Analyse statistique d'une expérience d'étude de l'éco-conduite : vers la conception d'un « éco index »

Cindie Andrieu

Master 2 de statistiques – 2008/2009

Responsable du master : Mr Deheuvels
Responsable de stage : Mr Saint Pierre

Remerciements

J’adresse mes respectueux remerciements à Jacques Ehrlich, directeur du LIVIC, pour m’avoir accueilli au LIVIC.

Mes sincères remerciements vont à mon responsable de stage Guillaume Saint Pierre pour sa confiance dans mon travail, sa permanente disponibilité et son aide tout au long de mon stage.

Je tiens à remercier Francis Dupin pour m’avoir accueilli dans son bureau, pour tous ses précieux conseils et pour m’avoir permis de travailler dans une ambiance agréable.

Je remercie Benoit Lusetti pour tout le travail effectué sur le véhicule, et qui nous a permis de recueillir les données nécessaires à cette étude.

Je remercie toute l’équipe du LIVIC qui a eu la gentillesse de participer à cette expérience sur l’éco-conduite, et particulièrement Benoit et Aurélien Cord pour avoir encadré certains participants.

J’adresse mes remerciements à Mr Nikssarian et à Mr Corrier, moniteurs d’éco-conduite de l’AFT-IFTIM, à Mr Poissonnet et Mr Chauveau du Ministère de la Défense, ainsi qu’à toutes les personnes extérieures au LIVIC qui ont participé à cette expérience.

Enfin je remercie mes camarades stagiaires Nezly et Vincent pour avoir contribué à créer une ambiance de travail agréable et pour m’avoir permis de trouver le temps moins long dans les transports.

Table des matières

I Introduction générale 9

II Présentation du LIVIC 11
1 Historique ……. 11
2 Démarche scientifique du LIVIC…….. 11
3 Plan d'accès ……. 12
4 Coordonnées utiles ……... 13

III Etat de l'art 14
1 Introduction …….. 14
2 Facteurs de réduction de consommation et éco-conduite…….. 15
3 L'éco-conduite ……. 17
3.1 Un style de conduite adaptée à la technologie moteur ……. 17
3.2 Principes fondamentaux de l'éco-conduite …….. 18
3.3 Avantages de l'éco-conduite …… 23
3.4 Inconvénients éventuels de l'éco-conduite …….. 23
4 Conclusion …… 24

IV Présentation de l'expérience 25
1 Description de l'expérience ……. 25
1.1 Description générale …… 25
1.2 Circuit …….. 25
1.3 Protocole d'expérience …… 28
2 Types de conduite ……... 28
2.1 Conduite normale ……. 28
2.2 Conduite économique ou « green driving » ……. 28
3 Considérations générales ……. 29
4 Sujets particuliers …… 30

V Données issues de l'expérience 31
1 Données issues du bus CAN ……... 31
2 Données issues du débitmètre ……. 32
3 Données supplémentaires …… 33
4 Prétraitement des données ……... 33
4.1 Reconstitution de la consommation cumulée issue du bus CAN …… 33
4.2 Problèmes observés avec le bus CAN ……. 34
4.2.1 Exemple de saut de la vitesse instantanée …… 34
4.2.2 Exemple de saut de la consommation …….. 35

4.2.3 Reconstitution de la distance parcourue issue du bus CAN et correction des sauts d'odomètre ……..... 36
4.3 Interpolation de la consommation issue du débitmètre …….. 38
5 Ajout de variables supplémentaires …… 39
5.1 Plages de régime moteur ……. 40
5.2 Rapports de boîte de vitesse …… 40

VI Indicateurs et base de données finale 42
1 Tableau des indicateurs ……... 42
2 Indicateurs supplémentaires liés à la consommation ……... 42
2.1 Temps passé à l'arrêt ……... 42
2.2 Temps passé en frein moteur ……... 43
2.3 Temps passé avec une vitesse non optimale …… 43
3 Indicateurs fortement liés à la consommation ……. 43
3.1 Positive Kinetic Energy (PKE) …… 43
3.2 Relative Positive Acceleration (RPA) ……. 43
3.3 Intégrale du carré de l'accélération …….. 44
4 Indicateurs relatifs au conducteur …… 44
5 Indicateurs relatifs au parcours …… 44
6 Etat du trajet ……. 45
7 Comparaison entre les deux mesures de consommation ……..... 45

VII Analyse descriptive des résultats 46
1 Répartition des conducteurs ……. 46
2 L'éco-conduite et la réduction de consommation …… 47
3 L'éco-conduite et la vitesse ……. 49
4 La mise en pratique des règles d'éco-conduite …… 53
4.1 Un régime moteur faible et un rapport de boîte élevé ……. 53
4.1.1 Utilisation du moteur …… 53
4.1.2 Répartition des rapports de boîte …….. 54
4.1.3 Analyse selon la vitesse de consigne …… 56
4.2 Anticipation du trafic ……... 60
4.2.1 Gestion de l'accélération ……... 60
4.2.2 Gestion des freinages …… 61
5 Conclusion …… 63

VIII Analyse statistique multivariée de l'ensemble des trajets 64
1 Suppression des trajets avec la climatisation ……... 64
2 Classification des variables …….. 65
2.1 Classification hiérarchique ascendante (CAH) …… 65
2.2 Sélection automatique des variables par régression linéaire multiple …….. 69
2.2.1 Généralités sur la régression linéaire ……. 69
2.2.2 Méthodes de régression …….. 71
2.2.3 Sélection automatique des variables avec la méthode stepwise ……. 72
3 Analyse en composantes principales ……. 73
3.1 Généralités sur l'ACP …… 73
3.2 Représentation graphique d'une ACP …… 75
3.3 Analyse en composantes principales de l'ensemble des trajets ……. 75

IX Modèles de prédiction 80
1 Prédiction de la consommation moyenne par régression linéaire …… 80
2 Modélisation de la probabilité d'être en éco-conduite par régression logistique …. 83
2.1 Généralités sur la régression logistique binaire ……. 84
2.2 Modèle logistique avec cinq variables explicatives …….. 85
2.2.1 Construction du modèle et critères de qualité …… 85
2.2.2 Interprétation des coefficients : odds-ratio ……. 87
2.2.3 Capacité prédictive du modèle …… 87
2.3 Modèle logistique avec une variable explicative …… 90
3 Conclusion …….. 95

X Conclusion générale 96

Bibliographie 98

A Tableau des indicateurs 100

B Tableau des corrélations pour les variables de chaque groupe obtenu par la classification hiérarchique ascendante 105

Liste des figures

II.1 Plan d'accès au LIVIC …… 12
II.2 Les bureaux du LIVIC …… 13

III.1 Logo éco-conduite ……. 15
III.2 Schéma du moteur sur une voiture à essence récente de catégorie moyenne …. 17
III.3 Logo conduite écolo …… 18
III.4 Courbes de consommation en fonction de la vitesse et du rapport de boîte …… 19
III.5 Consommation moyenne en fonction du rapport de boîte pour différentes cylindrées ……. 20
III.6 Surconsommation du véhicule en fonction de la pression des pneus …….. 21
III.7 Surconsommation du véhicule en fonction du chargement d'objets sur le toit 22
III.8 Avantages de l'éco-conduite ……. 23

IV.1 Image satellite du circuit ……... 26
IV.2 Carte de la zone du circuit …… 27
IV.3 Règles d'éco-conduite issues du projet Ecodrive de l'UE …… 29

V.1 Illustration du bus CAN ……. 31
V.2 Reconstitution de la consommation cumulée issue du bus CAN …….. 34
V.3 Superposition de la courbe initiale de la vitesse et de la courbe lissée …….. 35
V.4 Superposition de la courbe initiale de la consommation issue du bus CAN et de la courbe lissée ……... 36
V.5 Superposition de la courbe initiale, de la courbe représentant la distance parcourue reconstituée, et de la courbe lissée ……. 37

V.6 Superposition de la courbe initiale, de la courbe lissée et de la courbe représentant la distance parcourue reconstituée après lissage …… 38
V.7 Superposition de la courbe initiale, de la courbe interpolée et de la courbe lissée …… 39
V.8 Courbe initiale superposée à la courbe lissée et à la classification obtenue …... 41

VII.1 Répartition des conducteurs par sexe, âge, expérience, et connaissances de l'éco-conduite …… 47
VII.2 Boîtes à moustaches représentant la consommation moyenne (en l/100 km) selon le type de conduite. L'anova détecte une différence significative (DF=1, F=40.62, p<0.0001) entre les deux types de conduite …….. 48
VII.3 Boîtes à moustaches représentant la consommation moyenne (en l/100 km) des trajets effectués en conduite économique selon les connaissances du conducteur sur l'éco-conduite …….. 49
VII.4 Boîtes à moustaches représentant la vitesse moyenne (en km/h) des trajets selon le type de conduite. L'anova détecte une différence significative (DF=1, F=6.6, p=0.0137) entre les deux types de conduite …….. 50
VII.5 Répartition des vitesses instantanées selon le type de conduite …… 50
VII.6 Répartition des vitesses instantanées selon les vitesses de consigne pour chaque type de conduite ……. 51
VII.7 Boîtes à moustaches représentant le pourcentage de temps passé au-delà de la Vitesse de consigne selon le type de conduite. L'anova détecte une différence significative (DF=1, F=15.9, p=0.0002) entre les deux types de conduite …… 52
VII.8 Boîtes à moustaches représentant la durée moyenne du trajet en minutes selon le type de conduite. L'anova détecte une différence significative (DF=1, F=6.94, p=0.0116) entre les deux types de conduite …… 52
VII.9 Boîtes à moustaches représentant le régime moteur ainsi que le régime moteur moyen auquel s'effectue le changement de vitesse selon le type de conduite. L'anova détecte une différence significative dans les deux cas (DF=1, F=69812, p<0.0001 pour le 1er graphique et DF=1, F=21.5, p<0.0001 pour le 2ème) ……... 53
VII.10 Distribution des valeurs du régime moteur pour chaque type de conduite …….. 54
VII.11 Répartition des rapports de boîte de vitesse selon le type de conduite …… 54
VII.12 Distribution et distribution cumulée (en haut) des vitesses instantanées selon le rapport de boîte pour chaque type de conduite …… 55
VII.13 Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 50km/h …….. 57
VII.14 Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite, et pour les zones limitées à 50km/h. Les lignes de niveau sont identiques entre les graphiques …….. 57
VII.15 Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 70km/h ……. 58
VII.16 Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite, et pour les zones limitées à 70km/h. Les lignes de niveau sont identiques entre les graphiques …….. 58
VII.17 Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 90km/h …… 59
VII.18 Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite,

et pour les zones limitées à 90km/h. Les lignes de niveau sont identiques entre les graphiques …….. 59
VII.19 Distribution des valeurs d'accélérations/décélérations du véhicule pour chaque type de conduite …….. 60
VII.20 Boîtes à moustaches (a) et distribution (b) du pourcentage d'enfoncement de la pédale d'accélérateur selon le type de conduite …….. 61
VII.21 Boîte à moustaches représentant le nombre de freinages selon le type de conduite. L'anova détecte une différence significative (DF=1, F=20.94, p<0.0001) entre les deux types de conduite …… 61
VII.22 Boîte à moustaches représentant le pourcentage de distance parcourue en freinant selon le type de conduite. L'anova détecte une différence significative (DF=1, F=16.4, p=0.0002) entre les deux types de conduite …….. 62
VII.23 Boîte à moustaches représentant le pourcentage de temps de trajet effectué en frein moteur selon le type de conduite. L'anova ne détecte pas de différence significative (DF=1, F=0.5, p=0.482) entre les deux types de conduite……. 62
VII.24 Boîte à moustaches représentant le pourcentage de temps de trajet effectué en frein moteur en conduite économique selon les connaissances du conducteur sur l'éco-conduite. L'anova ne détecte pas de différence significative (DF=3, F=1.43, p=0.266) entre les quatre classes …….. 63

VIII.1 Boîtes à moustaches représentant la consommation moyenne (en l/100 km) selon l'utilisation ou non de la climatisation. L'anova détecte une différence significative au risque de 5% (DF=1, F=4.16, p=0.0473) ……. 65
VIII.2 Arbre de classification des 49 variables conservées dans la base de données finale. 68
VIII.3 Projection des points M_i sur la droite …….. 69
VIII.4 Décomposition des différents écarts …… 71
VIII.5 Résultats de la procédure REG du logiciel SAS : tableau récapitulatif des étapes de la méthode stepwise et graphique des résidus ……. 73
VIII.6 ACP de l'ensemble des trajets et classification ascendante hiérarchique ……..... 77
VIII.7 Premier plan principal de l'ACP de l'ensemble des trajets …….. 77
VIII.8 Diagrammes en bâtons représentant les moyennes des variables centrées-réduites pour chacun des groupes …….. 78

IX.1 Graphique des résidus standardisés en fonction de la valeur prédite $\hat{Y}$ …….. 83
IX.2 Courbe ROC pour le modèle avec cinq variables explicatives …… 89
IX.3 Graphique représentant les valeurs observées, les valeurs prédites pour un seuil s=0.6, ainsi que la probabilité estimée d'être en éco-conduite en fonction de la variable PKE …… 92
IX.4 Courbe ROC pour le modèle avec une variable explicative …… 94

Liste des tableaux

V.1 Variables mesurées par le bus CAN …… 32
V.2 Variables mesurées par le débitmètre …….. 32
V.3 Exemple de saut de la variable VInstantanee …….. 35
V.4 Exemple de saut de la variable ConsoCan …….. 36
V.5 Variables de repérage …….. 39

V.6 Répartition des plages de régime moteur ……. 40
V.7 Paliers correspondant aux changements de rapports de boîte selon la marque du véhicule …… 41

VI.1 Indicateurs relatifs au conducteur …… 44
VI.2 Indicateurs relatifs au parcours …… 45

VII.1 Répartition de la consommation en l/100 km pour l'ensemble des conducteurs selon le type de conduite …….. 48

VIII.1 Nouvelles variables représentant le pourcentage de temps passé dans chaque plage de vitesse et dans chaque plage d'accélérations/décélérations ……... 66
VIII.2 Variables sélectionnées d'après la classification ascendante hiérarchique …….. 67
VIII.3 Moyennes des différentes variables pour chacun des groupes ……. 79

IX.1 Tableau des corrélations entre les variables explicatives ……... 81
IX.2 Tableau récapitulatif des étapes de la méthode stepwise …….. 82
IX.3 Tableau des résultats estimés des paramètres ……... 82
IX.4 Résultats de la procédure LOGISTIC du logiciel SAS : critères de qualité du modèle à cinq variables explicatives …….. 86
IX.5 Résultats de la procédure LOGISTIC du logiciel SAS : estimation des paramètres et test de Wald pour le modèle à cinq variables explicatives …….. 86
IX.6 Résultats de la procédure LOGISTIC du logiciel SAS : odds-ratios et intervalles de confiance à 95% pour le modèle à cinq variables explicatives ……..... 87
IX.7 Proportion de bien classés et de mal classés parmi les 42 trajets pour un seuil s=0.6 …… 88
IX.8 Résultats de la procédure LOGISTIC du logiciel SAS : Table de classification pour différents niveaux de probabilité …… 89
IX.9 Récapitulatif des différentes étapes de la régression logistique avec la méthode stepwise ……... 90
IX.10 Résultats de la procédure LOGISTIC du logiciel SAS : critères de qualité du modèle à une variable explicative …… 90
IX.11 Résultats de la procédure LOGISTIC du logiciel SAS : estimation des paramètres et test de Wald pour le modèle à une variable explicative …….. 91
IX.12 Résultats de la procédure LOGISTIC du logiciel SAS : odds-ratios et intervalles de confiance à 95% pour le modèle à une variable explicative …….. 91
IX.13 Proportion de bien classés et de mal classés parmi les 42 trajets pour un seuil s=0.6 …….. 92
IX.14 Résultats de la procédure LOGISTIC du logiciel SAS : Table de classification pour différents niveaux de probabilité …… 94

Chapitre I

Introduction générale

Ce travail a été effectué au LIVIC (Laboratoire sur les Interactions Véhicules-Infrastructures-Conducteurs) dans le cadre d'un stage de fin d'études du master 2 de statistiques de l'université Pierre et Marie Curie. Il s'agit ici d'une analyse statistique appliquée au secteur de l'automobile, et plus particulièrement à l'évaluation des systèmes d'aide à la conduite. Mon stage a consisté à travailler sur toutes les étapes d'une expérience, de la collecte des données, en passant par le calcul d'indicateurs, pour aboutir à la mise au point d'un modèle.

Cette étude s'inscrit dans un contexte de développement d'une conduite économique en termes de consommation de carburant et de rejet de CO_2. La sensibilisation envers les impératifs écologiques et la nécessité d'un développement durable s'intensifie dans les pays industrialisés du monde entier. Le transport étant l'un des secteurs les plus polluants, de nombreuses recherches ont été effectuées ces dernières années afin de réduire la consommation de carburant et l'émission de gaz à effets de serre. Des études ont ainsi été réalisés afin d'étudier l'impact de certains systèmes d'aide à la conduite sur la consommation de carburant. Les chercheurs du LIVIC ont notamment étudié l'influence du limiteur de vitesse LAVIA sur la consommation. Cependant, de nombreuses études ont aussi montrées que le style de conduite avait un impact important sur la consommation de carburant. Ainsi, les nombreux développements relatifs à la technologie moteur des véhicules ont permis de voir apparaitre un nouveau style de conduite, appelée éco-conduite, permettant des progrès immédiats en termes de consommation et donc de réduction de rejets polluants. L'éco-conduite permettrait de réduire de près de 15% sa consommation de carburant.

Le LIVIC a donc organisé une expérience destinée à mesurer et à évaluer les paramètres d'une conduite économique, désignée par le terme « éco-conduite ». Ainsi une vingtaine de conducteurs ont effectués deux fois le même trajet : une première fois en conduisant « normalement », et une deuxième fois en respectant certains principes d'éco-conduite dans le but d'économiser du carburant. L'objectif de cette étude est d'approfondir les connaissances sur l'éco-conduite, et de déterminer à l'aide d'une analyse statistique multivariée les indicateurs caractéristiques d'une conduite économique. L'usage d'un véhicule du LIVIC équipé de multiples capteurs et d'un débitmètre a permis la réalisation de cette étude.

Nous commencerons par un état de l'art sur l'éco-conduite, puis nous décrirons l'expérience que nous avons mise en place. Les différents variables mesurées seront détaillées, ainsi que les variables et indicateurs supplémentaires que nous utiliserons dans nos analyses statistiques. Nous décrirons aussi les difficultés rencontrées lors du recueil des données, ainsi que les solutions pour y remédier. Ensuite, une première approche descriptive des résultats sera présentée afin d'étudier l'impact de l'éco-conduite sur la consommation de carburant et

sur la vitesse, et d'analyser la mise en pratique des principes de l'éco-conduite. Puis nous effectuerons une analyse statistique multivariée de l'ensemble des trajets afin de déterminer les indicateurs les plus pertinents pour caractériser l'éco-conduite. Enfin, le dernier chapitre sera consacré à la construction d'un modèle de prédiction de la consommation moyenne, et à la mise au point d'un modèle de régression logistique afin de déterminer la probabilité d'être dans une situation d'éco-conduite.

Chapitre II

Présentation du LIVIC

1 Historique

Le LIVIC (Laboratoire sur les Interactions Véhicules-Infrastructure-Conducteurs) est une unité mixte de recherche dont les tutelles sont :

- le LCPC (Laboratoire Central des Ponts et Chaussées)
- l'INRETS (Institut National de Recherche sur les Transports et leur Sécurité)

Le LIVIC a été crée en 1999 et il est composé d'une trentaine de personnes dont huit permanents (chercheurs, ingénieurs, techniciens, doctorants). De plus, il dispose d'un réseau de 7km de pistes d'essais et de véhicules instrumentés.
Ses activités ont pour finalité l'amélioration du fonctionnement des réseaux routiers par le développement de systèmes technologiques d'aides à la conduite.

2 Démarche scientifique du LIVIC

Les activités de recherche du LIVIC sont divisées en 4 orientations stratégiques :

- Rendre la conduite accessible à tous,
- Accroitre la sécurité, la mobilité et réduire l'impact environnemental,
- Développer les « routes automatisées » pour le TC, TI et le fret,
- Evaluer les performances, l'usage et l'impact des systèmes.

Pour les mettre en œuvre le LIVIC est organisé en 3 équipes :

1. Perception de l'environnement proche et lointain

   Sa mission est de développer les techniques de reconstruction de la scène routière dans l'environnement proche et lointain, de situer avec précision les objets mobiles dans cette scène et prévoir leur trajectoire y compris dans des situations environnementales dégradées (brouillard, nuit …).

2. Risque décision et action

   Sa mission est la modélisation du risque, la définition de trajectoire optimale sous contraintes (interactions, infrastructure, environnement, conducteur), l'observation par capteurs proprioceptifs et la conception d'automate d'assistance à la conduite permettant leur mise en œuvre dans différents modes de coopération avec le conducteur pouvant aller

jusqu'à l'automatisation complète. Les différentes stratégies d'optimisation devront être évaluées aussi du point de vue de leur impact sur le trafic.

3. Simulation, intégration test et évaluation

Sa mission est le développement des moyens d'essais et des méthodes d'évaluations de la conduite. Entrent dans son champ d'action :

- Le développement de moyens techniques : simulation, bancs de test. Ces systèmes sont principalement fondés sur les produits issus des deux autres équipes du LIVIC mais aussi sur les acquis de projets partenariaux notamment dans le domaine du handicap. Ceci inclus la construction et/ou l'intégration dans les véhicules et les équipements de bord de route de sous ensembles électroniques, le développement et/ou l'intégration logicielle selon les standards en vigueur (par exemple portage MATLAB vers RTMAPS)
- La conception de méthodes et d'outils pour l'évaluation : études méthodologiques, conception de plans d'expérience, réalisation d'expérimentations, recueil et analyse des données (statistiques).

L'équipe est également responsable de la gestion de l'Atelier Véhicule Instrumenté.

3 Plan d'accès

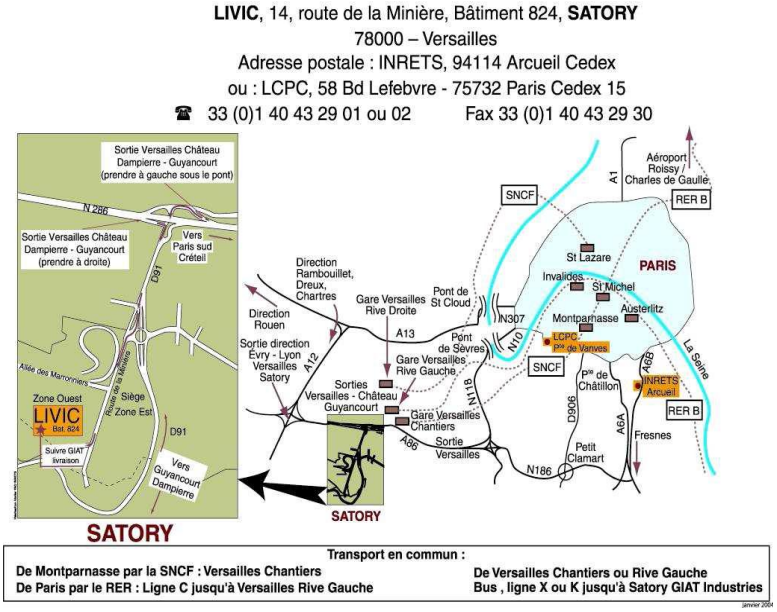


Fig. II.1 – Plan d'accès au LIVIC.

Fig. II.2 - Les bureaux du LIVIC.

4 Coordonnées utiles

LIVIC
14, route de la Minière – Bâtiment 824 – Satory
78000 Versailles
Tél.: +33 (0)1 40 43 29 01 ou 02 - Fax +33 (0)1 40 43 29 30
Site internet : http://www.inrets.fr/ur/livic/

Directeur : Jacques Ehrlich

Tél : +33 (0)1 40 43 29 03
e-mail : jacques.ehrlich@lcpc.fr

Responsable de stage : Guillaume Saint Pierre

Tél : +33 (0)1 40 43 29 33
e-mail : guillaume.saintpierre@lcpc.fr

Chapitre III

Etat de l'art

1 Introduction

Le transport est l'un des secteurs les plus polluants ; il est fortement responsable de l'accroissement de la consommation d'énergie ainsi que des émissions de gaz à effet de serre (26% des émissions de gaz à effet de serre, 34% des émissions de CO2). C'est le transport routier qui domine de manière écrasante le bilan énergétique. Il représente 80% de la consommation d'énergie, tous modes de transport confondus [16]. De nombreux efforts ont été réalisés dans ce domaine. En effet, les véhicules consomment moins mais cette diminution est compensée par l'accroissement du trafic routier.

On peut espérer à long terme, une baisse de la consommation d'énergie grâce au développement de nouvelles technologies : biocarburants, pile à combustible, véhicules hybrides et électriques. Mais de nombreux problèmes restent encore à résoudre dans ce domaine comme le stockage ou le recyclage des batteries pour les véhicules électriques.

Une autre piste est le développement des technologies utilisées pour créer des aides à la conduite efficaces à plusieurs niveaux (sécurité, fiabilité, économies …). Le développement des systèmes ISA (Intelligent Speed Adaptation), qui permettent d'adapter automatiquement le véhicule à la vitesse légale, a ainsi laissé penser à un effet sur la réduction de consommation de carburant. Mais des études ont montrées que si l'impact de ces systèmes sur la sécurité et l'accidentologie est maintenant bien connu, leur impact sur la réduction de consommation est moins avéré. Ainsi une étude réalisée par le LIVIC sur le limiteur de vitesse LAVIA a montré certains effets pervers liés au système (le conducteur « colle » à la limitation de vitesse et n'utilise pas un rapport de boîte adapté) [7]. La conduite induite par de tels systèmes ne peut donc pas être associée simplement à une conduite économique.
En revanche, de nombreuses études ont montrées un fort impact du style de conduite sur la consommation de carburant. Ainsi une conduite économique, souvent désignée par le terme éco-conduite, permettrait de réduire jusqu'à 25% la consommation de carburant. Il s'agit donc de faire prendre conscience aux automobilistes que leur comportement est la clé de progrès immédiats de consommation et par conséquent de réduction de rejets polluants. Au contraire, une conduite agressive augmente énormément la consommation de carburant et les émissions de gaz mais elle ne réduit que très peu le temps de parcours. Ainsi un essai réalisé en Europe a démontré que la conduite agressive, les départs précipités aux feux de circulation et les arrêts brusques réduisaient la durée de déplacement de seulement 4% (l'équivalent de 2,5 minutes sur un déplacement de 60 minutes). Cependant, la consommation de carburant augmentait de 37% et certaines émissions toxiques étaient 5 fois plus élevées [13].

2 Facteurs de réduction de consommation et éco-conduite

Il existe de nombreuses études détaillées concernant la consommation de carburant et les gains susceptibles d'être obtenus par diverses actions. Des programmes de formation des conducteurs et d'apprentissage de l'éco-conduite peuvent permettre d'espérer une réduction de la consommation moyenne de 10 à 15%. Cette conclusion provient d'un programme de formation organisé par l'organisme suisse Quality Alliance Eco-drive [12], et confirmée par plusieurs autres études Européennes. Des recherches ont également démontré que les conducteurs appliquant des techniques d'éco-conduite ont moins d'accidents et limitent l'usure de leur véhicule.

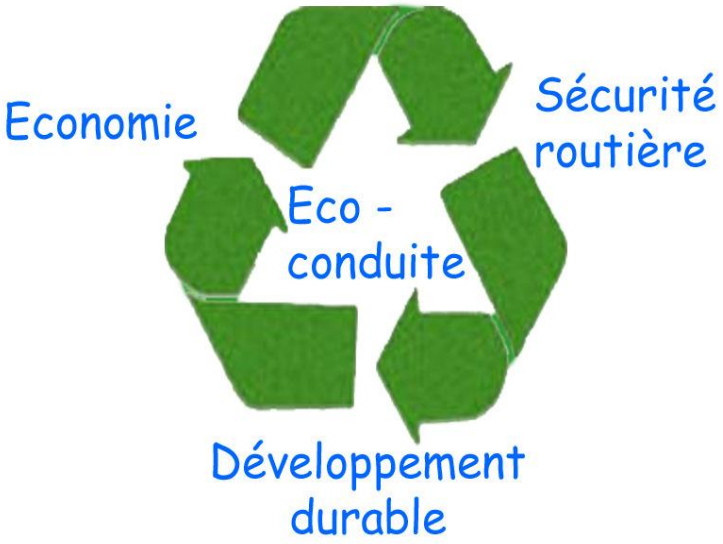


Fig. III.1 – Logo éco-conduite.

