}{n} \sum_{i=1}^n v_i \quad (2.1)$$

En fait, la vitesse moyenne des véhicules est une quantité scalaire généralement liée à la distance parcourue, ce qui signifie qu'il n'est pas obligatoire de la représenter par la grandeur et la direction du déplacement.

- La *vitesse instantanée* est le taux de mouvement d'un véhicule à un instant donné. La vitesse instantanée d'un véhicule peut être calculée en fonction de la « *vitesse moyenne* » sur une distance courte, et aussi un instant court, en utilisant cette approximation comme vitesse instantanée. Pour déterminer la vitesse instantanée à tout moment, il est important d'utiliser cette formule :

$$v_{SI} = \lim_{T \rightarrow t} \frac{dX}{dt} \quad (2.2)$$

La figure 2.1 ci-dessous, montre la différence entre la vitesse moyenne et la vitesse instantanée. La vitesse moyenne indique la distance parcourue (entre A et D) par un véhicule pendant un intervalle de temps, mais elle ne couvre pas le mouvement réel qui s'est produit. Tandis que la vitesse instantanée est la vitesse exacte au point B, à un instant donné.

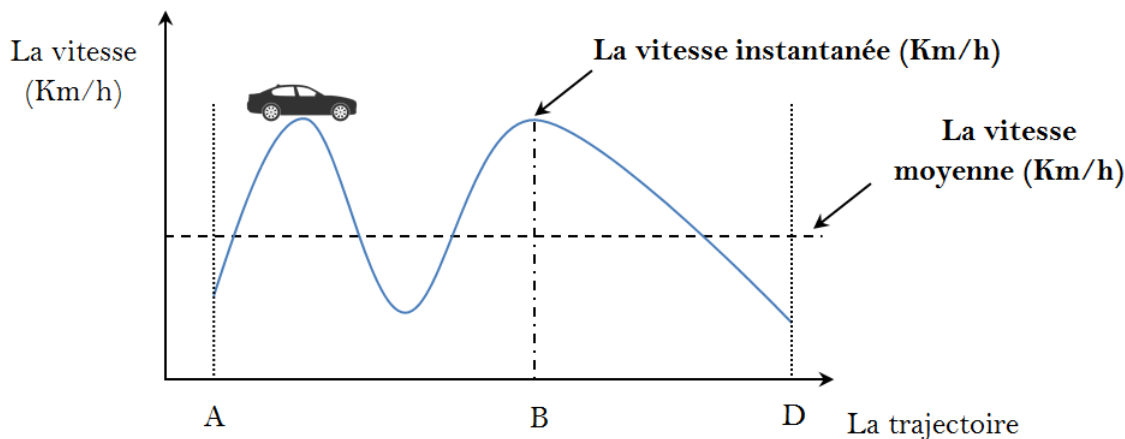


FIGURE 2.1 – Vitesse moyenne vs vitesse instantané.

**La densité ( $\rho$ )**

La densité du véhicule ( $\rho$ ) présente le nombre de véhicules par unité d'espace (noté aussi par Véhicule/km). Comme illustre la figure 2.2, la densité est le nombre des véhicules entre les points A et B diviser par la distance globale entre A et B.

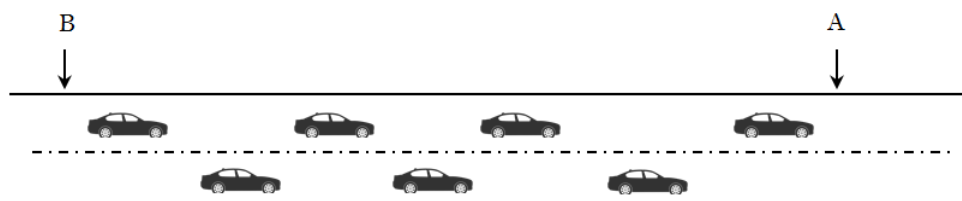


FIGURE 2.2 – la densité ( $\rho$ ).

Néanmoins, la densité est maximale entre A et B, ce qui implique que les véhicules individuels sont très proches les uns des autres, tandis que la densité minimale implique des grandes distances entre les véhicules. La densité peut être approximativement estimée en se basant sur l'équation suivante :

$$\rho = \frac{n}{L} \tag{2.3}$$

Où,  $n$  présente le nombre de véhicules dans un segment routier avec une longueur ( $L$ ).  $\rho$  peut-être aussi estimé à partir de la relation suivante :

$$\rho = \frac{q}{v} \tag{2.4}$$

Où, ( $q$ ) présente le nombre de véhicules par heure et  $v$  la vitesse moyenne du trafic. Cependant, il convient de noter que la densité du véhicule ( $\rho$ ) est liée aux véhicules sur une section routière d'une longueur donnée tandis que la vitesse du véhicule est calculée en fonction de l'intervalle de temps entre A et B.

## Le volume de trafic ( $q$ )

Le volume de trafic est simplement le nombre total de véhicules passant un point donné sur la route dans une période de temps spécifiée. Le volume de trafic est exprimé en nombre de véhicules par unité de temps (heure). Par exemple, la boucle inductive compte le nombre de véhicules qui passent un point sur la route pendant une heure qui est d'environ 2000 véhicules, (noté aussi 2000 véhicules / heure). Le volume de trafic est généralement converti directement en flux ( $q$ ), ce qui est un paramètre plus utile. Le comportement du flux de trafic ( $q$ ) est variable avec le temps, c'est-à-dire, le volume peut être estimé en des intervalles à court terme (ne dépassent pas quelques minutes ou quelques heures), quotidienne (de jour en jour), mensuel (du mois au mois), annuel (d'une année à une autre). Les variations peuvent également être observées d'une saison à l'autre. Pour mieux comprendre le comportement du flux de trafic, les relations entre le flux, la densité ( $\rho$ ) et la vitesse ( $V$ ) doivent être établies. Formellement, un modèle de flux de trafic générique peut être représenté comme suit :

$$q = \rho \cdot v \quad (2.5)$$

Le flux de trafic peut être calculé en fonction du temps d'avance ( $h$  : headway en anglais), comme présenté dans l'équation suivante :

$$q = 1/h \quad (2.6)$$

## 1.2 La modélisation du flux de trafic

Comme nous l'avons vu précédemment, la dynamique du trafic est au cœur de nombreux domaines de transport. Le plus impressionnant est que si le flux de trafic dépasse un nombre critique, la performance du trafic diminue avec la croissance de celui-ci. La modélisation du trafic d'un espace urbain composé de plusieurs routes devient un outil très puissant en théories du trafic. L'objectif de la modélisation est de simplifier la complexité de l'écoulement de trafic et prévoir son évolution.

Les théories des flux de trafic visent à démontrer mathématiquement la précision des interactions entre l'environnement des véhicules et l'infrastructure routière. En outre, étudier le comportement du flux de trafic représente une étape primordiale pour construire des modèles et des outils permettant d'optimiser l'exploitation de l'infrastructure. En règle générale, ces théories reposent sur trois niveaux d'écoulement (ou niveaux d'agrégation) qui facilitent l'analyse du flux de trafic. Ces niveaux sont classés en trois grandes catégories : niveau macroscopique, niveau microscopique et niveau mésoscopique.

Au niveau macroscopique, la dynamique individuelle des véhicules est agrégée en tenant en compte les relations entre les caractéristiques globales du flux (Francesco et al., 2015), telles que la densité, le débit, la vitesse moyenne, etc. A ce niveau, deux classes principales de modèles macroscopiques (Möller, 2016) peuvent être identifiées comme suit :

- Le *modèle discret de l'espace* décrit la propagation des flux à travers un lien dont les variables de trafic sont définies à chaque point de l'espace.
- Le *modèle continu de l'espace* dérive de l'analogie entre l'écoulement des véhicules et le flux continu mesuré par un support physique (ex., les cameras), dont les variables de trafic de base affectent la performance du lien, telles que le temps de déplacement, la densité ou la vitesse.

Contrairement aux modèles macroscopiques, les modèles microscopiques se concentrent sur les véhicules individuels et leur interaction avec les véhicules voisins, le changement de voie, le comportement aux points de fusion, de sorte que les variables dynamiques décrivent des phénomènes potentiels tels que la mobilité, la position et la vitesse des véhicules individuels. Par conséquent, le modèle fournit des informations sur lesquelles les modèles macroscopiques ne peuvent pas se reproduire (Möller, 2014). Les modèles microscopiques génèrent séparément pour chaque véhicule des informations individuelles, comme l'accélération et le freinage, et peuvent également inclure des nouvelles stratégies pour contrôler le routage des véhicules.

De plus, les modèles mésoscopiques modélisent le trafic comme des paquets de véhicules ou même des véhicules individuels, c.-à-d., au niveau de détail d'un seul véhicule mais toujours régis par les lois d'écoulement macroscopiques (Mohan et Ramadurai., 2013). Ce modèle offre une large gamme d'adaptabilité pour répondre aux demandes d'informations, aux informations disponibles et aux objectifs de simulation.

En fait, le déploiement de ces types de modèles dans le domaine des transports permet d'analyser efficacement les différents scénarios qui correspondent aux exigences essentielles de la croissance des transports. Ces modèles permettent également de choisir les solutions optimales pour la gestion, la planification, l'utilisation optimale des segments routiers et la réduction des coûts de développement pour l'infrastructure.

### 1.3 Le diagramme fondamental du trafic

Au cours des cinquante dernières années, les théories de trafic ont été proposées pour procéder à une étude approfondie sur des données empiriques. Ces théories cherchent à comprendre les relations fondamentales entre les paramètres du trafic permettant de distinguer les différentes phases de flux. Selon des théories antérieures, Reuben Smeed (1968) a modélisé le débit de trafic qui peut entrer dans une zone centrale de la ville en fonction de la superficie de la ville, de la fraction de la surface consacrée aux routes et de la capacité exprimée en véhicules par unité de temps et par unité d'espace. Malgré une réelle et forte intention à résoudre efficacement le problème de la congestion dans les villes, cette théorie ne juge pas ce qui arrive à la vitesse du trafic et du temps de déplacement lorsque la demande dépasse la capacité pendant les heures de pointe. L'idée d'explorer les relations entre la vitesse de trafic (kilomètres /heure) dépend du débit de trafic (véhicule /heure) en heure de pointe a été introduite par (Godfrey, 1969) et (McClenahan et Simkowitz, 1969). De telles relations sont importantes pour aborder les limitations des méthodes d'équilibrage du flux du trafic dans le réseau. Ainsi, les fonctions utilisées pour associer la vitesse de trafic, le temps de déplacement par unité d'espace et le flux de trafic sont souvent appelés « *fonction de capacité de lien* » (Branston, 1976). Le terme « *fonction de capacité de lien* » est utilisé pour impliquer une forme fonctionnelle qui relie la vitesse et le débit du trafic. L'ensemble des formes fonctionnelles est utilisé par la suite pour illustrer la dépendance entre les variables de trafic dans un diagramme fondamental. Ce diagramme s'appelle aussi modèle macroscopique de trafic. Mahmassani et al, (1984) étaient parmi les premiers qui ont introduit le concept. Ainsi, ce modèle peut être utilisé pour évaluer la capacité de l'infrastructure routière et contrôler le comportement du trafic dans le cas d'un processus d'équilibrage. Le modèle macroscopique de ces théories ne possède que la phase libre et la phase de congestion.

Boris Kerner, dans ces travaux (Kerner, 1998;1999), propose une théorie de trafic triphasé portant une phase libre et deux phases de congestion. Les trois phases peuvent être déterminées explicitement par le modèle macroscopique en examinant la linéarité entre les variables de trafic. Par conséquent, sur la base de ce principe, un Diagramme Fondamental Macroscopique (MFD) a été évalué pour la première fois à Yokohama (Japon) en tant que modèle de base fondé sur des données de trafic empiriques. Le modèle explique, via les nuages de points, la



dépendance entre trois variables de trafic qui sont : le flux de trafic, la vitesse et la densité. Dans la figure 2.3, deux nuages de points sont formés en se basant sur leurs distributions. Celui qui apparaît comme une pente linéaire positive exprime la phase libre et celui qui apparaît comme une distribution non-linéaire exprime la phase de congestion.

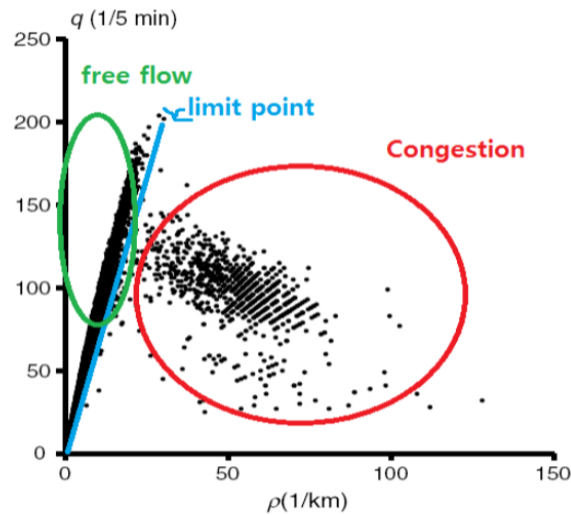


FIGURE 2.3 – Diagramme fondamental (modèle macroscopique). Cas d'étude dans une autoroute japonaise.

- **La phase libre** : le nuage de points qui exprime cette phase montre une corrélation positive entre la densité  $\rho$  (ou accumulation) et le flux de trafic  $q$ . De plus, la relation entre eux est proportionnelle jusqu'à ce que l'accumulation atteigne un seuil maximal. Une fois que l'accumulation de trafic atteint le seuil maximal, le débit de trafic aboutit également à son seuil maximal, ce débit maximum est appelé flux de capacité dénoté par  $q_{max}$  ou  $q_c$ . Là, la relation s'arrête et il est fort possible que l'état de trafic passe de l'état libre à l'état de congestion. On l'occurrence, les seuils maximaux dans le diagramme sont appelés des seuils critiques.
- **La phase capacité (synchronisé)** : cette phase est la transition du flux de trafic de la phase libre vers la première étape de la congestion où les véhicules se rapprochent les uns des autres dans l'autoroute. On l'occurrence, les variables de trafic montrent des relations faibles et la vitesse de celui-ci diminue de manière significative, mais il n'y a pas de changement remarquable dans le débit observé (Kerner, 2004). Ceci est possible lorsque la densité atteint le seuil maximum, dénoté par  $\rho_{max}$  ou  $\rho_c$ , de sorte que le produit de la vitesse de trafic et de la densité reste à peu près le même.
- **La phase congestionnée** : est une condition résultant des conditions de circulation saturées ou embouteillées. La congestion routière se produit lorsque la capacité routière devient insuffisante pour équilibrer la circulation. Cette phase est également caractérisée par l'apparition de notions de retard et de goulots d'étranglement. Dans ce cas, les conducteurs doivent être conscients, patients et se comporter calmement et correctement pendant le trafic.

## 2 Les techniques d'agrégation des données du trafic

Le déploiement et l'utilisation de la technologie des terminaux intelligents qui surveillent nos routes et détectent les données dynamiques de trafic peuvent aider à interpréter le comportement du trafic. Dans les systèmes de trafic intelligents conventionnels, certains capteurs de surveillance sont utilisés généralement en tant que source unique pour la collecte des informations de trafic. Cependant, en raison de la limitation de ces capteurs, certains challenges liés aux données de trafic apparaissent, tels que l'imperfection et l'exactitude. Développer des solutions en se basant sur la coopération entre les différentes entités telles que les véhicules, l'infrastructure routière et mieux exploiter les informations disponibles via les réseaux de capteurs déployés en tant qu'infrastructure devient crucial (Artuñedo et al., 2016).

En général, les données collectées à partir de sources multiples nécessitent un type de fonction pour identifier les corrélations entre les paramètres. L'agrégation de données est une fonction. Dans la littérature, les termes agrégation, fusion, intégration et combinaison de données peuvent généralement être utilisés comme synonymes. Des données qui se prêtent à l'agrégation sont destinées pour comprendre facilement les différentes conditions de circulation en assurant une meilleure gestion des réseaux routiers. De plus, l'information produite par la fonction d'agrégation est fondamentale pour la prise de décision.

Il existe plusieurs définitions pour l'agrégation de données :

Franklin (1991) a défini l'agrégation de données comme suit : « *un processus multi-niveaux traitant de l'association, de la corrélation, de la combinaison de données et d'informations provenant de sources uniques et multiples pour obtenir une position affinée, identifier des estimations et des évaluations complètes et opportunes des situations, des menaces et de leur signification* ».

Stanley et Steinhardt (2014) ont définie l'agrégation de données comme suit : « *L'agrégation de données est la collecte de données provenant de bases de données dans l'intention de préparer des ensembles de données combinés pour le traitement des données* ».

Selon Gary Palgon (2017), la définition de l'agrégation de données est : « *un processus au cours duquel les données sont recherchées, collectées et présentées dans un format réduit basé sur un rapport pour atteindre des objectifs des processus métier spécifiques et / ou effectuer une analyse humaine* ».

L'agrégation de données est la première étape de la génération de données. En effet, l'analyse de grands ensembles de données en temps réel, collectées à partir de sources multiples, nécessite des plates-formes puissantes telles que le Cloud Computing. Les techniques d'agrégation de données basées sur les relations entre les sources de données peuvent être classées en trois différents schémas de classification : agrégation complémentaire, redondante et coopérative (Abdelgawad et Bayoumi, 2012).

### 2.1 L'agrégation complémentaire

L'agrégation complémentaire peut être effectuée lorsque les nœuds sources obtiennent des données différentes qui représentent une partie de la scène, (voir la figure 2.4). Ces données peuvent être agrégées pour compléter la scène et avoir des résultats qui n'ont pas pu être obtenus à partir des données préalables. Par exemple, la vitesse des véhicules et la densité des véhicules d'un segment routier peuvent être agrégés pour obtenir le flux de trafic.

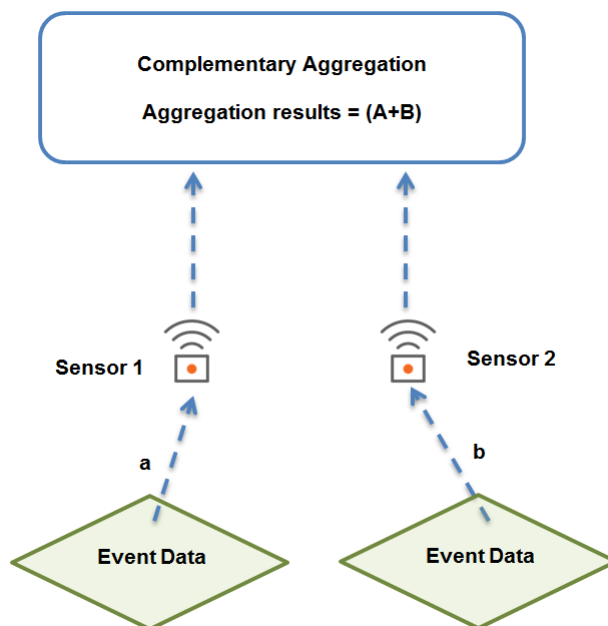


FIGURE 2.4 – Agrégation complémentaire.

## 2.2 L'agrégation redondante

Cette approche permet de garantir la fiabilité des données détectées par les capteurs dans le cas où deux capteurs sources partagent la même fonctionnalité, (voir la figure 2.5). En d'autre terme, les capteurs sont destinés à capturer le même type information et les informations détectées peuvent être redondantes. L'approche d'agrégation redondante va fusionner les données redondantes de sorte que les résultats de la fusion vont augmenter la qualité de données. Par exemple, un ensemble de capteurs de trafic est installé pour surveiller un segment routier. L'objectif est de détecter la même donnée de trafic comme la vitesse de véhicules. Les données de trafic dans ce cas sont considérées comme redondantes. L'approche d'agrégation redondante va fusionner ces données pour capturer une information précise sur le trafic actuel.

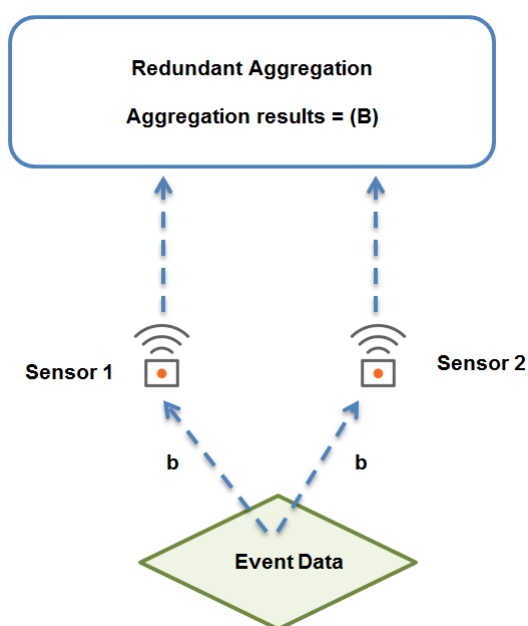


FIGURE 2.5 – L'agrégation redondante.

## 2.3 L'agrégation coopérative

Cette approche est appropriée pour combiner des informations ayant différentes valeurs fournies par des sources indépendantes (voir la figure 2.6). L'information produite sera de plus en plus complexe que l'information originale. Par exemple, la fusion de données multimodales (audio et vidéo) est considérée comme coopérative. Ce type d'agrégation est adapté aux nombreuses applications dans le traitement d'images (Sonka et al, 2014), les systèmes de récupération d'information (Wu et Crestani, 2015) et les systèmes décisionnels.

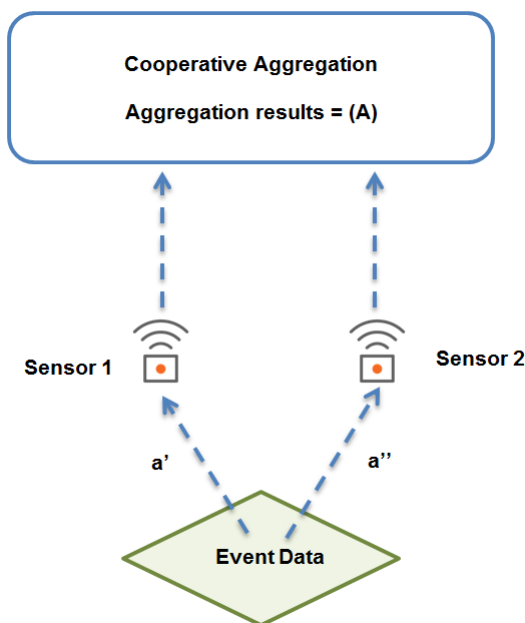


FIGURE 2.6 – L'agrégation coopérative.

## 3 Les fonctions d'agrégations

Plus généralement, les fonctions d'agrégation sont des fonctions mathématiques dont le but est de réduire un ensemble d'entrées en une seule sortie représentative. L'analyse des propriétés des données est un aspect important de l'étude des fonctions d'agrégation (Zimmermann 2011, Herrera et al., 1996). Dans ce qui suit, nous présentons deux classes de fonctions d'agrégation, qui sont :

### 3.1 Les fonctions linéaires

Les fonctions d'agrégation linéaire sont généralement connues dans les littératures par des fonctions additives ou classiques (Li, 2010; Li et Hang, 2010). L'opérateur de moyenne pondérée (OWA) est un exemple. Bien que nous nous concentrons dans cette thèse sur les fonctions d'agrégation non-linéaire, nous présenterons un bref aperçu des autres types d'agrégation linéaire et nous donnerons plusieurs définitions applicables aux fonctions d'agrégation en général.

**Définition 1 :** soit  $I \subseteq \mathbb{R}$ ,  $n \subseteq \mathbb{N}$  et  $\{x_1, x_2, \dots, x_n\} \in I^n$ .  $F$  est une fonction d'agrégation avec  $n$  variables telles que :  $F : I^n \rightarrow I$

- $F$  est non décroissant : pour toute  $x, x^* \in I^n$

$$x \leq x^* \Rightarrow F(x) \leq F(x^*)$$

- $F$  remplit les conditions aux limites :

$$\inf_{x \in I^n} F(x) = \inf I \text{ et } \sup_{x \in I^n} F(x) = \sup I$$

- $F^{(1)}(x) = x$  pour chaque  $x \in I$

Pour une fonction d'agrégation  $F : [0, 1]^n \rightarrow [0, 1]$ , les propriétés suivantes sont équivalentes :

$$(i) F(0, 0, \dots, 0) = 0, F(1, 1, \dots, 1) = 1$$

$$(ii) F(a_1, a_2, \dots, a_n) = a \text{ pour chaque } a \in [0, 1]$$

$$(iii) \min(x_1, x_2, \dots, x_n) \leq F(x_1, x_2, \dots, x_n) \leq \max(x_1, x_2, \dots, x_n)$$

**Définition 2 :** une fonction d'agrégation est une application étendue  $F : \bigcup_{n \geq 1} [0, 1]^n \rightarrow [0, 1]$  de sorte que la restriction de cette application à chaque  $[0, 1]^n$  est une fonction d'agrégation  $n$ -aire. L'entier  $n$  représente l'arité de la fonction d'agrégation, c'est-à-dire le nombre de ses arguments. Dans ce que suite, nous nous rappelons plusieurs exemples de la fonction d'agrégation étendue sur  $I$  :

**La moyenne arithmétique pondéré (WAM) :** soit un vecteur  $w$ , la moyenne arithmétique pondérée par  $w$  est la fonction :

$$WAM(x_1, x_2, \dots, x_n) = \sum_{i=1}^n w_i x_i$$

**La moyenne arithmétique pondérée ordonnée (OWA) :** l'opérateur OWA est introduit par Yager (2004) comme une bonne formule mathématique utilisée dans nombreuses applications statistiques. Il permet de déterminer le moyen d'un ensemble de nombres finis. La moyenne arithmétique pondérée ordonnée par un vecteur  $w$  est la fonction :

$$OWA(x_1, x_2, \dots, x_n) = \sum_{i=1}^n w_i x_{(i)}$$

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$  indique un vecteur obtenu à partir  $(x_1, x_2, \dots, x_n)$  en arrangeant ses arguments dans l'ordre décroissant  $(x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)})$

**Le minimum (Min) :**  $Min(x_1, x_2, \dots, x_n) = \bigwedge_{i=1}^n x_{(i)}$

**Le maximum (Max) :**  $Max(x_1, x_2, \dots, x_n) = \bigvee_{i=1}^n x_{(i)}$

**La somme ( $\Sigma(x)$ ) :**  $\Sigma(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_{(i)}$

**Le produit ( $\Pi(x)$ ) :**  $\Pi(x_1, x_2, \dots, x_n) = \prod_{i=1}^n x_{(i)}$

### 3.2 Les fonctions non-linéaires

Les fonctions d'agrégation non-linéaire sont généralement connues par des fonctions non-additives. Cette classe présente un outil très puissant d'un point de vue décisionnel. Dans les problèmes de prise de décision multi-attributs, une décision doit être choisie sur la base de plusieurs hypothèses généralement contradictoires. Les décisions sont évaluées en utilisant des attributs caractéristiques qui peuvent être exprimés numériquement (Beliakov et al, 2016).

Lors de l'achat d'une voiture, le choix d'un client dépend des attributs caractéristiques suivantes : le prix, la qualité, la consommation de carburant, la capacité de la voiture, la puissance, la marque, etc. Pour choisir une solution parmi un ensemble de solutions, les valeurs de ces attributs doivent être combinées. L'utilisation de l'intégrale de Choquet dans ce contexte en tant que fonction d'agrégation non linéaire peut aider les clients à trouver les meilleures voitures correspondant à leurs préférences. Cette fonction permet de modéliser les interactions entre plusieurs préférences afin de retourner les meilleurs choix. La puissance de cette fonction est la mesure floue  $\mu$ . Dont  $\mu$  est une application monotone sur un ensemble  $X$ . Supposons également que le prix de la voiture est fixé par un client. Évidemment, une sorte de dépendance monotone existe entre les attributs cités et le prix de la voiture.

La dépendance monotone entre les mesures observées et leurs réponses est au cœur de nombreuses applications et en particulier dans les applications d'apprentissage automatique. Il y a des raisons importantes pour lesquelles cette dépendance devrait être utilisée dans le processus d'apprentissage (Kotlowski and Slowinski, 2016). Premièrement, les caractéristiques monotones imposent des contraintes sur le processus de prédiction. Ces contraintes permettent de minimiser la complexité du modèle de prévision en réduisant l'espace des hypothèses, le risque structurel et en augmentant la précision de la prédiction. Deuxièmement, les connaissances des experts du domaine sont souvent intégrées dans l'évaluation du modèle de prédiction. Cela permet d'identifier les contradictions dans le modèle qui peuvent aider les experts à décider s'ils acceptent ou non le modèle, quelle que soit leur exactitude.

**Définition 3** : une fonction  $F : \bigcup_{n \geq 1} [0, 1]^n \rightarrow [0, 1]$  est monotone (croissante) si et seulement si :

$$\forall x, y \in (x_1, x_2, \dots, x_n) x \leq y \text{ alors } F(x) \leq F(y)$$

Nous utiliserons fréquemment la dominance  $x \leq y$ , ce qui signifie que, dans tous les cas, chaque valeur de  $x$  n'est pas plus grande que la valeur de sa correspondance  $y$ . La monotonie croissante peut être exprimée comme,  $x \leq y \Rightarrow F(x) \leq F(y)$ . En général, la monotonie dans tous les arguments et les conditions aux limites sont deux caractéristiques fondamentales qui caractérisent les fonctions d'agrégation (Beliakov et Sola, 2016). Nous ne pouvons pas considérer la fonction  $F : \bigcup_{n \geq 1} [0, 1]^n \rightarrow [0, 1]$  comme une fonction d'agrégation si l'une de ces caractéristiques précédentes n'est pas réalisable, car elle produira une sortie incohérente lorsqu'elle est utilisée.