En France, l'ADEME a de son coté entrepris des études complémentaires à plus petite échelle (conducteurs de poids lourds uniquement) qui confirment ces gains potentiels. L'ADEME a notamment aidé à faire connaître les résultats du projet BEET (Benchmarking Energy Efficiency in Transport), réalisé par le Département Etudes et Recherches de l'Institut Pédagogique du Transport et de la Logistique/Groupe AFT-IFTIM, en partenariat avec NEA et avec la collaboration de Renault Trucks, projet intégré avec d'autres sous la dénomination de ECLAB dans le cadre du projet européen SAVE. Les gains potentiels en termes de réduction de consommation obtenus grâce à des formations adéquates peuvent être alors très important.

Il est à noter qu'un phénomène d'érosion se produit au bout de 3 à 4 mois suivant la formation, et que le gain réel observé au terme de cette érosion se situe entre 80% et 20% du gain obtenu à l'issue de la formation, moyennant quoi, si les formations sont régulièrement et périodiquement réactualisées, on doit pouvoir atteindre à terme un niveau "permanent" de 75 à 80% de celui effectivement constaté à l'issue de la formation. Il s'ensuit que les gains potentiels qui peuvent être escomptés se situent dans des ordres de grandeur de 3.5 à 4 litres/100km au minimum, voire jusqu'à 5 à 6 litres/100km selon les activités et le champ géographique.

Les résultats obtenus dans le cadre du projet BEET, sur des populations indifférenciées et avant érosion, font apparaître que le gain moyen induit par les formations s'établit à 5.25 litres/100km pour l'ensemble de la population, et qu'il est de 8.64 litres/100km s'agissant de

conducteurs sélectionnés pour leur performances moyennes ou médiocres. Il s'agit là de gains qui se situent "au delà de ce que l'on pouvait raisonnablement espérer" selon les auteurs de l'étude.
Les différents points abordés lors de cette section trouvent un écho dans les conclusions de différentes études. La problématique des temps passés à l'arrêt, c'est à dire des congestions, est notamment abordée par le biais des réductions potentielles de consommation en ayant recours aux serveurs d'information de trafic. Cela peut il permettre de minimiser la consommation de carburant ou, au contraire, induit il un surcroît de consommation au motif que l'itinéraire alternatif, qui permet de contourner la perturbation, rallonge la distance, même si le temps pour arriver à destination est minimisé ?
Diverses expérimentations ont fait apparaître que l'utilisation d'une information de trafic peut induire non seulement des gains de productivité, mais engendrer, sous certaines conditions, des gains réels de consommation. Ce sera le cas notamment lorsque cette information permet d'éviter qu'un couple conducteur-véhicule soit pris dans une perturbation qui l'amène à passer du temps au ralenti et progresser sur le mode "Stop&Go" (c'est à dire que le conducteur va multiplier les accélérations -pour repartir- sans être en mesure d'atteindre un palier de couple optimal, c'est-à-dire entre 1300 et 1800 tr/min) sur une certaine distance et durant un certain temps.
Sachant qu'un véhicule consomme au ralenti 1 à 1.5 litres de carburant à l'heure, et que chaque accélération induit, selon que le conducteur est plus ou moins bien formé à la conduite économique, une consommation de 0.25 à 0.75 litres, alors que la consommation induite par la distance supplémentaire est de l'ordre de 28 à 30 litres/100km, il s'ensuit que théoriquement un allongement de parcours de 10km au motif de l'évitement de perturbation, représentera une consommation de carburant du même ordre de grandeur que celle d'un véhicule pris dans un bouchon durant 30 minutes et qui est amené à procéder à 5 accélérations.
Pour un même allongement du temps de parcours (en termes de distance), toute durée supplémentaire dans le bouchon, ainsi qu'un nombre plus important d'accélérations, donne l'avantage à l'itinéraire alternatif ... qui induit donc une consommation moindre.
De manière plus globale, ces études mettent en avant les principaux facteurs pouvant avoir une influence sur la consommation [16]. On peut schématiquement distinguer :

Les paramètres (fixes) relatifs au véhicule proprement dit.
– Silhouette du véhicule.
– Normes de pollution
– Aérodynamique.
– Roulement et résistance au roulement.
  Pneumatiques plus ou moins usés, taille basse, rechapés, pression des pneus, ...
– Electronique moteur.
  Régulateur, pompe à injection, ...
– Equipements complémentaires.
  Climatisation et chauffage notamment.

Les paramètres (variables) relatifs aux conditions opérationnelles d'exploitation.
– Charge.
– Entretien.
– Adaptation du véhicule à l'utilisation.
– Conditions de route et profils de parcours.
  Il s'agit du type de routes (autoroute, milieu urbain ou inter-urbain) et d'infrastructures (notamment revêtements). Mais aussi des déclivités, tracés et contraintes, conditions

de trafic, trajets de nuit, ...
– Conditions climatiques.
– Contraintes de conduite et vitesse commerciale.
Il est intéressant de noter que l'incidence de ces paramètres "sur-déterminants" peut représenter jusqu'à + ou - 20% des valeurs de consommation lorsque sont additionnés plusieurs facteurs de surconsommation (ou aussi + ou - 10 litres/100km).

3 L'éco-conduite

3.1 Un style de conduite adaptée à la technologie moteur

Ces dix dernières années, la technologie moteur et les performances des véhicules se sont considérablement améliorées. Les émissions polluantes de CO, d'HC et de NO_X ont diminuées grâce à l'introduction des pots catalytiques. La consommation de carburant et les émissions de CO_2 peuvent également être réduites grâce à des innovations technologiques mais ceci n'est que partiellement effectif dans les conditions réelles de circulation. En effet, la plupart des personnes ne conduisent pas leur véhicule comme l'exigeraient la conception et l'étalonnage du moteur. Les constructeurs étalonnent les moteurs de leur véhicule de manière à établir un équilibre délicat entre une maniabilité optimale et des niveaux d'émission d'échappement optimaux. Grâce à l'éco-conduite, il est possible de se rapprocher de ces conditions optimales [13].

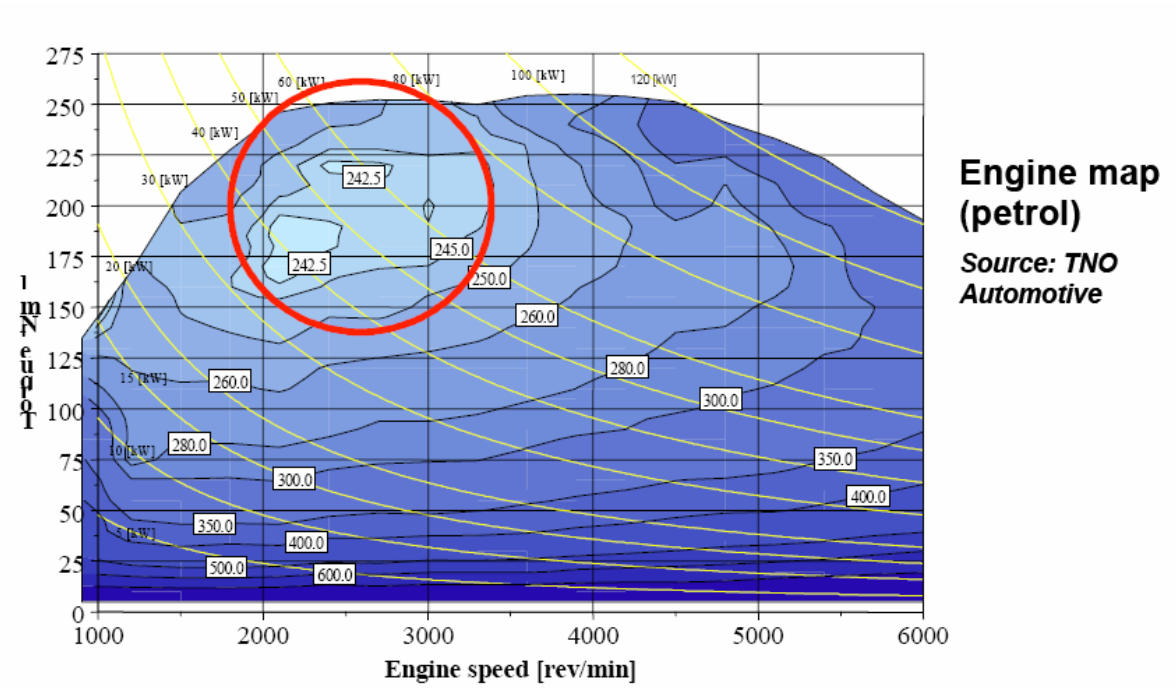


Fig. III.2 - Schéma du moteur sur une voiture à essence récente de catégorie moyenne.

L'axe x du graphique III.2 représente le régime moteur (en tr/min), et l'axe y indique le couple moteur, c'est-à-dire le « travail » que doit fournir le moteur. Pour les voitures à injection (diesel et essence), si une pression franche est exercée sur la pédale d'accélération à bas régime, le couple sera élevé. La zone entourée en rouge sur le schéma précédent indique

la configuration optimale pour consommer moins de carburant. Le conducteur doit donc adopter une conduite lui permettant de se rapprocher le plus possible de cette zone. Les principes d'une conduite économique vont dans ce sens : exploiter cette zone à efficacité énergétique tout en respectant la sécurité routière.

3.2 Principes fondamentaux de l'éco-conduite

Dans le contexte du projet Ecodrive de l'UE [11], les règles d'or de l'éco-conduite sont les suivantes :

1. Passer à la vitesse supérieure dès que possible.
Passer à la vitesse supérieure entre 2 000 et 2 500 tr/min.

2. Maintenir une allure constante.
Enclencher la plus haute vitesse possible et conduire avec un régime moteur faible.

3. Anticiper le trafic.
Regarder le plus loin possible et anticiper le trafic environnant.

4. Décélérer progressivement.
S'il faut ralentir ou s'arrêter, décélérer progressivement en relâchant l'accélérateur à temps et en laissant la voiture en prise (utilisation du frein moteur).

5. Vérifier régulièrement la pression des pneus.
Une pression 25% trop faible accroît la résistance au roulement de 10% et la consommation de carburant de 2%.

Fig. III.3 – Logo conduite écolo.

Revenons plus en détails sur ces principes de base :

1. Monter les rapports rapidement.
Une partie de l'énergie du moteur du véhicule est perdue par la friction mécanique. Ces pertes augmentent avec la vitesse du moteur (ou régime moteur). En conduisant à un régime moteur inférieur, ces pertes restent limitées. Ceci réduit par conséquent la consommation de carburant.

Ainsi monter les rapports de vitesse rapidement et conduire à des rapports de vitesse élevés même à faible vitesse engendre des réductions de la consommation de carburant comme l'illustre le graphique III.4.

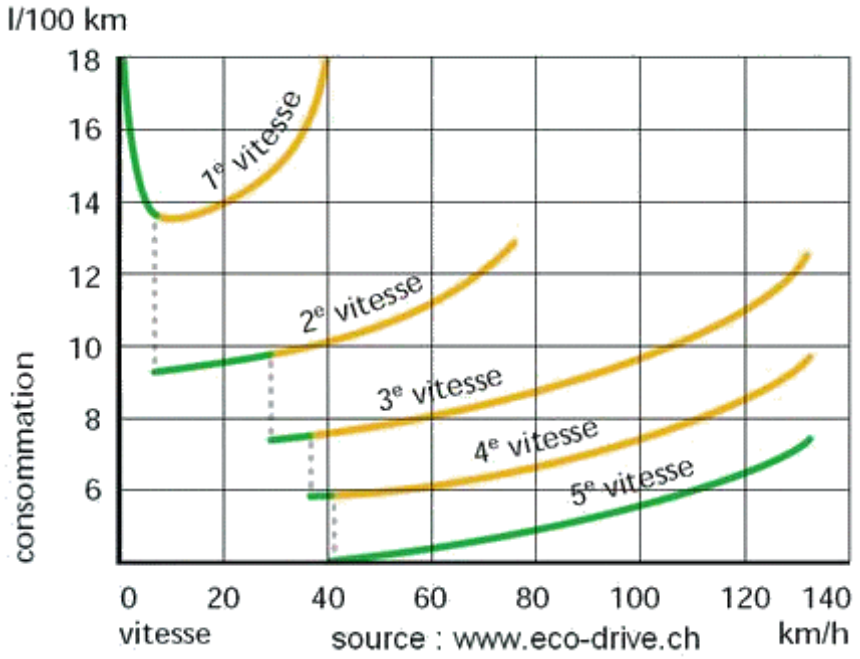


Fig. III.4 – Courbes de consommation en fonction de la vitesse et du rapport de boîte.

Afin d'utiliser l'efficacité du moteur de façon optimale, il est recommandé de passer à la vitesse supérieure à environ 2 000 tr/min en Diesel et 2 500 tr/min en Essence.

2. Maintenir une vitesse constante en utilisant le plus haut rapport possible.

A l'accélération, l'énergie du carburant est utilisée pour propulser le véhicule. Une partie de cette énergie est perdue au freinage (l'énergie perdue est dissipée en chaleur au niveau des disques et des plaquettes de freins). Par conséquent, des accélérations et des freinages répétées demandent beaucoup d'énergie et donc de carburant. Il est donc recommandé d'adopter un mode de conduite fluide et constant en évitant les freinages et accélérations inutiles.

De plus, en gardant un rapport de vitesse élevé et en maintenant une vitesse constante, le gain de consommation de carburant peut atteindre 38% suivant le type de véhicule par rapport à l'utilisation d'un rapport de vitesse plus faible (cf. figure III.5).

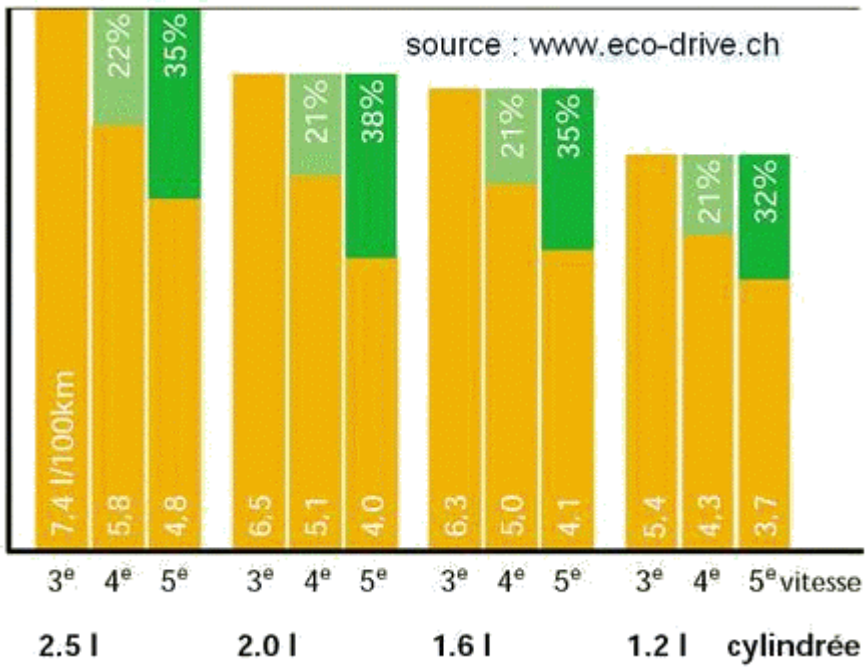


Fig. III.5 – Consommation moyenne en fonction du rapport de boîte pour différentes cylindrées.

3. Anticiper le trafic et conduire moins vite.

Afin de conduire à une vitesse aussi constante que possible, il est important d'anticiper le trafic environnant pour éviter les freinages et accélérations inutiles. Par exemple : à l'approche des feux de signalisation, à l'approche de cyclistes ou de véhicules agricoles, sur une autoroute ou une route chargée … Beaucoup de situations peuvent être anticipées en regardant loin devant.

Suivant les conditions de circulation, il n'est pas possible de conduire aussi vite que ce que permet la limite de vitesse; en adaptant la vitesse il est donc possible d'économiser du carburant. Ainsi sur un parcours de 20km, la différence de temps est de 1 minute et 40 secondes en conduisant à 90km/h au lieu de 80 km/h. Par contre suivant la motorisation du véhicule, l'économie est de 2 à 5% à 80km/h. Outre l'avantage économique de réduire la vitesse, il s'avère que conduire moins vite permet de prévenir beaucoup d'accidents (la distance de freinage est plus courte de 10m à 80km/h qu'à 90km/h).

4. Décélérer progressivement et utiliser le frein moteur.

Les véhicules diesels et essence construits depuis 1990 sont généralement équipés avec un système d'injection électronique qui coupe l'alimentation en carburant du moteur lors de l'utilisation du frein moteur (pas de pression sur la pédale d'accélérateur et vitesse engagée), autrement dit ce sont alors les roues qui entraînent le moteur. Cette technique désignée aussi par le terme « marche sur l'erre » consiste à utiliser l'élan accumulé, c'est-à-dire l'énergie cinétique du véhicule lorsqu'aucune pression n'est appliquée sur l'accélérateur, que l'embrayage n'est pas enclenchée et que la voiture est toujours en prise [1]. Il est important de noter que la marche sur l'erre en descente doit toujours s'effectuer en prise, car la conduite au point mort désactive le frein moteur et exerce une pression importante sur les freins (risque de surchauffe). L'avantage du frein moteur est qu'il peut être utilisé pour économiser du

carburant en relâchant l'accélérateur dès que possible lorsqu'il faut ralentir ou s'arrêter. Il permet aussi de réduire l'utilisation des freins et de réduire ainsi les coûts de maintenance.

5. Vérifier la pression des pneus.
Conduire avec un véhicule dont les pneumatiques sont sous-gonflés peut entrainer une surconsommation de carburant atteignant 8% (cf. figure III.6).

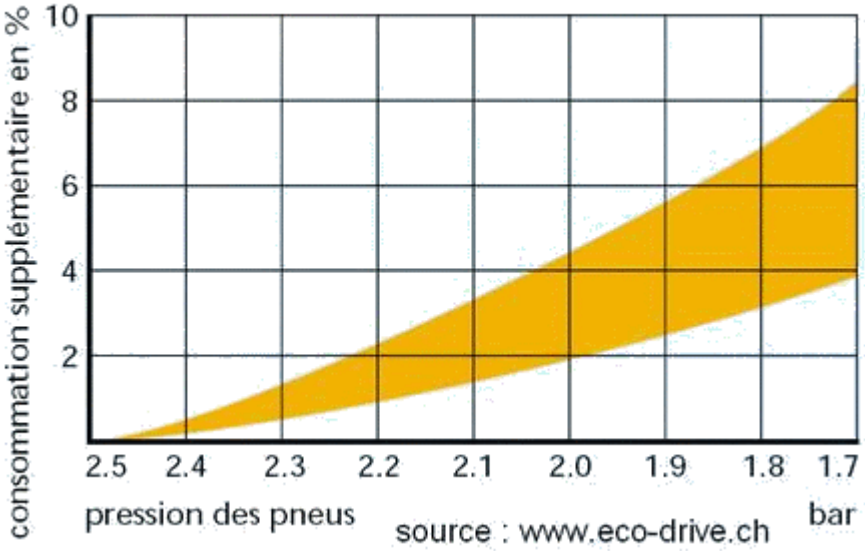


Fig. III.6 – Surconsommation du véhicule en fonction de la pression des pneus.

Citons par ailleurs d'autres principes permettant une conduite économique :

- **Enlever le surpoids inutile et les éléments nuisant à l'aérodynamisme du véhicule**

  Le poids additionnel sur un véhicule entraîne un accroissement de la résistance au roulement du véhicule (friction entre la roue et la route plus importante). Il faut donc éviter de surcharger son véhicule inutilement. Le graphique III.7 montre la surconsommation du véhicule en fonction du chargement d'objets sur le toit (la surconsommation est donnée pour une vitesse de 120km/h).

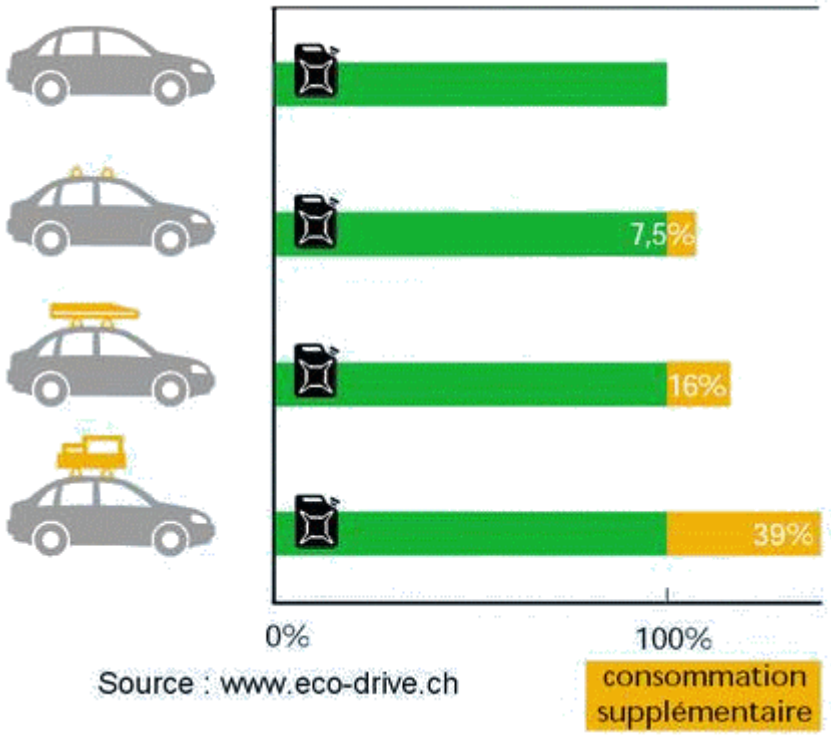


Fig. III.7 – Surconsommation du véhicule en fonction du chargement d'objets sur le toit.

- **Eviter d'utiliser la climatisation**
  Selon une étude de la Société des ingénieurs automobiles (SAE), dans une berline avec fenêtres baissées la consommation de carburant peut augmenter de 8% à 20%, la différence étant plus marquée au-delà des 80km/h. Cependant, la surconsommation moyenne avec la climatisation est de l'ordre de 20%, et selon les chiffres de l'ADEME, à 30° C avec un fort ensoleillement, la surconsommation peut atteindre 30% à 40% en ville, et 15% à 25% sur route-autoroute, si la température de consigne est de 26° C.

- **Entretenir régulièrement son véhicule**
  Un véhicule mal entretenu peut entraîner une surconsommation de carburant pouvant aller jusqu'à 25%. Par exemple, un filtre à air encrassé fait consommer 3% de plus et un moteur mal entretenu, encore plus.

- **Eviter les parcours inutiles et planifier ses déplacements**

- **Couper le moteur en cas d'arrêt ou d'embouteillage**
  Cependant ce principe s'avère difficile à mettre en œuvre étant donné les importantes différences entre les marques et modèles de voitures quant au meilleur moment pour couper le moteur : pour certaines voitures, il est judicieux de couper le moteur en cas d'arrêt de 10 secondes seulement tandis que pour d'autres, les émissions seront plus importantes si l'on coupe le moteur avant un délai de 8 minutes.

3.3 Avantages de l'éco-conduite

L'éco-conduite dispose d'un potentiel permettant de faire des économies de carburant considérables. Mais outre la réduction de consommation de carburant, les autres avantages de l'éco-conduite sont :

- **La réduction d'émissions de gaz à effets de serre** notamment en termes de CO2, lequel exerce un impact direct sur le réchauffement du climat.
- **La diminution des risques d'accidents** grâce à une meilleure anticipation et à une conduite plus souple (cependant cela n'a pas encore été prouvé dans le cadre de l'éco-conduite pour les apprentis-conducteurs ou les novices).
- **Le trajet est plus confortable** pour le conducteur et ses passagers grâce au style de conduite plus fluide.
- **Le véhicule est moins bruyant** grâce à une conduite à plus faible régime.

Fig. III.8 – Avantages de l'éco-conduite.

3.4 Inconvénients éventuels de l'éco-conduite

Dans le cadre du projet de la CIECA (Commission Internationale des Examens de Conduite Automobile) concernant l'intégration de l'éco-conduite dans la formation et l'examen des conducteurs de catégorie B, certains conflits potentiels entre l'éco-conduite et la conduite prudente ont été notés :

- Le déplacement lent dans les carrefours et leurs abords (afin de ne pas s'arrêter et donc d'économiser le carburant) suscite un risque pour la sécurité (véhicule trop proche des autres usagers de la route, capacités d'observation insuffisantes à proximité de passages pour piétons ou de carrefours à visibilité réduite).
- L'application trop précoce du frein moteur peut perturber les conducteurs qui suivent et accroître le risque de collision arrière.
- L'enclenchement et le maintien d'une vitesse élevée (économique en carburant) peuvent entraîner des manœuvres à trop vive allure (virage, etc…).
- La coupure du moteur en cas d'arrêt prolongé (plus de 1 minute) peut parfois entraîner un blocage du volant de sorte qu'il est difficile de redémarrer rapidement. Cette situation peut s'avérer assez stressante pour un apprenti-conducteur.

Il convient donc de conduire prudemment et d'appliquer les techniques d'éco-conduite dans la mesure du possible.

4 Conclusion

Plus qu'une technique, l'éco-conduite est avant tout une attitude, une nouvelle manière de se comporter sur la route. C'est un comportement de conduite citoyen permettant de diminuer le risque d'accident, de réduire sa consommation de carburant et de limiter l'émission de gaz à effet de serre, responsable du réchauffement climatique. L'éco-conduite est impulsé entre autres par le Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer. Ainsi, suite au Grenelle de l'environnement, un Comité opérationnel (Comop) a été mis en place sur le thème de l'éco-conduite. Par ailleurs, parmi les propositions du récent rapport sur la réforme du permis de conduire élaboré par Dominique Lebrun, inspecteur général de l'Equipement, figure l'enseignement de l'éco-conduite dès la formation théorique du permis de conduire. Il est en effet plus facile d'inculquer l'éco-conduite à des jeunes sans expérience que d'ôter certains mécanismes à des conducteurs chevronnés, qui ont l'habitude de conduire à leur façon. Certains pays, tels que l'Allemagne, la Finlande, la Suisse, les Pays-Bas et la Suède l'ont compris plus vite que d'autres (cf. projet interne de la CIECA concernant l'intégration de l'éco-conduite dans la formation et l'examen des conducteurs de catégorie B).

Chapitre IV

Présentation de l'expérience

Plusieurs études ont montré l'influence du comportement de conduite sur la consommation de carburant et les émissions de gaz à effet de serre [4, 5]. Mais il y a un certain manque de données précises relatives à la conduite induite par les consignes « économiques » ou « écologiques ». L'objectif de cette présente étude est ainsi d'analyser l'éco-conduite, c'est-à-dire une conduite économique en carburant, et de déterminer à l'aide d'une analyse statistique multivariée les indicateurs caractéristiques de l'éco-conduite.
Le LIVIC a donc souhaité lancer une étude interne en faisant appel à des volontaires acceptant de mesurer leurs performances sur un trajet précis. L'usage d'un véhicule du LIVIC équipé de multiples capteurs et d'un débitmètre devrait permettre de faire évoluer nos connaissances associées à ce type de conduite.

1 Description de l'expérience

1.1 Description générale

L'idée générale est d'effectuer une expérience interne au LIVIC utilisant des moyens déjà disponibles (véhicules, débitmètre, circuit bien connu). L'expérience s'est déroulée sur environ deux mois (juin et juillet 2009). Une vingtaine de participants ainsi que quatre moniteurs d'éco-conduite ont effectué deux fois le même parcours : une première fois sans aucune consigne et une autre fois en respectant des consignes destinées à économiser du carburant.

1.2 Circuit

Le circuit choisi est celui utilisé dans le cadre du projet ARCOS. En effet, ce circuit est bien connu du LIVIC, qui dispose d'informations cartographiques précises. Certaines mises à jour ont cependant été nécessaires à cause de travaux effectués sur le parcours.
Ce trajet est de type inter-urbain, plutôt rural, et comporte une variabilité intéressante des situations de conduite.

Le trajet

Distance totale : 14km
Lieu : Ponchartrain (Yvelines)

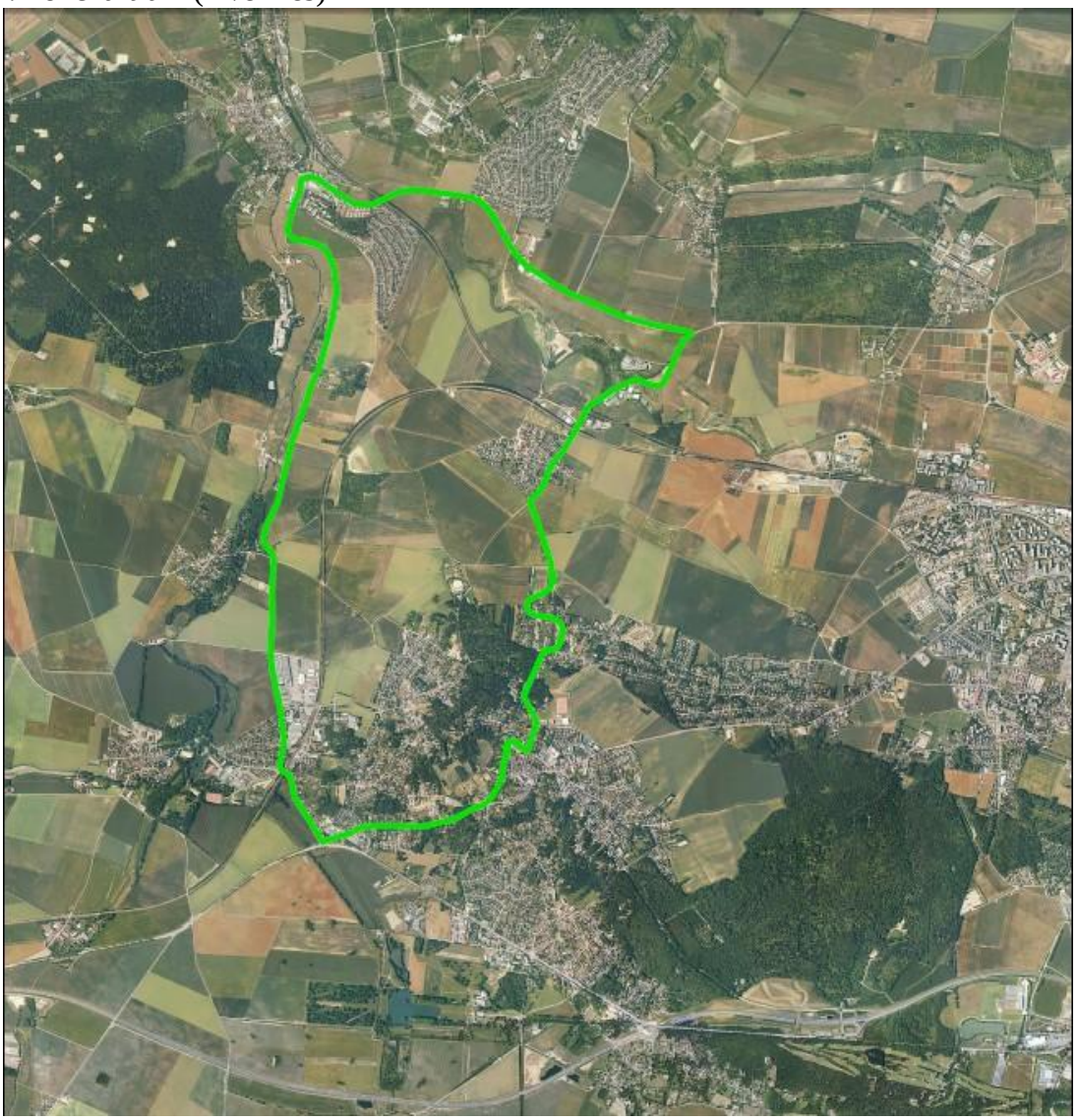


Fig. IV.1 – Image satellite du circuit.

Fig. IV.2 – Carte de la zone du circuit.

1.3 Protocole d'expérience

Les sujets sont des volontaires recrutés pour la plupart directement parmi le personnel du LIVIC, du LCPC ou de l'INRETS. La seule contrainte pour participer était d'avoir le permis depuis au moins 3 ans. La validité statistique de l'échantillon en pâtira certainement bien qu'une sélection des volontaires ait été effectuée afin de minimiser suffisamment les biais.
Le véhicule utilisé est une Renault Clio III essence équipée de multiples capteurs (cf. chap. V) et d'un débitmètre.

Les sujets sont pris en charge deux par deux par l'équipe expérimentale. Afin de limiter les effets d'apprentissage, une demi-journée d'expérimentation avec deux sujets se déroule de la façon suivante :

- Prise en main du véhicule : L'expérimentateur conduit les deux sujets jusqu'au circuit utilisé pour l'expérience. Pendant le trajet, chacun des deux sujets conduit le véhicule pendant environ 10 min afin de le prendre en main.
- 1er tour : L'expérimentateur conduit le véhicule avec une conduite « normale » sur tout le trajet, et les deux sujets le découvrent sans conduire le véhicule.
- 2ème tour : Le conducteur 1 effectue le trajet avec une conduite « économique ».
- 3ème tour : Le conducteur 2 effectue le trajet avec une conduite « normale ».
- 4ème tour : Le conducteur 1 effectue le trajet avec une conduite « normale ».
- 5ème tour : Le conducteur 2 effectue le trajet avec une conduite « économique ».

Il s'agit d'un plan d'expérience très simple destiné à contrebalancer les effets d'apprentissage.

2 Types de conduite

2.1 Conduite normale

Le conducteur aura pour instruction de conduire sans se presser, sans respecter aucune consigne particulière (excepté le code de la route bien entendu). La conduite désirée devrait être celle pratiquée en week-end, lors de trajets loisirs.
L'objectif est de pouvoir comparer la conduite en temps normal, avec la conduite lorsque des consignes précises sont données.

2.2 Conduite économique ou « green driving »

La priorité est ici donnée à l'économie de carburant. Les consignes données aux conducteurs sont les règles définies dans le contexte du projet Ecodrive de l'UE (cf. chapitre III-paragraphe 3.2) :

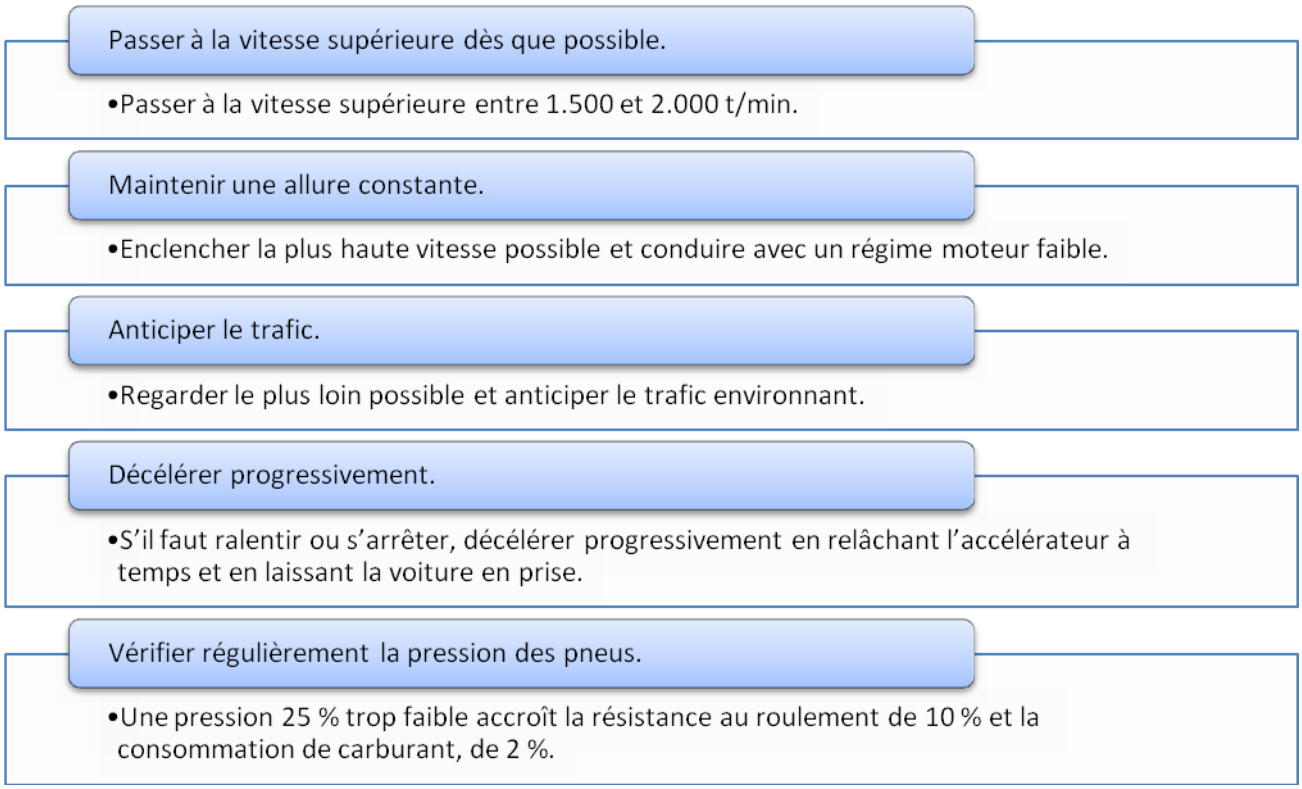


Fig. IV.3 – Règles d'éco-conduite issues du projet Ecodrive de l'UE.

La dernière consigne ne concerne évidemment pas le conducteur, le véhicule ayant été contrôlé au préalable.
Il est ainsi demandé au conducteur de suivre au mieux ces consignes afin de chercher à économiser du carburant, tout en respectant les consignes habituelles de sécurité.
D'autre part, il est important de noter que l'expérimentateur n'est pas un professionnel de l'éco-conduite. Il a donc été demandé au conducteur de ne lui poser aucune question concernant ces consignes afin de ne pas être induit en erreur et d'être seul juge dans l'interprétation de ces consignes.

3 Considérations générales

Le plan adopté pour cette expérience est un plan intra-sujets, ce qui signifie que les mêmes sujets sont exposés aux différentes conditions expérimentales (ici une conduite normale et une conduite économique). Afin de minimiser la variabilité entre les sujets, les mesures suivantes ont été prises :

- Tous les trajets ont été effectués avec le moteur chaud. En effet, un trajet de plus de 5km a été effectué pour se rendre sur le lieu du parcours.
- La charge du véhicule a été la même pour tous les trajets.
- Afin de garantir une charge identique, le véhicule a disposé du plein de carburant au départ de chaque trajet.
- Les trajets ont été effectués en conditions de trafic libre, c'est-à-dire à des heures où la circulation est fluide (10h-12h et 14h-16h).

- Tous les trajets ont été effectués dans des conditions climatiques correctes : beau temps ou légèrement couvert, route sèche, pendant la journée.
- Les trajets ont été effectués dans le même sens.
- Un même conducteur a effectué ses trajets dans la même demi-journée afin de respecter la constance des conditions climatiques.
- Afin d'éviter un « effet d'imitation », le conducteur 2 n'était pas présent dans le véhicule pendant que le conducteur 1 effectuait ses trajets et inversement. Seul l'expérimentateur était présent afin d'indiquer la direction à prendre.
- Aucun appareil ou système intégré pouvant influencer la conduite, la concentration, ou la consommation, n'a été utilisé. Ainsi la climatisation n'a pas été enclenchée, excepté pour 2 conducteurs, et la radio n'a pas été allumée.

D'autre part une inversion de l'ordre d'expérience des différents modes de conduite a été effectuée pour la moitié des sujets. Ainsi pour chaque demi-journée, l'un des conducteurs effectue le trajet ECO après son trajet NORMAL, alors que l'autre conducteur commence par le trajet ECO. Ce contrebalancement des effets permettra d'analyser un éventuel effet d'apprentissage pour les conducteurs ayant commencé par le trajet ECO : consommeront-ils moins que les autres lors de leur trajet NORMAL ? Appliqueront-ils inconsciemment certains principes d'éco-conduite effectués lors de leur trajet ECO ?

En revanche, même si le parcours est identique pour tous les sujets, il n'est pas possible de contrôler les conditions de circulations. Ainsi certains conducteurs ont pu être gênés dans la mise en œuvre des principes d'éco-conduite : véhicule forçant à ralentir, feux, présence de travaux, passage à niveau ferroviaire …
C'est pourquoi une feuille décrivant brièvement tous les événements particuliers relevés au cours de chaque trajet a été remplie. Ce point devra évidement être pris en compte lors de l'analyse des résultats.

4 Sujets particuliers

Parmi les sujets, quatre moniteurs d'éco-conduite (deux moniteurs de l'AFT-IFTIM, et deux moniteurs du Ministère de la Défense, Mission prévention et sécurité routières dans les armées) ont également participé à l'expérience en réalisant un trajet NORMAL et un trajet ECO. On pourra ainsi comparer les deux populations : des sujets « normaux » qui ont eu des consignes écrites, et des sujets professionnels pouvant servir de référence.

Chapitre V

Données issues de l'expérience

Les données recueillies pendant l'expérience proviennent en grande partie du bus CAN du véhicule utilisé. Ces données concernent le véhicule en lui-même, c'est-à-dire les paramètres du moteur ou du véhicule (régime, vitesse instantanée, accélération, distance parcourue…), ainsi que les actions du conducteur (freinage, essuie-glaces, clignotants, phares).
La consommation cumulée a aussi été mesuré à l'aide d'un débitmètre. Cette étude permettra ainsi de comparer l'information de consommation obtenue avec le bus CAN et le débitmètre.
Il a aussi été décidé d'enregistrer le trajet de chaque conducteur sous forme de suites d'images (jpeg). Ces informations seront précieuses pour l'analyse afin de déterminer avec certitude les interactions avec les autres usagers de la route (ou les obstacles).

1 Données issues du bus CAN

Le bus CAN est le réseau qui permet à tous les capteurs et actionneurs du véhicule de s'échanger leurs informations. Dans un câblage traditionnel, un capteur est relié à un autre (liaison point à point) ce qui fait un câblage très complexe. Dans un réseau CAN, tous les capteurs sont reliés entre eux par un même fil (le câblage est donc très simplifié comme l'illustre la figure V.1) et les informations sont échangées dans des trames avec une gestion de la priorité des trames et un contrôle d'erreurs poussé. Le bus CAN est universellement présent dans les véhicules depuis une dizaine d'années.

Fig. V.1 –Illustration du bus CAN.

Les variables mesurées à l'aide du bus CAN sont données dans le tableau V.1. Elles concernent le véhicule et les actions du conducteur.

Nom de la variable	Type	Description
Horaire	Entier	Horaire du système (en microsecondes)
RegimeMoteur	Réel	Régime moteur (en tours/min)
VInstantanee	Réel	Vitesse instantanée du véhicule (en km/h)
Distance	Réel	Distance parcourue depuis le point de départ (en m)
ConsoCanCumul	Réel	Consommation cumulée (en litres)
VitesseRotationAvD	Réel	Vitesse de rotation de la roue avant droite
VitesseRotationAvG	Réel	Vitesse de rotation de la roue avant gauche
VitesseRotationArD	Réel	Vitesse de rotation de la roue arrière droite
VitesseRotationArG	Réel	Vitesse de rotation de la roue arrière gauche
Acceleration	Réel	Accélération longitudinale du véhicule (en m/s²)
AccelerationLaterale	Réel	Accélération latérale du véhicule
VitesseLacet	Réel	Vitesse de lacet
BraquageVolant	Réel	Angle de braquage du volant
VitesseBraquageVolant	Réel	Vitesse de braquage du volant
CoupleConducteur	Entier	CoupleConducteur
PosAccelerateur	Réel	Pourcentage d'enfoncement de la pédale d'accélérateur
ContactPedaleFrein	Booléen	1 si usage du frein, 0 sinon
ClignotantDroit	Booléen	1 si usage du clignotant droit, 0 sinon
ClignotantGauche	Booléen	1 si usage du clignotant gauche, 0 sinon
EssuieGlace	Booléen	1 si usage des essuie-glaces, 0 sinon
MarcheArriere	Booléen	1 si usage de la marche arrière, 0 sinon
PointMort	Booléen	1 si usage du point mort, 0 sinon
FreinAMain	Booléen	1 si usage du frein à main, 0 sinon
Embrayage	Booléen	1 si usage de la pédale d'embrayage, 0 sinon

Tab. V.1 – Variables mesurées par le bus CAN.

Il est important de noter que l'intervalle de temps entre deux enregistrements successifs est de l'ordre de 50ms mais n'est pas constant.

2 Données issues du débitmètre

Le débitmètre utilisé pour notre étude est un DFL1x-5bar. Les variables mesurées par le débitmètre sont les suivantes :

Nom de la variable	Type	Description
Horaire	Entier	Horaire du système (en microsecondes)
ConsoCumulDebitmetre	Réel	Consommation cumulée (en litres)

Tab. V.2 – Variables mesurées par le débitmètre.

On obtient ainsi une autre mesure de la consommation cumulée.

3 Données supplémentaires

Un radar a été placé à l'avant du véhicule afin de connaitre l'environnement du conducteur et l'état du trafic, et une caméra a été installé sous le volant afin de filmer les pieds du conducteur et ainsi de savoir si un pied est posé sur chacune des pédales. Cependant, ces données n'ont pas été exploitées dans cette étude, mais pourront être étudiées par la suite afin de calculer des indicateurs relatifs au risque.

4 Prétraitement des données

Cette étape est nécessaire à tout travail de traitement statistique de données. Cela consiste notamment à nettoyer et vérifier la validité des données recueillies. Cependant le prétraitement peut-être très variable et dépend essentiellement de la qualité des données fournies.
Les principales étapes de ce prétraitement sont :

- Vérification et validation des données (recherches d'erreurs, de valeurs aberrantes, de données manquantes).
- Création/Transformation de variables.
- Recodages, transformations entre qualitatif et quantitatif.
- Mise en forme, stockage sous forme de fichiers nettoyés et accessibles facilement.

Ici les traitements suivants ont été appliqués aux données :

- reconstruction de certains trajets suite à des problèmes survenus pendant l'enregistrement des données.
- Correction des sauts de certaines variables issues du bus CAN en effectuant un lissage par médianes mobiles.
- Interpolation de la consommation issue du débitmètre.
- Troncature des trajets afin qu'ils soient identiques (même point de départ, même point d'arrivée et donc même distance).

4.1 Reconstitution de la consommation cumulée issue du bus CAN

La consommation issue du bus CAN correspond à la consommation cumulée modulo 20400 avec pour unité le mm^3. Ainsi la courbe de la consommation est en dents de scie comme l'illustre la figure V.2. La consommation cumulée a donc été reconstituée afin d'obtenir une courbe croissante sans retour à zéro.

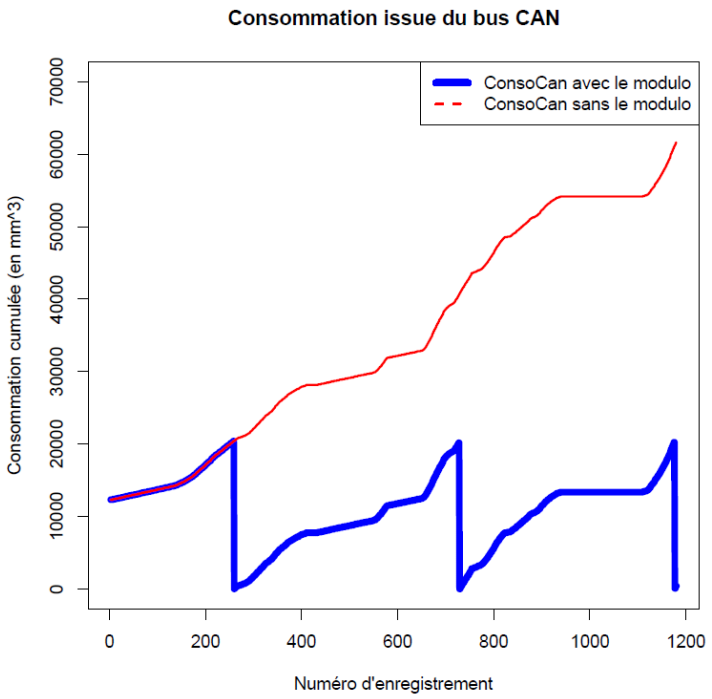


Fig. V.2 – Reconstitution de la consommation cumulée issue du bus CAN.

4.2 Problèmes observés avec le bus CAN

Au cours du prétraitement, des sauts ont été observés sur les données de certaines variables issues du bus CAN. Ces sauts ont été observés sur les variables correspondant à :

- la vitesse instantanée,
- le pourcentage d'enfoncement de la pédale d'accélérateur,
- le régime moteur,
- l'accélération latérale,
- la distance parcourue,
- la consommation

Afin de corriger ces sauts, un lissage par médianes mobiles a été effectué sur les variables citées.

4.2.1 Exemple de saut de la vitesse instantanée

Le tableau V.3 montre un exemple de saut de la vitesse à 57,43km/h à l'enregistrement 11457, alors que la vitesse enregistrée à l'enregistrement précédent et à l'enregistrement suivant est de l'ordre de 28km/h. Rappelons que l'intervalle de temps entre deux enregistrements successifs est de l'ordre de 50ms !

VInstantanee	NumEnr
28.29	11455
28.4	11456
57.43	11457

VInstantanee	NumEnr
28.85	11458
28.96	11459

Tab. V.3 – Exemple de saut de la variable VInstantanee.

La figure V.3 montre un exemple de sauts de la vitesse instantanée corrigés par le lissage.

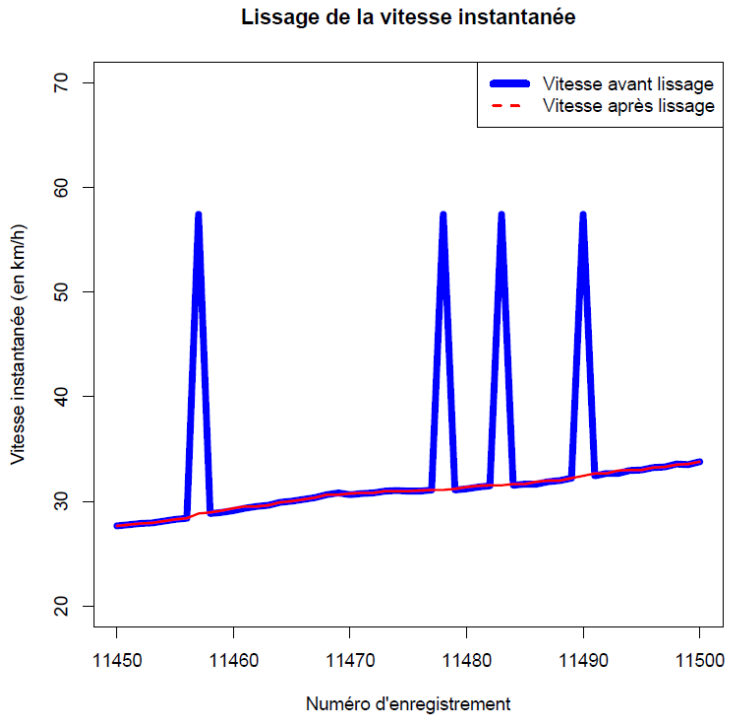


Fig. V.3 – Superposition de la courbe initiale de la vitesse et de la courbe lissée.

4.2.2 Exemple de saut de la consommation

Un lissage par médianes mobiles a aussi été effectué sur la consommation cumulée reconstituée (c'est-à-dire sans modulo) afin de corriger les sauts de cette variable. Par contre, tous les sauts n'ont pas pu être corrigés par le lissage. En effet, contrairement aux autres variables où les sauts ne durent que pendant un enregistrement, on a observé des sauts de la consommation sur plusieurs enregistrements successifs comme le montre l'exemple suivant :

ConsoCan	NumEnr
4000	10995
4160	10996
4160	10997
6800	10998
6800	10999
6800	11000

ConsoCan	NumEnr
6800	11001
4560	11002
4560	11003

Tab. V.4 – Exemple de saut de la variable ConsoCan.

En effet, dans le tableau V.4, on observe un saut de la consommation à 6800mm^3 pendant 4 enregistrements successifs. La figure V.4 montre que le lissage n'a pas corrigé tous les pics de la variable ConsoCan.

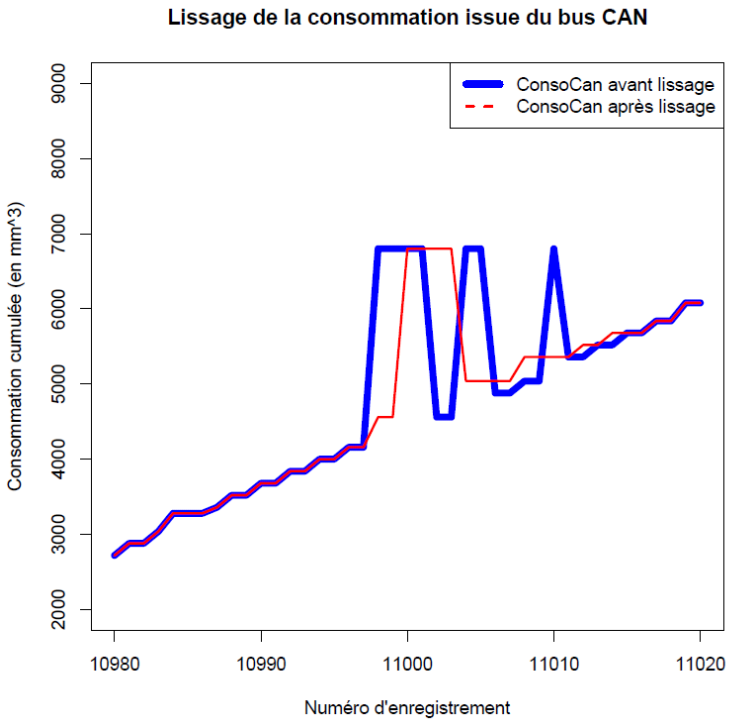


Fig. V.4 - Superposition de la courbe initiale de la consommation issue du bus CAN et de la courbe lissée.

Ainsi pour certains trajets, les pics de consommation non corrigés par le lissage ont été supprimés en remplaçant la valeur du saut (6800 dans l'exemple donné au tableau précédent) par la valeur de l'enregistrement qui précède le saut (4160 dans l'exemple donné au tableau précédent).

4.2.3 Reconstitution de la distance parcourue issue du bus CAN et correction des sauts d'odomètre

Comme pour la consommation, la courbe représentant la distance parcourue est en dents de scie. En effet, la distance issue du bus CAN est donnée modulo 6553.5 avec pour unité le mètre. La distance parcourue a donc été reconstitué de la même manière que pour la

consommation. D'autre part, comme indiqué au paragraphe précédent, des sauts d'odomètre ont été observés et ont été corrigés par un lissage par médianes mobiles. Cependant, contrairement à ce qui a été fait pour la consommation, le lissage a ici été effectué avant la reconstitution de la distance sans le modulo. En effet, certaines valeurs de saut sont proches de la valeur du modulo.

Ainsi, sur la figure V.5, on observe un retour à zéro dû au modulo (égal à 6553.5) puis un saut d'odomètre à l'enregistrement 2828 qui atteint quasiment la valeur du modulo. Dans cet exemple, on constate que si on effectue d'abord la reconstruction de la distance, la valeur du saut est comptabilisée dans le calcul de la distance parcourue, ce qui conduit à une erreur qui n'est pas rectifiée par le lissage.

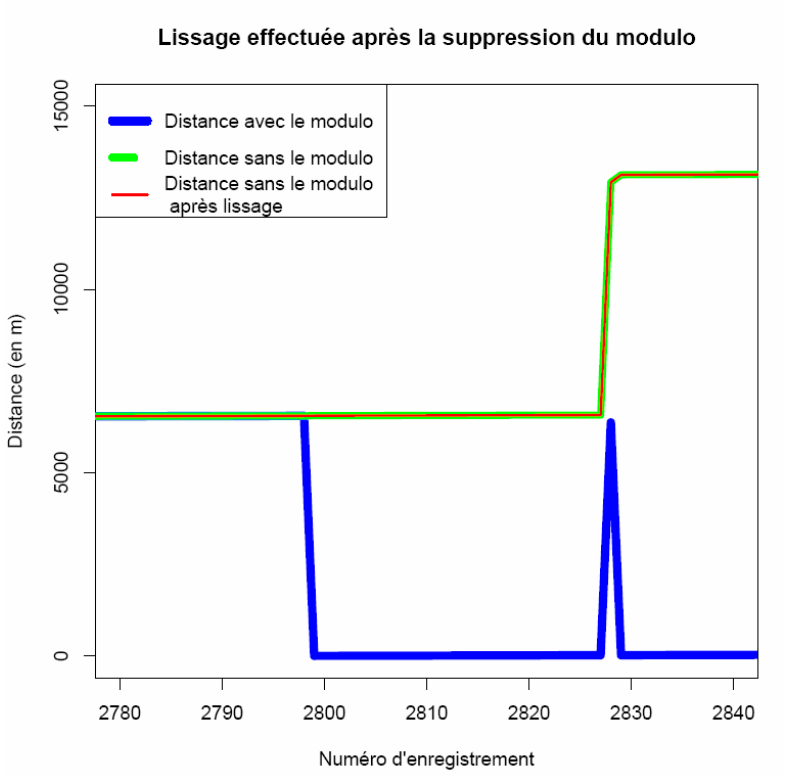


Fig. V.5 – Superposition de la courbe initiale, de la courbe représentant la distance parcourue reconstituée, et de la courbe lissée.