## 4 L'apprentissages automatique supervisé

Il existe plusieurs types de tâches d'apprentissage automatique. Dans cette thèse, nous nous concentrerons sur un type particulier, à savoir l'apprentissage d'une fonction prédictive qui peut être utilisée pour prédire des valeurs d'une classe continue. Comme cela a été déjà mentionné, la monotonie décrit une sorte de dépendance entre les mesures observées et celles attendues. Ainsi, il est lié à un problème d'apprentissage supervisé (AF Tehrani, 2014). L'apprentissage automatique peut être révolu de manière supervisée ou non supervisée. L'apprentissage supervisé, appelé aussi classification ou apprentissage inductif, est la tâche

d'apprentissage automatique consistant à déduire une fonction de prédiction à partir de données préalables (données d'entraînement sauvegardées dans une base de données) (Mohri et al, 2012). Un apprentissage supervisé permet d'analyser les données préalables et produire une fonction de prédiction qui peut être utilisée pour traiter de nouvelles instances, acquérir des nouvelles connaissances et généralement unifier l'intelligence.

## 4.1 Définition formel

Soit  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  un ensemble de données d'apprentissage, chaque modèle exemplaire dans  $D$  est un couples d'entrée/sortie  $(x_n, y_n)$  avec  $n \in \{0, 1, \dots, N\}$ ,  $x_n \in X$  et  $y_n \in Y$ .  $X = \{x_1, x_2, \dots, x_{|m|}\}$  typiquement est un vecteur des attributs, où  $|m|$  indique le nombre d'attributs ou la taille de l'ensemble  $X$ .  $Y = \{y_1, y_2, \dots, y_{|z|}\}$  est l'ensemble des classes où les attributs d'une classe  $y_i$  est un ensemble des valeurs discrètes avec  $i \in \{0, 1, \dots, z\}$ . Dans cette thèse, nous considérons  $Y$  est l'ensemble des valeurs continue qui représente la sortie désirée.

Comme cela a déjà été mentionné, le but de l'apprentissage est de produire une fonction de classification / prédiction pour associer les valeurs des attributs de  $X$  aux classes dans  $Y$ . La fonction est appelée un « *modèle de classification* » si on cherche à regrouper un ensemble d'attributs en des classes particulières. D'un autre côté, nous pouvons appeler cette fonction un « *régresseur* » si nous voulons prédire une valeur dans un ensemble continu.

## 4.2 Le problème d'apprentissage supervisé

Soit  $(x_1, x_2, \dots, x_n)$  un ensemble de prédicteurs,  $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n$  est son domaine.  $Y \in \mathcal{Y}$  variable aléatoire appelée réponse (output). La tâche de la régression et de la classification consiste à prédire  $Y$  en fonction de  $X$ , c'est-à-dire :

$$r(x) = E(Y | X = x) = \int yp(y | x)dx$$

Si  $y \in \mathbb{R}$  alors la tâche d'apprentissage est la régression, c.-à-d., à partir d'un ensemble de données, la tâche consiste à trouver la meilleure relation qui représente l'ensemble de données.

Si  $y = \{1, \dots, I\}$  est catégorique (i.e, nominal) alors la tâche d'apprentissage est la classification, c.-à-d., étant donné une relation connue, l'identification de la classe à laquelle les données appartiennent est un ensemble fini des valeurs de sortie.

**Définition 3 :** Le problème de classification binaire est défini par un ensemble de données  $D$  sur  $\mathcal{X} \times \mathcal{Y}$ , où  $Y = \{0, 1\}$ . Le but est de trouver un classificateur  $S : \mathcal{X} \rightarrow \mathcal{Y}$  en minimisant le taux d'erreur sur  $D$ .

$$r(S, (x, y)) = Pr_{(x,y) \in D}[S(x) \neq y]$$

**Définition 4 :** Un problème de régression est défini par un ensemble de données  $D$  sur  $X \times \mathbb{R}$ . Le but est de trouver une fonction  $f : X \rightarrow \mathbb{R}$  en minimisant la perte au carré suivant :

$$pt(S, (x, y)) = E_{(x,y) \in D}(f(x) - y)^2$$

## 4.3 Les fonctions de perte

Dans l'apprentissage automatique, à partir des données d'entraînement, le but est de déduire un modèle de classification / prédiction qui convient le mieux au modèle réel.

Cependant, un algorithme d'apprentissage essaye d'atteindre un écart-type minimum entre les observations et les données attendues. En fait, nous souhaitons tester la performance d'un tel modèle de classification / prédiction en quantifiant la perte (i.e., l'erreur) de la prédiction qui devient un problème d'optimisation. La fonction de perte, ou la fonction de risque, est une fonction permettant de mesurer le degré d'ajustement des modèles de classification/prédiction, dont les mesures en sortie montrent la différence entre les valeurs observées des données et les valeurs calculées.

- Une **fonction de perte binaire** (binary loss function) est définie comme suite :

$$L(y, f(x)) = \begin{cases} 0 & y = f(x) \\ 1 & \text{sinon} \end{cases}$$

L'espérance mathématique de cette mesure se définit comme :

$$E(t) = \begin{cases} \sum_{x \in \mathcal{X}} L(y, f(x)) P(x, y) & \text{Dans le cas de classification} \\ \int_{\mathcal{X} \times \mathcal{Y}} L(y, f(x)) dP(x, y) & \text{Dans le cas de régression} \end{cases}$$

En tous cas, on a :

$$E(f) = P(y \neq \hat{y})$$

- La **fonction du moindre carré** est définie comme suite :

$$l(y, \hat{y}) = (y - \hat{y})^2$$

Auquel,  $y$  est l'observation et  $\hat{y}$  est la valeur prédite.

- La **perte absolue** est une autre fonction de perte populaire pour évaluer la qualité de la prédiction :

$$l(y, \hat{y}) = |y - \hat{y}|$$

## 5 Les modèles d'apprentissages pour la prédiction du flux de trafic

En raison de la grande mobilité des véhicules sur les routes, les heures de pointe de la journée contribuent à près de la moitié de la congestion totale. Ces périodes sont caractérisées par des perturbations temporelles, une vitesse réduite et l'incapacité de l'infrastructure routière à soulager les routes. Par conséquent, il est impératif de prédire séquentiellement le flux de trafic en temps réel et de développer des contre-mesures appropriées pour réduire leurs effets négatifs.

Le débit de trafic peut être prévu pour un horizon de prédiction à court et/ou à long terme. Chaque horizon est pris en compte dans une période donnée. Compte tenu de l'horizon à court terme durant lequel la prédiction sera faite en quelques minutes dans le futur, différentes méthodes de prédiction ont été proposées telles que les séries temporelles multi variées (Ghosh et al, 2009), la méthode de filtrage de Kalman (Ji et al., 2010), le modèle de régression non paramétrique (Moretti et al, 2015 ; Williams et al, 1998). L'objectif principal de ces algorithmes



est de savoir anticiper un débit du trafic précis en se basant sur des modèles d'apprentissage. La structure des données d'apprentissage contient un ensemble de variables de trafic liées à deux dimensions, temporelles et spatiales.

Ces algorithmes d'apprentissage peuvent être classés en se basant sur des modèles statistiques (i.e., les régressions) ou des modèles artificiels comme les réseaux de neuronaux. En outre, le temps prévu d'arrivée des véhicules et le flux de trafic ( $q$ ) sont des paramètres de trafic qui ont attiré beaucoup d'attention dans les dernières années. Le temps d'arrivée peut généralement être utilisé pour évaluer l'exactitude de la prédiction du trafic (Lindveld et al, 2000). En effet, la précision de la prédiction de ces paramètres peut être réalisée par une collecte efficace de données et les instruments de surveillances en utilisant la technologie récente des détecteurs. Selon des travaux antérieurs, un ensemble de modèles de régression multilinéaires (Patnaik et al, 2004) est consacré à la prévision du temps d'arrivée à partir des données collectées par le compteur automatique de passagers (APC). Récemment, Feng et al (2015), ont proposé une solution de contrôle du trafic adaptatif en temps réel basée sur la collecte de données par véhicules connectés. Zhong et al (2015) ont utilisé, en temps réel, les données de trafic collectées en continu par une boucle inductive, tandis que la méthode de filtrage de Kalman a été utilisée pour prévoir les paramètres de trafic en se basant sur leurs historiques. D'un autre point de vue, la prédiction des paramètres de trafic est considérée comme un processus complexe affecté par d'autres facteurs. Bai et al (2015) estiment le temps de trajet à partir des données historiques de voyage en autobus en utilisant des machines vectorielles de support (SVM) et un algorithme basé sur le filtrage de Kalman, le modèle de prédiction considère les facteurs suivants : les périodes de pointe, la description routière, la moyenne pondérée du temps de déplacement et le temps de déplacement du dernier bus.

D'autre part, diverses observations permettent d'estimer le flux de trafic en fonction de la variation et la disponibilité des données. Dans cette thèse, plusieurs sources de données de trafic, y compris la variation des données, sont utilisées pour estimer le flux de trafic en temps réel. De toute évidence, diverses études s'accordent à dire que la prédiction du flux de trafic est un processus complexe nécessitant des entrées multiples où les sorties spécifient des niveaux de précision. En particulier, la plupart d'entre eux utilisent une seule source de données pour évaluer l'exactitude. Une approche non historique a été proposée par (Kriegel, 2008), permettant de prédire le flux de trafic en fonction de la densité attendue et de connaître le chemin le plus court du véhicule. Lindveld et al (2000) évaluent la précision de la prédiction en utilisant la prédiction du temps de déplacement du véhicule. Le modèle approximatif proposé par (Das, 2012) compare la vitesse du véhicule à la densité de la route pour construire un modèle approximatif de détection des conditions de circulation.

## 6 Discussion

Malgré la disponibilité dynamique des données de trafic en temps réel dans les travaux précédents, il convient de noter que l'erreur de prédiction affecte de manière significative la précision à court terme du flux de trafic, et devient une fonction objective à minimiser.

En fait, les modèles modernes d'apprentissage prédictif sont classés en modèles non paramétriques et paramétriques, ce qui suggère de nouvelles possibilités de réduire l'erreur de prédiction qui survient dans le processus de prédiction. Dans les modèles non paramétriques, où les réseaux neuronaux sont utilisés pour prédire le flux de trafic, les données historiques sont ensuite intégrées pour obtenir des performances cohérentes avec des erreurs de prédiction plus faibles (Kumar et al, 2015). D'autre part, les modèles d'apprentissage prédictif paramétrique tels que les séries temporelles (TSP) (Chu et Yang, 2016), l'analyse de régression linéaire (RA) (Zhu et al, 2016), la moyenne mobile intégrée autorégressive (ARIMA) (Xu et al, 2016), montrent de meilleures performances par rapport aux modèles d'apprentissage non

paramétriques.

Dans cette thèse, nous nous sommes intéressés à la dynamique du flux du trafic non-linéaire impactant plusieurs paramètres qui sont liés dans le temps et dans l'espace, tels que la vitesse, la densité, la distance prévue entre les véhicules. En examinant brièvement les littératures, la précision de flux du trafic dépend des fonctions d'agrégation utilisées dans le modèle de prédiction. Dans les travaux de (Fang et al 2010; Bachmann et al, 2013; Li et al, 2014), les auteurs utilisent l'intégrale floue de Choquet comme une fonction d'agrégation non-linéaire pour capturer les relations complexes entre les observations des capteurs. Tandis que, Qian et Eglese (2014) et Yasmin et al (2015) utilisent la moyenne pondérée en tant que fonction d'agrégation linéaire pour estimer la vitesse optimale du trafic. De même, en raison de la non linéarité du trafic véhiculaire dans les conditions congestionnées, nous proposons de généraliser la régression linéaire multiple par la fonction d'agrégation de l'intégrale de Choquet.

---

## CONCLUSION

Dans ce chapitre, nous avons commencé par présenter des théories pour analyser et modéliser la dynamique du trafic et aider à mettre en œuvre des outils et des techniques pour contrôler la fluidité. L'objectif était d'évaluer la qualité opérationnelle du trafic en prenant en compte les caractéristiques de base du trafic et la description des objets routiers tels que la conception du réseau routier (y compris les autoroutes, le pourcentage de poids lourds dans le trafic, etc.). En outre, ces théories visent à démontrer mathématiquement la précision des interactions entre l'environnement du véhicule et l'infrastructure routière. En général, elles sont basées sur trois niveaux d'agrégation : niveau macroscopique, niveau microscopique et niveau mésoscopique.

Nous avons ensuite présenté des techniques d'agrégation de données de trafic liées à la technologie des terminaux intelligents, tels que les détecteurs de trafic en tant que sources de données. Développer des solutions basées sur la coopération entre différentes entités telles que les véhicules, les infrastructures routières et mieux utiliser les informations disponibles via des réseaux de capteurs déployés en tant qu'infrastructure devient crucial. Nous avons discuté des principaux défis lors de l'exploitation de ces informations comme l'imperfection et la précision. Sur la base des relations entre les sources de données, trois schémas d'agrégation différents ont été présentés : agrégation complémentaire, redondante et coopérative.

En outre, nous avons mis en évidence deux classes de fonctions d'agrégation : les fonctions linéaires et les fonctions non linéaires. L'objectif était d'identifier quelle fonction serait la plus appropriée pour modéliser une telle interaction entre les sources de trafic. En raison de la non-linéarité des systèmes de trafic dans des conditions encombrées, nous préférons les fonctions d'agrégation non-linéaire pour la modélisation.

Des modèles d'apprentissage pour une prédiction précise du flux de trafic ont été également présentés. Nous avons exprimé l'ensemble des algorithmes qui raisonnent à partir des données de trafic empiriques. Par conséquent, le manque d'interactions fonctionnelles entre les entrées et les réponses présente un inconvénient majeur de ces modèles, ce qui entraîne un mauvais apprentissage et de grandes divergences des modèles.

Dans le chapitre suivant, nous présenterons un système de trafic intelligent pour préserver la sécurité routière en contrôlant les segments de route et en prédisant la probabilité d'incidents entrants. Basé sur des données empiriques, un modèle prédictif basé sur un modèle de régression logistique est adopté pour prédire si un incident se produira ou pas. L'objectif est de monter un type de dépendance monotone entre les variables de trafic et la réponse.

# Une évaluation précise des paramètres du trafic en utilisant la régression logistique

## SOMMAIRE: CHAPITRE 3

Introduction	49
1 Scénario d'étude	50
2 Estimation du temps de retard des véhicules	50
3 Un modèle de régression logistique pour la détection des incidents routiers inattendus	53
3.1 La vérification de la pertinence du modèle (Bonne adaptation)	54
3.2 Évaluation des variables expliquées	56
Conclusion	59

## INTRODUCTION

Anticiper les événements imprévus de la route révèle un grand problème dans les systèmes de transport intelligents à cause de la diversité des causes. Les incidents routiers n'offrent pas de conditions de circulation régulières et une prédiction précise de ces incidents en temps réel représente une tâche très compliquée et non définie jusqu'à présent.

Dans ce chapitre, une étude empirique est présentée pour développer un modèle prédictif permettant d'anticiper avec précision la probabilité d'occurrence d'un incident imprévu sur la route. L'objectif est de garantir une précision satisfaisante et un niveau d'assistance souhaité en préservant la sécurité routière et en surveillant les tronçons routiers ; le ralentissement soudain de la vitesse des véhicules est le cas pratique de l'étude. La tâche de classification de notre modèle prédictif inclut quatre variables expliquées, y compris la vitesse observée des véhicules, le temps de déplacement observé, le temps de déplacement prévu et le temps de retard estimé. La précision de la prédiction est prouvée en vérifiant la pertinence du modèle en fonction de la qualité de l'ajustement et de la signification statistique de chaque variable expliquée.

## 1 Scénario d'étude

Dans cette étude analytique, un modèle de trafic empirique fournit par le département de transport a été adopté dans le but de l'analyser. Les données enregistrées concernent un segment routier dans l'autoroute Est-Ouest, la route (A1) dans la figure 3.1 relie la ville de Constantine aux autres villes. Ce segment routier souffre parfois de congestion du trafic, en particulier dans les dernières heures de la journée. En règle générale, la congestion du trafic provient de l'apparition inattendue d'incidents tels que les accidents ou le ralentissement de la vitesse des véhicules.

Nous nous sommes intéressés à l'évaluation de l'état réel du trafic dans ce segment. Le modèle fourni est régénéré à l'aide d'un simulateur de mobilité urbaine, appelé Sumo (Krajzewicz, 2012). Le nouveau scénario de trafic comprend un ensemble de boucles inductives pour collecter les données entre le point de départ et le point final de ce segment. Selon les données enregistrées dans ce segment, les incidents routiers peuvent être principalement classés comme suit :

- Incidents causés par le ralentissement de la circulation des véhicules.
- Incidents causés par des conditions environnementales (i.e. les conditions météorologiques).

La première catégorie d'incidents est une catégorie particulière de prédiction de risque à court terme. Les incidents provoquent une mobilité lente quand le flux du trafic est affecté par les autres paramètres tels que la vitesse et la densité, i.e., les embouteillages et les goulots d'étranglement sont deux sortes d'incidents causés par la mobilité lente des véhicules. La deuxième catégorie d'incidents permet une prédiction de risque à long terme, comme le mauvais temps, où il n'est pas possible d'observer ses effets durant une courte durée, car elle peut persister pendant des heures, voire des jours ou des mois.

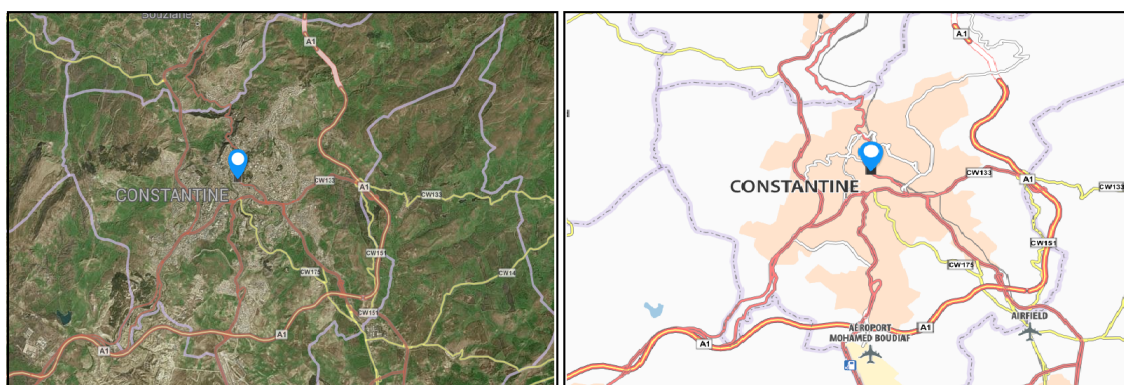


FIGURE 3.1 – L'autoroute Est-Ouest de Constantine.

## 2 Estimation du temps de retard des véhicules

Grâce à une étude approfondie sur les données enregistrées par le processus de simulation, nous avons constaté que pour examiner les facteurs de risque causés par le ralentissement de la vitesse, la recherche d'une relation entre le temps prévu de déplacement au deuxième point et ce qui a été enregistré est primordial. En fait, le temps de déplacement prévu est le moment idéal pour qu'un véhicule doive partir vers le deuxième point. En outre, le temps

d'enregistrement représente le temps de passage de ce véhicule au deuxième point. Les données enregistrées sont représentées dans le diagramme de la figure 3.2. Le diagramme comprend trois courbes de vitesse en fonction de l'axe temporel. La première courbe montre la vitesse moyenne enregistrée au premier point, les deux autres montrent la vitesse prévue et la vitesse observée au deuxième point.

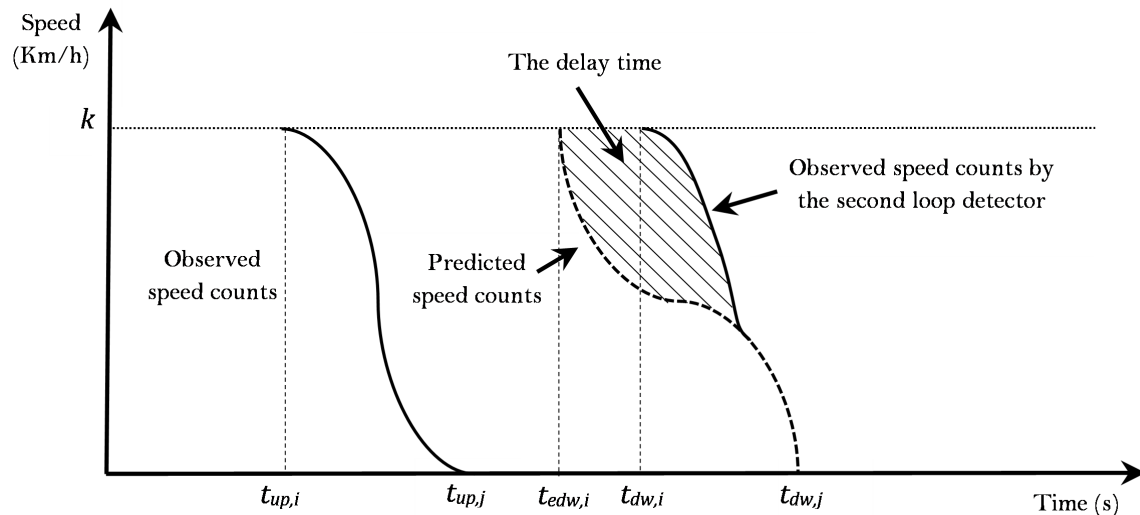


FIGURE 3.2 – Modéliser la vitesse du trafic entre un couple de boucles inductives.

En résumé, l'intervalle de temps  $[t_{up,i}, t_{up,j}]$  indique l'intervalle d'enregistrement dont la vitesse et le temps de départ observés sont collectés au premier point. L'intervalle de temps  $[t_{edw,i}, t_{dw,j}]$  indique l'intervalle prévue pour qu'un véhicule arrive au deuxième point, c.-à-d., la vitesse et le temps de déplacement prévus de ces véhicules sont calculés. En outre, l'intervalle de temps  $[t_{dw,i}, t_{dw,j}]$  représente la vitesse et le temps d'arrivée observés au deuxième point. La zone pointillée dans la figure 3.2 montre le temps de retard associé au risque d'un évènement qui pourrait se produire dans ce segment.

Il est intéressant de noter qu'à chaque fois qu'il y a eu une diminution notable de la vitesse du véhicule entre deux points, le temps nécessaire pour atteindre le deuxième point est étendu. Il est crucial ici de prédire l'apparition de l'incident de la route à partir des données enregistrées par les détecteurs de boucle. Cela justifie la nécessité d'utiliser les détecteurs de trafic connectés pour réduire les effets de l'incident prédit.

En fait, un temps de retard est causé par la non-arrivée à l'heure des véhicules au deuxième point. En outre, interpréter l'apparition du retard peut être expliqué en explorant les différentes relations entre les variables. À cet égard, la figure 3.3 et la figure 3.4 montrent le niveau de corrélation entre les variables de trafic avant et après l'incident. Avant l'incident, comme le montre la figure 3.3(a), le flux de trafic n'est pas affecté en raison de la forte corrélation entre la vitesse prévue et la vitesse observée. Après l'incident, la corrélation entre la vitesse prévue et la vitesse observée est influencée par l'incident qui se produit, comme le montre la figure 3.3(b). Ainsi, le temps de déplacement est également étendu. À partir de la figure 3.4, le temps de déplacement prévu du segment de la route est 8.71 minutes et le temps de déplacement observé par le détecteur de boucle est 11.58 minutes. Par conséquent, les véhicules ont passé environ 2.93 minutes pour traverser le segment de la route.

Pour quantifier le temps de retard, le temps de déplacement observé au deuxième point ainsi que le temps de déplacement prévu sont nécessaires. Dans ce qui suit, l'équation 3.1 permet de quantifier ce facteur (la zone pointillée de la figure 3.2) :

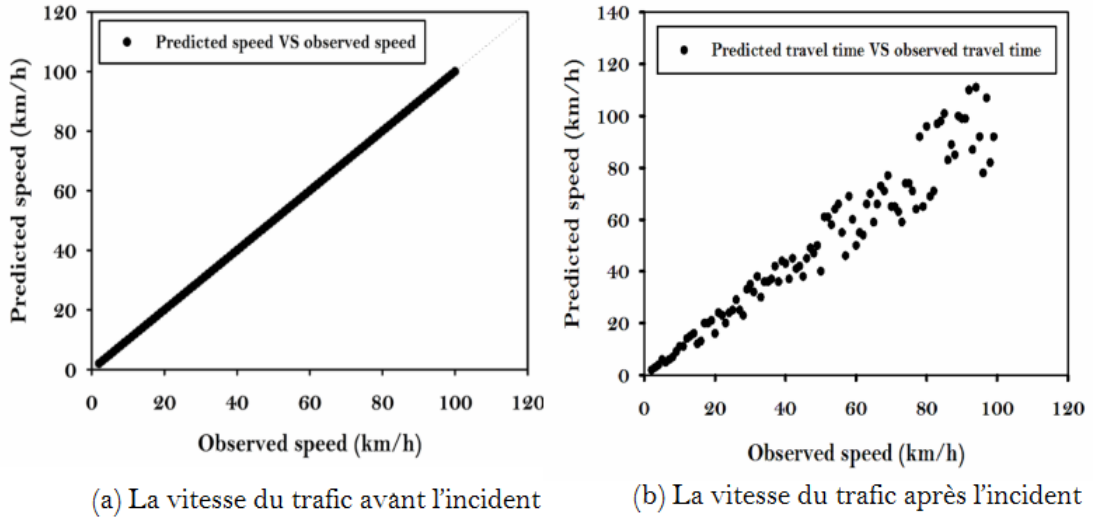


FIGURE 3.3 – La corrélation entre la vitesse observée et la vitesse prévue des véhicules.

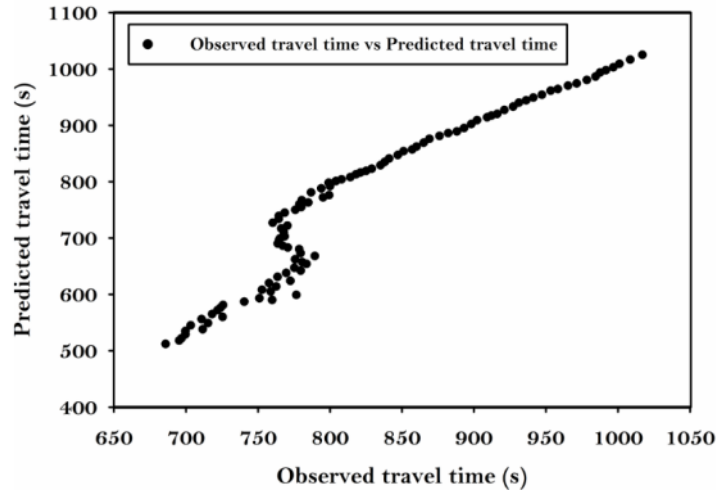


FIGURE 3.4 – Le niveau de corrélation entre le temps de déplacement observé et le temps de déplacement prévu après l'incident.

$$DelayTime = (x_k(t_{dw,j}) - x_k(t_{edw,i})) - \int_{t_{edw,i}}^{t_{dw,j}} (n_{dw,j} + n_{edw,i}) dt \quad (3.1)$$

Pour prouver l'équation précédente.  $S_2$  désigne la zone sous la courbe du temps de déplacement prévu.  $S_3$  désigne la zone sous la courbe du temps de déplacement observé. La différence entre ces deux zones indique le temps de retard qui pourrait apparaître pendant le ralentissement de la vitesse.

$$S_2 = \int_{t_{edw,i}}^{t_{dw,j}} (n_{edw,i}) dt$$

$$S_3 = \int_{t_{edw,i}}^{t_{dw,j}} (k - n_{dw,j}) dt$$

$$DT = S_3 - S_2 = \int_{t_{edw,i}}^{t_{dw,j}} (k - n_{dw,j}) dt - \int_{t_{edw,i}}^{t_{dw,j}} (n_{edw,i}) dt$$



$$\begin{aligned}
 &= k(t_{dw,j} - t_{edw,i}) - \int_{t_{edw,i}}^{t_{dw,j}} (n_{dw,j} + n_{edw,i}) dt \\
 &= (x_k(t_{dw,j}) - x_k(t_{edw,i})) - \int_{t_{edw,i}}^{t_{dw,j}} (n_{dw,j} + n_{edw,i}) dt
 \end{aligned}$$

### 3 Un modèle de régression logistique pour la détection des incidents routiers inattendus

En général, la régression logistique est plus adaptée lorsque les variables dépendantes sont binaires et ne prennent que deux valeurs. Elle est connue également comme modèle Logistique en raison de la fonctionnalité d'analyse qui explique la relation entre la variable dépendante et les variables expliquées. En exprimant cette relation mathématiquement, ces variables se rapportent potentiellement par une équation selon la régression linéaire ou les formes d'équation de régression non linéaire. Cependant, en utilisant le modèle de régression logistique, le but n'est pas seulement d'expliquer la variation des données détectées par les détecteurs, mais aussi de prédire la probabilité d'incidents entrants en prenant en compte ces variations. Dans notre cas, la prédiction est liée généralement à des paramètres statiques et dérivées fournis par ces détecteurs tels que la vitesse des véhicules observée, le temps de déplacement prévu, le temps de déplacement observé et le temps de retard. La régression logistique est définie selon l'équation suivante :

$$L = \ln\left(\frac{p_i}{1 - p_i}\right) = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k \quad (3.2)$$

Alors que  $p_i$  indique la probabilité de succès d'un incident, le succès de l'événement est symbolisé par (1). La probabilité  $(1 - p_i)$  est celle de non occurrence d'un incident, il est symbolisé par (0). Le modèle Logit ne nécessite pas des variables explicites (indépendantes) cohérentes par rapport au modèle de régression linéaire. Ces variables peuvent être continues, catégoriques et ne nécessitent aucun équilibre dans leur distribution. Comme mentionné précédemment, le modèle de régression logistique ne nécessite pas de variables indépendantes pour suivre une distribution régulière et par conséquent, la relation entre eux n'exprime pas une relation linéaire. Pour une classification efficace, cette méthode est également caractérisée par la possibilité d'estimer toutes les probabilités d'apparition d'une variable dépendante. Hosmer et al. (2013), estiment que le processus d'évaluation du modèle commence une fois que le modèle de régression logistique est bien ajusté. Cela veut dire simplement que l'étape d'identification des variables dépendantes et expliquées va servir comme étude. Selon King et Menard (2002), il existe deux façons pour vérifier la pertinence du modèle : (1) Vérifier la pertinence du modèle dans son ensemble (Bonne adaptation), (2) Examiner les preuves statistiques de chaque variable expliquée séparément. En revanche, à partir du tableau 3.1 et avant l'analyse des données, ces données souffrent d'un problème de multi-colinéarité, qui se réfère à l'existence d'une corrélation linéaire entre les variables explicatives du modèle de régression. En fait, cette corrélation s'explique principalement par de nombreuses raisons, y compris le changement par un intervalle de temps et parce qu'elles sont toutes affectées par les mêmes facteurs. Avant de présenter les résultats du modèle de régression logistique, les variables descriptives utilisées pour cette étude sont mentionnées comme suit :

- **X1** : (variable expliquée) indique la vitesse observée des véhicules.
- **X2** : (variable expliquée) indique le temps de déplacement prévu.