Maintenant sur le même exemple, le lissage a été effectué avant la reconstruction de la distance sans le modulo. On observe sur la figure V.6 que le lissage a corrigé le saut mais sans supprimer le modulo. On obtient ainsi après reconstruction, la distance parcourue réellement.

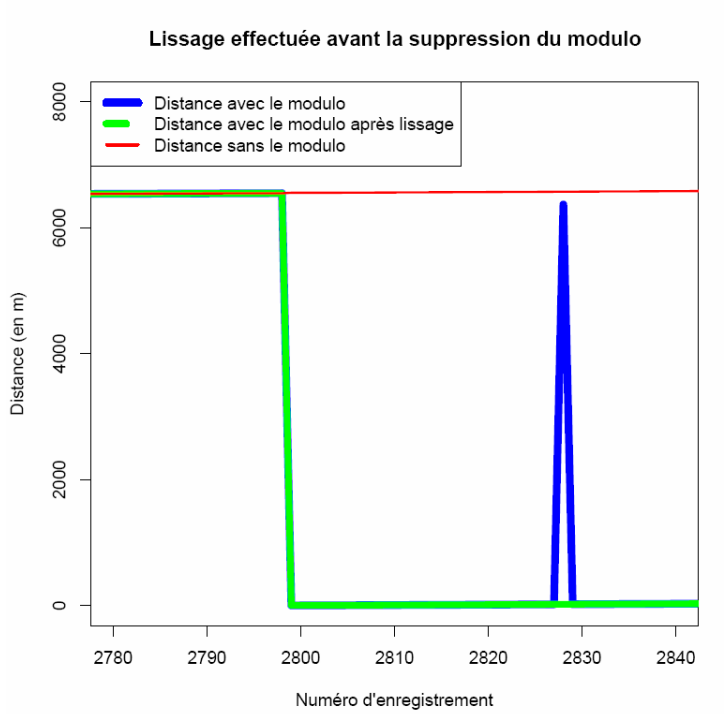


Fig. V.6 – Superposition de la courbe initiale, de la courbe lissée et de la courbe représentant la distance parcourue reconstituée après lissage.

4.3 Interpolation de la consommation issue du débitmètre

La consommation issue du débitmètre correspond à la consommation cumulée donnée en litres, sans modulo. La vitesse d'échantillonnage du débitmètre n'étant pas la même que celle du bus CAN, nous avons effectué une interpolation de la consommation issue du débitmètre en fonction de la variable **Horaire** du système. Cependant cette interpolation a conduit à des erreurs. En effet, sur la figure suivante, on observe que la courbe de la consommation cumulée après interpolation n'est pas monotone croissante. La courbe est légèrement décroissante à certains endroits, ce qui signifie que pour ces enregistrements la consommation instantanée est négative ! On a donc effectué comme précédemment un lissage par médianes mobiles de la consommation interpolée issue du débitmètre afin de réduire le nombre d'erreurs dû à l'interpolation.

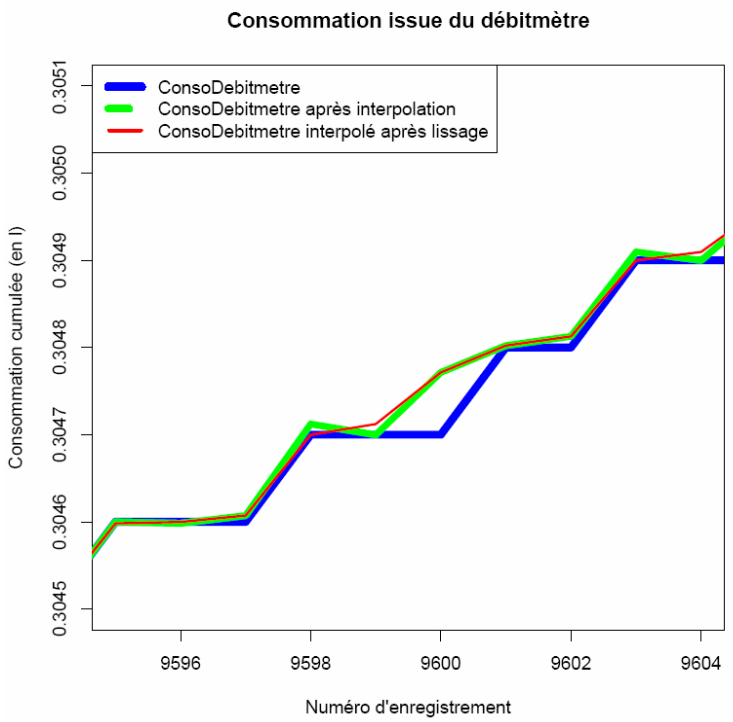


Fig. V.7 – Superposition de la courbe initiale, de la courbe interpolée et de la courbe lissée.

5 Ajout de variables supplémentaires

Suite à des modifications d'infrastructure sur le circuit ARCOS, les bases de données du LIVIC concernant les consignes de limitation de vitesse n'ont pas pu être utilisées. Cette information a donc été reconstruite « à la main ». Ainsi, les différentes limitations de vitesse ont été notées au cours d'un trajet puis ajoutés aux autres trajets en utilisant la variable **Distance** comme référence (les trajets étant tous identiques). Cette variable est appelée **ConsigneLimiteVitesse.** A partir de la vitesse instantanée et de la consigne de limitation, on peut aussi déduire le **dépassement de vitesse** à chaque instant.

On crée aussi la variable **OrdreTrajet** qui indique si le trajet effectué est le 1er ou le 2ème. Rappelons que la moitié des participants a commencé par le trajet en conduite normale, alors que l'autre moitié a commencé par le trajet en conduite économique. Cette variable servira à mettre en évidence un éventuel effet d'apprentissage.

Certaines variables dites de repérage ont aussi été ajouté.

Nom de la variable	Type	Description
NumEnr	Entier	Numéro d'enregistrement de l'observation
HoraireMS	Entier	Temps écoulé en ms depuis le point de départ
Heure	Horaire (HH:MM:SS)	Heure du système
Travia	Entier	Code permettant d'identifier le conducteur
Trajet	Entier	1 pour « conduite normale », 2 pour « conduite éco »

Tab. V.5 – Variables de repérage

5.1 Plages de régime moteur

On répartit le régime moteur en classe en créant la variable **RegimeClasse** définie de la manière suivante :

Régime moteur (en tr/min)	RegimeClasse
[0 ; 500[	1
[500 ; 1000[	2
[1000 ; 1500[	3
[1500 ; 2000[	4
[2000 ; 2500[	5
[2500 ; 3000[	6
[3000 ; 3500[	7
[3500 ; 4000[	8
[4000 ; 4500[	9
[4500 ; 5000[	10
[5000 ; 5500[	11
+ de 5500	12

Tab. V.6 – Répartition des plages de régime moteur.

Cette variable sera utilisée par la suite pour classer certaines variables en fonction de la plage de régime.

5.2 Rapports de boîte de vitesse

La reconstruction du rapport de boîte de vitesse a été réalisée en utilisant la même méthode que pour les données LAVIA. Elle consiste en un lissage par médianes mobiles du rapport Vitesse / Régime moteur, puis une classification par méthode des k-means.
En effet, le rapport (Vitesse / régime moteur) permet de reconstruire une fonction en plateaux donnant les rapports de boite du véhicule étudié. La courbe obtenue est bruitée et possède des pics importants lors des changements de rapports de vitesse (dû au fait que l'embrayage tourne une fraction de seconde dans le vide ce qui fait monter le régime moteur). Il est cependant possible d'appliquer un simple lissage par médianes mobiles puis d'effectuer une classification pour reconstruire simplement et avec très peu d'erreurs les rapports de boîte de vitesse. Un exemple de classification obtenue est donné à la figure V.8.

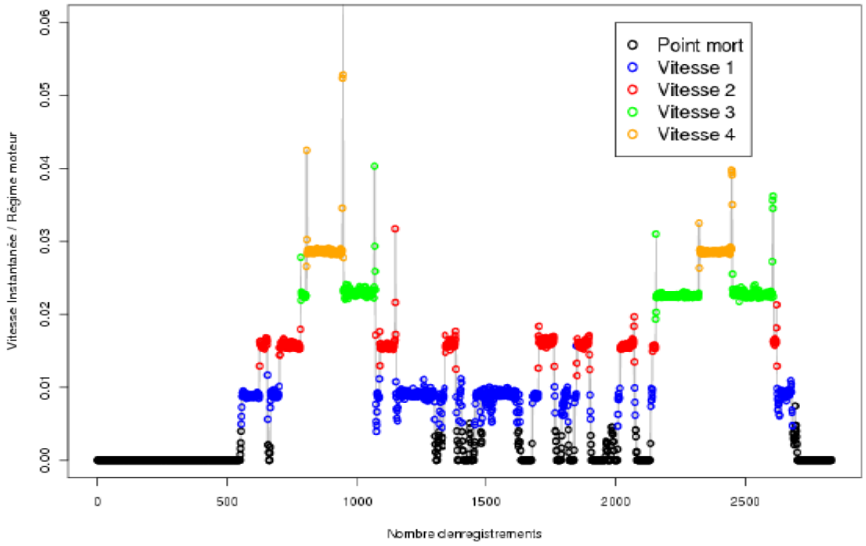


Fig. V.8 – Courbe initiale superposée à la courbe lissée et à la classification obtenue.

Les rapports de démultiplication représentés par les paliers sont constants pour un véhicule donné mais peuvent varier d'un modèle à l'autre. Avant de procéder à une classification de ces paliers, il est nécessaire de retrouver les centres des classes représentés par les rapports de démultiplication voulus par le constructeur du véhicule. Cela est possible expérimentalement en se basant sur des histogrammes permettant de repérer les zones d'accumulation correspondant aux paliers. Comme base à la classification nous utilisons les centres de classes suivants :

	Vitesse / Régime moteur (Renault)
Point mort	0.001
Vitesse 1	0.009
Vitesse 2	0.015
Vitesse 3	0.023
Vitesse 4	0.029
Vitesse 5	0.035

Tab. V.7 – Paliers correspondant aux changements de rapports de boîte selon la marque du véhicule.

Une fois que l'on a reconstitué cette variable nommée **RapBoite**, on crée une nouvelle variable **ChangeVitesse** qui prend la valeur 1 dès que le conducteur change de rapport et qui permettra par la suite de calculer le nombre de changement de vitesse effectué au cours d'un trajet.

Chapitre VI

Indicateurs et base de données finale

L'objectif de cette étude est de déterminer les indicateurs les plus caractéristiques d'une conduite économique ou non. C'est pourquoi la base de données finale utilisée dans les analyses statistiques décrites dans la suite de ce rapport est constituée d'un certain nombre d'indicateurs calculés pour chaque trajet. Un individu statistique est alors un trajet (« Normale » ou « Economique »). Le tableau final rassemble ainsi les caractéristiques de chaque trajet.

1 Tableau des indicateurs

Plusieurs variables définies à la section précédente ont été utilisées comme indicateurs de base : distance et durée du trajet, consommation cumulée mesurée par le bus CAN et le débitmètre (donnée en l/100 km). Des indicateurs classiques ont aussi été ajoutés : il s'agit des moyennes et écarts types de divers paramètres comme la vitesse, l'accélération, mais aussi le régime moteur. Enfin de nombreux indicateurs potentiels identifiés dans la littérature ont été calculées.
Dans la suite v désigne la vitesse instantanée, et a l'accélération. Le descriptif des indicateurs est donné à l'annexe A.

2 Indicateurs supplémentaires liés à la consommation

2.1 Temps passé à l'arrêt

Le **temps passé à l'arrêt** sur un trajet est un indicateur fortement lié à la consommation. En effet, le temps passé à l'arrêt est comptabilisé dans le calcul de la consommation d'essence mais correspond à une distance parcourue nulle. Ainsi lorsque l'on convertit la consommation en litres au 100km, les périodes passées à l'arrêt augmentent considérablement l'estimation de la consommation de carburant.
Le calcul du temps passé à l'arrêt pour chaque trajet nous permettra de comprendre des consommations élevées de certains trajets qui ne sont pas forcément liés au style de conduite proprement dit. Cet indicateur est ainsi très corrélé au **nombre d'arrêts aux feux** au cours du trajet.

2.2 Temps passé en frein moteur

Le temps passé en frein moteur est défini comme le temps passé avec une vitesse non nulle mais pas de pression sur la pédale d'accélérateur. Le véhicule est alors en prise, c'est-à-dire que l'alimentation en carburant du moteur est coupée et que ce sont les roues qui entrainent le moteur (cf. chap. III paragraphe 3.2). Cette technique est évidemment très représentative d'une conduite économique.

2.3 Temps passé avec une vitesse non optimale

Le temps passé avec une vitesse non optimale est défini comme le temps passé avec un régime moteur supérieur à 3000 tr/min et une pression exercée sur la pédale d'accélérateur (autrement dit, le régime moteur est élevé et on n'est pas en frein moteur).
Une partie de la puissance du moteur est perdue par friction interne et ces pertes augmentent avec le régime moteur. Ainsi le temps passé avec une vitesse non optimale contribue à une augmentation de la consommation de carburant car le régime moteur est trop élevé.

3 Indicateurs fortement liés à la consommation

Les trois indicateurs qui suivent sont plus techniques et difficiles à mettre en œuvre en pratique mais sont fortement liés à la consommation de carburant. Ils sont très représentatifs du style de conduite : plus la conduite est nerveuse, plus ces indicateurs sont élevés.

3.1 Positive Kinetic Energy (PKE)

Cet indicateur est défini de la manière suivante :
$\sum \left(v_f^2 - v_s^2 \right) x$ lorsque dv/dt>0, v_f =final speed, v_s =start speed, x=distance
Plus l'accélération du véhicule est constante, plus PKE est proche de 0. Cet indicateur indique en fait la capacité à garder l'énergie cinétique du véhicule, c'est-à-dire l'élan que possède le véhicule lorsqu'il est en mouvement. Ainsi une conduite nerveuse avec de nombreuses accélérations sera associée à un PKE élevé, et inversement une conduite plus fluide sera associée à un PKE proche de 0.

3.2 Relative Positive Acceleration (RPA)

RPA est défini ainsi :
$\frac{1}{x}\int va^{+}dt$ avec x=durée totale du trajet, v=vitesse du véhicule, a=accélération du véhicule
Cet indicateur augmente lorsque la conduite est caractérisée par de nombreuses accélérations.

3.3 Intégrale du carré de l'accélération

L'intégrale du carré de l'accélération est définie de la manière suivante :

$\frac{1}{n}\int a^2 dt$ avec n=nbre d'intervalles de temps

Cet indicateur permet de donner de l'importance aux grandes valeurs d'accélération.

Remarque :

Les données traitées étant discrètes, c'est-à-dire avec un nombre fini de valeurs (par opposition aux données continues), les intégrales contenues dans les deux indicateurs qui précèdent ont été calculées à l'aide de sommes de Riemann.

4 Indicateurs relatifs au conducteur

Au cours de l'expérience, certaines informations relatives à chaque conducteur ont été relevées : âge, sexe, nombre d'années de permis et connaissance préalable de l'éco-conduite.
Ces variables ont ensuite été transformées en variables qualitatives afin de pouvoir effectuer par la suite une analyse des correspondances entre les modalités de ces variables et les performances en termes d'éco-conduite.

Nom de la variable	Description	Modalités
AgeClasse	Age du conducteur	Moins de 30 ans 30-40 ans 40-55 ans Plus de 55 ans
Sexe	Sexe du conducteur	M (masculin) F (féminin)
Experience	Nombre d'années de permis du conducteur	Peu expérimenté (- de 10 ans de permis) Expérimenté (entre 10 et 30 ans de permis) Très expérimenté (+ de 30 ans de permis)
ConnaissanceEcoConduite	Connaissance préalable de l'éco-conduite	Aucune Théorique Pratique Moniteur d'éco-conduite

Tab. VI.1 – Indicateurs relatifs au conducteur.

5 Indicateurs relatifs au parcours

Des données relatives au parcours ont aussi été prises en compte afin de justifier certaines disparités entre les trajets, indépendamment du conducteur, comme la météo ou la congestion.

Ainsi le nombre d'arrêts aux feux a été noté sachant que le parcours comportait deux feux et que pendant une courte période de travaux un troisième feu a été installé. Certaines de ces variables qualitatives seront aussi étudiées lors de l'analyse des correspondances.

Nom de la variable	Description	Modalités
Meteo	Météo lors du parcours	Clair Nuageux Pluie légère
Temperature	Température moyenne lors du parcours	Moins de 15° 15-20° Plus de 20°
Adherence	Qualité de l'adhérence de la route lors du parcours	Bonne Moyenne Mauvaise
Fenetre	Etat des fenêtres et de la climatisation pendant le parcours	Entrouvertes Fermées Climatisation
NbArretFeu	Nombre d'arrêts aux feux pendant le parcours	0 1 2 3 (période de travaux)

Tab. VI.2 – Indicateurs relatifs au parcours.

6 Etat du trajet

Une dernière variable a été ajoutée concernant l'état du trajet. En effet, des problèmes d'enregistrements ou autres sont apparus au cours de certains trajets et ceux-ci n'ont donc pas pu être exploités par la suite. La variable **EtatTrajet** prend donc la modalité 'OK' si le trajet ne comporte pas de problème, et 'NOK' si des erreurs rendent le trajet inexploitable. Seuls les conducteurs ayant la modalité 'OK' pour leurs deux trajets (normal et économique) ont été pris en compte dans l'analyse.

7 Comparaison entre les deux mesures de consommation

La consommation a été mesurée de deux manières différentes : d'une part par le bus CAN, et d'autre part par le débitmètre. Nous avons vu à la section 4.2 du chapitre V que des erreurs sont survenues dans les données issues du bus CAN et que celles-ci ont été corrigées par un lissage. Les données issues du débitmètre ne comportent pas d'erreurs, mais nous avons noté deux inconvénients majeurs : il y a un retard dans l'acquisition des données par le débitmètre par rapport au bus CAN, et la vitesse d'échantillonnage n'est pas la même, ce qui nous a contraint d'interpoler les données issues du débitmètre. Nous avons donc utilisé la consommation issue du bus CAN pour la suite de notre étude. Cependant, nous avons utilisé la consommation issue du débitmètre afin de valider celle issue du bus CAN.

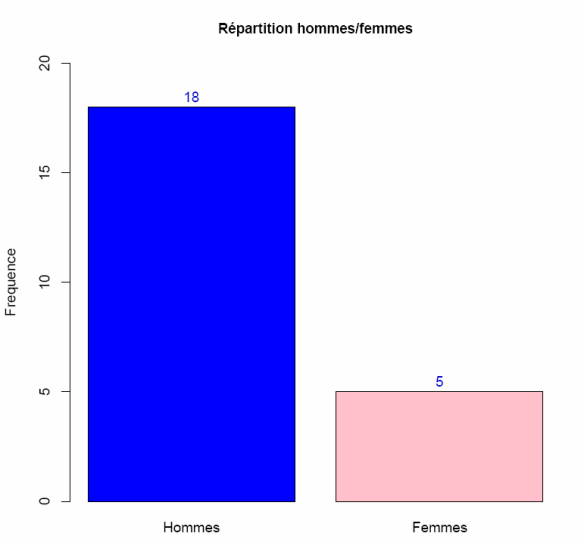
Chapitre VII

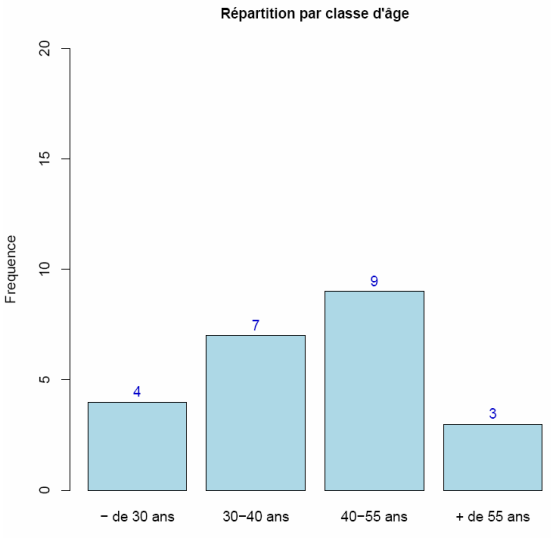
Analyse descriptive des résultats

Ce chapitre concerne l'analyse des résultats obtenus au cours de l'expérience. Il s'agit d'une approche descriptive qui tient compte de l'ensemble des trajets. La première section décrit la répartition des conducteurs ayant participé à l'expérience, puis nous étudierons l'impact de l'éco-conduite sur la consommation et sur la vitesse, et enfin nous analyserons la mise en pratique des principes de base de l'éco-conduite

1 Répartition des conducteurs

L'analyse a été effectuée sur 23 conducteurs ayant parcouru le même trajet deux fois : une fois en conduite normale, et une autre fois en conduite économique. La répartition des conducteurs en fonction de leur sexe, de leur âge, de leur nombre d'années de permis, ainsi que de leurs connaissances préalables de l'éco-conduite est donnée à la figure VII.1.





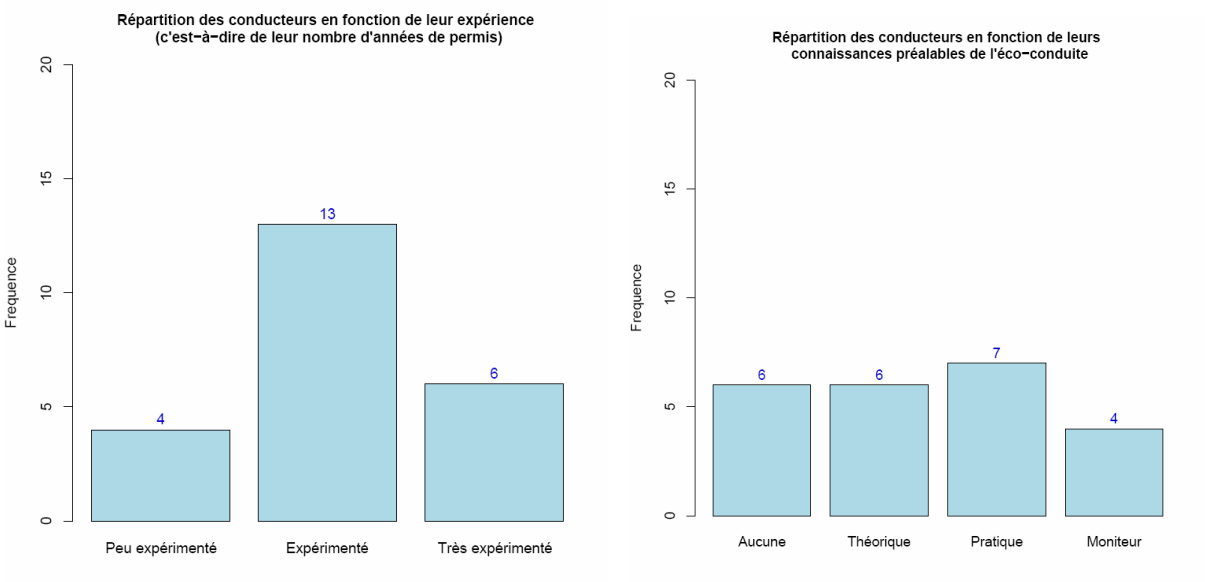


Fig. VII.1 - Répartition des conducteurs par sexe, âge, expérience, et connaissances de l'éco-conduite.

La majorité des conducteurs étant des hommes, la variable Sexe ne sera pas prise en compte dans la suite de l'analyse. D'autre part, on ne fera aucune comparaison de la consommation en conduite économique en fonction de l'âge ou de l'expérience des conducteurs. En effet, les quatre moniteurs d'éco-conduite ayant participé à l'expérience appartiennent à la classe d'âge 40-55 ans, ce qui biaise fortement les résultats : la classe 40-55 ans obtient de bonnes performances en terme d'éco-conduite grâce à la présence des moniteurs qui sont des spécialistes de l'éco-conduite ! Ainsi seule la variable correspondant aux connaissances de l'éco-conduite sera étudiée par la suite.

2 L'éco-conduite et la réduction de consommation

La figure VII.2 montre clairement l'intérêt de l'éco-conduite en termes de réduction de consommation. En effet, on observe une baisse de 13% en moyenne de la consommation de carburant entre une conduite normale et une conduite économique. Tous les conducteurs ayant participé à l'expérience ont moins consommé lors de leur trajet en conduite économique, cette baisse allant jusqu'à 26% pour certains. On constate également d'après le tableau VII.1 que la consommation maximale tous conducteurs confondus est de 8,3 l/100 km en conduite normale alors qu'elle n'est que de 6,7 l/100 km en conduite économique.
Ces résultats montrent qu'il semble assez simple de réduire sa consommation en appliquant certaines règles de base de l'éco-conduite. En effet, rappelons que la moitié des participants n'avaient que des connaissances théoriques voire nulles sur l'éco-conduite et n'ont suivi aucune formation spécifique avant l'expérience.

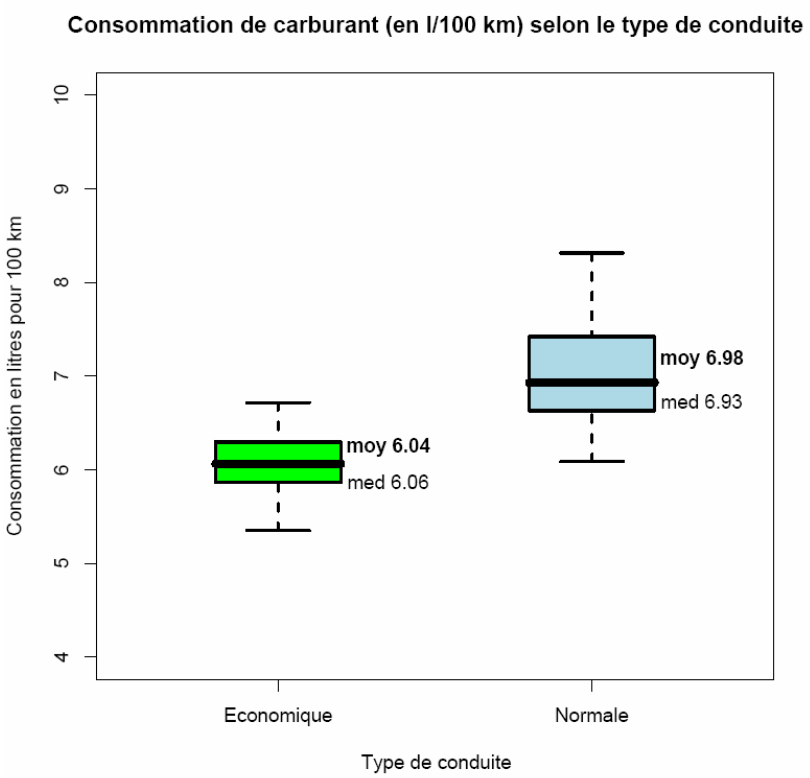


Fig.VII.2 - Boîtes à moustaches représentant la consommation moyenne (en l/100 km) selon le type de conduite. L'anova détecte une différence significative (DF=1, F=40.62, p<0.0001) entre les deux types de conduite.

	Minimum	Maximum	Moyenne	Ecart-type
Consommation moyenne en conduite normale (en l/100km)	6.086	8.312	6.977	0.590
Consommation moyenne en conduite économique (en l/100km)	5.350	6.715	6.038	0.388

Tab.VII.1 - Répartition de la consommation en l/100 km pour l'ensemble des conducteurs selon le type de conduite.