- **X3** : (variable expliquée) indique le temps de déplacement observé.
- **X4** : (variable dépendante) indique l'incident (oui, non).
- **X5** : (variable expliquée) est une variable binaire et sa valeur dépend au temps de retard calculé en fonction de l'équation 3.1. Avec cet échantillon de données, la grande surface sous la courbe qui correspond à la forte probabilité qu'un événement se produise est notée par (1). La plus petite surface sous la courbe qui correspond à la faible probabilité qu'il se produise est notée par (0).

Tableau 3.1 – Les paramètres de trafic sélectionnés pour la prédiction de l'incident.

X1	X2	X3	X4	X5
60.00	3.00	10.00	.00	.00
57.00	3.16	9.16	.00	1.00
56.00	3.21	4.21	.00	1.00
78.00	2.31	2.31	1.00	.00
40.00	4.50	5.50	.00	1.00
70.00	2.57	4.57	.00	.00
.		.		.
.		.		.
.		.		.
.		.		.
59.00	3.05	6.05	.00	1.00
45.00	4.00	6.00	.00	1.00
44.00	4.09	7.09	.00	1.00
65.00	2.77	6.77	.00	.00
47.00	3.83	8.83	.00	1.00
46.00	3.91	8.91	.00	1.00

Pour démontrer l'exactitude de la prédiction via l'utilisation de ce modèle, les résultats de la régression logistique sont analysés par l'outil statistique Spss (Field, 2013). En particulier, la pertinence du modèle est vérifiée comme suit :

### 3.1 La vérification de la pertinence du modèle (Bonne adaptation)

Grâce à cette propriété statistique, la précision du modèle peut être entièrement vérifiée par des mesures de qualité d'ajustement. Ces mesures de qualité résument la différence entre les valeurs observées et les valeurs attendues. En outre, cette différence explique la correspondance entre l'ensemble des observations et la réponse. Dans un premier temps, la signification statistique doit être testée pour vérifier l'adéquation du modèle en supposant qu'il soit entièrement approprié dans son ensemble. Ensuite, le problème de la multi-colinéarité sera traité avec succès. Pour cela, la démarche de la régression logistique est adoptée pour examiner les effets des variables expliquées (X1, X2, X3 et X5) sur la réponse (X4). En effet, le modèle détecte un ajustement parfait après une seule étape (étape 1). Sur la base de ces résultats, la signification statistique peut être obtenue grâce à l'utilisation des normes suivantes se rapportant à la qualité de l'ajustement :

### Le test Chi-squared

Le test du Chi-squared est un test d'hypothèse statistique visant à vérifier l'adéquation des séries de données à la distribution du chi-carré lorsque l'hypothèse nulle est correcte. En outre, il est principalement calculé sur la base de la somme des erreurs au carré.

Dans le tableau 3.2 présenté ci-dessous, le test Chi-Square montre l'ajustement parfait du modèle (étape 1 comme discuté précédemment). Il est à noter que  $\chi(1) = 22.915$ ,  $p - value = 0.000 < 0.05$ , ce qui démontre qu'il y a une signification statistique entre les variables expliquées (vitesse observée des véhicules, temps de trajet prévu, temps de trajet observé et temps de retard) et l'incident pouvant se produire. Grâce à cette signification, il devient clair que les variables indépendantes ont effectivement contribué à l'apparition d'un incident inattendu sur la route.

Tableau 3.2 – Le test Chi-squared pour les coefficients du modèle

<b>Step 1</b>	<b>Step</b>	22.915	4	.0000
	<b>Block</b>	22.915	4	.0000
	<b>Model</b>	22.915	4	.0000

### La mesure de la qualité de l'ajustement R-carré

Sur la base de la mesure du R-carré comme clé résultat pour la régression logistique, la mesure R-carré est appliquée pour identifier le taux de variance impliqué aux données qui peuvent être utilisées pour prédire l'incident. En fin de compte, les variables entrées au modèle ont montré qu'il y a encore d'autres variations causées par d'autres variables de trafic non inclus dans le modèle. En effet, comme le montre le tableau 3.3, les variables impliquées dans le modèle ont montré qu'elles étaient interprétées approximativement à 72% (en utilisant Cox et R-carré Snell) et à 100% (en utilisant le R-carré Nagelkerke).

Tableau 3.3 – Interprétation des variables impliquées dans le modèle (model summary)

Step	-2 Log likelihood	Cox Snell R-Square	Nagelkerke R-Square
1	.000	.720	1.000

### Test de l'efficacité du modèle de classification

Le test de l'efficacité du modèle de régression est considéré comme une étape clé pour évaluer l'exactitude de la réponse. Ce test explique l'exactitude du modèle prédictif pour la classification des cas. Les tableaux 3.4 et 3.5 ont montré la classification des tables du modèle de régression pour l'étape 0 et l'étape 1 respectivement. Comme observation, le pourcentage global de probabilité pour qu'un incident se produise atteint 66.7% à l'étape 0. A l'inverse, la dernière étape confirme que la possibilité qu'un incident se produise sera de 100%.

### Le test de Hosmer et de Lemeshow

Le test de Hosmer et Lemeshow (Hosmer et Lemeshow, 2013) a été adopté pour évaluer si l'hypothèse selon laquelle ce modèle est approprié pour la prédiction d'un incident avec

Tableau 3.4 – Tableau de classification à l'étape 0.

Observed		EventOccuring		Percentage Correct	
<b>Step 0</b>	EventOccuring	.00	0.00	6	0
		1.00	0.00	12	100.0
Overall Percentage					67.7

Tableau 3.5 – Tableau de classification à l'étape 1.

Observed		EventOccuring		Percentage Correct	
<b>Step 1</b>	EventOccuring	.00	6	0.00	100.0
		1.00	0.00	12	100.0
Overall Percentage					100.0

l'hypothèse alternative que ce modèle ne convienne pas à la prédiction. Les résultats du test Hosmer et Lemeshow sont présentés dans le tableau suivant :

Tableau 3.6 – Le test Hosmer et Lemeshow.

Step	Chi-square	df	Sig.
1	.000	5	1.000

En dernière étape, il ressort clairement du tableau que la valeur significative du modèle est plus proche de 1 (strictement supérieure à 0.05). Cela signifie qu'il n'y a pas assez de preuves pour rejeter l'hypothèse. En conséquence, nous concluons que le modèle final proposé est adapté aux paramètres de circulation.

### La courbe Roc

La caractéristique de fonctionnement du récepteur (Roc) présente un diagramme ou courbe de sensibilité / spécificité, qui vise à tester la précision de la classification et la sensibilité du modèle de régression. En effet, cette évaluation permet d'identifier les variables jouant un rôle important dans le processus de prédiction. Pour cela, toutes les variables indépendantes ont été testées pour l'incident en cours et les résultats sont présentés dans la figure 3.5.

À partir de la courbe de *sensibilité / spécificité* de la figure 3.5 et de la surface se trouvant sous la courbe présentée au tableau 3.7. Le paramètre de vitesse observé a une grande sensibilité sur le processus de prédiction en raison, d'une plus grande surface sous la courbe et une signification du modèle s'avérant complètement inférieure à 0.05. Néanmoins, le temps de déplacement/trajet observé a une signification sur le modèle de prédiction mais avec une sensibilité plus faible. D'autre part, les temps de déplacement prévus et le temps de retard ont été exclus du modèle de prédiction car il n'y a pas de signification cruciale et aucune valeur de sensibilité cruciale.

## 3.2 Évaluation des variables expliquées

Contrairement au test de qualité d'ajustement, cet examen attribue une importance statistique à chaque variable de trafic expliquée séparément. L'objectif principal de cet examen est de déterminer l'importance de chaque variable indépendante. Généralement, le test Wald

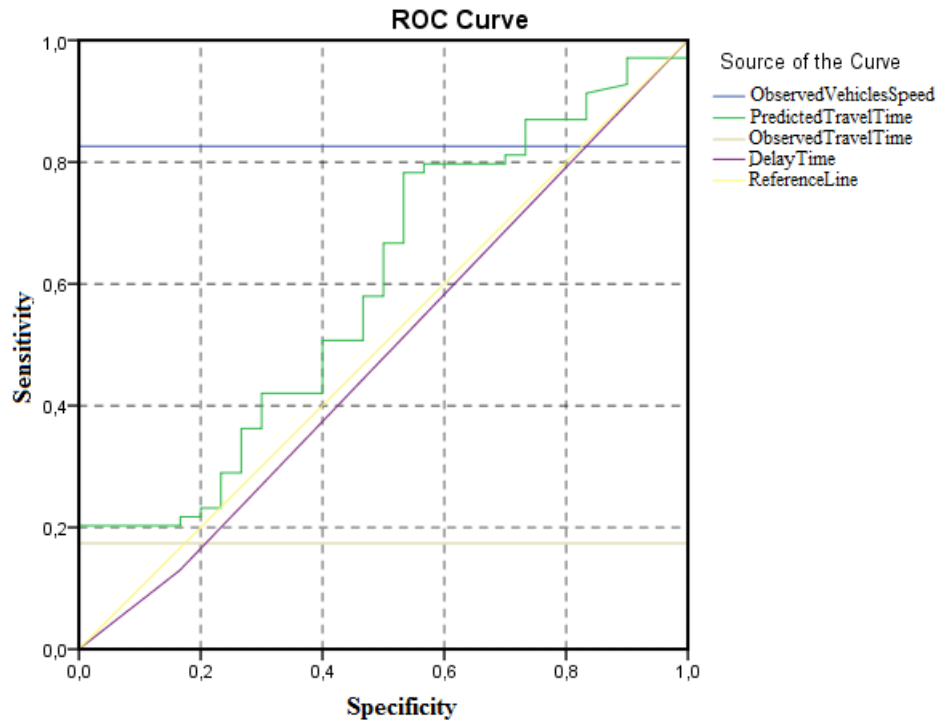


FIGURE 3.5 – La courbe Roc pour la précision de la classification et la sensibilité du modèle de régression.

Tableau 3.7 – La zone sous la courbe.

Test Result Variable(s)	Area	Std.Error	Asymptotic Sig.
PredictedArrivalTime	.826	.046	.000
RTArrivalTime	.593	.062	.144
ObservedRTArrival	.174	.046	.000
DelayTime	.482	.064	.775

vérifie l’hypothèse nulle selon laquelle le coefficient de régression logistique associé à la variable indépendante est égal à zéro (Cizek et Fitzgerald, 1999). Dans ce qui suit, les tableaux 3.8 et 3.9 montrent des variables de trafic indépendantes impliquées dans l’équation de prédiction d’incident pour l’étape 0 et l’étape 1 respectivement.

Tableau 3.8 – Variables dans l’équation à l’étape 0.

	B	S.E.	Wald	df	Sig.	Exp(B)
<b>Step 0</b> Constant	.693	.500	1.922	1	.166	2.000

Les variables indépendantes dans le tableau 3.9 peuvent être résumées dans l’équation de prédiction ci-dessous :

$$Y = -1584.387 + 10.013X_1 + 349.755X_2 - 8.188X_3 - 61.336X_5 \quad (3.3)$$

Il est évident que le modèle logistique est cohérent avec l’événement prévu à court terme. Par exemple, avec l’impact de stabilité des autres variables, l’augmentation de la vitesse des véhicules observée par 10 km / h entraîne une augmentation de la probabilité qu’un incident

Tableau 3.9 – Variables dans l'équation à l'étape 1.

		<b>B</b>	<b>S.E.</b>	<b>Wald</b>	<b>Sig.</b>	<b>Exp(B)</b>
<b>Step 1</b>	X1	10.01	4231.51	.00	.99	22324.724
	X2	349.75	108723.61	.00	.99	7.884E+151
	X3	-8.18	5145.80	.00	.99	.00
	X5	-61.33	42838.95	.00	.99	.00
	Constant	-1584.38	621201.19	.00	.98	2.000

surviene. En outre, l'augmentation du temps de déplacement prévu par 5.82 minute conduit à l'apparition d'un incident même avec la stabilité des autres variables. En outre, l'augmentation du temps de déplacement observé et le temps de retard feront atteindre le coefficient de Logit à la valeur 0, ce qui prouve la probabilité d'occurrence d'un incident. Malgré cette cohérence, le test de Wald prouve le non signification des paramètres estimés grâce aux valeurs de significations.

---

## CONCLUSION

Dans ce chapitre, en se basant sur le ralentissement de la vitesse des véhicules, un modèle prédictif basé sur un modèle de régression logistique est proposé afin de prédire si un incident va se produire ou non. En particulier, la vitesse des véhicules observée, le temps de déplacement prévu, le temps de déplacement observé et le temps de retard sont quatre variables de trafic expliquées visant à prédire un incident inattendu sur la route.

Pour vérifier la précision du modèle prédictif, nous avons utilisé le test de la qualité de l'ajustement et l'examen et l'évaluation des variables expliquées.

Dans le premier test, la diversité des significations statistiques (par exemple test du chi carré, la mesure en R-carré de la qualité de l'ajustement, l'efficacité du modèle de classification et les tests de Hosmer et Lemeshow) prouvent une forte compatibilité du modèle de régression attendu dans son ensemble avec les mesures observées données par les détecteurs de boucle.

Aussi, les mesures du R-carré de la qualité de l'ajustement montrent qu'il existe 72% des variables de trafic non incluses dans le modèle, ce qui prouve la complexité de la prévision de ce type d'incidents. Au stade de la classification, la probabilité d'occurrence d'un incident atteint 66.7% à la première étape et 100% à la dernière étape. Alors que la courbe de Roc montre que le processus de prédiction est très sensible à la vitesse des véhicules et le temps de déplacement observés.

Dans le second test, l'examen séparé de chaque variable expliquée montre une signification statistique faible mais de manière positive, il pourrait être utilisé afin de permettre de distinguer les conditions de conduite sûres de celles qui sont jugées dangereuses.

Dans ce chapitre, nous avons présenté un modèle prédictif sur la base d'un modèle de régression afin d'interpréter le phénomène de la congestion. Le modèle adopté permet d'interpréter de manière précise les dépendances entre les variables du trafic comme un modèle linéaire généralisé.

# Un système de fusion floue pour une prédiction précise du trafic

## SOMMAIRE: CHAPITRE 4

Introduction	61
1 Un système de fusion floue de données du trafic basé sur les données d'apprentissage	62
1.1 La source de données du trafic	62
1.2 Les niveaux de fusion des données	63
1.3 La phase d'évaluation	66
2 Modèle d'apprentissage floue	66
2.1 La régression linéaire multiple (MLR)	67
2.2 La mesure floue et l'intégral floue de Choquet	67
2.3 Le modèle de prédiction Choquet-MLR	69
3 Les prédicteurs du trafic	70
3.1 Le temps de passage du segment routier	71
3.2 Le prédicteur du temps de déplacement	73
4 Expérimentation et évaluation	74
4.1 Les performances d'agrégation	75
4.2 Les critères d'évaluation de la prévision	75
4.3 L'impact du volume de données sur le processus de prédiction	77
Conclusion	80



---

## INTRODUCTION

Ce chapitre traite de l'un des domaines les plus intéressants de la gestion du trafic, plus particulièrement, la prédiction précise du trafic à court terme dans les autoroutes. Cette prédiction permet d'assurer la robustesse des applications de transports intelligents. Cependant, il y a un grand consensus entre les méthodes de prédiction du fait que les performances de la précision dépendent principalement du niveau de corrélation linéaires entre les paramètres du trafic.

Nous avons présenté dans le chapitre précédent un modèle de prédiction du trafic en utilisant la régression logistique, les significations statistiques des variables montrent qu'il existe une interaction non linéaire entre ces variables. En effet, la modélisation de ce type d'interaction est rarement considérée avec les méthodes de prédiction des flux du trafic.

Dans ce chapitre, un système de fusion nommé Fuzzy Traffic Data Fusion (FTDF) basé sur un modèle d'apprentissage prédictif est proposé pour prédire des données précises de traçabilité du trafic à partir de plusieurs sources de données. Le modèle d'apprentissage intègre les attributs caractéristiques ajustés via l'intégrale de Choquet et la Régression Linéaire Multiple (MLR) afin de garantir la monotonie en attribuant des poids à l'ensemble des attributs. L'intégrale sélectionnée vise à modéliser les interactions entre les données du trafic pour ajuster la stabilité de ses paramètres dans le processus de prédiction.

# 1 Un système de fusion floue de données du trafic basé sur les données d'apprentissage

La fusion de données est largement adoptée dans différents domaines tels que la biologie (Gligorijevi et Prulj, 2015), civil (Haas, 2006), les applications de santé telles que la surveillance (Duro et al, 2016) et la reconnaissance de formes (Santos et al, 2016). Le but principal de la fusion de données est d'extraire les connaissances de domaine dans un environnement observé surtout lorsqu'il est difficile d'extraire des connaissances à partir des données brutes avec des méthodes de calcul classique et même rarement avec des modèles d'apprentissage. Dans les Systèmes de Transport Intelligents (STI), la fusion de données est beaucoup moins connue jusqu'ici, dans la prévision des flux du trafic.

Le but principal de cette étude est d'évaluer l'impact de la fusion floue de données en fonction des données d'apprentissage sur les systèmes du trafic urbains globaux avec une précision plus élevée. La figure 4.1 ci-dessous montre l'architecture du système proposé, nommé Fuzzy Traffic Data Fusion (FTDF). Le modèle structurel du système FTDF comprend la source des données, le traitement des données, l'adaptation et l'évaluation du modèle.

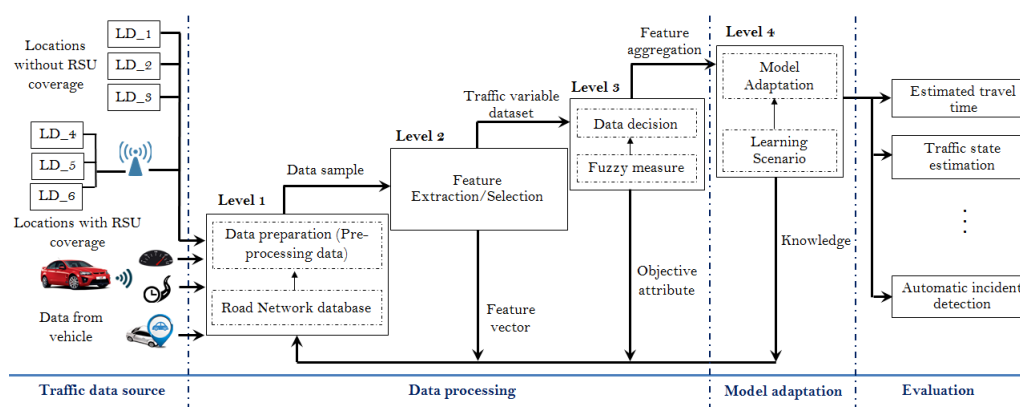


FIGURE 4.1 – L'architecture de base du système proposé pour la fusion floue de données à partir de plusieurs sources de données du trafic.

## 1.1 La source de données du trafic

La surveillance et la collecte des données du trafic provenant de diverses et multiples sources de données du trafic en temps réel sont considérées comme importantes pour la planification routière et la bonne gestion du réseau routier. Dans la littérature, il n'a pas été discuté de l'éventualité de fournir une décision précise sur un véhicule à partir de multiples sources de données du trafic.

Dans cette étude, les données d'entrées fournies par les sources de données à notre système de fusion sont collectées auprès de différentes entités routières (par exemple, véhicules, détecteurs, infrastructures routières, etc.). En particulier, les données d'entrées au système comprennent les données collectées à partir de détecteurs routiers intelligents installés dans le bitume tels que les détecteurs à boucle inductive connectés (cela peut inclure la vitesse et l'occupation), des données échangées entre les véhicules via une communication sans fil telles que Vehicule-à-Vehicule (V2V) et Vehicule-à-Infrastructure (V2I), lesquelles données sont capturées par les véhicules (par exemple, la vitesse, les données de trajet et la position).

Par conséquent, et en raison de la capacité limitée de ces entités en termes de stockage et de traitement, la grande variété des données du trafic et le traitement volumineux des données

collectées par ces entités en temps réel nécessitent des espaces de stockage et de traitement flexibles. Aux heures de pointe de la journée, où la masse des données du trafic est la plus volumineuse, la précision de l'information est sollicitée pour la prise de décision des véhicules. En effet, l'analyse efficace des données recueillies augmentera la précision et l'efficacité de l'information sur le trafic. Dans ce contexte, utiliser les technologies de réseau émergentes telles que le Cloud Computing est crucial dans ces cas pour améliorer la capacité des détecteurs en termes de stockage et de traitement (Savas et al, 2013).

La figure 4.2 ci-dessous montre l'architecture de notre système de collecte de données du trafic. En particulier, l'architecture utilise des réseaux Hertiens sans fil (HWN) et les stations de bases (RSU), Road Side Unit en anglais, comme base pour faciliter la collection des données et l'accès aux services Cloud. Dans le processus de collecte, les détecteurs de boucle sont utilisés pour détecter des informations routières en temps réel telles que la vitesse des véhicules, le passage des véhicules, l'occupation de la route, la densité de la route, le nombre de véhicules contribués, etc. Les données capturées par les détecteurs de boucle seront collectées dans des dispositifs de collections (DCE), Data Collection Equipment. Pour transférer les données du trafic vers un système Cloud, les périphériques UMTS (par exemple UC-7420) sont intégrés en tant que médiateurs de réseau pour connecter le DCE aux services Cloud. A cet égard, les détecteurs de boucle sont entièrement connectés et la tâche de la collection de données sera flexible.

Dans l'environnement Cloud, les données fournies par les détecteurs de boucle connectés sont utilisées comme entrées pour la prédiction des flux du trafic. En outre, le processus de prédiction intègre une base de données du trafic pour stocker les données capturées et augmenter la précision de la prédiction en fournissant des données historiques.

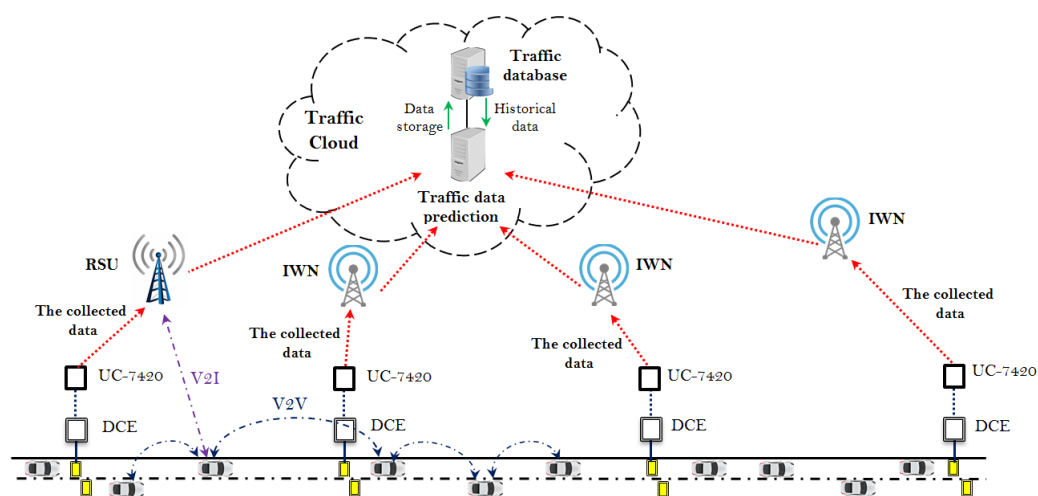


FIGURE 4.2 – La collection de données du trafic sur le Cloud pour la prévision des flux du trafic à court terme.

## 1.2 Les niveaux de fusion des données

La représentation illustrative de la fusion floue de données du trafic dans les STI est l'utilisation de différents objets routiers pour détecter typiquement l'information de trafic. L'avantage principal du système FTDF est de combiner les données provenant de sources de données multiples afin de produire des attributs de traçabilité cohérents, informatifs et synthétiques qui représentent idéalement un aperçu global des données disponibles. En règle générale, les modèles d'apprentissage peuvent ensuite utiliser ces caractéristiques à la place des données de traçage d'origine. Ceci est impératif pour améliorer les performances

de l'apprentissage lorsque la quantité de données du trafic disponibles devient grande pour récupérer les données de chaque source de données séparément.

Le système de fusion de données que nous proposons intègre quatre niveaux pour la fusion de données, y compris : la préparation de données (niveau 1), l'extraction / la sélection des attributs (niveau 2), la décision sur les données (niveau 3) et adaptation des modèles (niveau 4). Les trois premiers niveaux appartiennent au traitement de données et le niveau 4 appartient à l'adaptation des données.

### **Niveau 1 (la préparation des données)**

La préparation des données est une étape cruciale dans le traitement des données, qui doit être menée à l'avance pour fournir efficacement des ensembles de données fiables pour le modèle d'apprentissage. En fait, les données recueillies par les sources de données du trafic peuvent inclure diverses anomalies de données (par exemple, des données incomplètes, manquantes, bruitées et incohérentes). Ces anomalies diminuent la qualité des données et mènent à des résultats trompeurs d'apprentissage.

La préparation des données est combinée avec le système FTDF afin de bénéficier de ses avantages tels que le nettoyage des données brutes et l'amélioration de la qualité des données. Derrière la préparation des données, on peut supposer que les attributs caractéristiques entrées au modèle d'apprentissage ne contiennent pas de données incohérentes ou incomplètes. Dans notre système de fusion, une base de données de réseau routier est préparée à l'aide de multiples sources de données du trafic pour stocker des données cohérentes et générer des données de haute qualité. La base de données du réseau routier peut stocker des données statiques concernant le réseau routier telles que les nœuds (intersections), les arcs (liens), les segments routiers ou tronçons de route, la largeur des routes, la vitesse limitée, les panneaux d'arrêt, etc. En outre, la base de données peut également stocker des données dynamiques telles que les attributs caractéristiques produites par les autres niveaux, des statistiques sur les accidents de la route, les données sur les accidents de la route et le trafic quotidien.

### **Niveau 2 (Extraction / Sélection des attributs caractéristiques)**

L'objectif essentiel de ce niveau consiste à identifier le format pertinent des ensembles de données pour construire un modèle d'apprentissage précis avec un nombre minimum d'attributs caractéristiques. Cette étape commence par les jeux de données préparés, appelés ensemble de données, par le niveau 1 pour extraire/sélectionner les attributs caractéristiques utilisées comme entrées pour l'apprentissage.

L'extraction des attributs caractéristiques est sollicitée lorsque le modèle d'apprentissage nécessite des nouveaux attributs caractéristiques qui n'existent pas dans les jeux de données. Dans ce cas, les nouveaux attributs caractéristiques peuvent être dérivés à l'aide de processus de transformation des données.

La sélection des attributs caractéristiques est censée réduire les dimensions des jeux de données du trafic, en supprimant les attributs redondants et irréguliers et en réduisant également le temps d'exécution. Le processus d'extraction/sélection des attributs est invoqué lorsque les jeux de données du trafic de petites tailles deviennent grands. Avec les jeux de données de grosses tailles, les anomalies de données augmentent de façon exponentielle avec les dimensions des attributs caractéristiques. Dans ce cas, la variation des données peut provoquer des effets de sur-apprentissage (dit overfitting) dans les données d'entraînement. Un modèle sous effets de sur-apprentissage est un modèle qui, avec les données d'entraînement comprend l'incohérence et provoque aussi une mauvaise performance prédictive. Pour garantir une meilleure performance d'apprentissage, les effets de sur-apprentissage du modèle doivent

être réduits en sélectionnant les sous-ensembles des meilleurs attributs caractéristiques pour le fonctionner. Cela dépend du contexte où la fusion de données est utilisée. Dans ce cas, un vecteur de caractéristiques qui est un ensemble d'attributs caractéristiques qui peuvent être sélectionnés ou extraits, est produit et ensuite sauvegardé dans la base de données du réseau routier (niveau 1). L'objectif est de faciliter le traitement des données à l'avenir (prédiction à long terme).