En ce qui concerne la consommation en conduite économique selon les connaissances préalables du conducteur sur l'éco-conduite, la figure VII.3 montre une légère différence entre les moniteurs et les autres conducteurs. En effet, la consommation moyenne des moniteurs est de 5,73 l/100 km alors que celle des autres conducteurs est de 6,10 l/100 km. Cependant, les résultats d'une analyse de variance n'indiquent pas de différences significatives entre les classes (DF=3, F=1.07, p=0.387), bien qu'il semble difficile d'en tirer des conclusions car les quatre classes ne sont pas homogènes.

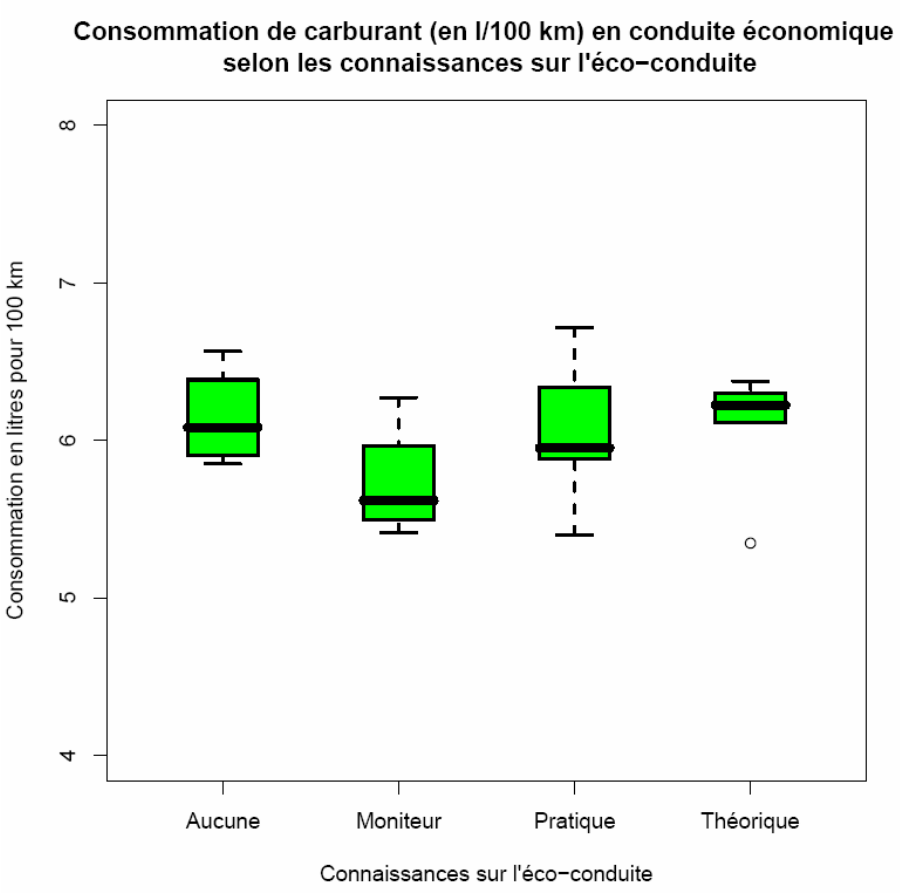


Fig.VII.3 - Boîtes à moustaches représentant la consommation moyenne (en l/100 km) des trajets effectués en conduite économique selon les connaissances du conducteur sur l'éco-conduite.

3. L'éco-conduite et la vitesse

La figure VII.4 représente la vitesse moyenne des conducteurs selon le type de conduite. On observe une vitesse moyenne moins élevée en conduite économique, cependant cela ne suffit pas à conclure que rouler lentement permet de réduire sa consommation, même si ce n'est pas faux. En effet, la vitesse moyenne est très sensible aux temps d'arrêts, ainsi un arrêt au feu de quelques minutes peut faire baisser la vitesse moyenne mais augmenter la consommation moyenne.

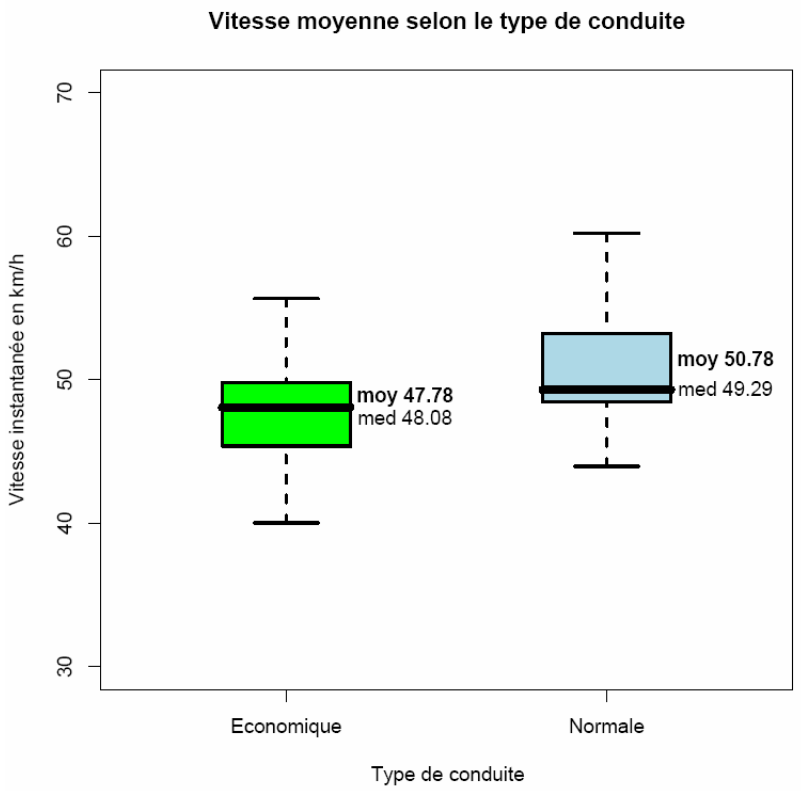


Fig.VII.4 - Boîtes à moustaches représentant la vitesse moyenne (en km/h) des trajets selon le type de conduite. L'anova détecte une différence significative (DF=1, F=6.6, p=0.0137) entre les deux types de conduite.

La figure VII.5 représentant la répartition des vitesses instantanées selon le type de conduite est plus intéressante car elle est moins biaisée par le problème du temps passé à l'arrêt. On observe un report vers des vitesses plus faibles en conduite économique, alors qu'en conduite normale il y a une plus forte concentration des vitesses supérieures à 70km/h.

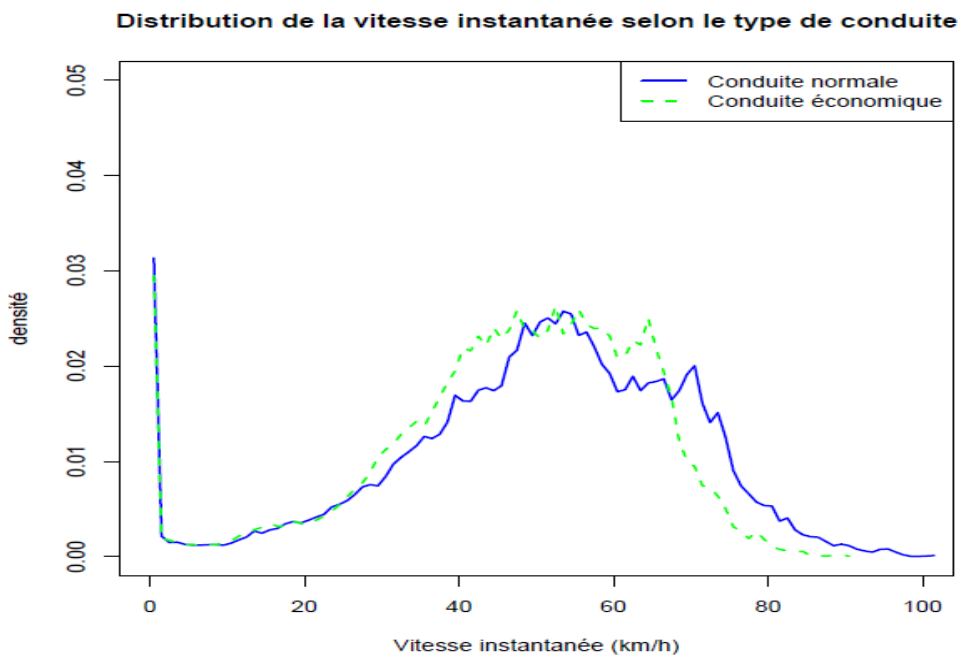


Fig.VII.5 - Répartition des vitesses instantanées selon le type de conduite.

Intéressons nous plus particulièrement au comportement du conducteur selon les vitesses de consigne. Les graphiques VII.6 montrent un pic des vitesses autour de 60km/h en conduite économique pour les zones limitées à 70km/h et 90km/h, alors que les vitesses sont plus élevées en conduite normale. On peut donc conclure qu'il y a bien une réduction de la vitesse mais qui se traduit par un respect des vitesses autorisées comme l'illustre la figure VII.7. En effet, le pourcentage de temps passé au-delà de la vitesse de consigne est nettement moins élevé en conduite économique. On peut donc penser que l'éco-conduite permet un gain en termes de sécurité. Ainsi avoir une conduite économique ne signifie pas rouler plus lentement mais adapter sa vitesse à la réglementation. De plus, la figure VII.8 montre qu'une conduite économique n'augmente que très peu le temps de parcours puisque la différence entre la durée moyenne du trajet selon les deux types de conduite n'est que d'une minute sur un trajet total de 18 minutes environ.

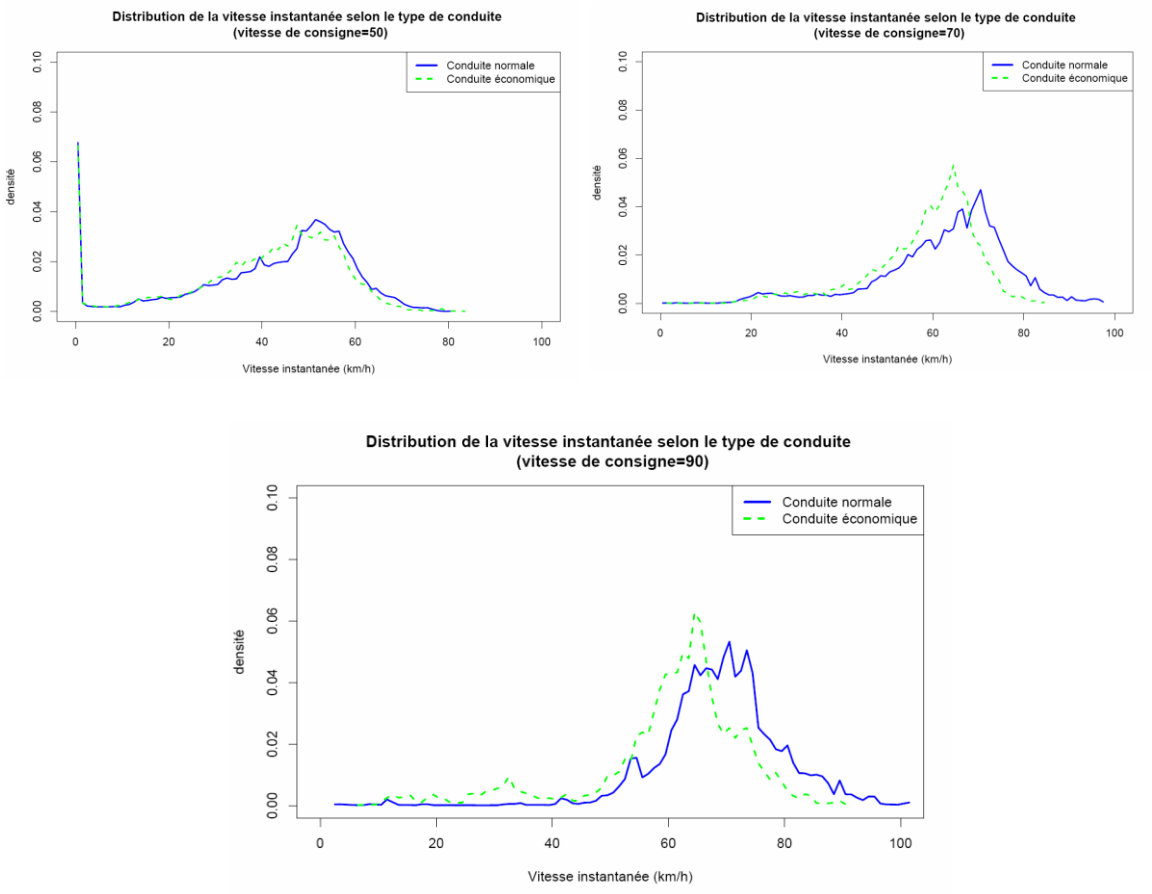


Fig.VII.6 - Répartition des vitesses instantanées selon les vitesses de consigne pour chaque type de conduite.

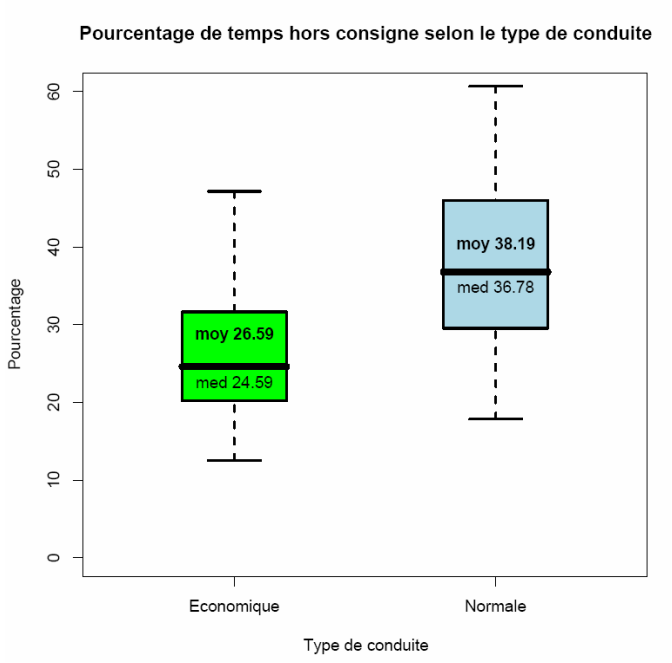


Fig.VII.7 - Boîtes à moustaches représentant le pourcentage de temps passé au-delà de la vitesse de consigne selon le type de conduite. L'anova détecte une différence significative (DF=1, F=15.9, p=0.0002) entre les deux types de conduite.

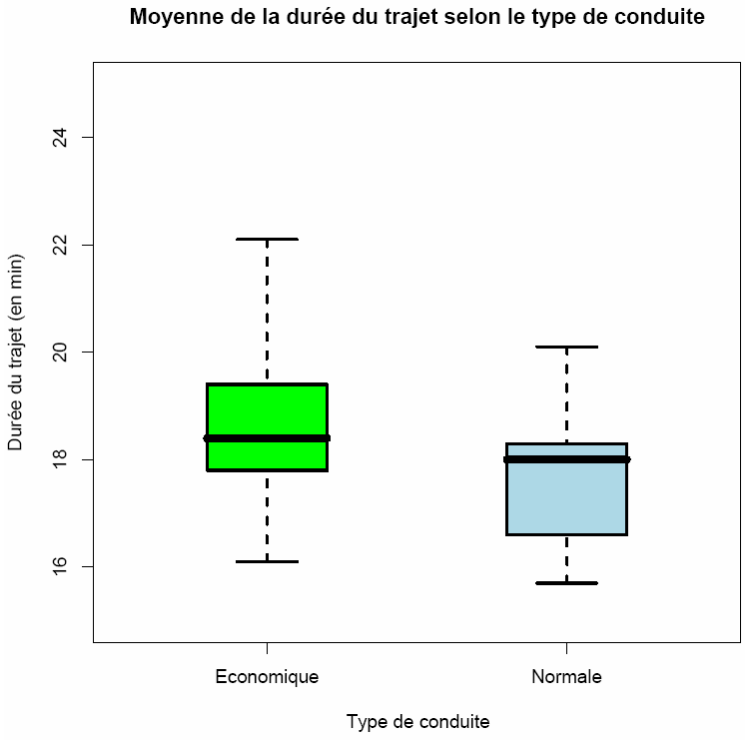


Fig.VII.8 - Boîtes à moustaches représentant la durée moyenne du trajet en minutes selon le type de conduite. L'anova détecte une différence significative (DF=1, F=6.94, p=0.0116) entre les deux types de conduite.

4 La mise en pratique des règles d'éco-conduite

Avant de parcourir le trajet en conduite économique, chaque conducteur a été informé des principes de bases de l'éco-conduite définies dans le contexte du projet Ecodrive. Il a ensuite été demandé au conducteur de suivre au mieux ces consignes afin de chercher à économiser du carburant. Cette section a pour objectif d'analyser la manière dont les conducteurs ont mis en pratique ces règles.

4.1 Un régime moteur faible et un rapport de boîte élevé

4.1.1 Utilisation du moteur

L'un des principes fondamentaux de l'éco-conduite est de conduire avec un régime moteur faible et de passer à la vitesse supérieure avec un régime moteur moins élevé. La figure VII.9 illustre cette technique : le régime moteur moyen passe de 2097 tr/min en conduite normale à 1822 tr/min en conduite économique, ce qui correspond à une baisse de 13%. De même le changement de vitesse s'effectue à un régime moteur moyen égal à 1590 tr/min en conduite économique contre 1806 tr/min en conduite normale.
Ces résultats sont confirmés à la figure VII.10 représentant la distribution des valeurs du régime moteur pour les deux types de conduite. On observe que le régime moteur ne dépasse pas les 2500 tr/min en conduite économique et que le pic situé aux environs de 2000 tr/min en conduite économique est translaté vers des régimes moteurs plus faibles en conduite économique.

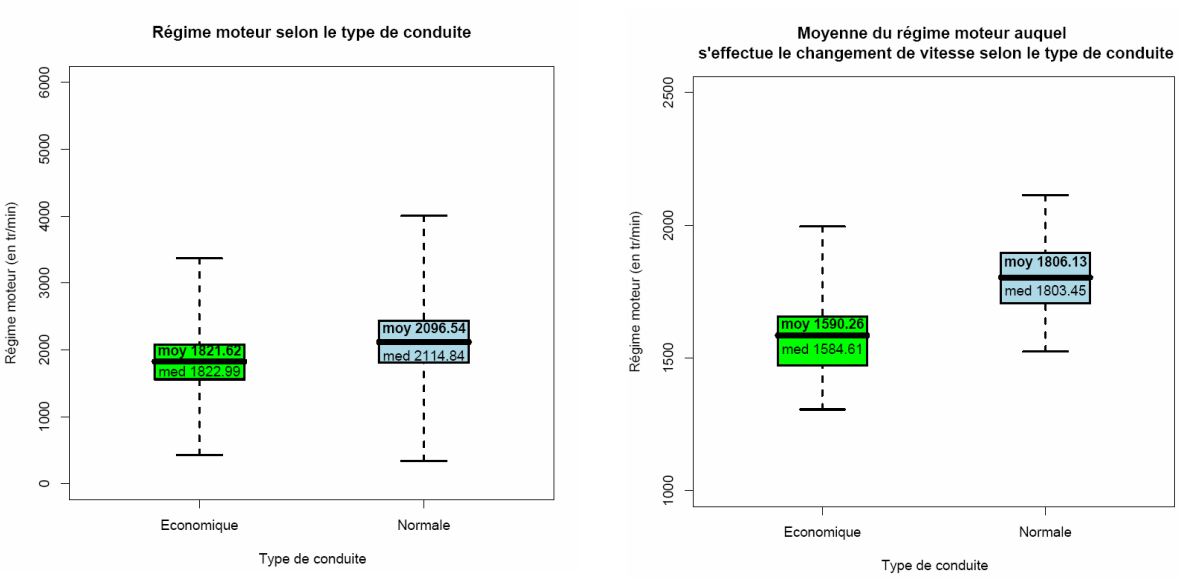


Fig.VII.9 - Boîtes à moustaches représentant le régime moteur ainsi que le régime moteur moyen auquel s'effectue le changement de vitesse selon le type de conduite. L'anova détecte une différence significative dans les deux cas (DF=1, F=69812, p<0.0001 pour le 1er graphique et DF=1, F=21.5, p<0.0001 pour le 2ème).

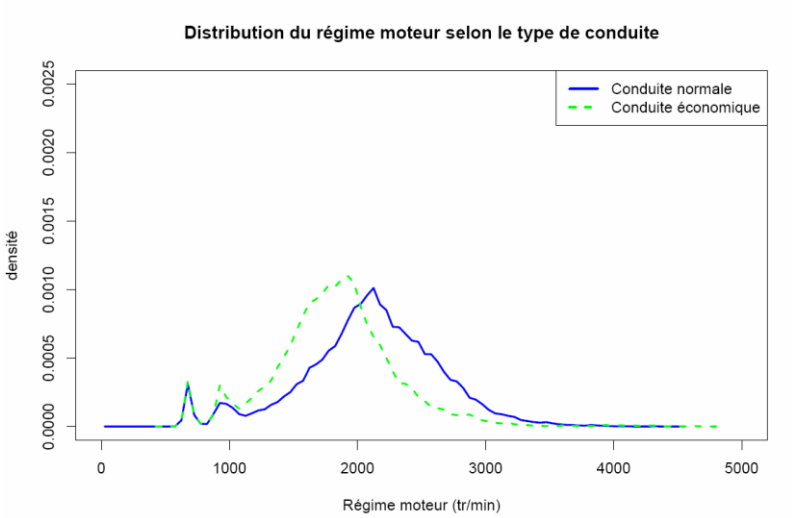


Fig.VII.10 - Distribution des valeurs du régime moteur pour chaque type de conduite.

4.1.2 Répartition des rapports de boîte

L'analyse des choix de rapport de boîte de vitesse des conducteurs selon le type de conduite est illustrée par les graphiques VII.11 et VII.12. En effet, un autre principe de l'éco-conduite est de monter les rapports de vitesse rapidement et de conduire avec un rapport de vitesse élevé même à faible vitesse. La répartition des rapports de boîte (figure VII.11) est très différente selon le type de conduite. La vitesse 5 est nettement plus représentée en conduite économique qu'en conduite normale (24.7% en conduite normale et 38.1% en conduite économique) alors que la vitesse 3 est beaucoup moins utilisée (30% en conduite normale et 20.2% en conduite économique). De plus, en conduite économique, la quatrième et la cinquième vitesse sont utilisées pour rouler à des vitesses beaucoup plus faibles (utilisation de la cinquième dès 40km/h) comme l'illustre les graphiques VII.12. En conduite économique, les conducteurs utilisent les vitesses 4 et 5 au lieu de la troisième vitesse et ceci même à faible vitesse.

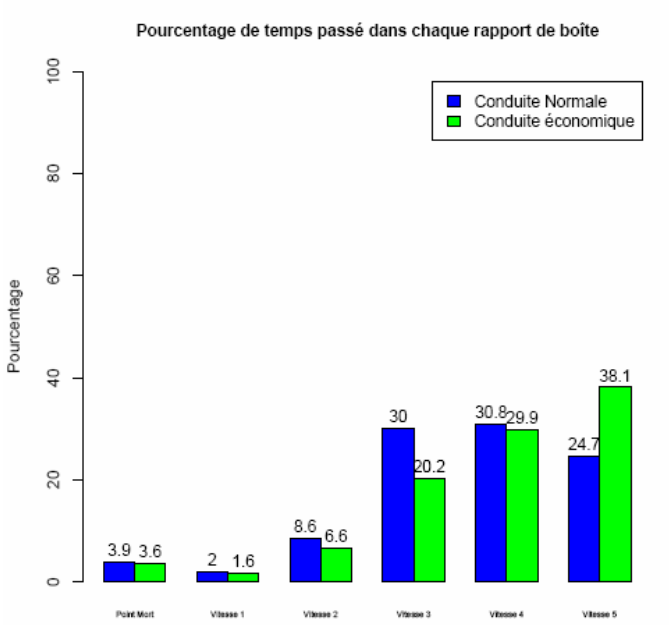


Fig.VII.11 - Répartition des rapports de boîte de vitesse selon le type de conduite.

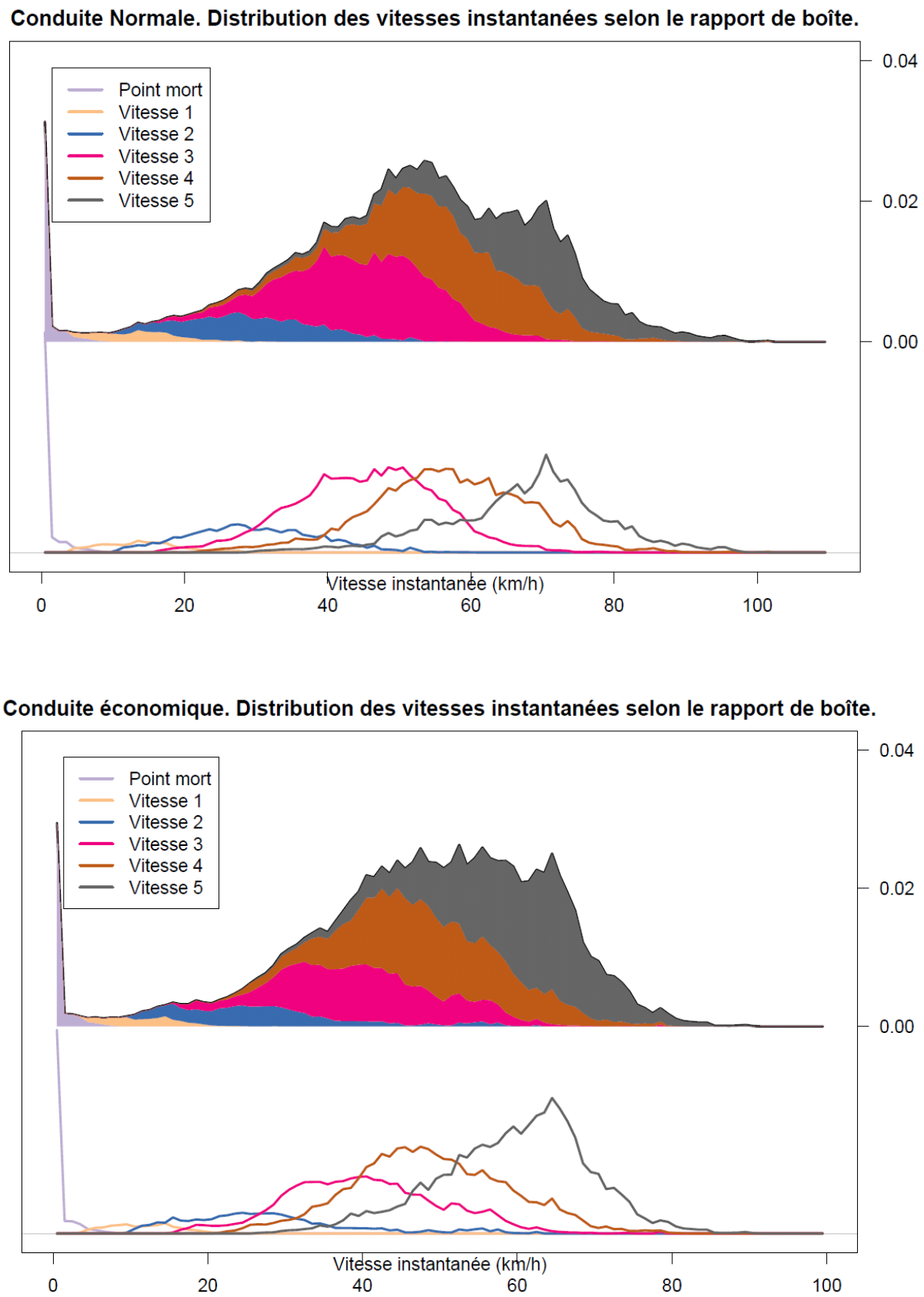


Fig.VII.12 - Distribution et distribution cumulée (en haut) des vitesses instantanées selon le rapport de boîte pour chaque type de conduite.

4.1.3 Analyse selon la vitesse de consigne

Nous abordons maintenant une analyse stratifiée selon les limitations de vitesse. Celle-ci se présente sous la forme de deux types de graphiques. Le premier donne la répartition des rapports de boîte de vitesse selon le type de conduite pour la limitation de vitesse en question. Le second présente la répartition des couples (Vitesse Instantanée $\times$ Régime Moteur). Il s'agit en fait d'une estimation de densité par la méthode des noyaux gaussiens, permettant de mettre en évidence les points d'accumulation qui seront associés à des pics délimités par les lignes de niveaux. La répartition de ces valeurs laisse apparaître les rapports de boîte de vitesse qui présentent des formes allongées.