Par exemple, en fonction du type d'application du trafic ainsi que des jeux de données fournis par les détecteurs de boucle connectés, la prédiction du flux de trafic serait évaluée. Les attributs sélectionnés et les attributs extraits sont deux façons d'obtenir une prédiction précise. En outre, concernant l'objectif de prédiction du trafic, un ensemble d'attributs caractéristiques de vitesse doit être sélectionné à partir des jeux de données originaux. En revanche, les caractéristiques extraites correspondent aux caractéristiques dérivées en utilisant les attributs de base. Par exemple, le temps de déplacement d'un segment routier peut être dérivé en fonction de l'heure de départ du véhicule et de la longueur de la route.

### **Niveau 3 (décision sur les données)**

Une fois que les attributs caractéristiques du trafic sont sélectionnés / extraits pour l'apprentissage, l'intégrale Choquet est utilisée pour prendre des décisions sur les attributs qui seront intégrés au processus d'apprentissage. La puissance d'agrégation de cette intégrale décrit les interactions fonctionnelles entre les attributs pour améliorer le stockage et le temps de traitement. Cet outil mathématique est nécessaire pour assurer que tous les attributs sélectionnés / extraits sont pertinents pour une décision précise. En fait, une bonne décision est basée sur l'importance attribuée à l'ensemble des attributs, connue sous le nom de mesure floue, pour confirmer que les attributs caractéristiques sélectionnés / extraits contiennent des données significatives pour le processus d'apprentissage. En outre, l'identification des mesures floues de l'intégrale de Choquet dépend principalement des données d'entraînement. Dans le cadre de notre étude, une décision est un attribut objectif qui représente des données informatives contenues dans multiples sources de données.

Pour créer un modèle prédictif basé sur l'agrégation des attributs, nous proposons la généralisation de la régression linéaire multiple (MLR) par la fonction d'agrégation de l'intégrale de Choquet pour ajuster les caractéristiques du trafic dans le processus de prédiction. Les détails du modèle Choquet-MLR seront décrits dans la section 2.

### **Niveau 4 (adaptation du modèle)**

Du fait de la continuité des données du trafic collectées par les entités routières en temps réel, il est impératif de Réappeler les ensembles de données originaux avec les jeux de données en temps réel pour reformer le modèle d'apprentissage, surtout si les jeux de données originaux sont de grande taille, consomment beaucoup de temps de traitement.

Par conséquent, la combinaison des jeux de données historiques du trafic avec des données variantes peut entraîner de grandes divergences dans le modèle d'apprentissage. Nous utilisons l'apprentissage par transfert pour réduire le temps de traitement en supprimant les variations liées aux données du trafic historiques. Comme le montre la figure 4.3, l'idée de base derrière l'utilisation de l'apprentissage par transfert est de construire une adaptation de modèle à partir des données originales où il n'est pas impératif d'appliquer des jeux de données originaux lorsque de nouveaux jeux de données sont combinés. En particulier, la connaissance extraite de l'adaptation du modèle original (domaine source) peut être utilisée pour résoudre le problème de divergence des données et produire une adaptation du modèle en temps réel (domaine cible). Pour adapter les jeux de données supplémentaires, il suffit de réformer l'adaptation

du modèle d'origine avec des jeux de données supplémentaires et un scénario d'apprentissage pour obtenir l'adaptation du modèle.

Par exemple, pour entraîner les données de flux du trafic pour le but de prédiction, les analyses des fluctuations des ensembles de données historiques du trafic comme les jeux de données mensuels entraînent des résultats d'apprentissage différents. Étant donné que les données du trafic sélectionnées peuvent être variantes dans des périodes spécifiques auxquelles, des gros embouteillages se produisent liés aux changements des conditions de circulation, aux mauvaises conditions météorologiques, aux événements sportifs, aux vacances, etc. Cela peut entraîner de grandes anomalies dans la prévision des données de circulation. L'adaptation du modèle dans ces circonstances peut être déployée par des systèmes du trafic intelligents, tels que Dynamic Intelligent Traffic Light Control (DITLC), pour adapter les scénarios d'apprentissage du trafic, contrôler les décisions des véhicules, réduire les accidents des véhicules, éviter la congestion routière, fournir des conseils fiables pour les véhicules, etc.

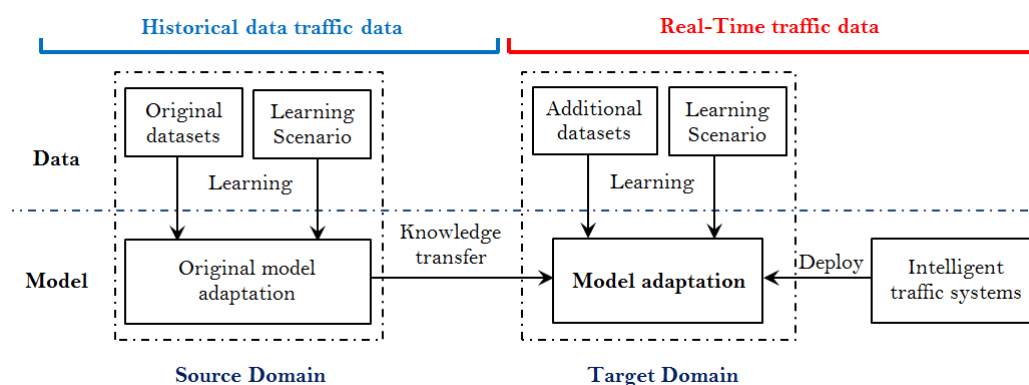


FIGURE 4.3 – L'adaptation du modèle de données du trafic.

### 1.3 La phase d'évaluation

L'évaluation des données est importante pour fournir des données du trafic de haute qualité aux voyageurs, telles que le temps de déplacement, l'état réel du trafic, les incidents routiers, etc. De plus, ces données peuvent être utilisées pour évaluer de manière fiable et précise les performances des applications en heures de pointe et améliorer la gestion du trafic.

## 2 Modèle d'apprentissage flou

Nous décrivons ici un modèle d'apprentissage flou nommé Choquet-MLR et montrons comment il peut être utilisé pour prédire les données du flux du trafic. Essentiellement, les données du trafic collectées à partir de sources de données multiples sont considérées dans un intervalle de temps court. Cet avantage permet de collecter des échantillons de données du trafic en fonction de leur passé récent. Par exemple, la vitesse collectée des véhicules en trois minutes présente le passé proche du paramètre vitesse. Ainsi, le comportement de chaque échantillon de données simule le comportement du flux du trafic.

Nous avons discuté dans les chapitres précédents de la complexité de la prédiction du flux du trafic qui varie selon l'influence d'autres paramètres du trafic. Evidemment, la précision potentielle de ces paramètres ajoutera une valeur de précision. Cela peut être fait en explorant les relations entre les données de chaque paramètre du trafic. En particulier, un modèle d'interaction entre les attributs caractéristiques de chaque échantillon de données présente



une relation clé pour détecter la précision potentielle de chaque paramètre du trafic. Cette perspective va créer des paramètres prédictifs pour le flux du trafic.

Cependant, les modèles traditionnels d'apprentissage prédictif du flux du trafic dépendent de la stabilité des paramètres du trafic. Par exemple, la stabilité du paramètre de vitesse indique une vitesse constante entre les véhicules. Cela conduit à une prédiction optimale du trafic. Dans des conditions synchronisées, les paramètres vitesse et le flux du trafic présentent une relation faible en raison de l'observation inconstante des détecteurs de boucle. La difficulté de la prédiction des flux du trafic dans ces circonstances provoque un taux d'erreur de prédiction plus élevée. Pour réduire l'erreur de prédiction, explorer la relation entre les prédicteurs de flux du trafic est cruciale pour une prédiction optimale. Dans ce qui suit, nous présentons la généralisation de l'analyse MLR par la fonction d'agrégation de l'intégrale Choquet pour ajuster les paramètres de flux du trafic et réduire l'erreur de la prédiction. Tout d'abord, nous présentons les concepts de base de l'analyse MLR et l'intégrale Choquet. Ensuite, nous présentons le modèle Choquet-MLR pour une prédiction précise des données du trafic.

## 2.1 La régression linéaire multiple (MLR)

L'analyse de la régression linéaire multiple (MLR) est une méthode statistique permet d'assurer la précision des solutions heuristiques. Grâce à leur utilisation optimale des données, la sortie de l'analyse interprètera les phénomènes. En particulier, l'analyse MLR utilise deux ou plusieurs variables indépendantes (également appelées variables prédictives) comme entrées pour prédire la variable dépendante (variable de réponse). Les variables indépendantes sont utilisées pour prédire les causes des phénomènes. La variable dépendante présente le phénomène à prédire. Dans l'analyse de la régression linéaire simple (SLR), la prédiction de la variable dépendante dépend d'une seule variable indépendante. Contrairement à l'analyse SLR, l'analyse MLR détecte les comportements complexes des phénomènes (c'est-à-dire la variable dépendante) à partir de deux ou plusieurs variables indépendantes. En fait, l'analyse MLR est largement utilisée dans les systèmes de transport intelligents pour prédire l'état réel du flux du trafic en fonction de variables indépendantes telles que le temps de déplacement du véhicule et la vitesse des véhicules (Bai et al, 2015). L'équation de prédiction générale de l'analyse MLR est définie comme suit :

$$\widehat{y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (4.1)$$

Où  $\widehat{y}$  désigne la variable dépendante;  $X_1, X_2, \dots, X_k$  les variables prédictives;  $k$  indique le nombre de variables prédictives;  $\alpha, \beta_1, \beta_2, \dots, \beta_k$  désigne les coefficients de corrélation;  $\epsilon$  désigne l'erreur de régression qui représente l'écart entre le niveau de corrélation entre les données prédites et les données exactes. Formellement, l'analyse MLR vise à déterminer une relation approximative entre  $\widehat{y}$  et  $X_1, X_2, \dots, X_k$ . Les variables indépendantes du modèle de régression  $X_1, X_2, \dots, X_k$  sont également connus en tant que variables prédictives, de variables expliquées ou de régresseurs. La variable dépendante ( $\widehat{y}$ ) est également appelée variable prédite ou variable de réponse.

## 2.2 La mesure floue et l'intégral floue de Choquet

L'intégrale floue Choquet est bien connue sous le nom d'une intégrale non linéaire pour combiner les informations provenant de plusieurs sources de données. L'intégrale Choquet intègre une fonction par rapport à une mesure floue  $\mu$ , où  $\mu$  est une fonction monotone sur un ensemble. La difficulté de l'intégrale est de savoir comment modéliser efficacement les interactions entre  $2^m - 2$  dimensions d'attributs caractéristiques. Cette complexité augmente

exponentiellement avec les dimensions des attributs caractéristiques. L'intégral de Choquet par rapport à une mesure floue  $\mu$  peut être adoptée dans différents contextes qui méritent une optimisation dynamique telle que la prise de décision multicritères (Mardani et al, 2015), les problèmes d'optimisation combinatoire multi-objectifs (Lust et Rolland, 2014) et la fusion de données (Su et al, 2011). L'originalité principale de l'intégrale est la notion de mesure floue et la généralisation des opérateurs d'agrégations classiques tels que l'opérateur moyen pondéré (WAO), la moyenne pondérée et les autres opérateurs arithmétiques (Yager 1988, 1993 et 1994). Avant de présenter la notation de l'intégrale Choquet, nous présentons ci-après les notions formelles de la mesure floue de l'intégrale Choquet.

**Définition 1 :** mesure floue

Soit  $X = \{x_1, x_2, \dots, x_n\}$  les attributs caractéristiques des mesures.  $\mathcal{P}(X)$  l'ensemble de puissance sur  $X$ . L'intégrale floue de Choquet utilise une fonction monotone, appelée aussi mesure floue,  $\mu : \mathcal{P}(X) \rightarrow [0, 1]$  pour représenter les interactions entre les attributs caractéristiques. Dans lequel, la définition de la mesure floue devrait satisfaire les axiomes suivants :

$$\begin{cases} \mu(\emptyset) = 0, \mu(X) = 1, & \text{Les conditions limites} \\ A \subseteq B \Rightarrow \mu(A) \leq \mu(B), & \text{Monotonie} \end{cases} \quad (4.2)$$

En particulier,  $A$  et  $B$  sont deux sous-ensembles et le poids  $\mu(A)$  représente l'importance de l'ensemble des attributs caractéristiques. La condition limite interprète que l'absence d'attributs caractéristiques (c.-à-d., ensemble vide) a une importance égale à 0 et la présence de tous les attributs a une importance maximale lorsqu'il est égal à 1. Cependant, la condition de monotonie garantit l'importance d'une combinaison ou d'un sous-ensemble où un nouvel attribut caractéristique est ajouté.

La mesure floue  $\mu(\cdot)$  comporte trois formes d'interactions : additif, superadditif et subadditif. *Additive*, si  $\mu(A \cup B) = \mu(A) + \mu(B)$  chaque fois que  $A \cap B = \phi$ . *Superadditive*, si  $\mu(A \cup B) \leq \mu(A) + \mu(B)$  et  $A \cap B = \phi$ . *Subadditif* si  $\mu(A \cup B) \leq \mu(A) + \mu(B)$  chaque fois que  $A \cap B = \phi$  et  $A \cap B = \phi$ .

Par exemple, la faculté d'éducation vise à évaluer les élèves en deux ensembles de sujets : l'ensemble scientifique { mathématique, physique } et l'ensemble langage { français }. Les bons élèves expriment que leurs scores étaient élevés dans un ensemble de sujets et bas pour l'autre ensemble. Le but de la faculté est de trouver les meilleurs étudiants qui ont la capacité de maîtriser, à la fois, les matières scientifiques et linguistiques, c'est-à-dire que leurs scores dans les deux ensembles ne sont pas faibles. Formellement, le but peut être exprimé par l'axiome *superadditif* de la mesure floue, puisque les mathématiciens (ou les physiciens) partagent un effet *superadditif* par rapport la langue. Par conséquent, l'importance de la combinaison devrait être égale ou supérieure à l'importance des attributs individuels. De même, les attributs caractéristiques de l'ensemble scientifique sont redondants (ou compétitifs) l'un de l'autre et interagissent de manière négative, cela peut être exprimé par l'axiome subadditif de la mesure floue. En outre, l'importance de la combinaison devrait être inférieure à la somme de l'importance individuelle des attributs caractéristiques.

**Définition 2 :** L'intégrale de Choquet

Soit  $\mu$  une mesure floue et  $f = \{f(x_1), f(x_2), \dots, f(x_n)\}$  sont des valeurs de mesure non négatives sur  $X$ . L'équation générale du l'intégrale de Choquet est définie comme suit :

$$\int f d\mu = \int_{-\infty}^0 [\mu(F_\alpha) - \mu(X)] d\alpha + \int_0^{+\infty} \mu(F_\alpha) d\alpha \quad (4.3)$$

$F_\alpha = \{x | f(x) \geq \alpha, x \in X\}$ ,  $\alpha \in [0, \infty)$  est appelé un fonction d'ensemble  $\alpha$ -cut, une fois que



la mesure floue  $\mu(x)$  est bien définie, l'intégrale discret de Choquet (CI) peut être appliquée pour agréger les mesures non négatives  $f(x_1), f(x_2), \dots, f(x_n)$ , comme indiqué dans l'équation suivante :

$$\int f d\mu = \sum_{i=1}^n (f(x_{(i)}) - f(x_{(i-1)}))\mu(X_{(i)}) \quad (4.4)$$

Où  $\{x_1, x_2, \dots, x_n\}$  indique une permutation sur  $X$ , de sorte que les valeurs mesurables doivent être commandées comme suit :  $f(x_1) \leq f(x_2) \leq \dots \leq f(x_n)$  and  $f(x_0) = 0$ .

### 2.3 Le modèle de prédiction Choquet-MLR

Dans cette section, nous proposons un modèle de prédiction Choquet-MLR pour améliorer la performance de prédiction de l'analyse MLR en intégrant la fonction d'agrégation de CI. Le nouveau modèle de prédiction va ajuster les paramètres de flux du trafic dans le processus de prédiction et va augmenter le niveau d'approximation entre la réponse et les variables prédictives, réduisant ainsi l'erreur (résiduelle) de la régression. En particulier, la puissance de la mesure floue est intégrée pour représenter l'interaction non linéaire entre les variables prédictives et la réponse. Dans notre modèle de prédiction, soit  $x_1, x_2, \dots, x_n$  représentent les variables prédictives (les attributs caractéristiques) et  $y$  désigne la réponse (l'attribut objectif). En outre, les échantillons de données collectés au moment donné  $j$  sont connus sous le nom de jeux de données (ou jeux d'entraînement), où chaque jeu de données se compose de  $n$  paramètres prédictifs,  $k$  attributs de  $x_1, x_2, \dots, x_n$  et de réponse  $y$ . Les attributs des jeux de données d'entraînement sont structurés selon le format suivant :

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n & y \\ f_{11} & f_{12} & \dots & f_{1n} & y_1 \\ f_{21} & f_{22} & \dots & f_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{k1} & f_{k2} & \dots & f_{kn} & y_k \end{pmatrix}$$

Où,  $f_{i1}, f_{i2}, \dots, f_{in}$  désignent les mesures (les observations) du  $i$ -ème attribut de l'ensemble de données d'entraînement;  $K$  désigne la taille de données d'apprentissage qui doit être supérieure au nombre d'attributs de fonctionnalité.

Pour modéliser les interactions entre les attributs caractéristique et représenter la dépendance non linéaire entre la variable de réponse ( $y$ ) et les paramètres prédictifs  $x_1, x_2, \dots, x_n$ . Nous proposons d'étendre le modèle MLR en remplaçant la fonction linéaire  $x \rightarrow \beta^T x$  du modèle MLR par la fonction d'agrégation de l'intégrale de Choquet. Comme le montre l'équation suivante :

$$y_j = c + \int (\beta \cdot f) d\mu + \epsilon \quad (4.5)$$

Où  $y_j$  représente l'attribut objectif au moment  $j$ ;  $c$  est une constante;  $f d\mu$  est la fonction intégrale de Choquet par rapport à la mesure floue  $\mu$ .  $\epsilon$  est généralement l'erreur de la régression.  $\beta (b_1, b_2, \dots, b_n)$  est un vecteur de  $n$  paramètres de régression associés aux attributs de la fonction. Le vecteur  $\beta$  indique les valeurs d'importance des variables prédictives dans le processus de prédiction, c'est-à-dire les paramètres de régression du vecteur présentent les effets marginaux des variables prédictives sur la réponse  $y_j$ . Dans cette étude, les paramètres de régression sont contrôlés par l'intégrale de Choquet et utilisés pour interagir avec les attributs caractéristiques des données collectées. En fait, le but principal des paramètres de régression est de minimiser la somme de l'erreur carrée entre les données exactes et les

données attendues. Avec la présence des valeurs précises des paramètres de régression, le nouveau modèle prédictif garantira une meilleure approximation entre les données exactes et les données attendues. En d'autres termes, ceci peut être fait en déterminant les valeurs des paramètres de régression qui résolvent le problème de minimisation suivant :

$$\begin{aligned}
 & \text{minimize} && \sum_{i=1}^n (y_i - c - \int (\beta \cdot f) d\mu)^2 \\
 & \text{subject to} && 0 < \mu(\{x_1, x_2, \dots, x_k\}) \leq 1 \quad k = 1, \dots, 2^n - 2, \\
 & && -1 \leq b_j \leq 1 \quad j = 1, \dots, n, \\
 & && b_j \geq b_{j+1} \quad j = 1, \dots, n-1, \\
 & && c \in R
 \end{aligned} \tag{4.6}$$

L'algorithme 1 dans ce qui suit vise à minimiser la fonction objective présentée dans l'équation 4.6 et explique pourquoi le Choquet-MLR est proposé. L'algorithme comporte des entrées et six étapes où les données de sortie indiquent la réponse. L'entrée présente les données du trafic collectées à partir de plusieurs sources de données (c'est-à-dire des détecteurs de boucle connectés).

La collection de données se compose de plusieurs variables du trafic (dimensions) telles que la vitesse des véhicules, l'occupation de la route, le nombre de véhicules contribués, etc. D'abord, les dimensions des attributs de chaque collection de données vont contribuer à construire un échantillon de données individuel pour chaque variable. Une fois que les échantillons de données sont générés pour chaque variable du trafic, les données se fragmentent simplement en des ensembles de données d'entraînement et des ensembles de données de test. Les ensembles de données de test visent à identifier le comportement d'apprentissage du chaque variable. La troisième étape de l'algorithme consiste à identifier séparément les mesures floues de chaque échantillon. Dans cette étude, l'identification des meilleures mesures floues repose sur l'algorithme d'apprentissage proposé par (Grabisch, 1995). L'algorithme d'apprentissage sélectionné prend les échantillons de données du trafic et l'ensemble de données de test comme entrées pour identifier les mesures floues qui appartiennent à l'intervalle  $[0, 1]$ . Ensuite, l'intégrale de Choquet est appliquée pour déterminer l'attribut objectif de chaque échantillon. Dans la cinquième étape, l'ensemble de données de test de Choquet peut être formé en combinant les attributs objectifs des variables du trafic dans un seul ensemble de données de test. Une fois que l'ensemble de données test de Choquet est formé, l'analyse Choquet-MLR peut être appliquée pour prédire les données du trafic. L'analyse prend l'ensemble de données de test de Choquet et les ensembles de données d'entraînement des instants  $(t-1)$  et  $(t-2)$  comme entrées pour prédire la réponse. Ensuite, les résultats en sortie sont comparés aux données réelles afin de contrôler les paramètres de régression et réduire l'erreur la régression. Si ces derniers montrent une erreur de prédiction élevée, des ensembles de données d'entraînement supplémentaires sont incorporés par la suite pour ajuster les paramètres de régression. Ce processus sera répété jusqu'à ce qu'une erreur de prédiction minimale soit atteinte.

La figure 4.4 ci-dessous montre l'organigramme proposé pour l'ajustement des paramètres du trafic et la prédiction des données du trafic.

### 3 Les prédicteurs du trafic

Supposons qu'un conducteur, pendant son voyage, souhaite connaître le flux de circulation actuel (q) et son niveau de sécurité, le processus de prédiction regroupe les données observées qui stimulent ensuite pour prédire le flux du trafic. En outre, le tableau 4.3 énumère les

**Algorithm 1** La prédiction de la réponse basée sur les données d'apprentissage

**Require:** Les données du trafic recueillies à partir de  $n$  sources de données où les dimensions de chaque collecte de données représentent  $m$  variables du trafic.

- 1: Créer un échantillon de données du trafic de chaque variable du trafic en combinant le  $i$ -ème attribut caractéristique de chaque collection de données.
- 2: Diviser l'échantillon de données de chaque variable du trafic en plusieurs ensembles de données d'entraînement et un ensemble de données de test.
- 3: Identifier les meilleures mesures floues de chaque échantillon de données en fonction de l'ensemble de données d'entraînement et l'ensemble de données de test.
- 4: Une fois que les mesures floues sont bien définies par l'étape précédente, l'intégrale Choquet peut être appliquée pour chaque échantillon de données afin de déterminer son attribut objectif.
- 5: Une fois que les attributs objectifs sont bien définis pour tous les échantillons de données, ces attributs se combinent pour former l'ensemble de données de test de Choquet.
- 6: Prévoir la réponse en utilisant l'analyse Choquet-MLR. L'analyse prend les données d'entraînement des instants  $(t - 1)$  et  $(t - 2)$  et les jeux de test de Choquet comme entrées. Si les résultats montrent une erreur de prédiction élevée, le MLR ajustera les paramètres de régression en intégrant des ensembles de données d'entraînement supplémentaires. Ce processus sera répété jusqu'à ce que l'erreur de prédiction optimale soit atteinte.

abréviations abrégées des paramètres macroscopiques utilisés pour décrire la collecte de données. En outre, la description des données utilisées dans cette étude traite uniquement des véhicules légers. En particulier, la masse brute du véhicule ne dépasse pas  $3 \times 10^3$  kg et moins de 3 mètres de largeur maximale.

Tableau 4.1 – les termes qui décrit la description de la collecte de données.

Abréviation	Description
$q$	Le trafic du flux.
$Sp_i$	La vitesse individuelle du véhicule au détecteur de boucle.
$L_v$	La longueur du véhicule.
$d$	La longueur de la route.
$Sd$	La distance de sécurité.
$AvgSpeed$	La vitesse moyenne.
$SegLengh$	La longueur du segment.
$Csp$	La vitesse agrégée.
$h_{v1,v2}^{(t)}$	Le temps « headway ».
$\Delta t_p$	L'heure d'arrivée agrégée.
$TimeSlot$	L'intervalle de temps.

### 3.1 Le temps de passage du segment routier

Le temps de progression  $h_{v1,v2}^{(t)}$  est défini comme le temps requis qui doit s'écouler entre deux véhicules consécutifs,  $v1$  et  $v2$  respectivement. Au moment du temps  $j$ , nous désignons

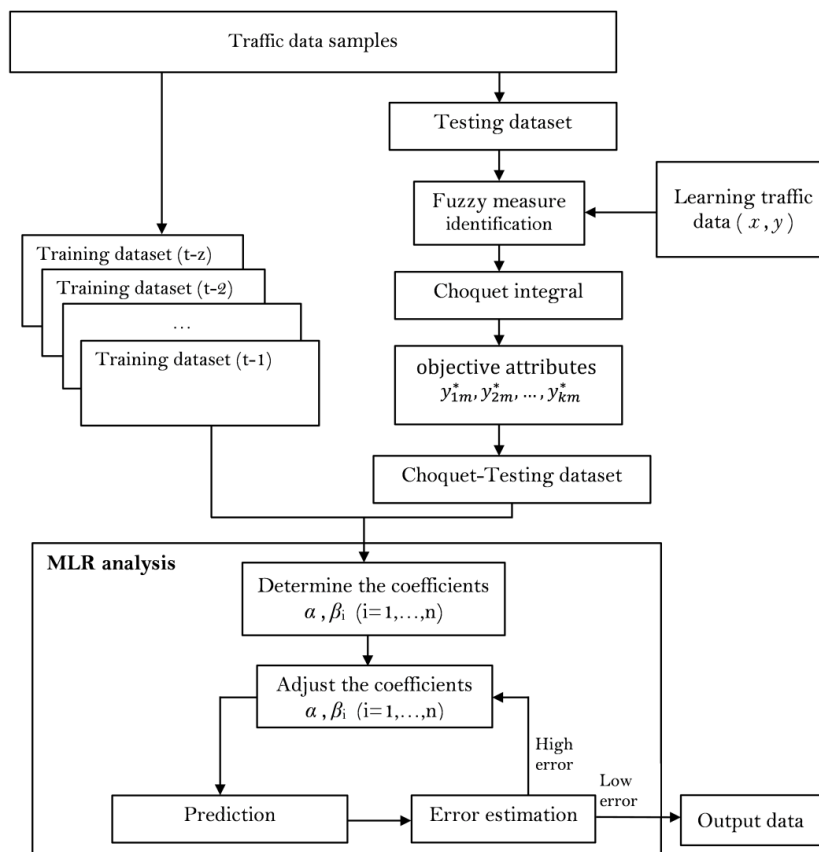


FIGURE 4.4 – L’organigramme du processus de prédiction de la réponse en utilisant Choquet-MLR.

$Sp_j = \{Sp_{1j}, Sp_{2j}, \dots, Sp_{nj}\}$  l’ensemble de la vitesse des véhicules observés via les détecteurs de boucle et que l’on veut également l’agrégés.  $Csp_j(Sp_{1j}, Sp_{2j}, \dots, Sp_{nj})$  désigne l’attribut objectif de la vitesse des véhicules en utilisant l’opérateur intégral Choquet et satisfaisant l’éq. 4.7.

$$Csp_j(Sp_{1j}, Sp_{2j}, \dots, Sp_{nj}) = \sum_{i=1}^n (Sp_{(i)j} - Sp_{(i-1)j}) \mu(\Delta Sp_{(i)j}) \quad (4.7)$$

Où,  $\mu(\Delta Sp_{(i)j})$  est un poids d’importance entre deux vitesses consécutives,  $Sp_{(i)}$  et  $Sp_{(i-1)}$ . Pour estimer le temps de progression, la distance de sécurité et la vitesse moyenne sont nécessaires. En outre, selon le décret français *no. R412-12 (23 novembre 2001)*, la distance de sécurité doit être évaluée en fonction de la vitesse du véhicule, car elle fournit un délai suffisant pour que les conducteurs réagissent à des situations anormales. Par exemple, les véhicules doivent maintenir cette distance pour éviter le risque d’accident et pour détecter une décélération soudaine. En particulier, ce temps de réaction est spécifiquement autour de 2 secondes. En outre, 5/9 de la vitesse du véhicule représentent la distance de sécurité qui doit être maintenue devant le véhicule, comme indiqué dans l’équation suivante :

$$Sd_j = Csp_j(Sp_{1j}, Sp_{2j}, \dots, Sp_{nj}) * 5/9 \quad (4.8)$$

Sur la base de la distance de sécurité calculée conformément à l’équation 4.8, le temps de progression  $h_{Seg(i)}^{(j)}$  du segment de la route au moment  $j$  est effectué par l’équation suivante (Eq .4.9) :

$$h_{Seg(i)}^{(j)} = TimeSlot/Sd_j \quad (4.9)$$

### 3.2 Le prédicteur du temps de déplacement

Pour conduire avec moins de risque, détecter des conditions de circulation plus sûres, on se réfère à la valeur critique de l'importance de la sécurité qui dépend des conditions de circulation spatiotemporelles observées. A titre d'exemple, le fait de ne pas maintenir la distance de sécurité et opérer par une décélération soudaine des véhicules sur la même ligne de la route sont deux conditions de circulation affectant le trafic. Pour cette raison, le temps de déplacement du véhicule présente un paramètre clé permettant au processus de prévision de considérer l'instabilité du flux du trafic. En outre, l'agrégation du temps de déplacement du véhicule de chaque segment en prédicteur  $\Delta t_{p(j)}$  pour tous les segments de la route permettra de détecter quel segment sera influencé par la sécurité routière. De même, en adoptant l'opérateur intégral de Choquet, l'équation Eq.4.10 suivante fournit le niveau de cette précision.

$$\Delta t_{p(j)}(t_{1j}, t_{2j}, \dots, t_{nj}) = \sum_{i=1}^n (t_{ij} - t_{(i-1)j}) \mu(\Delta t_{ij}) \quad (4.10)$$

Lorsque  $t_{1j}, t_{2j}, \dots, t_{nj}$  correspond aux temps d'arrivée observés des véhicules passés au moment  $j$ .  $\mu(\Delta t_{ij})$  implique un poids d'importance entre deux temps de déplacement du véhicule,  $t_{ij}$  et  $t_{(i-1)j}$ , de véhicules consécutifs. En outre, l'agrégation du temps d'arrivée est censée à prédire l'aspect temporel de la circulation.

Indépendamment aux équations Eq.4.9 et Eq.4.10, des nouvelles conditions de circulation en temps réel peuvent être identifiées en comparant le temps d'arrivée des véhicules et le temps de progression. En particulier, si  $\Delta t_{p(j)}(t_{1j}, t_{2j}, \dots, t_{nj})$  est supérieur à  $h_{Seg(i)}^{(j)}$ , les véhicules sont incorporés dans une mobilité plus sûre, sinon le mouvement des véhicules est associé à un facteur de risque pouvant perturber le flux du trafic actuel. Comme le montre l'équation suivante :

$$SegCond_{(i)j} = \begin{cases} \text{if } \Delta t_{p(j)}(t_{1j}, t_{2j}, \dots, t_{nj}) \geq h_{Seg(i)}^{(j)}, & SafeFlow \\ \text{Otherwise,} & UnsafeFlow \end{cases} \quad (4.11)$$

En résumé, cette section définit trois prédicteurs pour le trafic, à savoir la vitesse du véhicule, le temps de progression et le temps de déplacement. En fait, L'identification de ces prédicteurs du trafic est basée sur des données fournies récemment par une source de données multiples. Néanmoins, selon la définition du flux du trafic de (Kerner et Klenov, 2010), ces prédicteurs ne suffisent pas à définir une prédiction précise car ils ne définissent que l'aspect temporel du flux du trafic. En outre, il est difficile de détecter le comportement d'un groupe de véhicules à partir de données observées par les détecteurs de boucle (c'est-à-dire le comportement macroscopique du modèle du trafic actuel) du fait que ces données observées visent uniquement à détecter le comportement de chaque véhicule séparément. Pour compléter ce point de vue, l'aspect spatial du flux du trafic est nécessaire, où le prédicteur spatial de la densité de la route (c'est-à-dire le nombre de véhicules par unité d'espace) est incorporé pour spécifier les modèles (c'est à dire le mouvement libre, la capacité d'écoulement ou la congestion) sont conçus par la prédiction. Ensuite, l'aspect spatio-temporel de la prédiction de flux du trafic sera complété en tenant compte des prédicteurs spatiaux et temporels.

## 4 Expérimentation et évaluation

Dans cette section, une étude expérimentale est faite pour évaluer la méthodologie proposée avec un réseau routier incluant des conditions complexes du trafic. L'étude se concentre principalement sur les ensembles de données empiriques du trafic sauvegardées dans une base de données fournie par un département du transport. En particulier, les informations numériques sont collectées toutes les 5 secondes à partir de détecteurs de boucles (sources de données) installés dans le segment routier (A1) de 3 km sur l'autoroute Est-Ouest, Constantine, Algérie.

Dans cette étude, les données du trafic recueillies aux heures de pointe du matin sont utilisées pour les expérimentations. Les véhicules préfèrent l'autoroute Est-Ouest pour optimiser leur temps de déplacement où ils se déplacent librement, car il n'y a aucun obstacle pour ralentir la vitesse. Les obstacles sont causés par des incidents routiers inattendus et de panneaux d'arrêt temporaires. Pendant leurs voyages, le régime libre de la circulation demande toujours l'augmentation de la vitesse. De même, la densité et le débit du trafic ne dépasse pas la capacité routière. Dans ces conditions normales de circulation, la relation entre le flux du trafic et la densité de la route est positive.

Le but de cette étude expérimentale est de prédire le flux du trafic à partir de sources de données multiples pendant les conditions de trafic complexes. Dans cas là, les variables du trafic ont des relations faibles en raison de leurs comportements instables. Tester la méthodologie de prédiction proposée dans ces conditions est nécessaire pour démontrer et vérifier l'efficacité de la méthode Choquet-MLR. La figure 4.5 représente la vitesse des véhicules et le flux du trafic capturé à partir de trois sources de données. Le modèle de vitesse de la figure 4.5.a indique que la vitesse moyenne du flux du trafic est variable dans chaque source. Le modèle de flux du trafic dans la figure 4.5.b montre le nombre de véhicules contributeurs à 08h00. Selon la description précédente, nous adoptons notre méthodologie de prédiction dans le modèle expérimental où la prédiction précise des données aura une valeur importante pour aider les conducteurs à connaître les conditions actuelles du trafic et les orientations rapides.

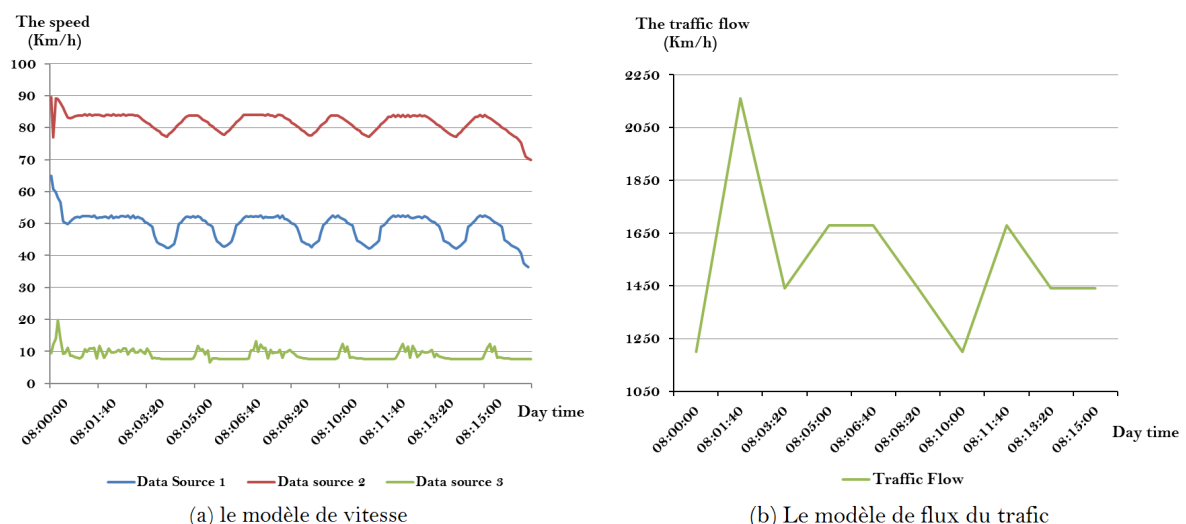


FIGURE 4.5 – La description du modèle de flux du trafic.

Selon les trois dimensions des caractéristiques, toutes les mesures floues pour la vitesse ( $S_p$ ), le temps de déplacement  $\delta t_p$  et les ensembles de données de la densité de la route ( $\rho$ ) sont attribuées dans le tableau 4.3. Les coalitions des coefficients sont affichées dans la première colonne et la valeur du coefficient de la vitesse, le temps de déplacement et la densité de la

route sont indiqués dans les colonnes 2 à 4.

Tableau 4.2 – Coefficients pour agréger les variables du trafic.

Coefficients	$Sp$	$\delta t_p$	$\rho$
$\mu(\{\phi\})$	0	0	0
$\mu(\{x_1\})$	0.4	0.25	0.3
$\mu(\{x_2\})$	0.45	0.25	0.4
$\mu(\{x_3\})$	0.4	0.2	0.1
$\mu(\{x_1, x_2\})$	0.9	0.9	0.9
$\mu(\{x_1, x_3\})$	0.6	0.35	0.45
$\mu(\{x_2, x_3\})$	0.7	0.35	0.5
$\mu(\{x_1, x_2, x_3\})$	1	1	1

#### 4.1 Les performances d'agrégation

Dans ce qui suit, nous comparons les performances d'agrégation des prédicteurs de flux du trafic en utilisant l'agrégation non linéaire (opérateur Choquet) et en utilisant l'interaction linéaire telle que l'opérateur moyen pondéré (WAO). La fonction d'agrégation linéaire de l'opérateur moyen pondéré (WAO) est conforme à la forme suivante :

$$WAO(y_1, y_2, \dots, y_n) = \sum_{i=1}^n w_i y_i \quad (4.12)$$

Nous choisissons la vitesse des véhicules comme prédicteur de référence dans la comparaison numérique. La puissance d'agrégation de chaque opérateur est présentée dans le tableau 4.3. L'attribut objectif calculé par WAO est le plus proche de l'attribut de la dimension DS2. Il indique que le processus de prédiction va utiliser l'attribut caractéristique de DS2 comme référence pour la précision des données. L'attribut objectif calculé par l'IC indique que l'interaction entre les attributs caractéristiques DS1 et DS2 est importante pour la précision des données. Pour montrer la contribution de chaque attribut objectif, la figure 4.6 montre l'influence de chaque attribut objectif sur la précision de la prédiction.

#### 4.2 Les critères d'évaluation de la prévision

Cette étude sert à juger l'exactitude de la prédiction par la méthode Choquet-MLR proposée. Les performances de la méthode sont évaluées selon les trois critères les plus couramment utilisés, c'est-à-dire l'erreur de pourcentage absolue moyenne (MAPE) (Chen, 2001), Root Mean Square Error (RMSE) (Wei, 2007) et EC Index (Wang, 2005). En fait, ces mesures sont couramment utilisées pour évaluer l'erreur absolue relative à la prédiction du flux du trafic à court terme. Les mesures sont définies comme suit :

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{N} \right| \quad (4.13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y(t) - \hat{Y}(t))^2} \quad (4.14)$$



Tableau 4.3 – Comparaison entre la puissance d'agrégation du prédicteur de vitesse à l'aide d'opérateurs CI et WAO.

	DS1	DS2	DS3	CI	WAO
01	50.43	81.55	10.98	62.04	47.65
02	50.21	81.04	9.64	61.56	46.96
03	49.61	80.31	7.73	60.77	45.88
04	48.98	79.78	7.93	60.27	45.56
.				.	.
.				.	.
.				.	.
105	45.83	79.15	7.78	58.68	44.25
106	44.39	78.63	7.71	57.84	43.57
.				.	.
.				.	.
138	46.95	81.66	7.59	60.369	45.4
139	49.5	82.36	7.59	61.73	46.48
140	50.05	83.16	7.59	62.35	46.93

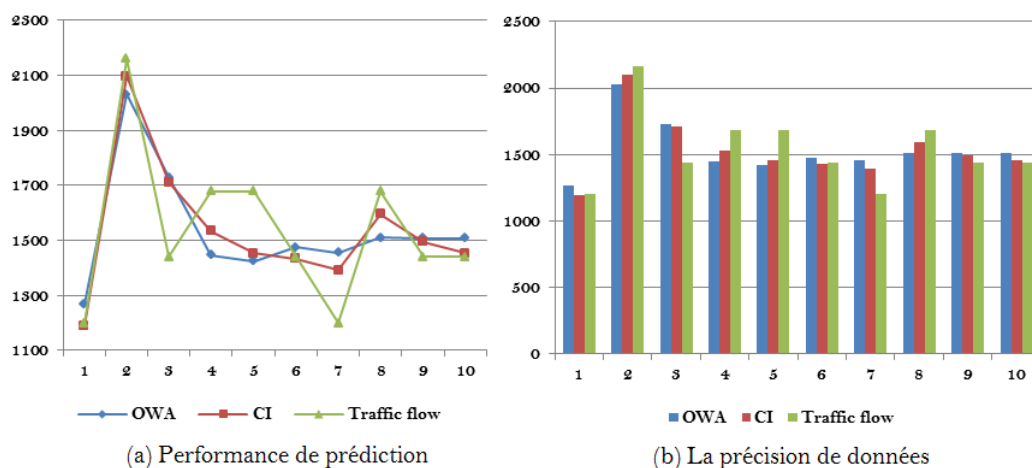


FIGURE 4.6 – La prévision du flux du trafic en utilisant l'agrégation non linéaire (IC) par rapport à l'agrégation linéaire (WAO).

$$EC = 1 - \frac{\sqrt{\sum_{i=1}^n (Y_i(t) - \hat{Y}_i(t))^2}}{\sqrt{\sum_{i=1}^n Y_i^2 + \sum_{i=1}^n \hat{Y}_i^2}} \quad (4.15)$$

Où  $Y_i$  désigne le nombre de véhicules observés;  $\hat{Y}_i$  fait référence aux valeurs prédites. Le tableau 4.4 ci-dessous montre la comparaison entre les performances de précision obtenues par la méthode Choquet-MLR proposée et les performances obtenues par les autres méthodes de prédiction telles que ARIMA (Autoregressive integrated moving average), les séries temporelles à base réseau de neurones multicouches (TSP-MNN) et l'agrégation linéaire à base régression multivariée (MLR-LA).

Pour comparer les performances, ARIMA, TSP-MNN et MLR-LA sont également appliquées aux mêmes données utilisées pour évaluer la méthode Choquet-MLR. Il est clair que la



précision obtenue par la méthode Choquet-MLR dépasse les autres méthodes de prévision à l'aide des metrics MAPE, RMSE et EC. En particulier, l'erreur de prédiction de la méthode Choquet-MLR est comprise entre (5.14 à 10.67) et (117.69 à 228.82) en termes de MAPE et de RMSE respectivement. Par conséquent, les performances sont meilleures dans presque tous les cas par rapport l'erreur de prédiction obtenue par ARIMA (MAPE : 6.04 à 32.76 - RMSE : 141.41 à 702.67), TSP-MNN (MAPE : 20.53 à 37.95 - RMSE : 754.46 à 1025.56) et MLR-LA (MAPE : 6,28 à 15,71 - RMSE : 168,51 à 257,98).

En utilisant la méthode Choquet-MLR, l'erreur de prédiction est réduite 25% par la metric MAPE et de 16% par la metric RMSE. En outre, on remarque que l'erreur de prédiction de la méthode Choquet-MLR avec les métriques précédentes est cohérente dans tous les cas de prédiction par rapport aux autres méthodes.

De plus, selon la métrique EC, le taux d'approximation des données prédites aux données exactes de flux du trafic obtenues par la méthode Choquet-MLR appartient à l'écart (98,91% - 99,71%). Donc, il est cohérent par rapport aux autres écarts d'approximation telles que ARIMA avec l'écart (90,64% - 99,58%), TSP-MNN avec l'écart (43,2% à 99,53%) et MLR-LA avec l'écart (98,61% - 99,56%). En conséquence, la combinaison de l'analyse MLR avec l'opérateur intégral Choquet présente une amélioration remarquable par rapport aux méthodes actuelles de prédiction des flux du trafic et des performances fiables dans des conditions du trafic complexes par rapport à d'autres méthodes de prédiction à court terme.

### 4.3 L'impact du volume de données sur le processus de prédiction

Dans cette section, nous voulons montrer l'impact du volume de données du trafic sur la performance de prédiction par le modèle Choquet-MLR. En particulier, nous cherchons à tester la complexité de la méthode proposée dans les heures de pointe pour modéliser efficacement les interactions entre  $2^n + n$  attributs caractéristique. Cette complexité augmente exponentiellement avec les dimensions des attributs de la fonction.

Pour gérer la complexité de la méthode Choquet-MLR, nous devons définir les entrées, les intervalles de prévision, les jeux de données et la définition des mesures floues pour le modèle. La méthode de prédiction utilise les données du trafic collectées provenant de plusieurs sources de données comme entrées. Les données du trafic sont collectées et agrégées avec deux courts intervalles de 60 secondes et 180 secondes pour chaque source de données. Pour prédire le flux du trafic à l'instant  $T$ , les ensembles de données du trafic des instants  $(T - 1)$  et  $(T - 2)$  sont sélectionnés comme ensembles de données d'entraînement et l'ensemble de données de l'instant  $T$  est sélectionné comme ensemble de données de test.

En conséquence, les figures 4.7, 4.8 et 4.9 illustrent les performances de prédiction en utilisant la méthode proposée et en fonction des intervalles de temps courts d'agrégation (c'est-à-dire 60s et 180s). La méthode est utilisée pour prédire le flux du trafic en 60 minutes et notamment le flux du trafic quotidien. La figure 4.7 montre le flux du trafic prévu pour une période ne dépassant pas 60 minutes. Les performances de prédiction avec un intervalle de temps de 60s (figure 4.7.a) est plus précise par rapport aux performances en 180s d'intervalle de temps (figure 4.7.b). Cette preuve confirme l'adoption de la méthode proposée pour prédire les flux du trafic à court terme. De même, les performances montrées dans la figure 4.8 rapportent un niveau plus proche des données en temps réel pour la prévision quotidienne. En outre, la figure 4.9 représente le niveau de corrélation entre le flux du trafic observé et son flux du trafic prévu. Les performances présentées dans la figure (4.9.a) montrent un niveau d'approximation plus élevé entre les données observées et les données attendues en 60s d'intervalle de temps. Les résultats obtenus sont plus précis par rapport au niveau d'approximation dans la figure (4.9.b), où les données sont attendues en intervalle de temps de 180s. Cette forte corrélation indique la validité de l'hypothèse selon laquelle l'analyse

Tableau 4.4 – Comparaison entre la méthode Choquet-MLR proposée avec les autres méthodes de prédiction.

Prediction interval	ErrorType	MLR – LA	TSP – MNN	ARIMA	Choquet – MLR
08 :00 :00 -	MAPE	15.71	37.95	11.63	<b>10.67</b>
08 :01 :40	RMSE	181.34	1025.56	279.93	<b>140.85</b>
	EC (%)	99.31 %	43.20 %	96.82 %	<b>99.59%</b>
08 :01 :40 -	MAPE	11.12	22.26	10.63	<b>5.16</b>
08 :03 :20	RMSE	234.44	775.27	237.85	<b>124.12</b>
	EC (%)	98.87 %	85.69 %	98.82 %	<b>99,68 %</b>
08 :03 :20 -	MAPE	8.54	22.87	32.76	<b>5.82</b>
08 :05 :00	RMSE	206.11	847.88	702.67	<b>149.83</b>
	EC (%)	98.93 %	99.53 %	90.64 %	<b>99.53 %</b>
08 :05 :00 -	MAPE	7.79	24.5	9.41	<b>5.79</b>
08 :06 :40	RMSE	182.48	866.49	216.69	<b>135.67</b>
	EC (%)	99.52 %	82.94 %	99.03 %	<b>99.63 %</b>
08 :06 :40 -	MAPE	10.16	20.53	11.18	<b>8.58</b>
08 :08 :20	RMSE	236.13	754.46	242.57	<b>206.94</b>
	EC (%)	98.81 %	86.35 %	98.75 %	<b>99.09%</b>
08 :08 :20 -	MAPE	7.79	22.98	8.79	<b>5.7</b>
08 :10 :00	RMSE	195.5	847.89	194.41	<b>136.18</b>
	EC (%)	99.07 %	82.62 %	99.22 %	<b>99.61 %</b>
08 :10 :00 -	MAPE	7.88	24.32	14.76	<b>5.74</b>
08 :11 :40	RMSE	179.09	851.92	451.66	<b>132.11</b>
	EC (%)	99.56 %	83.26 %	95.69 %	<b>99.65 %</b>
08 :11 :40 -	MAPE	12.18	21.46	<b>6.04</b>	10.37
08 :13 :20	RMSE	257.98	782.98	<b>141.41</b>	228.82
	EC (%)	98.61 %	85.48 %	<b>99.58%</b>	98.91 %
08 :13 :20 -	MAPE	6.28	22.89	32.76	<b>5.14</b>
08 :15 :00	RMSE	176.25	847.82	702.67	<b>117.69</b>
	EC (%)	99.27 %	82.37 %	90.64 %	<b>99.71 %</b>
08 :15 :00 -	MAPE	6.88	23.50	7.51	<b>6.06</b>
08 :16 :40	RMSE	168.51	855.30	195.59	<b>142.17</b>
	EC (%)	99.48 %	84.70 %	99.30 %	<b>99.63 %</b>

Choquet-MLR est adoptée pour prédire le trafic en temps réel.

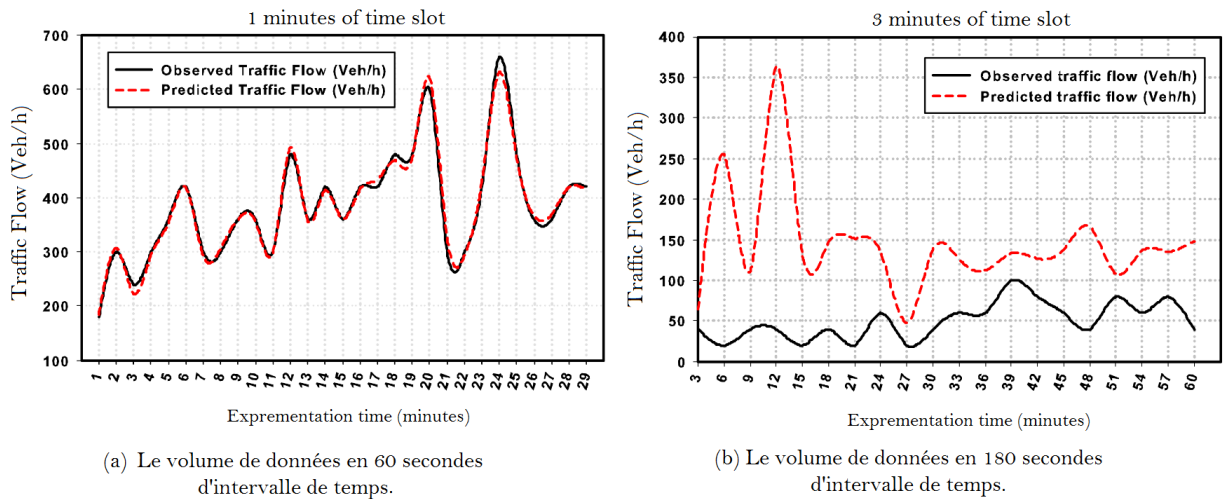


FIGURE 4.7 – La prévision à court terme du trafic.

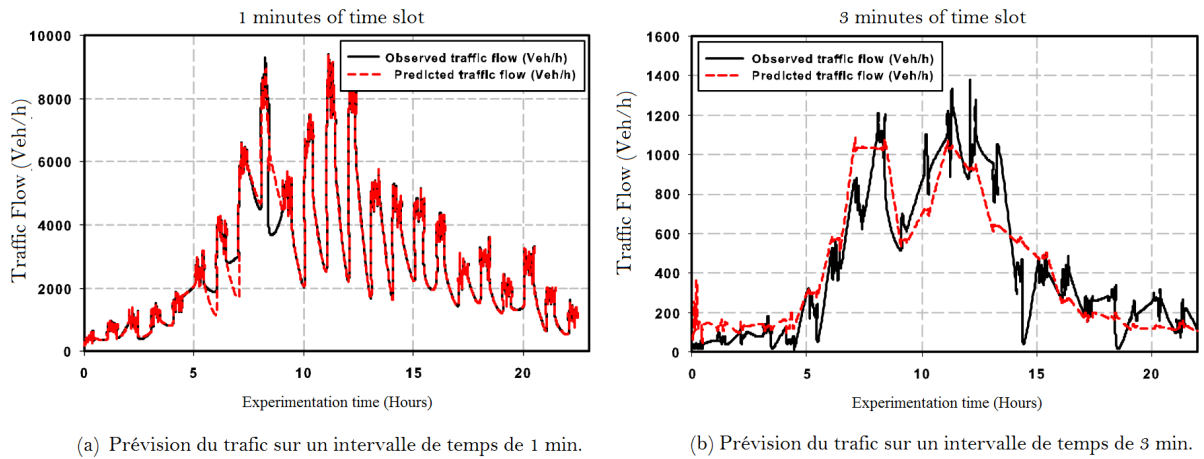


FIGURE 4.8 – Prévision quotidienne du trafic.

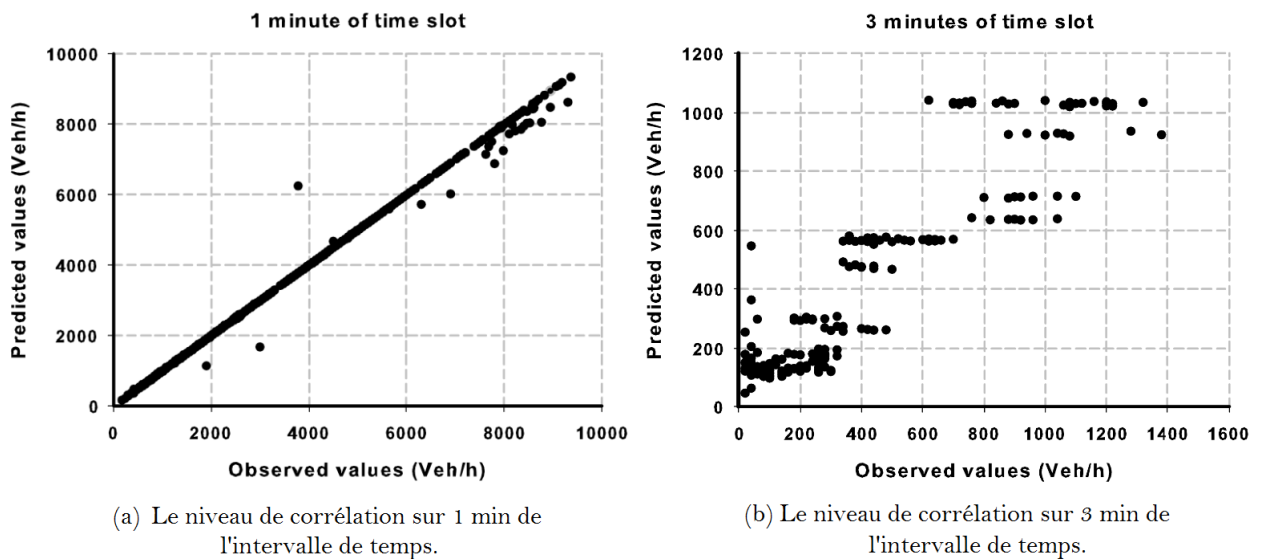


FIGURE 4.9 – Les données du trafic observé versus les données prédites en 1 min et 3 min de temps.

---

## CONCLUSION

Dans ce chapitre, nous avons proposé un système de fusion de données basé sur un modèle d'apprentissage flou pour une prédiction précise à court terme du flux du trafic. La fusion est procédée en fonction de plusieurs sources de données de trafic. Le système proposé comprend quatre niveaux de fusion : préparation de données, extraction / sélection de caractéristiques, la prise de décision et adaptation de modèle. Les niveaux de fusion visent à réduire les anomalies de données (i.e., erroné, bruyant, incomplet, etc.) qui peuvent diminuer la qualité du modèle prédictif, éviter le sur-apprentissage du modèle et améliorer le temps de traitement. L'objectif principal de ces niveaux est de fournir des ensembles de données fiables en tant qu'entrées du modèle d'apprentissage.

Notre modèle prédictif est basé sur la généralisation de la Régression Linéaire Multiple (MLR) par la fonction linéaire de l'intégrale de Choquet. Dans l'étape d'évaluation, nous avons évalué le modèle prédictif proposé en fonction de données du trafic empirique fournie par le département du transport, L'objectif est de prédire le flux du trafic dans des conditions de circulation complexes (les congestions). Les performances montrent des améliorations par rapport aux autres modèles existants d'apprentissage prédictif. La précision de la prédiction de chaque modèle est évaluée en fonction des critères de performances tels-que MAPE, RMSE et EC. Par conséquent, nous soulignons que le modèle d'apprentissage proposé (Choquet-MLR) prouve son adoption pour la prévision précise du trafic en temps réel.