Pour les zones limitées à 50km/h, le graphique VII.13 montre que la cinquième vitesse est nettement plus utilisée en conduite économique et que la troisième l'est beaucoup moins. Les graphiques de la figure VII.14 présentent une répartition spatiale très différente selon le type de conduite. En conduite normale, la troisième est plus souvent choisie, et on observe un pic autour de 2700tr/min (cf. figure VII.14 (a)), ce qui augmente la consommation de carburant. La figure VII.14 (b) montre un report des vitesses 3 et 4 vers les vitesses 4 et 5 en conduite économique, ce qui entraîne un régime moteur plus faible et donc une conduite plus économique en termes de consommation de carburant.

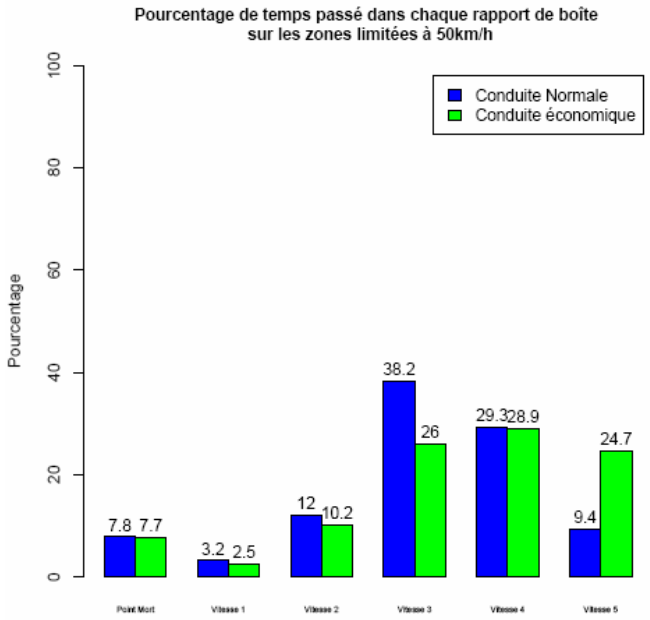
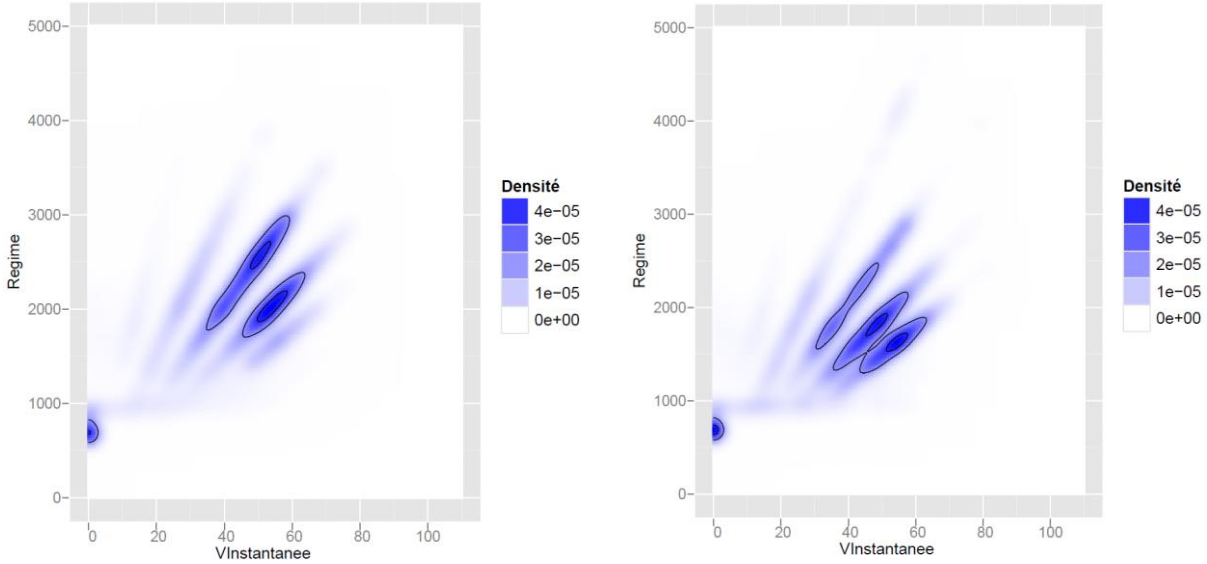


Fig.VII.13 - Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 50km/h



(a) Conduite normale,
Vitesse de consigne = 50km/h

(b) Conduite économique,
Vitesse de consigne = 50km/h

Fig.VII.14 - Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite, et pour les zones limitées à 50km/h. Les lignes de niveau sont identiques entre les graphiques.

On observe les mêmes résultats que précédemment pour les zones limitées à 70km/h avec une cinquième vitesse plus représentée en conduite économique (cf. figure VII.15). De plus les graphiques VII.16 montrent que la vitesse 5 est utilisée avec un régime moteur plus élevé en conduite normale. En effet, en conduite normale, on observe un pic en cinquième autour de la vitesse de consigne associé à un régime moteur supérieur à 2000 tr/min (cf. graphique VII.16 (a)), alors qu'en conduite économique ce pic apparait pour des vitesses plus faibles et un régime moteur inférieur à 2000 tr/min (cf. graphique VII.16 (b)).

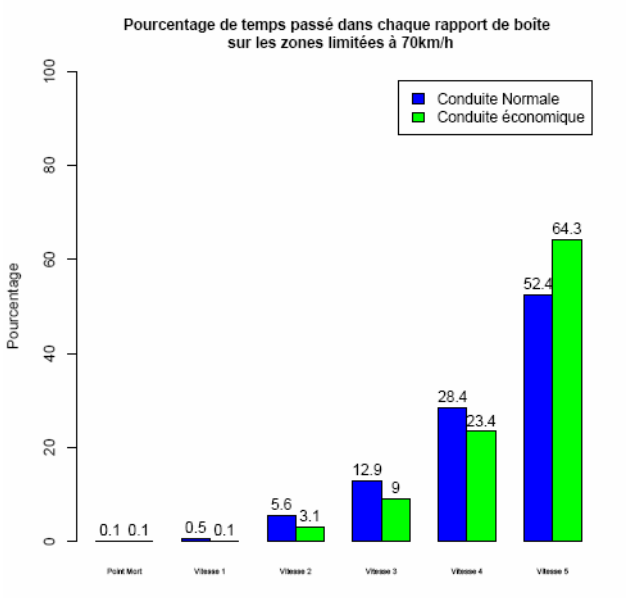
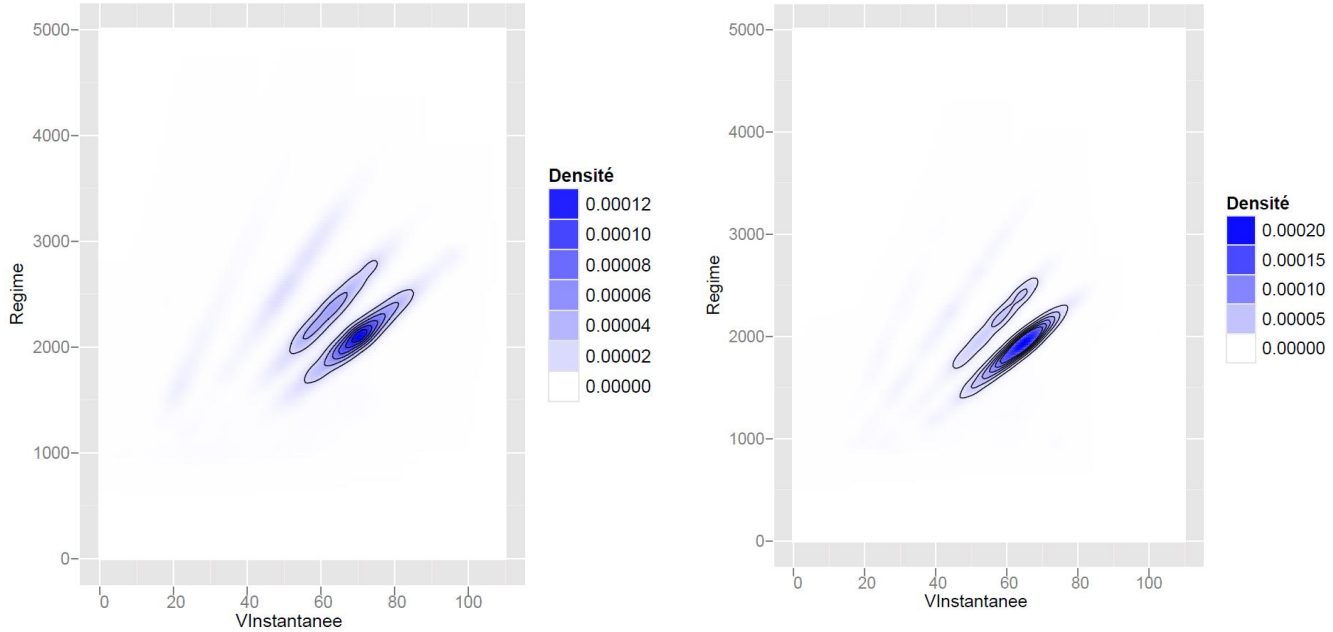


Fig.VII.15 - Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 70km/h.



(a) Conduite normale,
Vitesse de consigne = 70km/h

(b) Conduite économique,
Vitesse de consigne = 70km/h

Fig.VII.16 - Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite, et pour les zones limitées à 70km/h. Les lignes de niveau sont identiques entre les graphiques.

Pour les zones limitées à 90km/h, on observe un choix plus fréquent de la vitesse 5 en conduite économique à la place de la vitesse 4 (cf. graphique VII.17). La répartition spatiale des couples (Vitesse Instantanée × Régime Moteur) présentée figure VII.18 confirme ce résultat. De plus, la figure VII.18 (a) correspondant à une conduite normale montre un pic en vitesse 4 autour de 70km/h pour des valeurs du régime moteur supérieures à 2500 tr/min, alors qu'en conduite économique la vitesse 5 est nettement plus représentée. Le choix de la vitesse 5 en conduite économique induit un régime moteur plus faible et donc une réduction de la consommation de carburant.

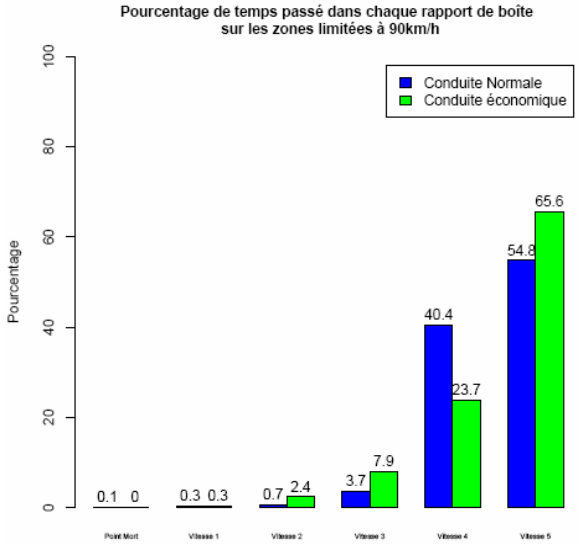
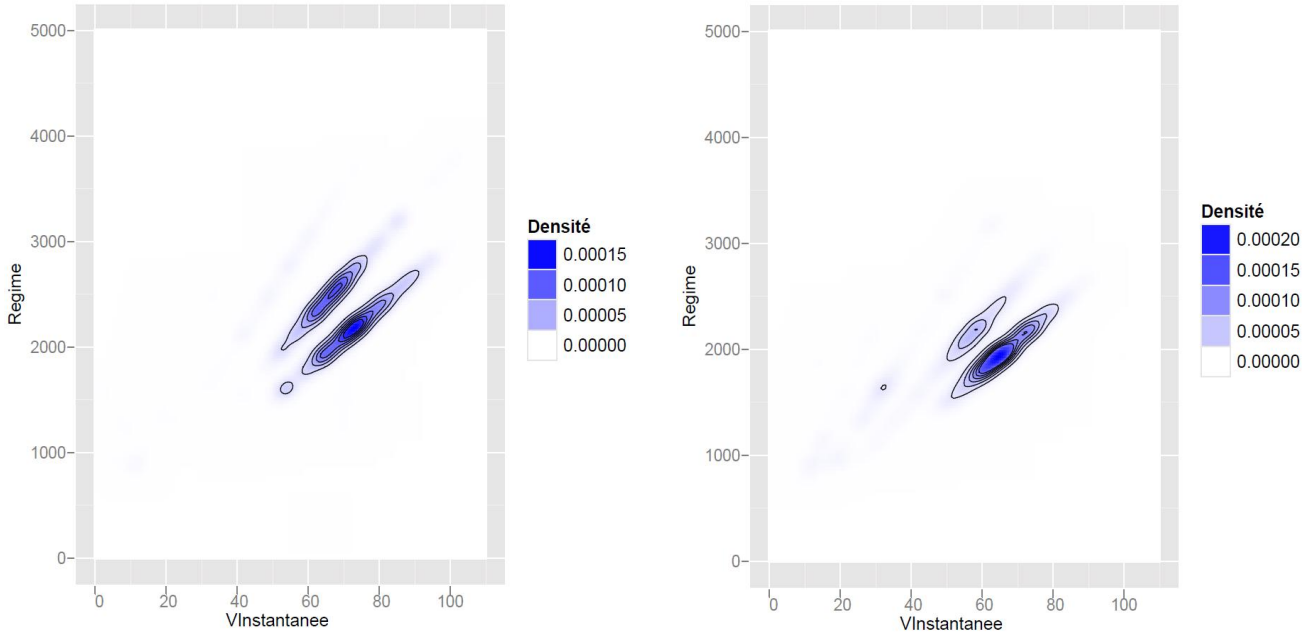


Fig.VII.17 - Répartition des rapports de boîte de vitesse selon le type de conduite pour les zones limitées à 90km/h.



(a) Conduite normale,
Vitesse de consigne = 90km/h

(b) Conduite économique,
Vitesse de consigne = 90km/h

Fig.VII.18 - Estimation de densité non paramétrique par noyaux gaussiens de la répartition des couples (Vitesse Instantanée × Régime Moteur), pour chaque type de conduite, et pour les zones limitées à 90km/h. Les lignes de niveau sont identiques entre les graphiques.

4.2 Anticipation du trafic

Un deuxième principe fondamental de l'éco-conduite est l'anticipation du trafic afin de maintenir une vitesse constante et d'éviter les accélérations et freinages inutiles.

4.2.1 Gestion de l'accélération

La figure VII.19 présente la distribution de l'accélération du véhicule selon le type de conduite. On observe une plus forte concentration des valeurs proches de zéro en conduite économique, ce qui signifie que les accélérations et décélérations du véhicule sont moins importantes qu'en conduite normale. Lors d'une conduite économique, les conducteurs évitent les accélérations inutiles qui entrainent une augmentation de la consommation. La conduite est ainsi plus souple comme le montre les graphiques VII.20 relatifs au pourcentage d'enfoncement de la pédale d'accélérateur. La moyenne du pourcentage d'enfoncement de la pédale d'accélérateur est plus faible en conduite économique : 15.8% en conduite normale, et 14% en conduite économique (cf. graphique VII.20 (a)). De plus, les conducteurs enfoncent très peu la pédale d'accélérateur de plus de 40% en conduite économique comme l'illustre la figure VII.20 (b).

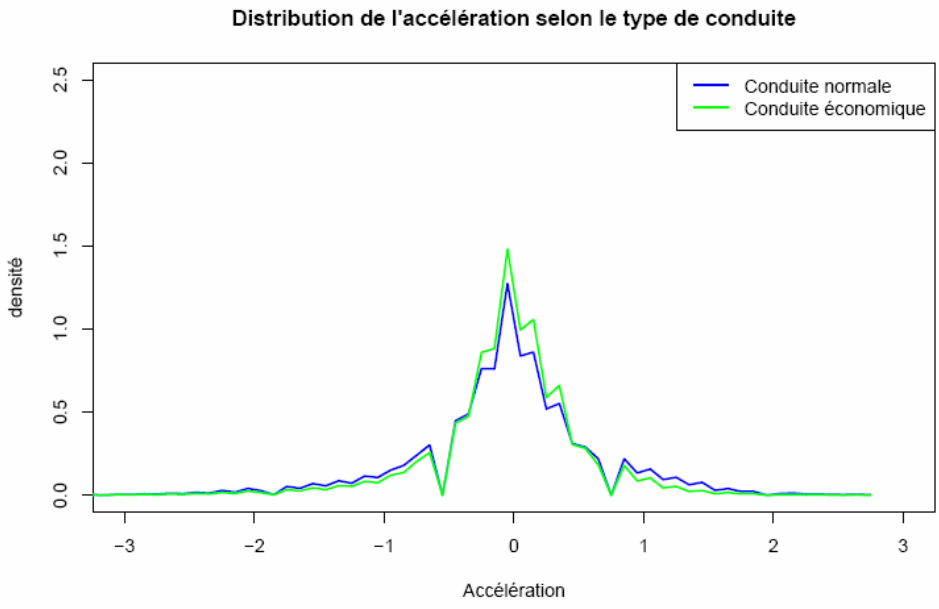
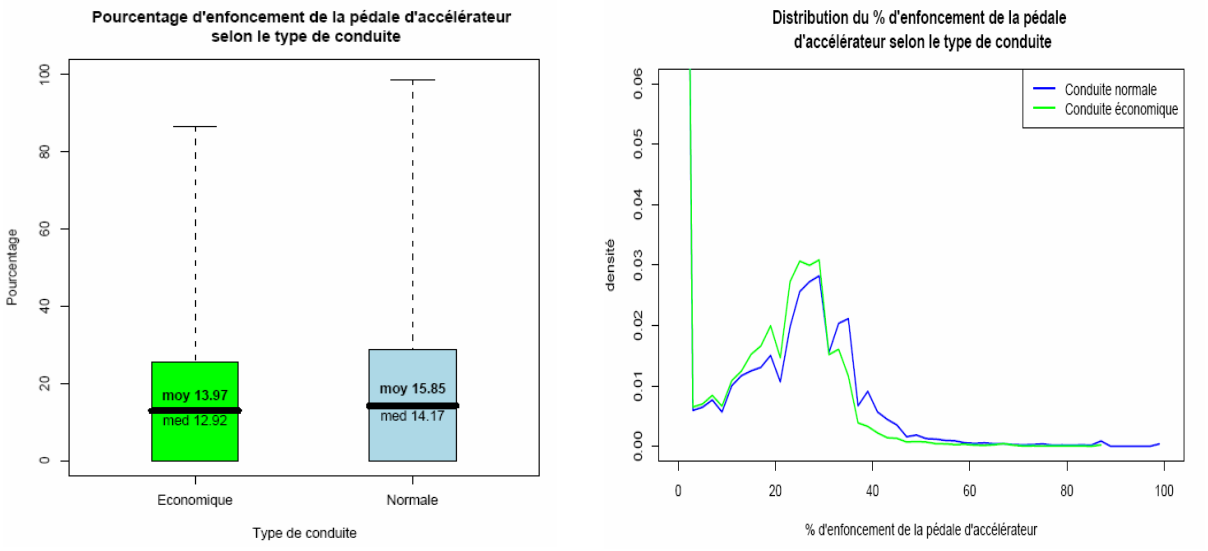


Fig.VII.19 - Distribution des valeurs d'accélérations/décélérations du véhicule pour chaque type de conduite.



(a) Boîtes à moustaches *(b) Distribution*

Fig. VII.20 - Boîtes à moustaches (a) et distribution (b) du pourcentage d'enfoncement de la pédale d'accélérateur selon le type de conduite.

4.2.2 Gestion des freinages

La gestion des freinages en conduite économique est elle aussi très représentative d'une conduite plus souple. En effet, le graphique VII.21 montre que le nombre de freinages a baissé de 30% en moyenne entre une conduite normale et une conduite économique (41 freinages en moyenne en conduite normale, et 29 en conduite économique). Le pourcentage de distance parcourue en freinant a aussi diminué : 14% en conduite normale, et 9% en conduite économique (cf. graphique VII.22). Ainsi l'éco-conduite est caractérisée par une meilleure anticipation du trafic et donc une réduction des freinages, ce qui induit un gain en termes de consommation.

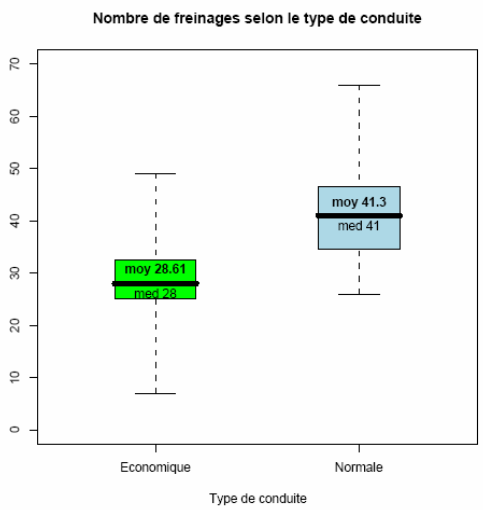


Fig. VII.21 – Boîte à moustaches représentant le nombre de freinages selon le type de conduite. L'anova détecte une différence significative (DF=1, F=20.94, p<0.0001) entre les deux types de conduite.

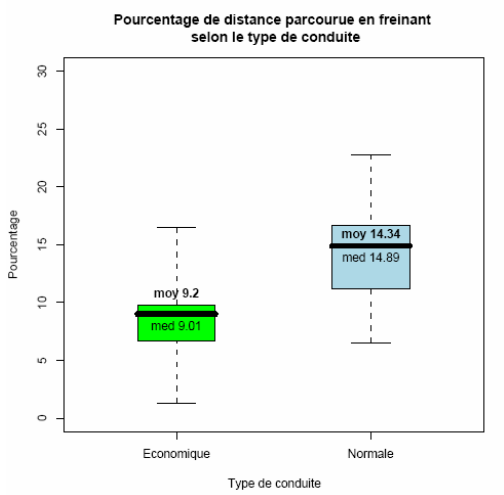


Fig. VII.22 – Boîte à moustaches représentant le pourcentage de distance parcourue en freinant selon le type de conduite. L'anova détecte une différence significative (DF=1, F=16.4, p=0.0002) entre les deux types de conduite.

Une technique associée à l'éco-conduite et liée à l'anticipation du trafic est l'utilisation du frein moteur. Rappelons que lorsque le véhicule est en frein moteur, l'alimentation en carburant du moteur est coupée. Le pourcentage de temps passé en frein moteur a légèrement augmenté en conduite économique comme l'illustre le graphique VII.23 mais la différence est assez faible. Les conducteurs ne sont pas assez formés à cette technique comme le montre le graphique VII.24. Les moniteurs ont davantage utilisé le frein moteur lors de leur trajet en conduite économique que les autres conducteurs : 41% en moyenne pour les moniteurs, 36% pour les conducteurs ayant des notions pratiques de l'éco-conduite, et 35% pour ceux n'ayant que des notions théoriques.

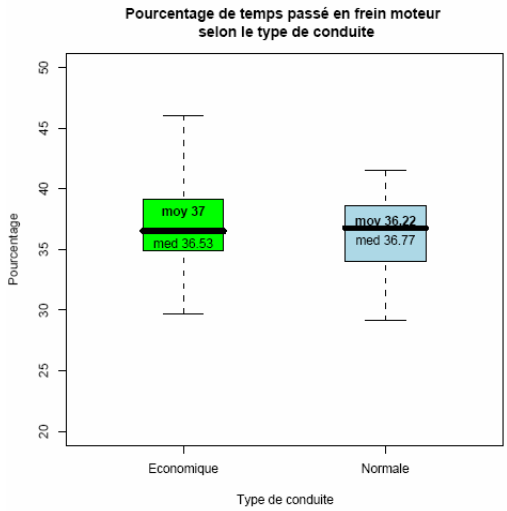


Fig. VII.23 – Boîte à moustaches représentant le pourcentage de temps de trajet effectué en frein moteur selon le type de conduite. L'anova ne détecte pas de différence significative (DF=1, F=0.5, p=0.482) entre les deux types de conduite.

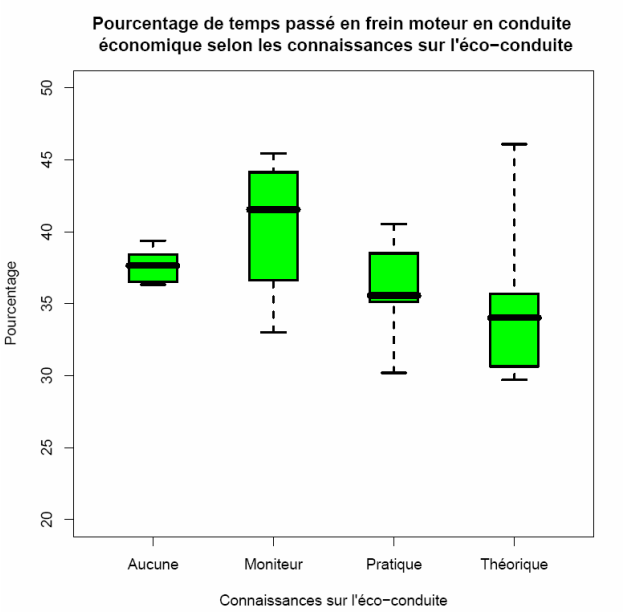


Fig. VII.24 – Boîte à moustaches représentant le pourcentage de temps de trajet effectué en frein moteur en conduite économique selon les connaissances du conducteur sur l'éco-conduite. L'anova ne détecte pas de différence significative (DF=3, F=1.43, p=0.266) entre les quatre classes.

5 Conclusion

La mise en pratique des principes de l'éco-conduite a montré un réel gain en termes de consommation de carburant, et ceci pour tous les conducteurs. L'éco-conduite semble aussi être associée à une conduite plus respectueuse des limitations de vitesses mais sans impliquer de réelle augmentation du temps de parcours. D'autre part, cette analyse a montré que les conducteurs n'ont pas eu de difficultés à mettre en œuvre les principes de base de l'éco-conduite tels que conduire avec un régime moteur faible en enclenchant la plus haute vitesse possible et maintenir une vitesse constante par anticipation du trafic. Seule l'utilisation du frein moteur semble nécessiter une formation d'ordre pratique.

Chapitre VIII

Analyse statistique multivariée de l'ensemble des trajets

Les données prises dans leur ensemble étant trop complexes, il est nécessaire de synthétiser au mieux l'information. En effet, notre base de données contient une centaine d'indicateurs, ce qui rend la comparaison entre conduite normale et conduite économique difficile. L'objectif de ce chapitre est donc de simplifier notre base de données en perdant le moins d'information possible.
Dans une première partie, nous effectuerons une classification des variables afin de sélectionner celles qui sont le plus pertinente pour notre étude. Puis nous effectuerons une analyse en composantes principales de l'ensemble des trajets afin de détecter les variables discriminant le mieux les trajets et de repérer d'éventuels groupes d'individus.

1 Suppression des trajets avec la climatisation

Parmi les 23 conducteurs ayant participé à l'expérience, deux conducteurs ont effectué leurs deux trajets avec la climatisation. La figure VIII.1 montre que l'utilisation de la climatisation entraine une augmentation de la consommation de l'ordre de 10% en moyenne. Cependant cette augmentation n'est pas liée au style de conduite du conducteur, c'est pourquoi nous ne tiendrons pas compte de ces quatre trajets dans la suite de l'analyse afin de ne pas biaiser les résultats.

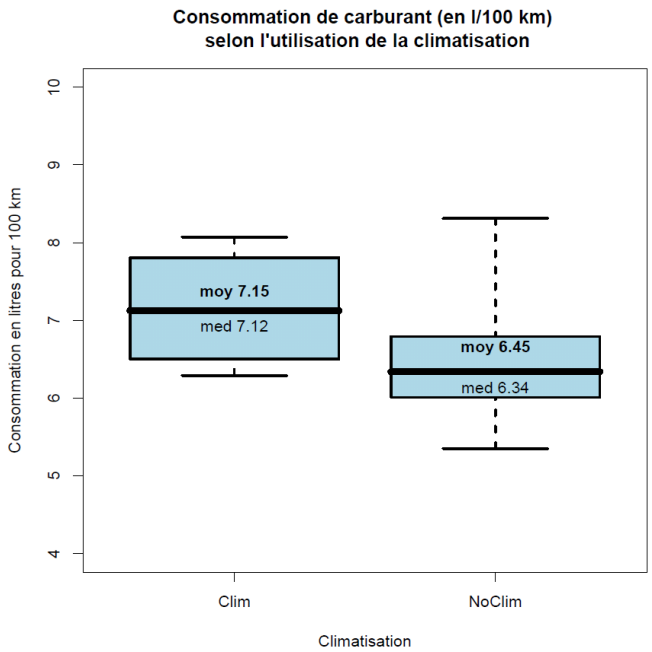


Fig.VIII.1 - Boîtes à moustaches représentant la consommation moyenne (en l/100 km) selon l'utilisation ou non de la climatisation. L'anova détecte une différence significative au risque de 5% (DF=1, F=4.16, p=0.0473).

2 Classification des variables

Notre base de données finale étant composée d'une centaine de variables, correspondant aux indicateurs donnés dans l'annexe A, il est nécessaire d'effectuer un tri. Une première sélection a été effectuée en supprimant les variables jugées les moins pertinentes. Notre nouvelle base de données comporte ainsi 49 variables. L'un des objectifs étant de prédire la consommation moyenne à partir de certains indicateurs, la variable ConsoCanL100, représentant la consommation moyenne en l/100 km donnée par le bus CAN, a été dissociée de l'ensemble des autres variables. Beaucoup de variables étant encore disponibles, il est nécessaire d'effectuer une classification de celles-ci afin de résumer l'information apportées par les données. De nombreuses méthodes permettent de classer et de regrouper les variables les plus corrélées. Nous présentons dans ce paragraphe une méthode bien connue qui est la classification hiérarchique ascendante, et nous avons aussi tenté de sélectionner les variables les plus liées à la consommation en effectuant une régression linéaire multiple.

2.1 Classification hiérarchique ascendante (CAH)

L'objectif de cette méthode est de créer des groupes de variables de manière à rassembler les variables les plus corrélées, c'est-à-dire portant les mêmes informations, et dissocier celles qui expriment des informations complémentaires. Le principe consiste à créer, à chaque étape, une partition obtenue en agrégeant deux à deux les éléments les plus proches. L'algorithme est le suivant :

- A l'étape initiale, chaque individu (correspondant ici aux variables) forme une classe. Il y a donc *n* classes.

- A la première étape, on regroupe les deux individus les plus proches qui vont former une nouvelle classe (utilisation de critères d'agrégation). On obtient une nouvelle partition en $n - 1$ classes.
- A la deuxième étape, on regroupe les deux classes les plus proches en utilisant le critère d'agrégation. On obtient une nouvelle partition en $n - 2$ classes.
- Et ainsi de suite jusqu'à l'étape $n - 1$ où le regroupement des deux classes de la partition de l'étape précédente conduit à une seule classe.

Ainsi l'algorithme ne fournit pas une partition en q classes d'un ensemble de n individus mais une hiérarchie de partitions se présentant sous la forme d'arbres appelés dendogrammes et contenant $n - 1$ partitions.

L'arbre de classification obtenu est présenté à la figure VIII.2. Chaque groupe réunissant les variables qui sont similaires en termes de corrélation, il peut y avoir redondance d'information entre les variables d'un même groupe. On cherchera donc à synthétiser et simplifier l'analyse en ne retenant que les variables les plus significatives mais aussi les plus simples à manipuler. Pour chaque groupe, les tables des corrélations des variables sont données en annexe B. Notons que les variables représentant le pourcentage de temps passé dans chaque plage de vitesse et dans chaque plage d'accélérations/décélérations ont été regroupées en six nouvelles variables (cf. tableau VIII.1).

Nom de la variable	Type	Description
TempsVit30	Réel	Pourcentage de temps de trajet passé avec une vitesse inférieure à 30km/h
TempsVit70	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 30 et 70km/h
TempsVit100	Réel	Pourcentage de temps de trajet passé avec une vitesse supérieure à 70km/h
TempsAccNeg	Réel	Pourcentage de temps de trajet passé avec une accélération comprise entre -10 et -1 m/s²
TempsAccMed	Réel	Pourcentage de temps de trajet passé avec une accélération comprise entre -1 et 1 m/s²
TempsAccPos	Réel	Pourcentage de temps de trajet passé avec une accélération comprise entre 1 et 10 m/s²

Tab. VIII.1 – Nouvelles variables représentant le pourcentage de temps passé dans chaque plage de vitesse et dans chaque plage d'accélérations/décélérations.

L'arbre de la figure VIII.2 permet de distinguer dix principaux groupes de variables :

- Le premier groupe est associé à la notion de vitesse moyenne, reliée au pourcentage d'enfoncement de la pédale d'accélérateur et au pourcentage de temps passé au delà de la vitesse de consigne. De ce groupe nous ne conserverons que les variables *MeanPosAcc* et *TempsHorsConsigne* qui sont fortement corrélées aux autres variables et qui semblent être les plus significatives (cf. Annexe B).
- Le deuxième groupe ne contient qu'une seule variable, *SumChangeVitesse*, que nous conserverons.
- Le troisième groupe représente la notion d'accélération. Nous ne conserverons que les variables *StdAccel* et *PKE* qui sont très fortement corrélées aux autres variables et qui semblent être les plus discriminantes (cf. Annexe B).

- Le quatrième groupe ne contient que deux variables et représente la notion de freinage. Nous ne retiendrons que la variable *SumFreinage*.
- Nous regrouperons les cinquième et sixième groupes qui contiennent les variables associées à une conduite standard, c'est-à-dire le pourcentage de temps passé dans chaque rapport de boîte ainsi que le pourcentage de temps passé avec un régime moteur médian. Nous ne conserverons aucune des variables de ces deux groupes car elles ne semblent pas significatives et présentent un faible pouvoir discriminant.
- Le septième groupe représente la notion de régime moteur, et plus particulièrement le temps passé avec un régime moteur élevé. De ce groupe, nous ne conserverons que la variable *MeanRegime* fortement corrélée aux autres variables du groupe ainsi que la variable *TempsVitesseNonOptimale* (cf. Annexe B).
- Le huitième groupe représente la notion de temps passé à l'arrêt. Nous ne retiendrons que la variable *TempsArrete*.
- Le neuvième groupe ne contient que deux variables, *TempsVAcc19* et *TempsFreinMoteur*. Pour des facilités d'interprétation nous ne conserverons que la deuxième.
- Le dernier groupe ne contient qu'une seule variable *TempsRapBoite26* que nous ne conserverons pas par la suite car elle semble peu significative.

Nous conserverons donc pour la suite de l'étude les 10 variables données dans le tableau VIII.2.

Nom de la variable	**Description**
MeanPosAcc	Moyenne du pourcentage d'enfoncement de la pédale d'accélérateur
TempsHorsConsigne	Pourcentage de temps de trajet passé au-delà de la vitesse de consigne
SumChangeVitesse	Nombre de changement de rapports de boîte
StdAccel	Ecart-type de l'accélération du véhicule
PKE	Positive Kinetic Energy
SumFreinage	Nombre de freinages
MeanRegime	Moyenne du régime moteur
TempsVitesseNonOptimale	Pourcentage de temps de trajet passé avec une vitesse non-optimale
TempsArrete	Pourcentage de temps de trajet passé à l'arrêt
TempsFreinMoteur	Pourcentage de temps de trajet passé en frein moteur

Tab. VIII.2 – Variables sélectionnées d'après la classification ascendante hiérarchique.

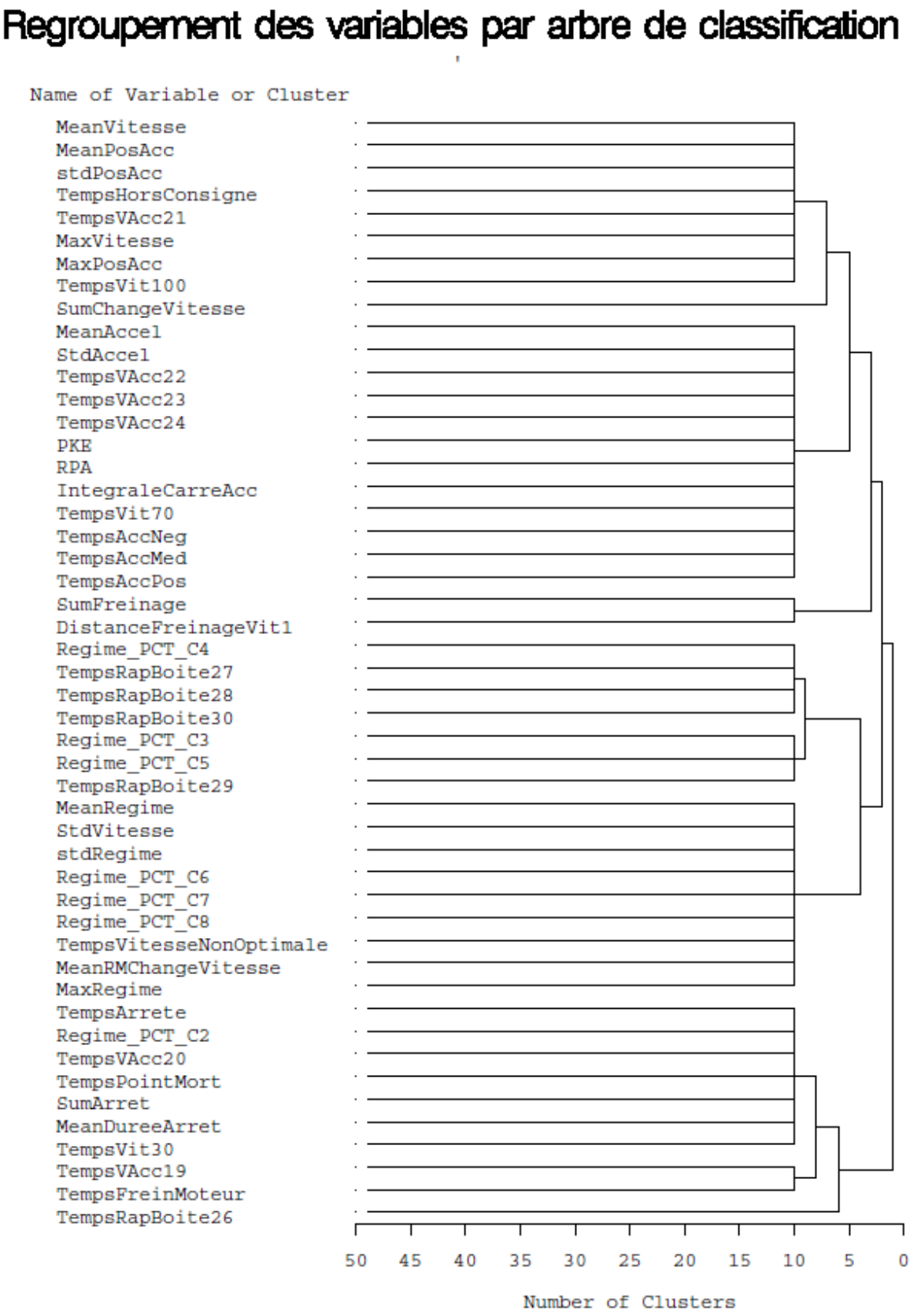


Fig. VIII.2 – Arbre de classification des 49 variables conservées dans la base de données finale.

2.2 Sélection automatique des variables par régression linéaire multiple

2.2.1 Généralités sur la régression linéaire

Description

La régression linéaire est une méthode permettant d'étudier la liaison linéaire entre une variable à expliquer (ou dépendante) Y et une ou plusieurs variables explicatives (ou indépendantes). Toutes les variables considérées doivent être numériques.
Dans le cas de la régression linéaire simple, il n'y a qu'une seule variable explicative. Le modèle s'écrit alors :

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

où Y est la variable « réponse »,
X est la variable « régresseur »,
ε est un terme d'erreur (c'est une variable aléatoire centrée suivant une loi normale).
L'objectif est alors de chercher la droite qui passe « au plus près » de tous les points observés. Le critère d'ajustement est le « critère des Moindres Carrés Ordinaires » (MCO) qui consiste à minimiser la somme des carrés des écarts (SCE) des points observés à la droite solution. Les écarts sont calculés en projetant les points observés M_i parallèlement à l'axe des Y (cf. figure VIII.3).

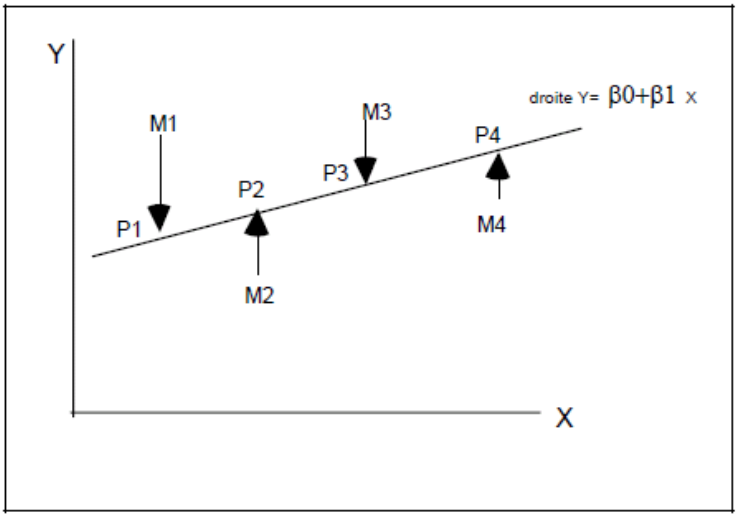


Fig. VIII.3 – Projection des points M_i sur la droite.

La somme des carrés des écarts (SCE) est donnée par :

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

La droite obtenue, appelée droite de régression, a alors pour équation : $\hat{Y} = b_0 + b_1 X$

où $\hat{Y}$ est l'estimation de Y obtenue à partir de l'équation de régression (prédicteurs),
b_0 et b_1 sont les estimateurs des moindres carrés des paramètres inconnus β_0 et β_1.

L'écart entre Y_i observé et son estimation $\hat{Y}_i$ est appelé le résidu. C'est une estimation de l'erreur inobservable ε_i.

La régression linéaire simple se généralise au cas de plusieurs variables explicatives. On parle alors de régression linéaire multiple, et l'on écrit :

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \varepsilon$$

en ajoutant une hypothèse supplémentaire : l'indépendance linéaire des X_i.

La régression linéaire simple cherchait à faire passer une droite "au mieux" (c'est à dire au sens des Moindres Carrés) à travers le nuage de points représentant les données dans le plan (*x*, *y*). Il va en être de même pour la régression linéaire multiple, si ce n'est que la représentation visuelle devient maintenant impossible puisque l'on cherche l'hyperplan de dimension *k* minimisant la somme des carrés des écarts (mesurés parallèlement à l'axe y) entre les points représentant les données et l'hyperplan.

Formule de décomposition

La notion de liaison entre *X* et *Y* signifie qu'une variation de *X* entraîne une variation de *Y*. La formule de décomposition suivante est fondamentale car elle permet de préciser la part de la variation de *Y* expliquée par la variation de *X* :

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

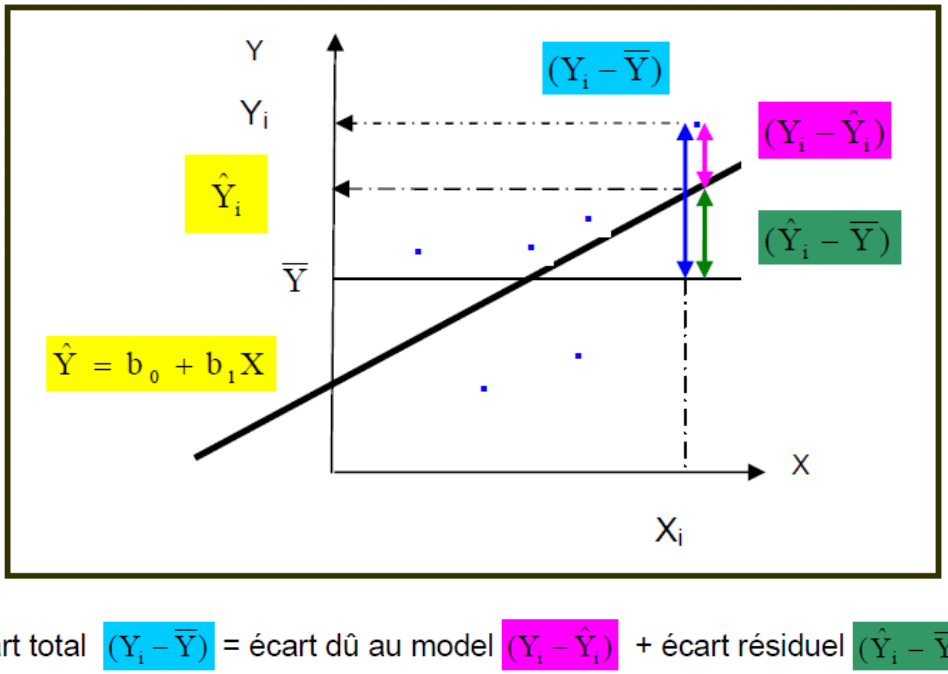
Le terme de gauche $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ s'appelle la somme des carrés totale et mesure la variation globale des Y_i autour de leur moyenne $\overline{Y}$.

Le premier terme $\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$ du membre de droite s'appelle la somme des carrés expliquée par la variable *X*. C'est la partie de la variation de *Y* expliquée par la variable *X*.

Le deuxième terme $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ du membre de droite s'appelle la somme des carrés résiduelle. C'est la partie de la variation de *Y* non expliquée par la variable *X*.

La formule précédente se note ainsi :

SCE totale = SCE modèle + SCE erreur



Ecart total $(Y_i - \overline{Y})$ = écart dû au model $(Y_i - \hat{Y}_i)$ + écart résiduel $(\hat{Y}_i - \overline{Y})$

Fig. VIII.4 –Décomposition des différents écarts.

Une mesure d'appréciation globale de la régression : le coefficient de détermination R²

Les deux quantités SCE totale et SCE modèle sont des sommes de carrés donc toujours positives ou nulles et telles que SCE Modèle $\leq$ SCE Totale. Le rapport $\frac{SCE\ Modèle}{SCE\ Totale}$ est donc compris entre 0 et 1. On appelle ce rapport le coefficient de détermination et on le note R². Il représente la part expliquée par la régression de la somme des carrés des écarts à la moyenne. Autrement dit, il mesure le pouvoir explicatif du modèle : plus R² est proche de 1, meilleure est l'adéquation du modèle aux données. Un coefficient R² égal à 1 correspond au cas particulier où tous les points Y_i observés sont alignés sur une droite.

2.2.2 Méthodes de régression

Lorsque le nombre de variables explicatives est important, il devient nécessaire de disposer de méthodes de sélection automatique des variables. Les méthodes les plus utilisées sont les suivantes :

- Méthode FORWARD (ascendante) : c'est une méthode itérative consistant à sélectionner à chaque étape la variable explicative maximisant le R² de *Y* avec toutes les variables sélectionnées aux étapes précédentes et la nouvelle variable choisie, tant que l'apport de cette dernière est significative.
- Méthode BACKWARD (descendante) : on part cette fois du modèle complet, et on élimine à chaque pas la variable la moins significative.
- Méthode STEPWISE (progressive) : c'est une combinaison FORWARD/BACKWARD. On effectue une sélection FORWARD, en laissant la possibilité de faire sortir du modèle à chaque pas une des variables devenues non significative.

2.2.3 Sélection automatique des variables avec la méthode stepwise

Afin de sélectionner les variables les plus liées à la consommation, nous avons effectué une régression linéaire multiple avec la méthode stepwise. Nous avons ainsi pris la variable ConsoCanL100 comme variable réponse Y, et les 49 variables pré-selectionnées comme variables explicatives. Les résultats de la procédure REG du logiciel SAS sont donnés à la figure VIII.5. On observe qu'après neuf étapes, sept variables ont été conservées dans le modèle :
MeanAccel, StdPosAcc, Regime_PCT_C4, TempsFreinMoteur, SUmFreinage, TempsVit30 et TempsAccMed. Le R^2 est de 0.96, ce qui signifie que le modèle est bon d'un point de vue mathématique. Si l'on compare cette sélection avec la classification ascendante hiérarchique effectuée précédemment, on remarque que ces sept variables appartiennent quasiment toutes à des groupes différents :

- StdPosAcc appartient au groupe 1,
- MeanAccel et TempsAccMed appartiennent au groupe 3,
- SumFreinage appartient au groupe 4,
- Regime_PCT_C4 appartient au groupe 5,
- TempsVit30 appartient au groupe 8,
- TemposFreinMoteur appartient au groupe 9.

Les deux méthodes semblent donc cohérentes mais les variables sélectionnées par la régression linéaire sont moins pertinentes et plus difficiles à interpréter. Nous n'utiliserons donc pas les résultats obtenus par cette méthode et nous conserverons les dix variables sélectionnées d'après la classification hiérarchique ascendante pour la suite de l'étude.

Étape	Variable entrée	Variable supprimée	Libellé	Nombre de variables dans le modèle	R carré partiel	R carré du modèle	Valeur F	Pr > F
1	TempsAccMed		% de temps passé avec une accélération comprise entre -1 et 1 m/s²	1	0.8582	0.8582	242.08	<.0001
2	TempsFreinMoteur		% de temps passé avec le frein moteur	2	0.0271	0.8853	9.22	0.0043
3	stdPosAcc		Ecart-type du % d'enfoncement de la pédale d'accelerateur	3	0.0309	0.9162	14.02	0.0006
4	TempsVit30		% de temps passé avec une vitesse inférieure à 30km/h	4	0.0137	0.9299	7.24	0.0106
5	TempsRapBoite28		% de temps passé en vitesse 3	5	0.0119	0.9418	7.34	0.0103
6	MeanAccel		Acceleration moyenne	6	0.0072	0.9490	4.93	0.0329
7	SumFreinage		Nombre de freinages	7	0.0085	0.9575	6.77	0.0136
8	Regime_PCT_C4		% de temps de trajet passé avec un régime moteur compris entre 1500 et 2000 tr/min	8	0.0058	0.9633	5.22	0.0289
9		TempsRapBoite28	% de temps passé en vitesse 3	7	0.0003	0.9629	0.31	0.5840

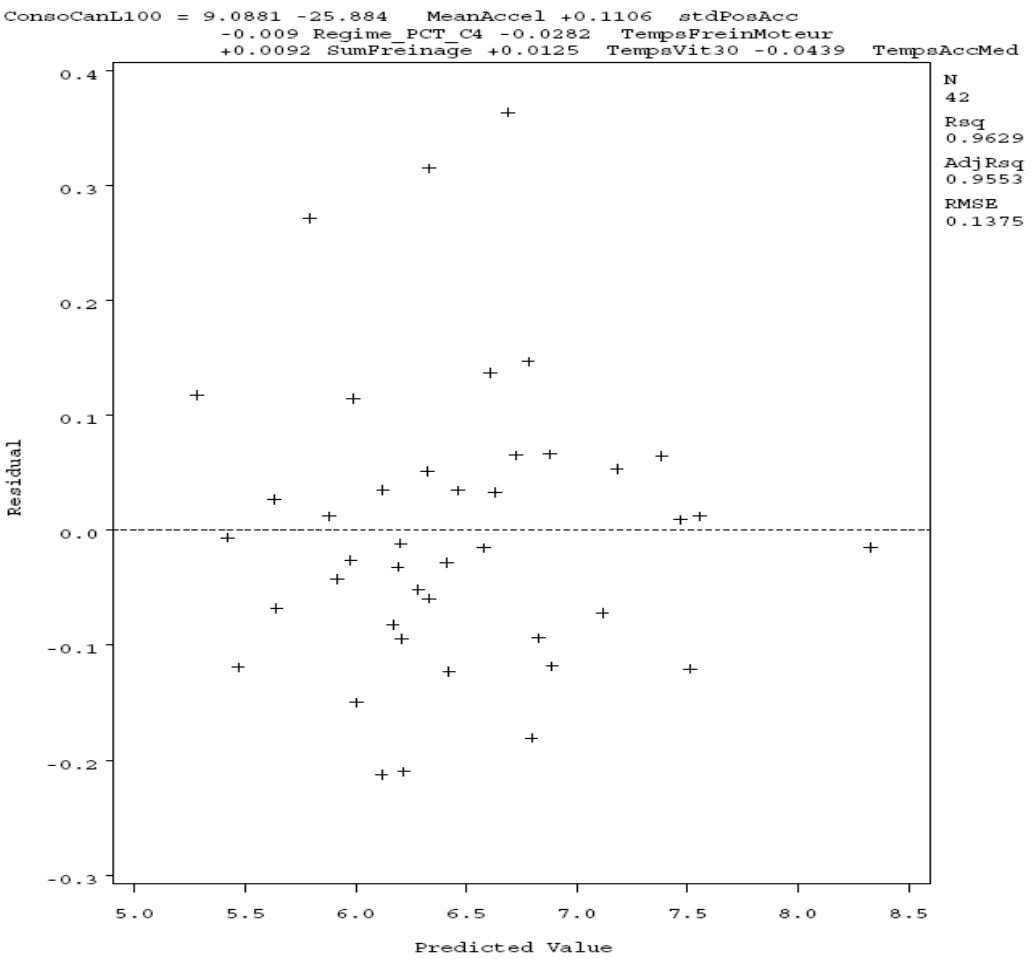


Fig. VIII.5 – Résultats de la procédure REG du logiciel SAS : tableau récapitulatif des étapes de la méthode stepwise et graphique des résidus.

3. Analyse en composantes principales

3.1 Généralités sur l'ACP

L'analyse en composantes principales (ACP) est une méthode d'analyse de données qui permet, à la fois de représenter en 2 ou 3 dimensions, le plus fidèlement possible, les individus d'une population, et aussi de détecter les liaisons entre les variables ainsi que les variables séparant le mieux les individus. On cherche à synthétiser l'information contenue dans un tableau croisant des individus et des variables quantitatives.

Quand les p variables décrivant les n individus sont toutes numériques, chaque individu peut être représenté par un point dans un espace R^p à p dimensions. L'ensemble des individus est un « nuage de points ». Quand $p \leq 2$, l'observation du nuage montre bien les distances entre individus, mais cette observation devient plus difficile quand p=3, et elle est impossible dès que le nombre de variables excède 3. Il est donc naturel de vouloir réduire l'espace R^p à R^2 ou R^3 en projetant le nuage de points de l'espace à p dimensions sur un espace de faible dimension. Le problème est qu'une telle projection diminue systématiquement les distances

entre les individus : il faut évidemment chercher à les diminuer le moins possible afin que les distances entre les points projetés soient aussi proches que possible des distances entre les points dans l'espace d'origine. Ainsi l'ACP permet de projeter le nuage d'individus sur des sous-espaces de dimension inférieure en respectant au mieux les distances entre individus.
La détermination de ces sous-espaces se fait axe par axe. La dispersion du nuage N autour de son centre de gravité *g* est mesurée à l'aide de l'inertie du nuage N par rapport à *g*. Cette inertie, appelée inertie totale du nuage, est définie par une formule qui généralise la notion de variance :

$$I(N,g) = \frac{1}{n}\sum_{i=1}^{n} d^2(x_i, g) \text{ où } x_i \text{ est le vecteur colonne associé à l'individu } i.$$

Notons qu'afin de rendre les données homogènes, on transforme les variables d'origine en variables centrées-réduites (c'est-à-dire de moyenne 0 et d'écart-type 1). Le nouveau nuage de points étudié devient $N^* = \{x_1^*, ..., x_n^*\}$ et son centre de gravité est alors égal à 0. Le but de l'ACP consiste alors à trouver l'axe permettant de maximiser l'inertie projetée sur cet axe où l'inertie projetée sur un axe est définie par :

$$\frac{1}{n}\sum_{i=1}^{n} v_i^2 \text{ où } v_i \text{ est la coordonnée de l'individu } i \text{ sur l'axe.}$$

Cela revient à minimiser l'écart entre chaque individu et sa projection, c'est-à-dire allonger au maximum la projection du nuage d'individus sur l'axe. On cherche ensuite un deuxième axe qui, parmi tous les axes orthogonaux au premier sera celui qui maximise l'inertie projetée sur ce deuxième axe, celle-ci étant par construction inférieure à celle projetée sur le premier axe. Plusieurs axes factoriels sont ainsi successivement déterminés avec des inerties projetés décroissantes. En raison de leur orthogonalité, l'inertie totale du nuage d'individus se décompose en la somme des inerties projetées sur chaque axe.

Dans le nuage R^n des variables, on cherche, comme dans le nuage des individus, à maximiser l'inertie projetée. Toutefois, il ne s'agit pas ici de rechercher la direction d'allongement maximale du nuage, puisque toutes les variables ont même norme 1 et sont donc sur l'hypersphère de rayon 1. Il s'agit de maximiser les sommes des cosinus carrés des angles que forment l'axe cherché et les variables, autrement dit de maximiser les carrés des coefficients de corrélation des variables avec cet axe (qui définit une nouvelle variable C puisque l'on est dans l'espace des variables).