# Une plateforme Vanet-Cloud pour la gestion des informations du trafic

## SOMMAIRE: CHAPITRE 5

Introduction	83
1 Une plateforme VANET-CLOUD pour la gestion du trafic	84
1.1 Une couche de collecte des données du trafic	84
1.2 Infrastructure en tant que couche de service (IaaS)	85
1.3 Une plate-forme en tant que couche de service	85
1.4 Une couche de service VANET-Cloud	85
2 Modèle d'apprentissage non linéaire monotone en utilisant la régression de Choquet	86
3 Mécanisme de diffusion de données avec la communication V2C	87
3.1 Le mécanisme d'échange de données	89
3.2 Les algorithmes de traitement des messages	90
3.3 L'organigramme de flux de données	90
4 Simulation et évaluation	92
4.1 Une forme étendue de la plateforme iCanCloud pour évaluer les performances des architectures VANET-Cloud	95
4.2 Taux de livraison de données	96
4.3 La haute performance de communication	97
4.4 La latence du service Cloud	97
Conclusion	99

<b>Conclusion générale</b>	100
<b>Perspectives</b>	102
<b>Liste Des Publications de Nos Travaux</b>	103
<b>Bibliographie</b>	104

---

---

## INTRODUCTION

Anticiper les informations sur le trafic est une tâche indispensable pour réduire la congestion routière et diminuer le temps perdu durant les heures de pointe. Dans ce chapitre, une couche de services Cloud est proposée pour augmenter la sécurité routière, améliorer les informations du trafic et optimiser les performances du réseau dans des conditions d'encombrement. En particulier, l'architecture de gestion que nous proposons repose principalement sur l'infrastructure Cloud et la prolifération du réseau de capteurs connectés autour des villes. Le réseau de capteurs est utilisé pour collecter des données du trafic, alors que le système du Cloud est utilisé pour fournir deux types de services de trafic : des services à la demande (ex : temps de déplacement du véhicule) et les services automatiques (ex : système de signalisation intelligent). Les services du trafic utilisent le mécanisme de prédiction de données, proposé dans le chapitre précédent, pour prévoir les données du trafic en temps réel et le mécanisme d'échange de données pour diffuser les données prédites. Le mécanisme d'échange de données est exploité via la communication entre les véhicules et le Cloud (V2C). Dans l'étape d'évaluation, en se basant sur une description d'un scénario, les résultats de la simulation montrent l'efficacité de la couche proposée en termes de taux de livraison, la haute performance de communication (HPC) et la latence du service Cloud par rapport aux réseaux VANET traditionnels.

# 1 Une plateforme VANET-CLOUD pour la gestion du trafic

Cette section a pour but de décrire notre infrastructure VANET-Cloud pour la gestion du trafic, comme le montre la figure 5.1. L'architecture proposée vise à mieux utiliser l'infrastructure routière et à réduire le nombre d'anomalies associées à la détection des données et la qualité des données diffusées. De plus, l'architecture intègre des détecteurs de trafic connectés pour capturer les données du trafic en temps réel. Il montre l'évolution des architectures VANET classiques vers un modèle de services multicouches. En particulier, l'architecture mentionne trois couches principales comprenant, la couche de collecte de données du trafic, l'infrastructure en tant que couche de service et la couche VANET-Cloud.

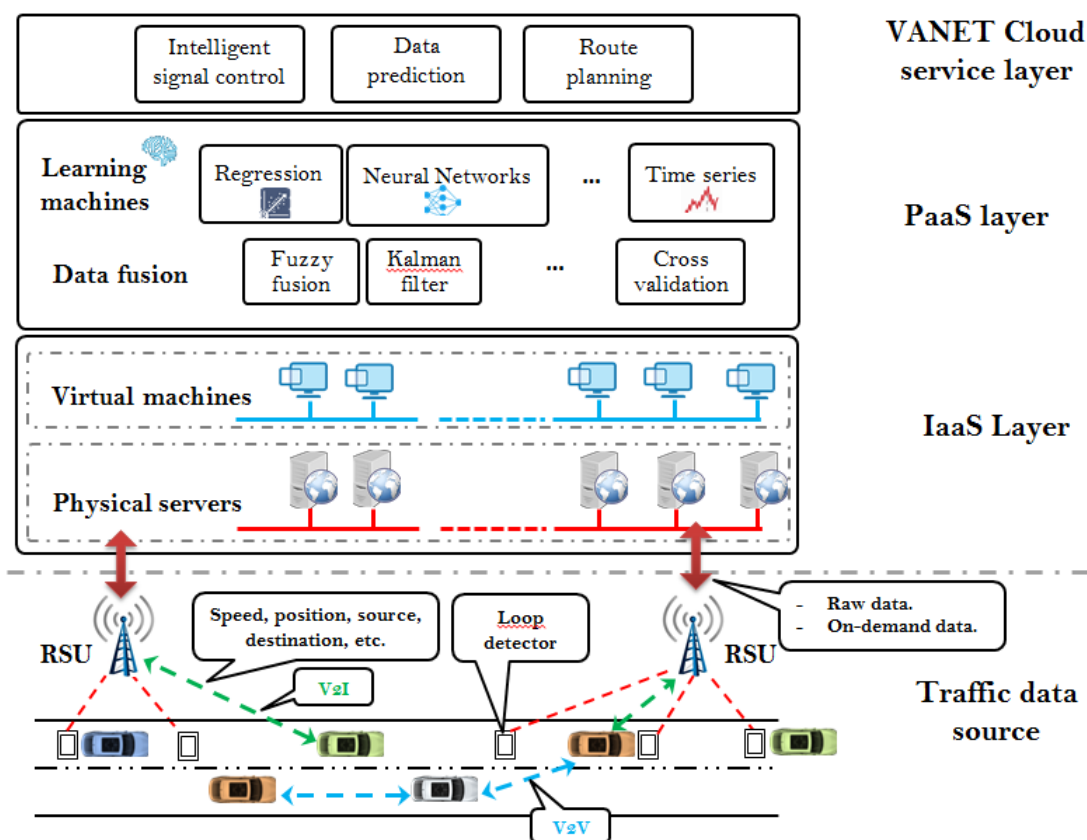


FIGURE 5.1 – Une plateforme VANET-Cloud pour la gestion du trafic.

## 1.1 Une couche de collecte des données du trafic

La couche de collecte de données du trafic comprend des données décrivant le contexte des réseaux du trafic. Cependant, la couche permet de fournir des données du trafic en temps réel adaptées à plusieurs contextes des services du trafic dans un système vaste et complexe. L'objectif principal de cette couche est de réduire le nombre d'anomalies rencontrées par l'acquisition de données du trafic et les mécanismes de collecte de données. En outre, il permet de mieux comprendre la stabilité des paramètres du trafic en temps réel (vitesse des véhicules, densité des routes et circulation), de détecter les régions d'encombrement, de contrôler les segments routiers et la mobilité des véhicules, de fournir des données historiques, et s'adapte automatiquement aux changements imprévus dans les routes. En particulier, la prolifération de détecteurs du trafic connectés autour des points spécifiques de la ville augmentera la



précision de détection de la congestion du trafic. D'autre part, l'interface de communication utilise la connexion Internet comme base pour fournir un accès aux services du trafic en temps réel. Les véhicules utilisent les réseaux cellulaires (3G, 4G, etc.) comme base pour envoyer et recevoir des données du trafic prédites par le service de Cloud (par exemple, le statut réel du flux du trafic). En outre, ces données prédites offrent des avantages non seulement pour les véhicules individuels, mais la précision de ces données va optimiser également les performances des autres applications VANET telles que la planification optimale des itinéraires et les applications de congestion du trafic.

## 1.2 Infrastructure en tant que couche de service (IaaS)

Les éléments clés de la plateforme proposée comprennent l'infrastructure comme couche de service, qui, est à son tour, contient deux sous-couches : la sous-couche des ressources physiques et la sous-couche virtuelle. La première sous-couche contient un ensemble de serveurs physiques regroupés en fonction de leurs caractéristiques (stockage ou calcul). Les serveurs de stockage sont destinés à stocker les données du trafic en temps réel et à fournir un accès aux bases de données du trafic en ligne, tandis que les serveurs de calcul sont dédiés pour assurer des performances informatiques élevées. La sous-couche virtuelle est conçue pour réduire la dégradation des performances rencontrées par les applications VANET traditionnel. Par exemple, pendant les heures de pointe, la congestion routière augmente de façon exponentielle le nombre de demandes d'utilisateurs pour les tronçons routiers, Dans lequel, les véhicules passent au serveur de la région de leurs demandes. Les réponses du serveur peuvent prendre un certain temps dans ces périodes, à cause de la surcharge et des ressources disponibles presque toutes occupées. L'accès aux ressources du serveur devient beaucoup plus facile avec l'avènement du Cloud Computing, où l'élasticité de l'infrastructure offrira la possibilité de s'adapter aux demandes des véhicules le plus rapidement possible.

## 1.3 Une plate-forme en tant que couche de service

La plate-forme en tant que couche de service vise à fournir diverses techniques d'apprentissage automatique et de fusion de données à la couche supérieure. Dépend à des scénarios de trafic, l'intégration de deux ou plusieurs techniques est importante pour résoudre les défis causés par l'augmentation des problèmes de transport. Par exemple, l'absence de la gestion efficace du trafic et de services de trafic critiques aux heures de pointe est un problème intéressant. De plus, les ressources du véhicule sont limitées pour effectuer des tâches d'apprentissage. D'une part, fournir des outils intelligents pour analyser les données est nécessaire pour produire des services de trafic avec un temps de réponse du service précis aux conducteurs. D'autre part, les experts en embouteillage peuvent simplement visualiser la traçabilité du flux de trafic en temps réel et identifier la méthode qui accomplit la tâche d'apprentissage.

## 1.4 Une couche de service VANET-Cloud

La couche de service VANET-Cloud vise à traiter les informations du trafic en temps réel et à optimiser les performances des applications VCC en prédisant les données du trafic en temps réel (par exemple, le temps de déplacement). L'opération de gestion s'étend à partir de technique de collection et d'agrégation de données vers une prédiction précise. Le mécanisme de prédiction peut être adopté dans divers contextes méritent une optimisation dynamique telle que la communication véhiculaire et les pratiques de déplacements quotidiens tout en tenant en compte les conditions encombrées. En particulier, la couche proposée contient un

ensemble de services de sécurité qui utilisent essentiellement les données capturées par un groupe de véhicules et les détecteurs du trafic. La prédiction précise du trafic présent une étape de haute primauté vers une gestion efficace du trafic et la prise de décision fiable. Les données prédites diffusées par le service Cloud incluent des informations en temps réel sur l'état prédit comme le temps de trajet prévu, la vitesse du trafic, les routes bloquées, etc. En outre, le processus de prédiction peut incorporer les données historiques du trafic. Les détails de l'agrégation de données et le mécanisme de prédiction seront expliqués dans les sections suivantes.

Dans ce qui suit, la figure 5.2 représente notre couche de service VANET-Cloud proposée pour le système de gestion du trafic. La couche supplémentaire fournit un accès à plusieurs fonctionnalités pour les conducteurs et le centre d'information routière, comme la possibilité de contrôler la mobilité du véhicule en temps réel, d'améliorer le temps de réponse aux heures de pointe, de réduire le temps de déplacement et le coût des déplacements.

Le cœur de la couche proposée est la collecte de données et la prédiction des données du trafic. La collecte de données est basée sur les données fournies par les détecteurs de route tels que les détecteurs de boucle (LD). La prolifération des détecteurs du trafic dans la ville est destinée à surveiller et capturer le flux précis de circulation dans les sections. Les détecteurs de boucle visent à capturer des données brutes du trafic telles que le flux du trafic, la vitesse des véhicules et l'occupation de la route. Les données brutes sont collectées dans un équipement de collecte de données (DCE), puis transmises à la couche VANET-Cloud via la série UC-7420. L'UC-7420 est un périphérique réseau qui permet de connecter des ETCD et des feux de circulation (TL) via des systèmes UMTS. Le rôle principal du dispositif UC-7420 est de connecter les détecteurs de route (LD) et les feux de circulation (TL) à la couche VANET-Cloud via le réseau cellulaire hétérogène (HCN). Cependant, la couche VANET-Cloud est basée sur l'infrastructure de Cloud (IaaS) pour fournir des fonctionnalités d'évolutivité pour différents services du trafic.

En particulier, la couche proposée fournit deux types de services Cloud : les services à la demande et les services automatiques. Les services à la demande comprennent la planification optimale des itinéraires, la demande d'itinéraire, le système d'information sur le temps de trajet, etc. Les services automatiques sont déclenchés lorsqu'un événement est détecté par les détecteurs de route ou par les capteurs embarqués intégrés dans les véhicules. Par exemple, lorsqu'un incident inattendu (par exemple, le ralentissement de la vitesse du véhicule) est détecté par les détecteurs de boucle (LD), un système automatique de contrôle des feux de circulation peut être déclenché pour contrôler les feux de circulation. D'autre part, le véhicule peut être averti par une réponse suite à une demande ou par des informations du trafic fournies périodiquement.

## 2 Modèle d'apprentissage non linéaire monotone en utilisant la régression de Choquet

Dans cette étude, le processus d'agrégation et de prédiction des données repose principalement sur le mécanisme de prédiction proposé dans le chapitre précédent. Comme il n'a pas encore été défini dans la littérature, le processus fait partie intégrante de la plateforme proposée. L'agrégation de données vise à résoudre les problèmes de précision des paramètres du trafic, ajuster la stabilité des paramètres du trafic dans le processus de prédiction. En tant que caractéristique clé, le processus d'agrégation identifie la précision des paramètres de flux du trafic tels que la vitesse, la densité et le temps de déplacement. Ces paramètres sont utilisés ensuite par le processus de prédiction et les services Cloud. Plus précisément, nous utilisons l'opérateur intégral Choquet pour agréger les données brutes recueillies dans

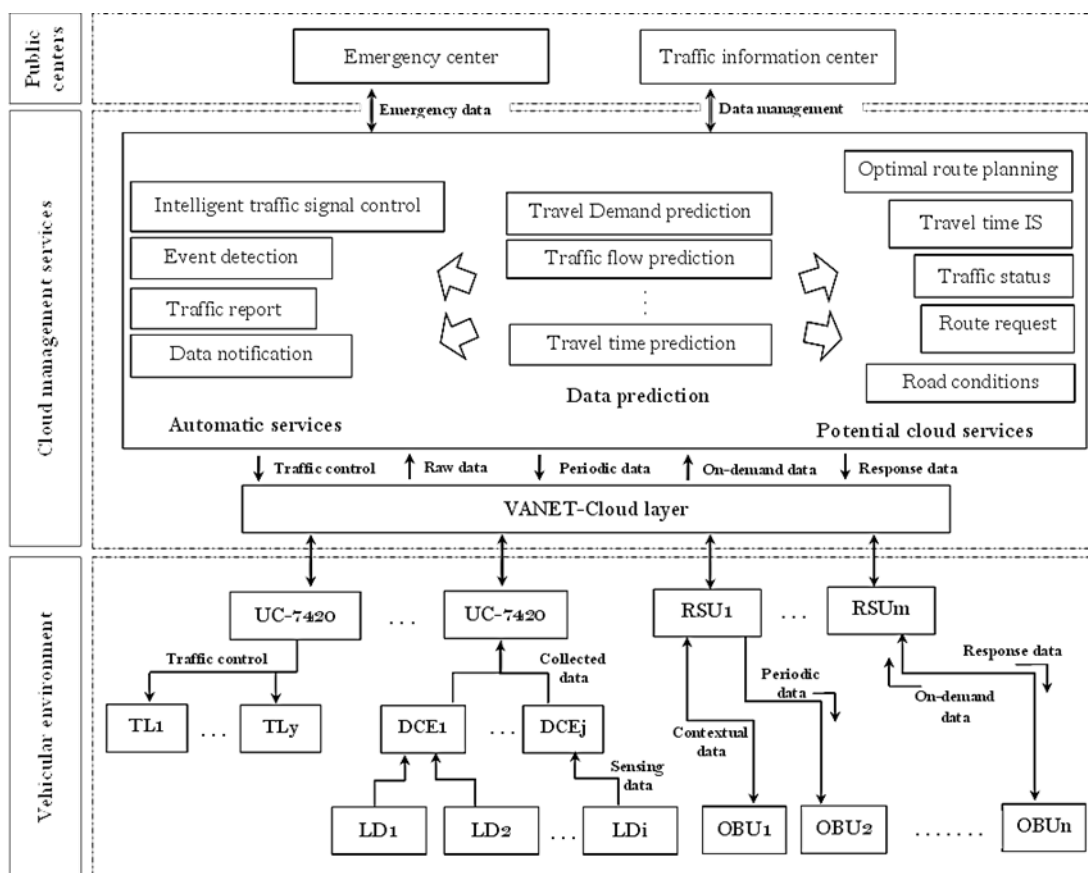


FIGURE 5.2 – Une couche de service VANET-Cloud pour les services à la demande et automatiques.

un court intervalle de temps, où la mesure floue  $\mu : g(X) \rightarrow [0, 1]$  est la puissance de l'opérateur où il représente l'interaction entre les mesures.

Le processus de prédiction utilise les mesures agrégées pour prédire des paramètres du trafic précis. En particulier, la précision des données du trafic peut être réalisée en explorant la relation entre les paramètres du trafic. Par exemple, pour prédire le temps de trajet du véhicule, le niveau de corrélation entre la vitesse du trafic et l'occupation de la route pourrait être étudié. Pour atteindre ce but, l'analyse de la régression est appliquée pour détecter la corrélation entre les paramètres du trafic.

### 3 Mécanisme de diffusion de données avec la communication V2C

Une fois que des données fiables sur le trafic sont prédites, ces données peuvent être utilisées pour contrôler la mobilité des véhicules en diffusant automatiquement ces données entre les véhicules. Dans cette étude, nous supposons que tous les véhicules sur la route ne sont pas équipés de connexion internet, ces véhicules doivent donc être avertis en cas d'incident sur la route, tel qu'un accident ou une congestion routière. Dans ces circonstances, les données du trafic prédites doivent disséminer avec les communications VANET traditionnelles (V2I et V2V). Pour contrôler la mobilité des véhicules, trois modèles de diffusion de données sont proposés : les modèles réactifs, proactifs et hybrides. Les modèles proposés sont exploités sur des modèles de communication VANET traditionnels et la communication Cloud-Véhiculaire

(V2C). Dans chaque modèle, l'échange des données du trafic est effectué entre les véhicules et les services de Cloud, et vice versa. Cependant, l'utilisation de chaque modèle de diffusion dépend sur des données du trafic fournies par les services Cloud. En particulier, les modèles de diffusion de données sont définis comme suit :

**Modèle de diffusion des données réactives :** Le modèle de diffusion des données réactives (comme le montre la figure 5.3.a) permet de recevoir, gérer, collecter les données du trafic et doit aussi permettre aux véhicules de demander des services à la demande. Ce modèle est associé principalement au mécanisme demande/réponse où les requêtes des véhicules sont considérées individuellement. Le modèle réactif permet à chaque véhicule de mettre à jour ses données locales pour devenir plus efficace et plus robuste sur la route.

**Modèle de diffusion des données proactive :** Le modèle de diffusion proactive des données est utilisé lorsque les informations sur le trafic concernent tous les véhicules sur la route. Dans le modèle proactif, comme le montre la figure 5.3.b, l'information du trafic est diffusée sans tenir compte des besoins des véhicules et est diffusée selon un mécanisme périodique. Les véhicules utilisent ce modèle pour se tenir au courant et partager les informations de la circulation telles que les embouteillages, l'emplacement le plus proche (gare, restaurant et parking), les événements imprévus, les conditions routières, etc.

**Modèle de diffusion de données hybride :** Comme le montre la figure 5.3.c, le modèle de diffusion hybride présente la combinaison des modèles précédents où les véhicules sont impliqués dans la collecte des données du trafic, détectent les incidents routiers et collaborent avec les services Cloud pour prévenir les autres véhicules. Par exemple, lorsqu'un événement inattendu se produit, les véhicules envoient aux services de Cloud un message de description révélant un événement ou des conditions d'encombrement. Lorsque ce message est reçu par les services Cloud, le modèle de diffusion hybride est utilisé pour prévenir et informer un groupe de véhicules sur les nouvelles informations du trafic.

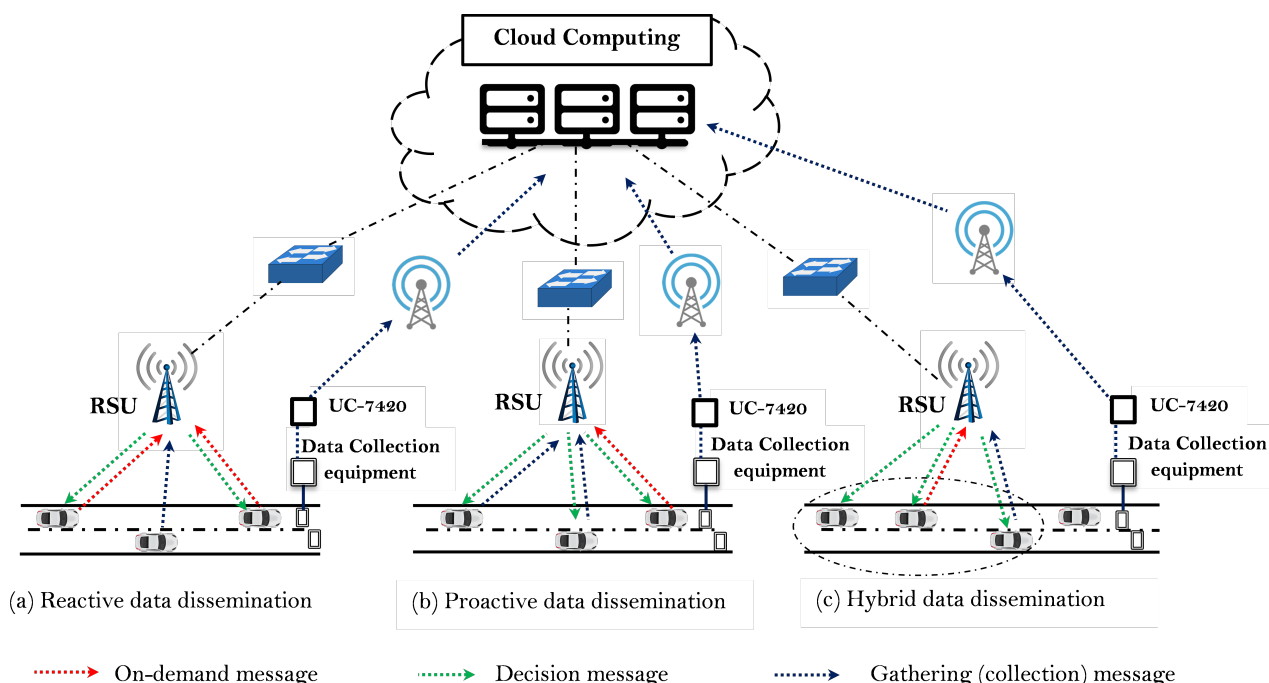


FIGURE 5.3 – Les modèles de diffusion de données avec la communication V2C.

### 3.1 Le mécanisme d'échange de données

Pour permettre la diffusion des données du trafic dans chaque modèle, nous proposons plusieurs types de messages à échanger entre les véhicules et les services Cloud. L'objectif principal des données échangées est de fournir un mécanisme simple permettant aux véhicules d'éviter la congestion du trafic, efficace dans différents scénarios du trafic pendant les heures de pointe ou soudainement lorsqu'un événement se produit sur la route. Cependant, nous définissons plusieurs types de messages supportés par les modèles de diffusion précédents :

#### 1) Message de description d'événement (EDM)

Message utile émis par des véhicules témoins ou par les détecteurs du trafic de témoins dans le voisinage pour signaler un événement imprévu sur la route. Leur description comprend l'emplacement de l'événement, la description des données d'événement, le temps de déclenchement de l'événement. Les données de localisation de l'événement sont identifiées par le système GPS. La description de données d'événement se réfère à des données relatives à des problèmes survenus subitement sur la route (par exemple, la vitesse de ralentissement) ou qui se produisent à l'intérieur du système du véhicule. Par exemple, le dispositif enregistreur de données d'événement installé dans les véhicules est destiné à enregistrer des problèmes de données spécifiques liés aux performances du moteur, au changement de vitesse, etc. Ce type de message informera le service Cloud des événements soudains. En particulier, chaque message de description d'événement (EDM) doit contenir les fonctionnalités de base suivantes pour annoncer les informations d'événement en cours :

- *< VehicleIndex >*: Le champ Index du véhicule inclut l'ID de l'expéditeur du véhicule connecté unique.
- *< EventPosition >*: l'emplacement de l'événement inattendu actuel.
- *< EvtStartTime >*: le temps typique de déclenchement de l'événement.
- *< EvtDescription >*: La description de l'événement en cours.

#### 2) Message à la demande (ODM)

Est un message envoyé par le véhicule pour demander des informations sur la circulation routière aux services Cloud, chaque message à la demande comporte les informations du véhicule et ses requêtes. Un message ODM comprend les champs suivants : *< VehicleIndex >*, *< On-demandData >* et *< VehPosition >*.

#### 3) Message de collecte de données (DCM)

Les services de Cloud offrent ce type de message pour collecter des informations de véhicule individuel (par exemple, source, destination, vitesse, etc.), augmentent la précision des informations du trafic prévues. Les messages de collecte de données sont diffusés dans le réseau véhiculaire via des communications V2V et V2I. Pour collecter les données du véhicule, deux types de DCM sont utilisés :

- **Message de collecte de données en liaison descendante (DDC)** : est un message de demande diffusé par le service de Cloud pour collecter des informations sur les véhicules ou un groupe de véhicules. Le DDC est utilisé pour contrôler la mobilité du véhicule lorsque des événements prédits par le service Cloud sont détectés.

- **Message de collecte de données en liaison ascendante (UDC)** : c'est un message de réponse envoyé par les véhicules au service de Cloud. Le message contient des informations sur le véhicule individuel telles que : la vitesse moyenne, la position actuelle, la source, la destination, la ligne, etc.

#### 4) Message d'informations sur la circulation (TIM)

C'est un message contenant les informations du trafic pour les véhicules. Ce message est annoncé par le service Cloud en vue de maintenir la sécurité des véhicules individuels. Il comporte : l'index du véhicule (VI), la position du véhicule (VehPosition) et les données essentielles du trafic (ETD) (par exemple, changement de route, changement de ligne, etc.). De plus, les véhicules qui reçoivent ce message doivent ajuster leur trajectoire de déplacement en fonction de ces données. Dans le cas où l'information du trafic nécessite le déclenchement d'un système intelligent, les services intelligents de Cloud sont incorporés pour contrôler les feux de circulation.

#### 5) Message périodique (PM)

C'est un message qui est envoyé périodiquement par le service de Cloud lorsque l'événement en cours persiste. Ce type de message contient des informations sur le trafic, le taux de congestion, les nouvelles directions, la vitesse moyenne du trafic, etc.

#### 6) Event-Driven Message (EDvM)

Ce message est diffusé par les services potentiels du Cloud pour intégrer les services intelligents de Cloud afin de contribuer à la sécurité du trafic.

## 3.2 Les algorithmes de traitement des messages

Les algorithmes de traitement des messages, décrits dans cette section, visent à fournir une réaction précise du véhicule dans les régions congestionnées détectées. Les régions d'encombrement sont détectées et identifiées grâce à la description d'événement envoyé par les véhicules ou détecté par les détecteurs de boucle. Dans les conditions du trafic synchronisé, le statut réel du flux du trafic peut être changé en flux encombré. Dans ces conditions, seuls les véhicules témoins sont autorisés à diffuser le message de description de l'événement. Si un véhicule reçoit un message de description d'événement, il envoie le message au service de Cloud via les véhicules les plus proches par une communication V2V, une communication V2I ou par des véhicules agissant comme des passerelles vers le service Cloud.

## 3.3 L'organigramme de flux de données

La figure 5.4 montre l'organigramme qui traite les messages entrants. L'organigramme illustre les interactions entre le service de nuage et les véhicules où les sorties visent à diffuser des informations du trafic pour le véhicule individuel, le groupe de véhicules, le déclenchement de systèmes intelligents ou la génération de rapports du trafic. De toute évidence, le rapport de circulation est généré automatiquement lorsqu'un accident se produit. Dans ce cas, le service Cloud va activer un canal vocal entre les véhicules témoins et le système d'urgence sanitaire pour spécifier une prévention précise de la situation. Dans le modèle de diffusion réactive, le service Cloud envoyé les informations du trafic sur le port du véhicule spécifique. Dans un modèle proactif, où les informations du trafic sont destinées à toutes les portes de nuages, il y a une diffusion rapide si ces données sont pratiques pour le temps écoulé de diffusion et autre temps de diffusion dans lequel ces données ne conviennent pas pour le temps écoulé. Si les informations sur le trafic sont liées à des événements, le service déclenche un système intelligent pour contrôler la mobilité du véhicule. Sinon, les informations sur



---

**Algorithm 2** Cloud service handle message algorithm

---

- 1: **Handle Message : Received Event Description Message**
- 2: check and cast < EDM >;
- 3: Go to line 10;
- 4: **Handle Message : Received On-demand Message**
- 5: check and cast < ODM >
- 6: Go to line 10;
- 7: **Handle Message : Received Uplink Data Collection**
- 8: check and cast < UDD >
- 9: Go to line 10;
- 10: **Handle Message : Predict the Traffic Information**
- 11: **if** the traffic information require data collection from the vehicles
- 12: Send( DDC, all cloud gates);
- 13: **end if**
- 14: **if** the collected data are available
- 15: Predict the traffic information;
- 16: Go to line 18;
- 17: **end if**
- 18: **Handle Message : Send Traffic Information**
- 19: **if** the traffic information are convenient for On-demand model
- 20: Send(TIM, vehicle gate);
- 21: **end if**
- 22: **if** the traffic information are convenient for all vehicles
- 23: wait until the elapsed time of the broadcast = true
- 24: Send(PM, all cloud gates);
- 25: **end if**
- 26: **if** the traffic information are convenient for EVENT-DRIVEN
- 27: Send EDvM to intelligent cloud services to control the vehicles mobility.
- 28: **end if**
- 29: **if** the traffic information is convenient for serious situations
- 30: Send traffic report to the traffic management center.
- 31: **end if**

---

---

**Algorithm 3** Vehicle handle message algorithm

---

```
1: Handle Message : Send On-Demand Message
2: Create and initiate < ODM >;
3: ODM.VI = getVehicleID();
4: ODM.VehPosition = getVehGPSPosition();
5: ODM.On-demandData = getOn-demandData();
6: Send (ODM, vehicle gate);
7: Handle Message : Send Event Description Message
8: Create and initiate < EDM >
9: EDM.VehicleIndex = getVehicleID();
10: EDM.EventPosition = getEvtPosition();
11: EDM.EvtStartTime = getEvtStartTime();
12: EDM.EvtDescription = getEvtDescriptionData();
13: Send (EDM, vehicle gate);
14: Handle Message : Receive Downlink Data Collection Message
15: Create and initiate < UDC >
16: UDC.VehicleIndex = getVehicleID();
17: UDC.VehicleSource = getVehSource();
18: UDC.VehicleDestination = getVehDestination ();
19: UDC.VehiclePosition = getVehGPSPosition();
20: Send(UDC, vehicle gate);
21: Handle Message : Receive Traffic Information Message
22: check and cast < TIM >
23: Update local data;
24: Handle Message : Receive Periodic Message
25: check and cast < PM >
26: Update local data
```

---

le trafic seront envoyées sous forme de rapport du trafic au centre d'informations sur la circulation.

## 4 Simulation et évaluation

Au cours des dernières décennies, le nombre et l'ampleur des embouteillages ont augmenté de façon exponentielle en Algérie. Un simple voyage peut prendre des heures, surtout dans les grandes villes et les environs. En outre, notre scénario analytique vise à atteindre l'objectif du département des transports pour résoudre le problème de la congestion routière quotidienne qui concerne la route principale de Tébessa (N10, longueur : 19,1 km). Aux heures de pointe, les conditions d'encombrement sur cette route conduisent à augmenter le temps de déplacement des véhicules. Le but du département des transports est de détecter les régions congestionnées de la route principale et d'améliorer le flux de circulation dans les voies artérielles. Pour cela, nous étendons la plateforme de simulation iCanCloud (Nunez et al., 2012) pour évaluer la performance proposée en intégrant le mécanisme de prédiction et d'échange de données. De cette manière, la description du scénario est basée sur l'hypothèse montrée dans la figure 5.5.

Les véhicules sont équipés d'une large gamme de dispositifs embarqués, tels que des capteurs connectés (RFID, caméra connectée, détection d'événement de véhicule), des réseaux cellulaires récents (3G, GPRS et LTE) pour fournir une connexion Internet souple. Ils sont aussi équipés de dispositifs embarqués tels que le GPS, EDR (événement Data Recorder), OBU, et les autres dispositifs de communication sans fil. Ainsi, notre scénario comprend deux intersections qui relient la route principale aux voies artères (N1 et N2).

La figure 5.6 montre le modèle de flux du trafic utilisé pour évaluer le modèle de prédiction, où les données sont enregistrées tout au long de la journée en utilisant des détecteurs de boucle



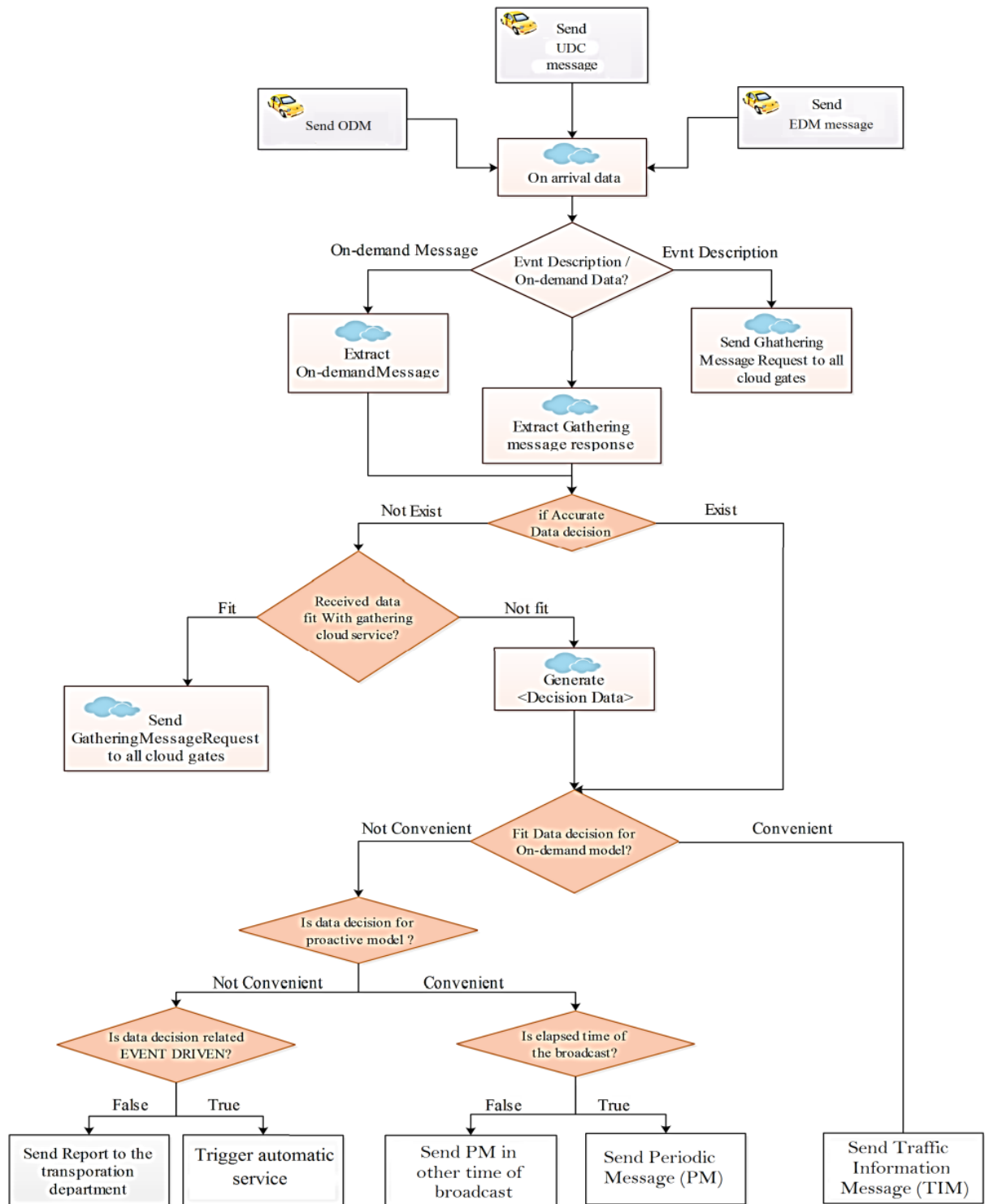


FIGURE 5.4 – Organigramme de flux d’informations sur le trafic des véhicules.

connectés.

Dans le scénario précédent, le mécanisme d’échange de données utilise une suite d’événements pour éviter la congestion détectée, i.e., à partir d’événement de détection jusqu’à l’événement de décision. Le premier événement (Evt1 : message de description d’événement, EDM) est déclenché par les véhicules lorsqu’une vitesse plus lente est détectée par les véhicules ou une variation de données est capturée par les détecteurs du trafic. Les applications des unités embarquées des véhicules témoins peuvent déclencher immédiatement ce message en



FIGURE 5.5 – Description du scénario.

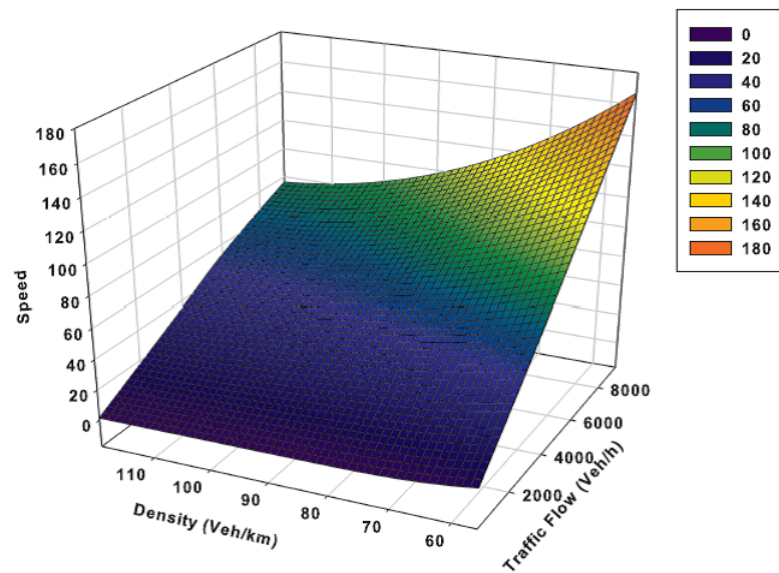


FIGURE 5.6 – Le modèle de flux du trafic.

vue de localiser l'emplacement de l'événement (EventPosition), le temps de déclenchement de l'événement (EvtStartTime) et la description de l'événement (EvtDescription). Une fois le message EDM reçu au service Cloud, l'événement 2 (Evt2 : Send Downlink Data Collection Message) est déclenché pour prédire l'état réel du réseau du trafic en utilisant les données collectées par les véhicules et les détecteurs. Cependant, les véhicules impliqués dans l'événement 3 (Evt3 : Sending Uplink Data Collection Message) vont générer chacun un message UDC qui comprend les données contextuelles suivantes : VehicleIndex, VehicleSource, VehicleDestination et VehiclePosition. Lors de l'arrivée de l'événement 4 (Evt4 : Sending Traffic Information), le service Cloud va diffuser les données prédites de chaque véhicule en utilisant le message d'information sur la circulation (TIM).

## 4.1 Une forme étendue de la plateforme iCanCloud pour évaluer les performances des architectures VANET-Cloud

iCanCloud (Nunez et al, 2012) est une plateforme de simulation basée sur la plateforme INET (Varga, 2015) en vue de simuler les architectures de type Cloud Computing évolutives. L'un des principaux objectifs est de prévoir le coût et les performances d'une application donnée dans un environnement Cloud.

Pour modéliser et simuler la performance des architectures Cloud Véhiculaires, nous étendons la plateforme de simulation iCanCloud par les composants clés de l'architecture proposée (voir la figure 5.7). Du point de vue conceptuel d'intégration, les véhicules sont équipés d'interfaces Ultran ou d'interfaces DSRC (IEEE 802.11 p). Grâce aux cartes DSRC, les véhicules peuvent se connecter aux réseaux traditionnels VANET via les communications V2V et V2I. D'autre part, les interfaces Ultran permettent aux véhicules de se connecter via des réseaux sans fil hétérogènes tels que les réseaux 3G et 4G. De toute évidence, les véhicules équipés par doubles interfaces (Ultran et DSRC) agissent comme des passerelles Internet pour les véhicules non-connectés. Des détecteurs de trafic sont installés pour chaque segment de la route où ils sont directement connectés au Cloud via la communication I2I ou la norme IEEE 802.11. Le segment routier est relié par deux intersections et caractérisé par le trafic en temps réel, tel que les données du trafic, la densité de route actuelle, la vitesse moyenne et l'heure d'arrivée estimée. La prédiction de flux du trafic est un service accessible via SaaS (i.e., accessible par Internet) et exécuté en machine virtuelle. En outre, le service est appelé en vue de prédire le flux du trafic à partir des données brutes. « L'hyperviseur » est un logiciel installé dans les RSU, qui vise à créer pour chaque segment de la route un système virtuel pratique et véhiculer des données capturées vers une instance de machine virtuelle. La collaboration entre ces entités conduit à construire un système VANET-Cloud efficace qui prédit un flux du trafic précis. Grâce à des données de détection en temps opportun, les services Cloud peuvent diffuser et choisir quelles informations peuvent influencer l'avenir de la mobilité des véhicules.

Notre évaluation a été réalisée sur la base de trois scénarios en se basant sur les trois modèles de diffusion précédents, dans lesquels, pour chaque modèle de diffusion, nous estimons le taux de livraison des données, les performances de communication, la latence du service Cloud. De plus, nous testons la précision de la prédiction de données du service Cloud. Le premier scénario vise à évaluer le modèle de diffusion réactive (le modèle à la demande) où seuls les messages réactifs sont impliqués par le mécanisme d'échange. Le deuxième scénario vise à évaluer le modèle de diffusion proactif où les informations du trafic sont destinées à informer un groupe de véhicules. Dans ce modèle, seuls les messages proactifs sont pris en compte par le mécanisme d'échange de données. Le troisième scénario vise à tester le modèle hybride du mécanisme d'échange de données en impliquant à la fois des modèles réactifs et proactifs. Au lieu de cela, ce modèle, intentionnellement, est dédié à assigner automatiquement des informations du trafic de véhicule à un modèle réactif, à un modèle proactif ou aux deux.

Tableau 5.1 – Paramètres de simulation.

Parameters	Value
Road length	2 km
Frequency Band	5.9 GHZ
Rsu/vehicle data rate	16 Mbps
Rsu/vehicle delay	20ms
Cloud/Rsu data rate	10 Gbps
Cloud/Rsu delay	1.25E-4us
Transmission range	20 km
Max number of vehicles/h	2200

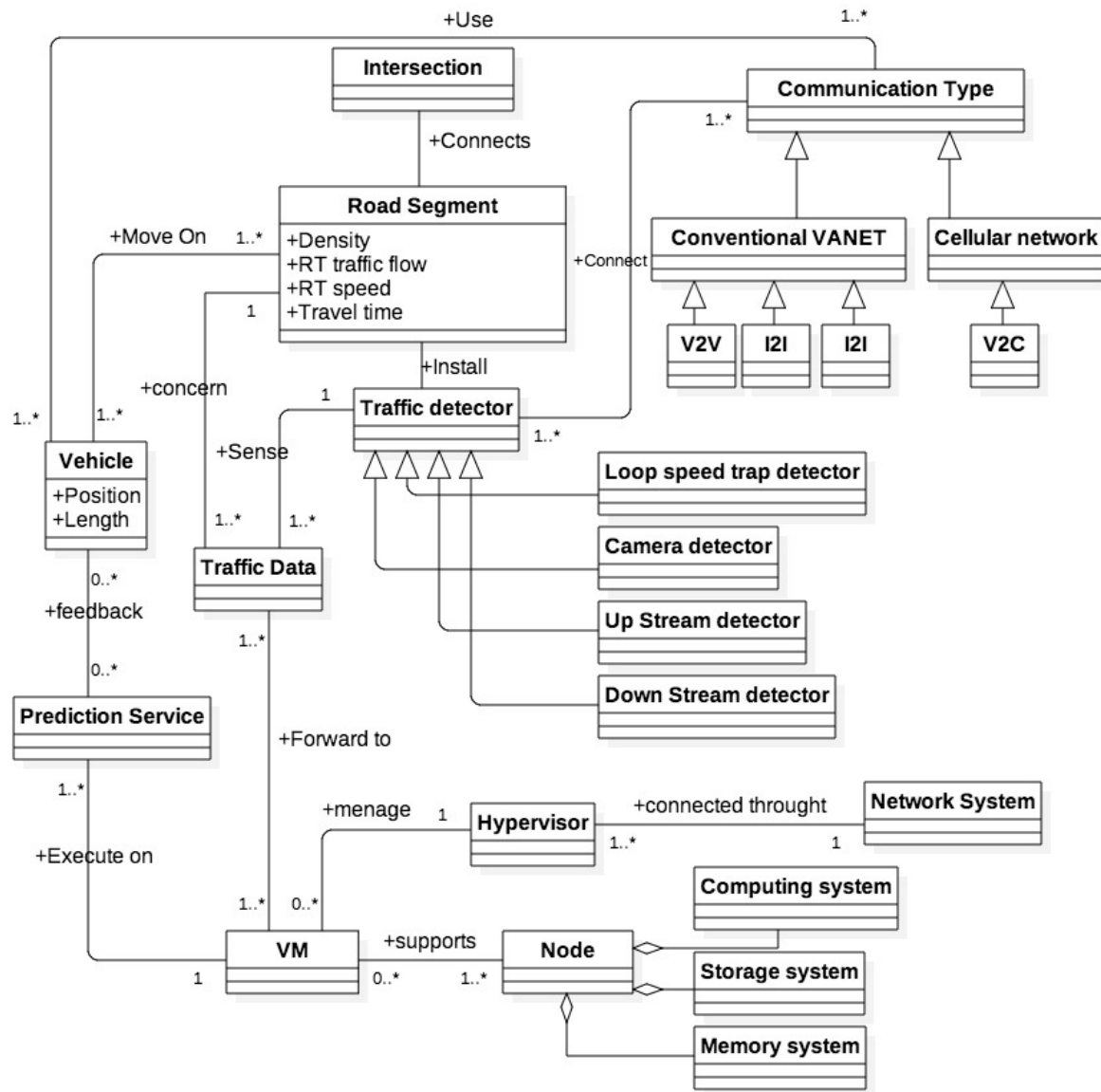


FIGURE 5.7 – Vue conceptuelle de l'intégration.

## 4.2 Taux de livraison de données

Afin d'analyser les résultats, le même nombre de véhicules contribue à simuler chaque scénario séparément. Les paramètres de simulation sont définis dans le tableau 5.1. La figure 5.8 montre la performance des trois scénarios proposés, où chaque résultat concerne un modèle spécifique et chaque point de la figure représente le taux de réussite de la livraison des données. Le temps de simulation disposé sur l'axe X est indiqué avec le débit de livraison. Dans le scénario réactif, nous utilisons la formule (5.1) pour estimer le taux de messages réactifs échangés. Les résultats de performance de ce scénario montrent que le taux de réussite des données de livraison est stable pendant la simulation. Cela signifie que le taux de réussite de la livraison restera inchangé même si le nombre de messages à la demande augmente. Dans le scénario proactif, nous utilisons la formule (5.2) pour estimer le taux de livraison des données par le service cloud. Cependant, nous avons constaté que le taux de livraison réussie a commencé avec 90%. Les résultats de performance du modèle proactif descendent sans cesse avec le temps de simulation. Dans lequel, le taux de livraison des données est influencé par

la collecte de données et la diffusion des données. Dans le modèle hybride, le taux de réussite de la transmission des données atteint 91%. La performance de simulation de ce modèle est meilleure par rapport au performance obtenue par le scénario proactif. En conséquence, les performances de simulation des modèles précédents prouvent le potentiel de notre approche pour distribuer, de manière automatique et sûre, les informations du trafic par le service Cloud.

$$ReactiveDR = \frac{\sum ReactiveDataMsgs}{\sum RecievedMsgs} * 100 \quad (5.1)$$

$$ProactiveDR = \frac{\sum ProactiveDataMsgs}{\sum RecievedMsgs} * 100 \quad (5.2)$$

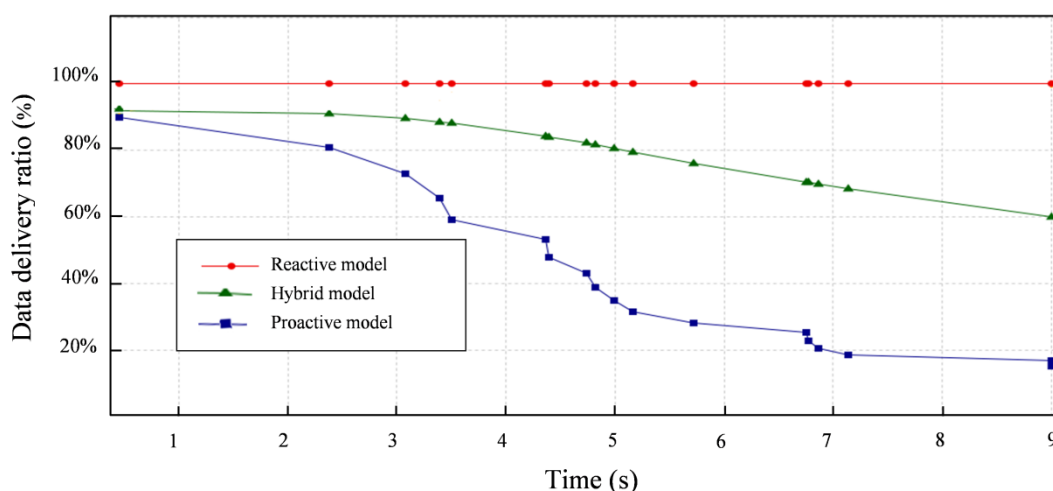


FIGURE 5.8 – Débit de transmission de données par le Cloud pour chaque modèle.

### 4.3 La haute performance de communication

En outre, nous estimons la bande passante entre l'infrastructure de Cloud et l'infrastructure routière. Dans lequel, la connectivité de bout en bout fonctionne sur 10 Gbps de transmission de données. Sur la base des performances de collecte de données et de la taille du message, la figure 5.9 montre la bande passante requise pour chaque modèle en spécifiant la taille du message en (Kilo-octets). Le modèle proactif nécessite 3.7 Gbps de bande passante pour la taille des messages longs (5 Ko) et 0.9 Gbps pour la taille des messages inférieure (1 Ko). Le modèle réactif nécessite une taille de message de 30.8 Mbps (5 Ko) et une taille de message de 8.817 Mbps (1 Ko). Dans le modèle hybride, nous évaluons 1.6 Gbps sur une taille de message longue (5 Ko) et 40 Mbps sur une taille de message inférieure (1 Ko).

### 4.4 La latence du service Cloud

La latence du service de Cloud est le délai entre la demande du conducteur et la réponse du service de prédiction. La latence du mécanisme d'échange de données proposé a été examinée en se basant sur deux normes de communication : l'IEEE 802.11p et l'évolution à long terme (LTE), (voir la figure 5.10). Cependant, la latence sur le LTE est tracée par rapport à la latence sur la communication IEEE 802.11p, qui varie de 1 Ko à 16 Ko de taille de message. La



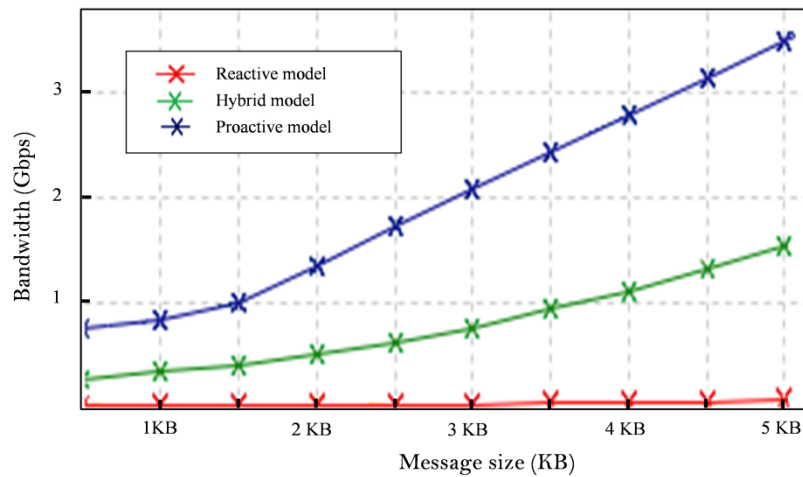


FIGURE 5.9 – Les performances de communication des services Cloud pour les trois modèles de diffusion de données.

simulation est réalisée afin d’expérimenter le scénario du modèle hybride précédent. Nous observons que la taille des paquets a un impact important sur le temps de latence sur les standards de communication précédents. De plus, nous remarquons que les performances de latence du modèle hybride utilisant le standard IEEE 802.11p surpassent les performances de latence en utilisant le LTE. Par conséquent, nous concluons que les performances du mécanisme d’échange de données proposé sur LTE offrent une meilleure performance plutôt que la norme IEEE 802.11p.

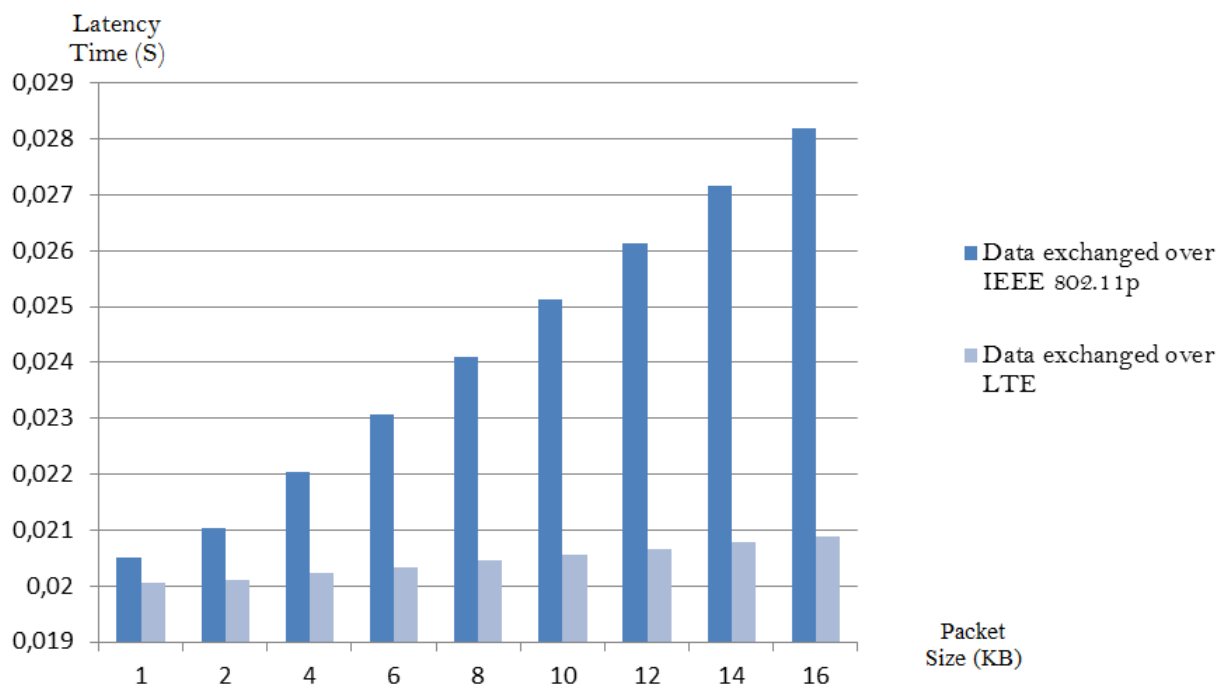


FIGURE 5.10 – La latence des services Cloud avec les normes V2C et IEEE 802.11p.

---

## CONCLUSION

Dans ce chapitre, nous avons défini une couche VANET-Cloud qui vise à fournir de nombreuses fonctionnalités aux conducteurs en prédisant et en diffusant les informations du trafic en temps réel, telles que : le temps de déplacement, la vitesse des véhicules et le statut réel du trafic. La couche proposée est destinée à fournir deux types de service du trafic routier : les services potentiels (services à la demande) et les services automatiques. Les services potentiels utilisent le mécanisme de prédiction de trafic pour résoudre les problèmes inhérents à la précision des données, tel que l'ajustement des paramètres du trafic dans le processus de prédiction. De plus, les services du trafic utilisent le mécanisme d'échange de données où plusieurs types de messages ont été proposés pour diffuser les informations du trafic entre les véhicules et les services de Cloud. Les informations sur le trafic sont diffusées selon trois modèles : les modèles réactifs, proactifs et hybrides..

Le modèle de données réactif est intéressant lorsque les véhicules nécessitent une décision à la demande, tandis qu'un modèle de données proactif utilise des messages périodiques pour découvrir l'arrivée de nouveaux véhicules sur une zone de couverture du Cloud. Les performances de simulation prouvent que le modèle hybride de diffusion de données est utile pour plus d'élasticité en impliquant à la fois des modèles réactifs et proactifs sur des situations où un événement attendu se produit. Selon les résultats de la simulation, les services VANET-Cloud favorisent, dans l'état congestionné, des performances réseau optimales en termes de livraison de données et d'utilisation de la bande passante et de latence. De plus, les performances de la couche proposée en utilisant le mécanisme d'échange de données (V2C) surpassent les performances d'un réseau sous la norme IEEE 802.11p.

# Conclusion Générale

Ce chapitre conclut la thèse en donnant un bilan du travail effectué et les perspectives envisageables au terme de cette recherche. Nous rappelons l'objectif principal de cette thèse qui était de fournir des solutions innovantes dans le domaine de transport intelligent (STI) parallèlement à la croissance des capacités Internet. Ces solutions rationalisent l'utilisation des infrastructures routières et préservent la sécurité routière en réduisant le nombre et l'ampleur des embouteillages en augmentant la mobilité des conducteurs, et en améliorant les performances de nombreuses applications du trafic routier inscrit dans ce domaine. Le véhicule informatique en nuage est une solution pour résoudre certains problèmes de STI tels que la congestion du trafic.

Dans ce contexte, notre objectif était de se concentrer sur un domaine de recherche particulier du monde du Cloud Computing Véhiculaire et de remédier aux défis pertinents dans ce domaine. En particulier, le traitement de l'information du trafic en temps réel et son impact potentiel sur la gestion dynamique du flux de trafic.

Dans un premier temps, nous avons mis en évidence le paradigme du Cloud Computing Véhiculaire en présentant ses fondements théoriques, ses architectures et ses applications innovantes. L'objectif de ce nouveau paradigme est de construire un cadre global de communication et de traitement de l'information accessible sur internet, ce qui rend cet élément essentiel dans notre travail de recherche. En outre, nous avons discuté de la flexibilité de ses applications pour la gestion quotidienne de la dynamique du trafic.

Ensuite, nous avons présenté les théories du trafic et les niveaux d'agrégation des flux qui nous permet d'analyser et de modéliser efficacement la dynamique du trafic. Les observations expérimentales de la circulation automobile montrent des comportements dynamiques de plus en plus complexes provenant de plusieurs sources : la structure de la route, le comportement du conducteur et les conditions météorologiques. Nous concluons que le flux de trafic n'a pas de modèle universelle qui le représente et que sa dynamique aux heures de pointe est non-linéaire.

Dans nos premières contributions, en se basant sur la description du trafic fournie par le département de transport, nous avons présenté une étude empirique pour développer un modèle de régression prédictif afin d'interpréter le phénomène de congestion. En particulier, le modèle développé est capable de prédire avec précision la probabilité d'un incident imprévu se produisant sur cette route. L'objectif est d'assurer une précision satisfaisante et un niveau de sécurité routière souhaité en surveillant de manière efficace les tronçons routiers. Le ralentissement soudain de la vitesse des véhicules est le cas pratique de notre étude. La tâche de classification de notre modèle prédictif comprend quatre variables expliquées, y compris la vitesse du véhicule observée, le temps de déplacement observé, le temps de déplacement prévu et le temps de retard estimé. La précision de la prédiction est prouvée en vérifiant la pertinence du modèle en fonction de la qualité de l'ajustement et de la signification statistique de chaque



variable expliquée. Nous trouvons que le modèle adopté permet d'interpréter avec précision les dépendances entre les variables de trafic en tant que modèle linéaire généralisé.

Dans cette perspective, nous avons présenté également un système de fusion floue pour modéliser les interactions entre les variables de trafic. L'objectif de ce système est d'améliorer l'apprentissage et d'avoir des prédictions précises. Le système comprend quatre niveaux de fusion : préparation des données (niveau 1), extraction / sélection des attributs (niveau 2), décision sur les données (niveau 3) et adaptation du modèle (niveau 4). Le modèle d'apprentissage est un modèle prédictif floue basé sur la convergence de la régression linéaire multiple (MLR) par l'intégrale de Choquet pour garantir l'interaction entre les variables. L'évaluation montre une prédiction précise aux heures de pointe par rapport à d'autres modèles d'apprentissage telles que ARIMA (Autoregressive integrated moving average), les séries temporelles à base réseau de neurones multicouches (TSP-MNN) et l'agrégation linéaire à base régression multivariée (MLR-LA). Au stade de l'évaluation, une description réaliste d'un scénario a été montrée pour tester l'efficacité du modèle d'apprentissage prédictif dans des conditions d'encombrement. Les résultats montrent que le modèle d'apprentissage Choquet-MLR est efficace s'agissant de minimiser la fonction objective. De plus, les performances de prédiction de cette analyse surpassent les autres modèles de prédiction à court terme en utilisant des métriques d'erreur telles que MAPE, RMSE et EC.

Finalement, nous avons proposé une couche VANET-Cloud pour augmenter la sécurité du trafic, améliorer les informations sur le trafic et optimiser les performances du réseau dans des conditions d'embouteillage. En particulier, l'architecture qui permet de gérer le trafic repose principalement sur une infrastructure Cloud et la prolifération du réseau de capteurs connectés autour des villes, y compris les détecteurs de boucles. Le réseau de capteurs est utilisé pour collecter des données de trafic et le système de Cloud est utilisé pour fournir des services à la demande (ex : temps de déplacement d'un véhicule) et des services intelligents (ex : un système de signalisation intelligent). Ces services utilisent notre modèle d'apprentissage pour fournir aux véhicules des informations fiables et précises sur le trafic. Sur la base d'une analyse de scénario, les résultats de la simulation illustrent l'efficacité de la gestion du trafic proposée pour réduire considérablement le temps de trajet des véhicules et les performances optimales par rapport aux réseaux VANET traditionnels.

Les perspectives envisageables en prolongement direct de cette thèse concernent trois objectifs importants :

- Parmi les contraintes que doit satisfaire un système est la forte mobilité des nœuds dans un Vanet. Nous envisageons intégrer le module de mobilité orientée dans le simulateur OMNET++ afin de maîtriser mieux la modélisation et la simulation des Vanets.
- Traitement du problème de covoiturage quotidien en tant que modèle de déplacement, tout en assurant une flexibilité élevée et en maximisant les gains des utilisateurs.
  - Incorporer notre service de prédiction et de communication pour résoudre le problème de covoiturage quotidien
  - Avoir à la fois satisfaire les objectifs des conducteurs, des passagers et des agences de covoiturage. Il est possible d'intégrer plusieurs critères de déplacement comme (la distance du trajet, le temps d'arriver, l'état de trafic, etc.).
- Une perspective peut être lointaine mais des plus intéressantes serait d'utiliser les réseaux d'apprentissage automatique afin de fournir des décisions plus exactes.

• REVUES INTERNATIONALES AVEC COMITÉ DE LECTURE

- 1) Abdelatif, S., Makhlouf, D., Roose, P. (2018). Smart Traffic Management System for Anticipating Unexpected Road Incidents in Intelligent Transportation Systems. International Journal of Grid and High Performance Computing (IJGHPC), 10(4).

• CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE

- 1) Sahraoui Abdelatif, Derdour Makhlouf, Philippe Roose – Extended iCanCloud simulation framework for VANET-Cloud architectures – 3rd International Conference on Networking and Advanced Systems – 13-14 December 2017, Annaba, Algeria.
- 2) Abdelatif, S., Makhlouf, D., Roose, P., & Becktache, D. (2016, August). Loop Speed Trap Data Collection Method for an Accurate Short-Term Traffic Flow Forecasting. In International Conference on Mobile Web and Information Systems (pp. 56-64). Springer International Publishing.
- 3) Abdelatif, S., Makhlouf, D., Roose, P., & Becktache, D. – An Intelligent Architecture for VANET-Cloud Computing – PAIS'2016 – Tébessa, Algeria

• WORKSHOP ET JOURNÉES DOCTORALES AVEC COMITÉ DE LECTURE

- 1) Abdelatif, S., Makhlouf, D., & Becktache, D. "Réseau pour l'informatique de nuage véhiculaire". Journées Ouvertes sur les Mathématiques et l'Informatique (JOMI'2015). University of Tébessa-Algeria.

- 
- Federal Communications Commission. "FCC 99-305." FCC Report and Order (1999).
- Brickley, O., Koubek, M., Rea, S., & Pesch, D. (2010, March). A network centric simulation environment for CALM-based cooperative vehicular systems. In Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (p. 75). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- INRIX (October 14, 2014). UK Summer Traffic Predictions : INRIX Traffic Congestion Report. Url : <http://inrix.com/press-releases/americans-will-waste-2-8-trillion-on-traffic-by-2030-if-gridl>
- Li, Y. J. (2010, November). An overview of the DSRC/WAVE technology. In International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (pp. 544-558). Springer, Berlin, Heidelberg.
- DSRC Implementation Guide – A guide to users of SAE J2735 message sets over DSRC. SAE International, 2010
- Eichler, S., Schroth, C., & Eberspächer, J. (2006). Car-to-car communication.
- Morgan, Y. L. (2010). Notes on DSRC & WAVE standards suite : Its architecture, design, and characteristics. IEEE Communications Surveys & Tutorials, 12(4), 504-518.
- Hartenstein, H., & Laberteaux, K. (Eds.). (2009). VANET vehicular applications and inter-networking technologies (Vol. 1). John Wiley & Sons.
- Jiang, D., & Delgrossi, L. (2008, May). IEEE 802.11 p : Towards an international standard for wireless access in vehicular environments. In Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE (pp. 2036-2040). IEEE.
- D31 European ITS Communication Architecture Overall Framework Proof of Concept Implementation.
- Bossom, R., Brignolo, R., Ernst, T., Evensen, K., Frötscher, A., Höfs, W., ... & Kulp, I. (2009). European ITS Communication Architecture—Overall Framework—Proof of Concept Implementation.
- Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.
- Al Mamun, M. A., Anam, K., Onik, M. F. A., & Esfar-E-Alam, A. M. (2012, October). Deployment of cloud computing into vanet to create ad hoc cloud network architecture. In Proceedings of the world congress on engineering and computer science (Vol. 1, pp. 24-26).
- Mandal, M., Landge, C., Gaikwad, P., Nagaraj, U., & Abhale, A. (2014). Implementing Storage as a Service in VANET using Cloud Environment. International Journal of Advance Foundation and Research in Computer, 1(5), 70-76.
- Sedjelmaci, H., Senouci, S. M., & Abu-Rgheff, M. A. (2014). An efficient and lightweight intrusion detection mechanism for service-oriented vehicular networks. IEEE Internet of Things Journal, 1(6), 570-577.
- Zhang, W., Chen, Y., Yang, Y., Wang, X., Zhang, Y., Hong, X., & Mao, G. (2012). Multi-hop connectivity probability in infrastructure-based vehicular networks. IEEE Journal on Selected Areas in Communications, 30(4), 740-747.
- Boskovich, S., & Barth, M. (2013, October). Vehicular network rerouting autonomy with a V2V, I2V, and V2I communication matrix classification. In Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on (pp. 172-177). IEEE.
- Karthikeyan, T., & Bhuvaneshwari, N. S. (2012). An Agent based approach towards building a co-operative vehicular network. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(6), pp-62.
- Talib, M. S., Hussin, B., & Hassan, A. (2017). Converging VANET with Vehicular Cloud
-

---

Networks to reduce the Traffic Congestions : A review. *International Journal of Applied Engineering Research*, 12(21), 10646-10654.

He, W., Yan, G., & Da Xu, L. (2014). Developing vehicular data cloud services in the IoT environment. *IEEE Transactions on Industrial Informatics*, 10(2), 1587-1595.

Lee, E., Lee, E. K., Gerla, M., & Oh, S. Y. (2014). Vehicular cloud networking : architecture and design principles. *IEEE Communications Magazine*, 52(2), 148-155.

Godfrey, K. R. (1969). The theory of the correlation method of dynamic analysis and its application to industrial processes and nuclear power plant. *measurement and control*, 2(5), T65-T72.

McClenahan, J. W., & Simkowitz, H. J. (1969). The effect of short cars on flow and speed in downtown traffic : A simulation model and some results. *Transportation Science*, 3(2), 126-139.

Branston, D. (1976). Link capacity functions : A review. *Transportation research*, 10(4), 223-236.

Mahmassani, H., & Herman, R. (1984). Dynamic user equilibrium departure time and route choice on idealized traffic arterials. *Transportation Science*, 18(4), 362-384.

Varga, A. (2001). Discrete event simulation system. In *Proc. of the European Simulation Multiconference (ESM'2001)*.

Henderson, T. R., Lacage, M., Riley, G. F., Dowell, C., & Kopena, J. (2008). Network simulations with the ns-3 simulator. *SIGCOMM demonstration*, 14(14), 527.

Chang, X. (1999). Network simulations with OPNET. In *Simulation Conference Proceedings, 1999 Winter (Vol. 1, pp. 307-314)*. IEEE.

Mussa, S. A. B., Manaf, M., Ghafoor, K. Z., & Doukha, Z. (2015, October). Simulation tools for vehicular ad hoc networks : A comparison study and future perspectives. In *Wireless Networks and Mobile Communications (WINCOM), 2015 International Conference on (pp. 1-8)*. IEEE.

Baldo, N., Miozzo, M., Guerra, F., Rossi, M., & Zorzi, M. (2010). Miracle : the multi-interface cross-layer extension of ns2. *EURASIP Journal on Wireless Communications and Networking*, 2010, 26.

Varga, A. (2015). INET framework for the OMNeT++ discrete event simulator, 2010.

Núñez, A., Vázquez Poletti, J. L., Caminero, C., González Castañé, G., Carretero Pérez, J., & Llorente, I. M. (2012). iCanCloud : A flexible and scalable cloud infrastructure simulator.

Todincă, D., Graja, H., Perry, P., & Murphy, J. (2004). Novel admission control algorithm for GPRS/EGPRS based on fuzzy logic.

Abramson, N. (1970, November). THE ALOHA SYSTEM : another alternative for computer communications. In *Proceedings of the November 17-19, 1970, fall joint computer conference (pp. 281-285)*. ACM.

Chen, Y., Guizani, M., Zhang, Y., Wang, L., Crespi, N., & Lee, G. M. (2017). When Traffic Flow Prediction Meets Wireless Big Data Analytics. *arXiv preprint arXiv :1709.08024*.

White, F. E. (1991). *Data fusion lexicon*. JOINT DIRECTORS OF LABS WASHINGTON DC.

Stanley, J., & Steinhardt, B. (2014). Bigger monster, weaker chains : The growth of an American Surveillance Society. In *Ethics and Emerging Technologies (pp. 269-284)*. Palgrave Macmillan UK.

Gary Palgon. (2017). *Data Aggregation : Definition and Importance to Life Sciences Researchers*. Machine Intelligence Research Institute (MIRI). [http :](http://)

---

[//intelligency.org/big\\_data\\_transfer.php](http://intelligency.org/big_data_transfer.php)

Abdelgawad, A., & Bayoumi, M. (2012). Data fusion in WSN. In *Resource-aware data fusion algorithms for wireless sensor networks* (pp. 17-35). Springer US.

Smeed, R. J. (1968). Traffic studies and urban congestion. *Journal of Transport Economics and policy*, 33-70.

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Wu, S., & Crestani, F. (2015). A geometric framework for data fusion in information retrieval. *Information Systems*, 50, 20-35.

Aiello, G., Giovino, I., Vallone, M., Catania, P., & Argento, A. (2017). A decision support system based on multisensor data fusion for sustainable greenhouse management. *Journal of Cleaner Production*.

Kotlowski, W., & Slowinski, R. (2009, June). Rule learning with monotonicity constraints. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 537-544). ACM.

Beliakov, G., Sola, H. B., & Sánchez, T. C. (2016). Review of Aggregation Functions. In *A Practical Guide to Averaging Functions* (pp. 1-53). Springer International Publishing.

Beliakov, G., Sola, H. B., & Sánchez, T. C. (2016). *A practical guide to averaging functions* (Vol. 329). Heidelberg : Springer.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Ali Fallah Tehrani, V. (2014). *learning nonlinear monotone classifiers using the choquet integral*, thèse de doctorat, Université du Philipps-Universität, Marburg.

Ghosh, B., Basu, B., & O'Mahony, M. (2009). Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE transactions on intelligent transportation systems*, 10(2), 246-254.

Fang, H., Rizzo, M. L., Wang, H., Espy, K. A., & Wang, Z. (2010). A new nonlinear classifier with a penalized signed fuzzy measure using effective genetic algorithm. *Pattern recognition*, 43(4), 1393-1401.

Li, W., Liu, P., Wang, Y., & Ni, H. (2014). Multifeature fusion vehicle detection algorithm based on choquet integral. *Journal of Applied Mathematics*, 2014.

Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C : Emerging Technologies*, 43, 50-64.

Mannini, L., Carrese, S., Cipriani, E., & Crisalli, U. (2015). On the short-term prediction of traffic state : an application on urban freeways in Rome. *Transportation Research Procedia*, 10, 176-185.

Agarwal, S., Kachroo, P., & Regentova, E. (2016). A hybrid model using logistic regression and wavelet transformation to detect traffic incidents. *IATSS Research*, 40(1), 56-63.

Yin, Y., & Shang, P. (2016). Forecasting traffic time series with multivariate predicting method. *Applied Mathematics and Computation*, 291, 266-278.

Li, L., Su, X., Wang, Y., Lin, Y., Li, Z., & Li, Y. (2015). Robust causal dependence mining in big data network and its application to traffic flow predictions. *Transportation Research Part C : Emerging Technologies*, 58, 292-307.

- 
- Kerner, B. S. (2009). Introduction to modern traffic flow theory and control : the long road to three-phase traffic theory. Springer Science & Business Media.
- Davis, L. C. (2004). Multilane simulations of traffic phases. *Physical Review E*, 69(1), 016108.
- Kerner, B. S. (1998). Experimental features of self-organization in traffic flow. *Physical review letters*, 81(17), 3797. Kerner, B. S. (1999). The physics of traffic. *Physics World*, 12(8), 25.
- Kerner, B. (1999). Congested traffic flow : Observations and theory. *Transportation Research Record : Journal of the Transportation Research Board*, (1678), 160-167.
- Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams : Some experimental findings. *Transportation Research Part B : Methodological*, 42(9), 759-770.
- Moretti, F., Pizzuti, S., Panzieri, S., & Annunziato, M. (2015). Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing*, 167, 3-7.
- Krajzewicz, D., Erdmann, J., Behrisch, M., & Bieker, L. (2012). Recent development and applications of SUMO-Simulation of Urban MObility. *International Journal On Advances in Systems and Measurements*, 5(3&4), 128-138.
- Kriegel, H. P., Renz, M., Schubert, M., & Ze, A. (2008, April). Statistical density prediction in traffic networks. In *Proceedings of the 2008 SIAM International Conference on Data Mining* (pp. 692-703). Society for Industrial and Applied Mathematics.
- Lindveld, C. D., Thijs, R., Bovy, P., & Van der Zijpp, N. (2000). Evaluation of online travel time estimators and predictors. *Transportation Research Record : Journal of the Transportation Research Board*, (1719), 45-53.
- Kumar, K., Parida, M., & Katiyar, V. K. (2013). Short term traffic flow prediction for a non-urban highway using artificial neural network. *Procedia-Social and Behavioral Sciences*, 104, 755-764.
- Kumar, K., Parida, M., & Katiyar, V. K. (2015). Short term traffic flow prediction in heterogeneous condition using artificial neural network. *Transport*, 30(4), 397-405.
- Das, A. K., & Asundi, J. (2012). A simple explicit model approximating the relationship between speed and density of vehicular traffic on urban roads. *International Journal of Critical Infrastructures* 4, 8(2-3), 195-204.
- Chu, Y., & Yang, A. (2016). Study on the Time Series of Traffic Flow Based on Algorithm Complexity Measure and Approximate Entropy. *Journal of Computational and Theoretical Nanoscience*, 13(7), 4730-4734.
- Zhu, Z., Peng, B., Xiong, C., & Zhang, L. (2016). Short term traffic flow prediction with linear conditional Gaussian Bayesian network. *Journal of Advanced Transportation*, 50(6), 1111-1123.
- Xu, C., Li, Z., & Wang, W. (2016). Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming. *Transport*, 31(3), 343-358.
- Abdelatif, S., Makhoul, D., Roose, P., & Bechtache, D. (2016, August). Loop Speed Trap Data Collection Method for an Accurate Short-Term Traffic Flow Forecasting. In *International Conference on Mobile Web and Information Systems* (pp. 56-64). Springer International Publishing.



---

Savas, O., Jin, G., & Deng, J. (2013, May). Trust management in cloud-integrated wireless sensor networks. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 334- 341). IEEE.

Bai, C., Peng, Z. R., Lu, Q. C., & Sun, J. (2015). Dynamic bus travel time prediction models on road with multiple bus routes. *Computational intelligence and neuroscience*, 2015, 63.

Mardani, A., Jusoh, A., & Zavadskas, E. K. (2015). Fuzzy multiple criteria decision-making techniques and applicationsTwo decades review from 1994 to 2014. *Expert Systems with Applications*, 42(8), 4126-4148.

Lust, T., & Rolland, A. (2014, July). 2-additive Choquet optimal solutions in multi objective optimization problems. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 256-265). Springer International Publishing.

Su, K. L., Jau, Y. M., & Jeng, J. T. (2011). Modeling of Nonlinear Aggregation for Information Fusion Systems with Outliers Based on the Choquet Integral. *Sensors*, 11(3), 2426-2446.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on systems, Man, and Cybernetics*, 18(1), 183-190.

Yager, R. R. (1993). Families of OWA operators. *Fuzzy sets and systems*, 59(2), 125-148.

Yager, R. R., & Filev, D. P. (1994). Parameterized and-like and or-like OWA operators. *International Journal of General Systems*, 22, 297316.

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European journal of operational research*, 89(3), 445-456.

Grabisch, M. (1995). Fuzzy integral in multicriteria decision making. *Fuzzy sets and Systems*, 69(3), 279-298.

Grabisch, M. (1995, March). A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE Int* (Vol. 1, pp. 145-150). IEEE.

Wei, C. H., & Lee, Y. (2007). Sequential forecast of incident duration using Artificial Neural Network models. *Accident Analysis & Prevention*, 39(5), 944-954.

Wang, J., Shi, Q., & Lu, H. (2005, June). The study of short-term traffic flow forecasting based on theory of chaos. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* (pp. 869-874). IEEE.

Chen, M., & Chien, S. (2001). Dynamic freeway travel-time prediction with probe vehicle data : Link based versus path based. *Transportation Research Record : Journal of the Transportation Research Board*, (1768), 157-161.

Gligorijevi, V., & Prulj, N. (2015). Methods for biological data integration : perspectives and challenges. *Journal of the Royal Society Interface*, 12(112), 20150571.

Haas, C. (2006). A model for data fusion in civil engineering. *Intelligent Computing in Engineering and Architecture*, 315-319.

Duro, J. A., Padget, J. A., Bowen, C. R., Kim, H. A., & Nassehi, A. (2016). Multi-sensor data fusion framework for CNC machining monitoring. *Mechanical systems and signal processing*, 66, 505-520.

Santos, J. P., Cremona, C., Orcesi, A. D., & Silveira, P. (2016). Early damage detection based on pattern recognition and data fusion. *Journal of Structural Engineering*, 143(2), 04016162.

Bachmann, C., Abdulhai, B., Roorda, M.J., Moshiri, B. : A comparative assessment of



---

multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling. *Transportation research part C : emerging technologies* 26, 33–48 (2013)

Feng, Y., Head, K.L., Khoshmaghani, S., Zamanipour, M. : A real-time adaptive signal control in a connected vehicle environment. *Transportation Research Part C : Emerging Technologies* 55, 460–473 (2015)

Ji, H., Xu, A., Sui, X., Li, L. : The applied research of kalman in the dynamic travel time prediction. In : *Geoinformatics, 2010 18th International Conference on*. pp. 1–5. IEEE (2010)

Patnaik, J., Chien, S., Bladikas, A. : Estimation of bus arrival times using apc data. *Journal of public transportation* 7(1), 1 (2004)

Qian, J., Eglese, R. : Fuel emissions optimization in vehicle routing problems with time-varying speeds. *European Journal of Operational Research* 248(3), 840–848 (2016)

Williams, B., Durvasula, P., Brown, D. : Urban freeway traffic flow prediction : application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record : Journal of the Transportation Research Board* (1644), 132–141 (1998)

Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A. : An ordered fractional split approach for aggregate injury severity modeling (2015)

Zhong, J.t., Ling, S. : Key factors of k-nearest neighbours nonparametric regression in short-time traffic flow forecasting. In : *Proceedings of the 21st International Conference on Industrial Engineering and Engineering Management 2014*. pp. 9–12. Springer (2015)

Li, D. F. (2010). Multiattribute decision making method based on generalized OWA operators with intuitionistic fuzzy sets. *Expert Systems with Applications*, 37(12), 8673-8678.

Zimmermann, H. J. (2011). *Fuzzy set theory—and its applications*. Springer Science & Business Media

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1996). Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and systems*, 79(2), 175-190.

Li, D. F., Huang, Z. G., & Chen, G. H. (2010). A systematic approach to heterogeneous multiattribute group decision making. *Computers & Industrial Engineering*, 59(4), 561-572.

Yager, R. R. (2004). Generalized OWA aggregation operators. *Fuzzy Optimization and Decision Making*, 3(1), 93-107.

Whaiduzzaman, M., Sookhak, M., Gani, A., & Buyya, R. (2014). A survey on vehicular cloud computing. *Journal of Network and Computer Applications*, 40, 325-344.

D. Nurmi, R. Wolski, C. Grzegorzczak, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov. The eucalyptus open-source cloud-computing system. In *Proceedings of the IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID)*, pages 124–131, May 2009.

Olariu, S., & Weigle, M. C. (Eds.). (2009). *Vehicular networks : from theory to practice*. Crc Press.

Al Mamun, M. A., Anam, K., Onik, M. F. A., & Esfar-E-Alam, A. M. (2012, October). Deployment of cloud computing into vanet to create ad hoc cloud network architecture. In *Proceedings of the world congress on engineering and computer science* (Vol. 1, pp. 24-26).

Yan, G., Rawat, D. B., & Bista, B. B. (2012, July). Towards secure vehicular clouds. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on* (pp. 370-375). IEEE.

Eltoweissy, M., Olariu, S., & Younis, M. (2010). Towards autonomous vehicular clouds. *Ad*

---

hoc networks, 1-16.

Hussain, R., Son, J., Eun, H., Kim, S., & Oh, H. (2012, December). Rethinking vehicular communications : Merging VANET with cloud computing. In *Cloud Computing Technology and Science (CloudCom)*, 2012 IEEE 4th International Conference on (pp. 606-609). IEEE.

Bitam, S., Mellouk, A., & Zeadally, S. (2015). VANET-cloud : a generic cloud computing model for vehicular Ad Hoc networks. *IEEE Wireless Communications*, 22(1), 96-102.

Arif, S., Olariu, S., Wang, J., Yan, G., Yang, W., & Khalil, I. (2012). Datacenter at the airport : Reasoning about time-dependent parking lot occupancy. *IEEE Transactions on Parallel and Distributed Systems*, 23(11), 2067-2080.

Olariu, S., & Weigle, M. C. (Eds.). (2009). *Vehicular networks : from theory to practice*. Crc Press.

Al Mamun, M. A., Anam, K., Onik, M. F. A., & Esfar-E-Alam, A. M. (2012, October). Deployment of cloud computing into vanet to create ad hoc cloud network architecture. In *Proceedings of the world congress on engineering and computer science (Vol. 1, pp. 24-26)*.

Kenney, J. B. (2011). Dedicated short-range communications (DSRC) standards in the United States. *Proceedings of the IEEE*, 99(7), 1162-1182.

Brickley, O., Koubek, M., Rea, S., & Pesch, D. (2010, March). A network centric simulation environment for CALM-based cooperative vehicular systems. In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (p. 75)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Li, Y. J. (2010, November). An overview of the DSRC/WAVE technology. In *International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (pp. 544-558)*. Springer, Berlin, Heidelberg.

DSRC Implementation Guide – A guide to users of SAE J2735 message sets over DSRC. SAE International, 2010.

Morgan, Y. L. (2010). Notes on DSRC & WAVE standards suite : Its architecture, design, and characteristics. *IEEE Communications Surveys & Tutorials*, 12(4), 504-518.

Hartenstein, H., & Laberteaux, K. (Eds.). (2009). *VANET vehicular applications and inter-networking technologies (Vol. 1)*. John Wiley & Sons.

Jerbi, Moez. *Protocoles pour les communications dans les réseaux de véhicules en environnement urbain : Routage et GeoCast basés sur les intersections*. Diss. Evry-Val d'Essonne, 2008.

IEEE P1609.1 SWG, IEEE Trial-Use Standard for Wireless Access in Vehicular Environments (WAVE), Resource Manager, IEEE Computer Society, 2006.

IEEE P1609.2 SWG, IEEE Trial-Use Standard for Wireless Access in Vehicular Environments (WAVE), Security Services for Applications and Management Messages, IEEE Computer Society, 2006.

IEEE 1609.3 SWG, IEEE Standard for Wireless Access in Vehicular Environments (WAVE), Networking Services. IEEE Computer Society, 2010.

Jiang, Daniel, and Luca Delgrossi. "IEEE 802.11 p : Towards an international standard for wireless access in vehicular environments." *Vehicular Technology Conference, 2008. VTC Spring 2008*. IEEE. IEEE, 2008.

Bossom, R., Brignolo, R., Ernst, T., Evensen, K., Frötscher, A., Höfs, W., ... & Kulp, I. (2009). *European ITS Communication Architecture—Overall Framework—Proof of Concept Implementation*.

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

Gerla, M. (2012, June). Vehicular cloud computing. In Ad Hoc Networking Workshop (Med-Hoc-Net), 2012 The 11th Annual Mediterranean (pp. 152-155). IEEE.

Seredynski, M., Arnould, G., & Khadraoui, D. (2013, November). The emerging applications of intelligent vehicular networks for traffic efficiency. In Proceedings of the third ACM international symposium on Design and analysis of intelligent vehicular networks and applications (pp. 101-108). ACM.

Leng, Y., & Zhao, L. (2011, August). Novel design of intelligent internet-of-vehicles management system based on cloud-computing and internet-of-things. In Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference On (Vol. 6, pp. 3190-3193). IEEE.

B. S. Kerner, S.L. The Physics of Traffic Springer (2004)

Artuñedo, A., del Toro, R. M., & Haber, R. (2016, November). Consensus-Based Cooperative Control Approach Applied to Urban Traffic Networks. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 1, No. 2, p. 29).