En résumé, l'objectif de l'ACP est de trouver l'axe qui :
- dans le nuage des variables, maximise $\sum_i r(X_{i,}C)^2$
- dans le nuage des individus, maximise $\frac{1}{n}\sum_i v_i^2$, ce terme étant la variance de C.

On cherche donc la variable, appelée composante principale, combinaison linéaire des variables analysées, qui est à la fois la plus corrélée à l'ensemble des variables analysées et qui est de variance maximale. Autrement dit, l'ACP peut aussi être présentée comme la recherche d'un ensemble réduits de variables non-corrélées, combinaisons linéaires des variables initiales résumant avec précision les données.

3.2 Représentation graphique d'une ACP

Il y a diverses manières de présenter les résultats d'une ACP. Nous utilisons ici un graphique complexe mis au point par Eric Lecoutre qui présente non seulement les résultats de l'ACP, mais aussi les résultats d'une classification ascendante hiérarchique sur les individus. Les résultats de l'ACP sont illustrés par la figure principale représentant l'un des plans factoriels (c'est-à-dire un sous-espace engendré par deux axes factoriels). La couleur de représentation des individus correspond à leur groupe d'appartenance, et leur taille varie selon leur poids dans la définition des axes. A côté, figure le cercle des corrélations des variables (toujours sur le même plan factoriel) accompagné d'un diagramme en bâtons illustrant les inerties de chaque axe. La partie en bas à gauche montre un arbre de classification accompagné d'un diagramme en bâtons illustrant la répartition des individus. Enfin, un dernier graphique présente la répartition sous forme de boîte à moustaches évoluée (car représentant aussi la densité des observations).

Représentation des variables

Dans l'espace des variables, l'intersection de la sphère unité, où se trouvent toutes les variables (supposées réduites donc de norme 1), avec le plan factoriel est appelé cercle des corrélations. Les variables sont d'autant plus proches du cercle des corrélations, qu'elles sont bien représentées par le plan factoriel. Le cosinus de l'angle entre deux variables est égal au coefficient de corrélation linéaire entre ces deux variables. Ainsi deux variables proches l'une de l'autre sont corrélées positivement, deux variables diamétralement opposées sont corrélées négativement, et deux variables perpendiculaires sont non-corrélées.

3.3 Analyse en composantes principales de l'ensemble des trajets

Cette section concerne l'analyse en composantes principales de l'ensemble des trajets avec les dix variables sélectionnées précédemment par la classification hiérarchique ascendante. La figure VIII.6 présente les résultats de l'ACP et de la classification (cf. section 3.2). Les deux premiers axes de l'ACP expliquent 43% de la variabilité des données. Le 1er axe correspond aux variables MeanRegime, StdAccel, PKE, MeanPosAcc et TempsHorsConsigne. Il oppose ainsi les trajets avec de fortes accélérations et un régime moteur élevé, aux trajets avec une conduite plus souple.
On retrouve cette opposition avec la classification : on distingue clairement deux groupes.
Les caractéristiques de chaque groupe par rapport aux valeurs moyennes de l'ensemble de l'échantillon sont illustrées par la figure VIII.8. Nous utilisons ici les valeurs moyennes des variables centrées et réduites. Une valeur élevée et positive pour une variable précise signifie que le groupe en question a des caractéristiques plus élevées que la moyenne pour cette variable. Ainsi, une barre très haute pour la variable MeanRegime correspond à un groupe avec un régime moteur moyen plus élevé que la moyenne. A l'inverse, une valeur négative sera associée à un groupe avec un régime moteur plus faible que la moyenne. Enfin, des valeurs proches de zéro seront associées à un comportement « normal » vis-à-vis de l'échantillon global considéré.
La figure VIII.8 ainsi que le tableau VIII.3 donnant les valeurs moyennes des différentes variables montre que le groupe 1, correspondant aux trajets situés le plus à gauche sur le 1er axe factoriel, représente des trajets avec un régime moteur élevé, de fortes accélérations, de nombreux freinages et donc une consommation élevée en carburant. A l'inverse, le groupe 2, correspondant aux trajets situés le plus à droite sur le 1er axe factoriel, représente des trajets

avec un régime moteur peu élevé, de faibles accélérations, peu de freinages et donc des trajets moins consommateurs en carburant. La figure VIII.7 représente le 1er plan factoriel où chaque individu est représenté par le numéro du trajet, les chiffres impairs correspondant à une conduite normale, et les chiffres pairs correspondant à une éco-conduite. On observe clairement que le groupe 1 correspond à des trajets effectués en conduite normale (16 trajets « normal » et 6 trajets « eco »), alors que le groupe 2 est plutôt associé à des trajets effectués en éco-conduite (15 trajets « eco » et 3 trajets « normal »).
Le 2ème axe oppose quand à lui le pourcentage de temps passé à l'arrêt au pourcentage de temps passé en frein moteur. Ainsi les trajets situés vers le haut correspondent à des trajets avec un temps passé à l'arrêt élevé, et les trajets situés vers le bas correspondent à des trajets où le frein moteur a été particulièrement utilisé.
Notons aussi que la classification a fait apparaître deux groupes constitués chacun d'un seul trajet. Le groupe 3 correspond à un trajet avec un régime moteur élevé mais peu de changement de rapport de boîte de vitesse, ce qui caractérise une conduite non optimale. Le groupe 4 a les mêmes caractéristiques que le groupe 3, avec en plus de très fortes accélérations et un temps passé avec une vitesse non optimale particulièrement élevé, ce qui caractérise une conduite nerveuse. Ces deux groupes correspondent naturellement à des trajets très consommateurs en carburant.

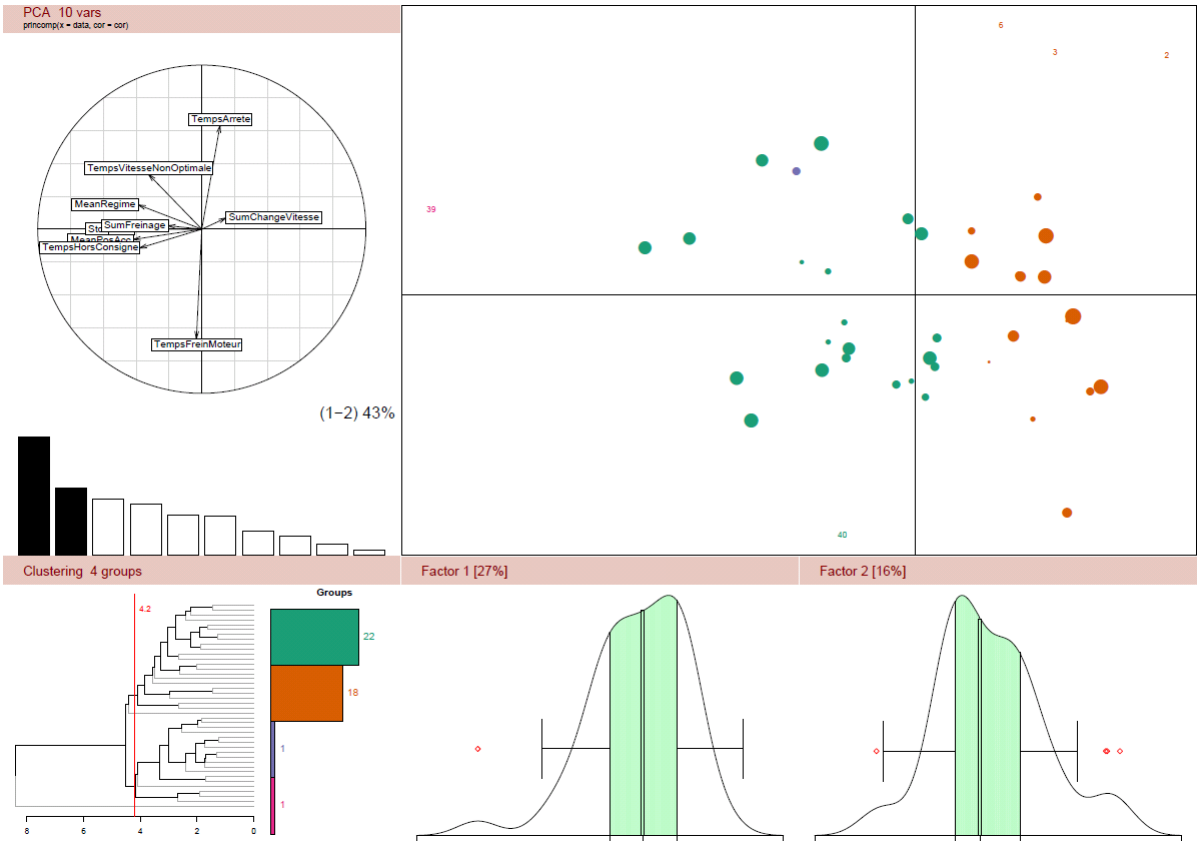


Fig. VIII.6 – ACP de l'ensemble des trajets et classification ascendante hiérarchique.

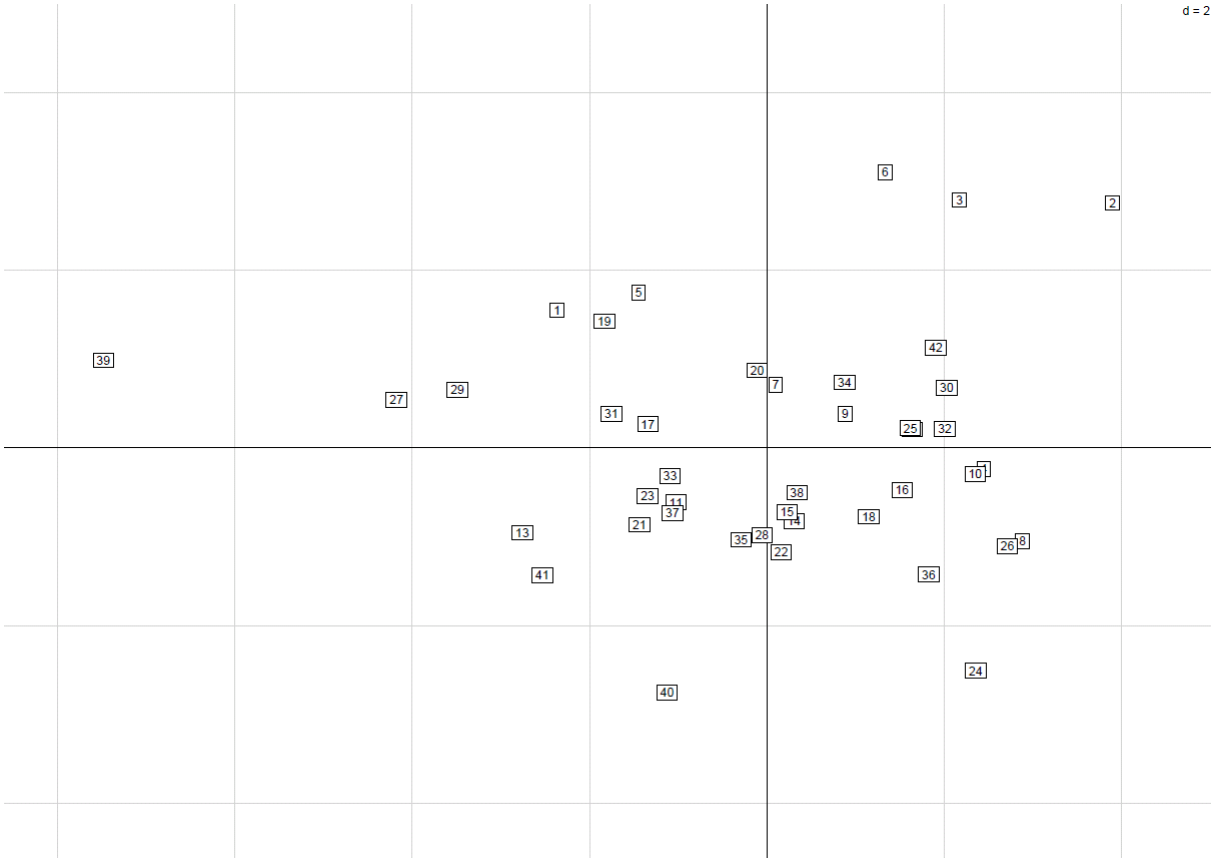


Fig. VIII.7 – Premier plan principal de l'ACP de l'ensemble des trajets.

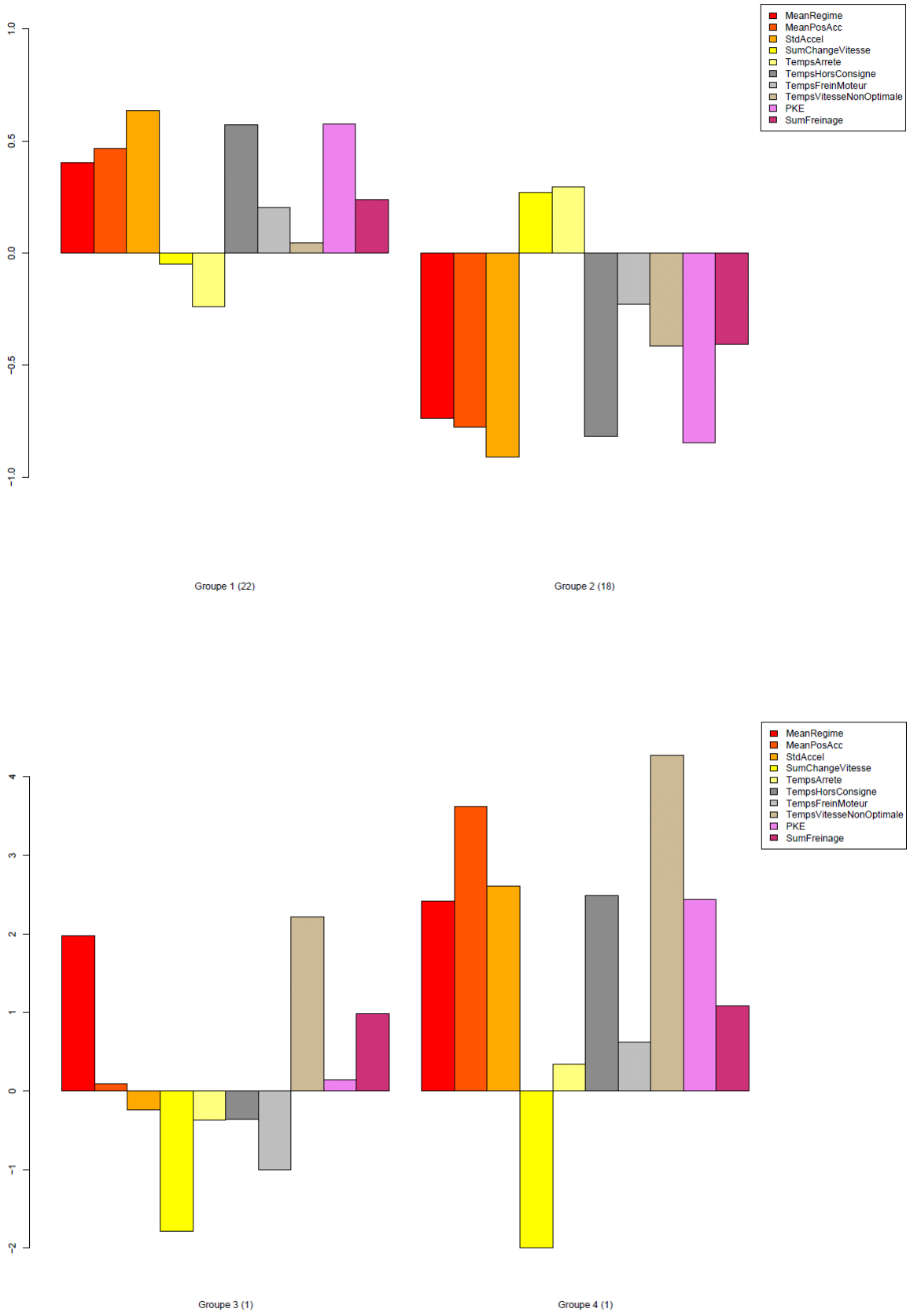


Fig. VIII.8 – Diagrammes en bâtons représentant les moyennes des variables centrées-réduites pour chacun des groupes.

Variables	Groupe 1 (22)	Groupe 2 (18)	Groupe 3 (1)	Groupe 4 (1)
MeanRegime	2056.93	1789.30	2424.66	2528.86
MeanPosAcc	15.75	13.39	15.03	21.78
StdAccel	0.70	0.54	0.61	0.91
SumChangeVitesse	111.86	122.06	56	49
TempsArrete	2.21	3.40	1.93	3.51
TempsHorsConsigne	38.46	22.31	27.61	60.69
TempsFreinMoteur	37.37	35.70	32.67	39
TempsVitesseNonOptimale	1.44	0.26	6.99	12.24
PKE	0.34	0.23	0.31	0.49
SumFreinage	37	30	45	46

Tab. VIII.3 – Moyennes des différentes variables pour chacun des groupes.

Chapitre IX

Modèles de prédiction

L'objectif de ce chapitre est, dans une première partie, de construire un modèle de prédiction de la consommation moyenne par régression linéaire, puis dans une deuxième partie, de construire un modèle de prédiction de la probabilité d'être dans une situation d'éco-conduite en utilisant la régression logistique.

1. Prédiction de la consommation moyenne par régression linéaire

On souhaite expliquer la consommation moyenne en l/100km à partir des 10 variables sélectionnées par la classification ascendante hiérarchique (cf. Tab. VIII.2). On effectue pour cela une régression linéaire multiple dont la méthode a été détaillée à la section 2.2.1 du chapitre VIII. Cependant si l'on observe le tableau des corrélations entre les variables explicatives (cf. Tab. IX.1), on constate que certaines variables sont très corrélées (par exemple, StdAccel et PKE). Ce problème de colinéarité rend les résultats de la régression très instables :

- coefficients de régression très sensibles aux fluctuations même faibles de la variable réponse Y,
- écarts-types élevés pour les coefficients de régression,
- dégradation de la précision des prédictions.

Afin de corriger ces effets nocifs de la colinéarité, nous avons effectué une sélection des variables par la méthode stepwise. Cette technique a déjà été exposée à la section 2.2 du chapitre VIII. Les résultats obtenus par la procédure REG du logiciel SAS figurent dans les tableaux IX.2 et IX.3. On obtient le modèle à 5 variables suivant :

$$\begin{aligned}\text{ConsoCanL100} = 3.571 &+ 0.1167 \times \text{MeanPosAcc} + 0.0396 \times \text{TempsArrete} \\ &- 0.0212 \times \text{TempsFreinMoteur} + 5.0048 \times \text{PKE} + 0.0098 \times \text{SumFreinage}\end{aligned}$$

Ce modèle montre une forte influence de la variable PKE sur la variable ConsoCanL100 (le coefficient associé est égal à 5.0048) : une légère augmentation de la variable PKE entraine une forte augmentation de la consommation de carburant. Notons que le modèle est bon puisque le coefficient R^2 est égal à 0.9344, ce qui signifie que les cinq variables régresseurs expliquent 93,44% de la variance de la variable réponse Y. D'autre part, le graphique des résidus standardisés (cf. fig. IX.1) montre que seulement deux valeurs semblent aberrantes (outliers). Selon, les critères habituels, des valeurs sont considérées comme aberrantes si leur résidu standardisé est supérieur à 2 en valeur absolu. Enfin, il est important de noter que le problème de colinéarité entre les variables explicatives a été résolu grâce à la méthode

stepwise. En effet, on a mesuré la colinéarité entre les cinq variables régresseurs à l'aide d'un indice appelé « tolérance ». La tolérance d'une variable est définie comme la proportion de la variance de cette variable qui n'est pas expliquée par les autres variables (c'est-à-dire 1-R^2, où R est la corrélation entre la variable testée et les autres variables explicatives). On considère qu'elle doit être supérieure à 0,2 ou au moins 0,1. Or, d'après le tableau IX.3, la tolérance de chacune des variables est bien supérieure à 0,2, ce qui indique que les cinq variables régresseurs sélectionnées pour le modèle sont peu corrélées, et donc que le modèle est stable.

	MeanRegime	**MeanPosAcc**	**StdAccel**	**SumChange Vitesse**	**TempsArrete**
MeanRegime	1.000	0.740	0.733	-0.161	-0.186
MeanPosAcc	0.740	1.000	0.838	-0.329	-0.306
StdAccel	0.733	0.838	1.000	-0.189	-0.163
SumChangeVitesse	-0.161	-0.329	-0.189	1.000	0.048
TempsArrete	-0.186	-0.306	-0.163	0.048	1.000
TempsHorsConsigne	0.682	0.910	0.788	-0.265	-0.259
TempsFreinMoteur	-0.132	0.036	0.090	-0.130	-0.480
TempsVitesseNonOptimale	0.738	0.562	0.589	-0.303	0.121
PKE	0.707	0.721	0.951	-0.184	-0.142
SumFreinage	0.221	0.302	0.477	-0.036	-0.110

	TempsHors Consigne	**TempsFrein Moteur**	**TempsVitesse NonOptimale**	**PKE**	**SumFreinage**
MeanRegime	0.682	-0.132	0.738	0.707	0.221
MeanPosAcc	0.910	0.036	0.562	0.721	0.302
StdAccel	0.788	0.090	0.589	0.951	0.477
SumChangeVitesse	-0.265	-0.130	-0.303	-0.184	-0.036
TempsArrete	-0.259	-0.480	0.121	-0.142	-0.110
TempsHorsConsigne	1.000	0.117	0.420	0.620	0.111
TempsFreinMoteur	0.117	1.000	-0.168	0.148	0.011
TempsVitesseNonOptimale	0.420	-0.168	1.000	0.592	0.265
PKE	0.620	0.148	0.592	1.000	0.577
SumFreinage	0.111	0.011	0.265	0.577	1.000

Tab. IX.1 – Tableau des corrélations entre les variables explicatives.

Étape	Variable entrée	Variable supprimée	Libellé	Nombre de variables dans le modèle	R carré partiel	R carré du modèle	C(p)	Valeur F	Pr > F
1	StdAccel		Ecart-type de l'acceleration	1	0.8409	0.8409	48.0305	211.39	<.0001
2	SumFreinage		Nombre de freinages	2	0.0349	0.8758	31.1629	10.96	0.0020
3	TempsFreinMoteur		% de temps passé avec le frein moteur	3	0.0266	0.9024	18.7740	10.36	0.0026
4	TempsVitesseNonOptimale		% de temps passé avec une vitesse non optimale	4	0.0124	0.9148	14.0825	5.37	0.0261
5	PKE		PKE	5	0.0056	0.9203	13.0749	2.51	0.1216
6	MeanPosAcc		Moyenne du % d'enfoncement de la pédale d'accelerateur	6	0.0077	0.9280	10.9037	3.75	0.0608
7		StdAccel	Ecart-type de l'acceleration	5	0.0018	0.9263	9.8723	0.87	0.3570
8		TempsVitesseNonOptimale	% de temps passé avec une vitesse non optimale	4	0.0032	0.9231	9.6009	1.56	0.2196
9	TempsArrete		% de temps de trajet passé à l'arrêt	5	0.0113	0.9344	5.4696	6.22	0.0173

Tab. IX.2 - *Tableau récapitulatif des étapes de la méthode stepwise.*

Résultats estimés des paramètres							
Variable	**Libellé**	**DF**	**Résultat estimé des paramètres**	**Erreur std**	**Valeur du test t**	**Pr > |t|**	**Tolérance**
Intercept	Intercept	1	3.57095	0.50038	7.14	<.0001	.
MeanPosAcc	Moyenne du % d'enfoncement de la pédale d'accelerateur	1	0.11669	0.02388	4.89	<.0001	0.37201
TempsArrete	% de temps de trajet passé à l'arrêt	1	0.03958	0.01587	2.49	0.0173	0.62281
TempsFreinMoteur	% de temps passé avec le frein moteur	1	-0.02118	0.00873	-2.43	0.0203	0.67297
PKE	PKE	1	5.00483	0.64430	7.77	<.0001	0.29339
SumFreinage	Nombre de freinages	1	0.00979	0.00333	2.94	0.0056	0.60833

Tab. IX.3 – Tableau des résultats estimés des paramètres.

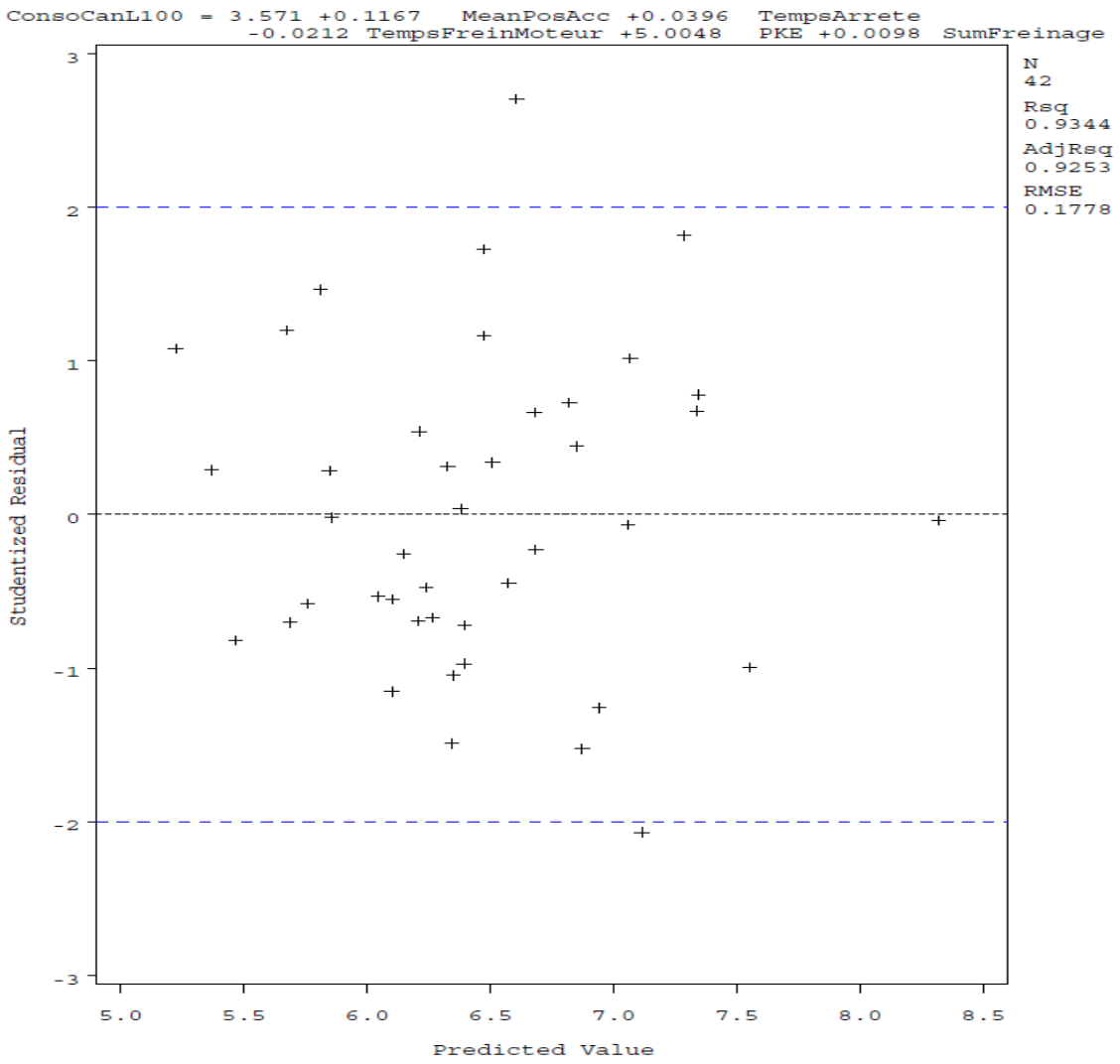


Fig. IX.1 – Graphique des résidus standardisés en fonction de la valeur prédite $\hat{Y}$.

2. Modélisation de la probabilité d'être en éco-conduite par régression logistique

A la section précédente, nous avons construit un modèle permettant de prédire la consommation moyenne sur un trajet à partir de cinq indicateurs. Dans cette section, on ne cherchera pas à prédire une variable, mais à prédire la probabilité d'un état. On cherche ici à modéliser la probabilité qu'un trajet ait été effectué avec une conduite économique. Pour cela, nous utiliserons un cas particulier de la régression, la régression logistique.

2.1 Généralités sur la régression logistique binaire

Nous avons vu à la section 2.2.1 du chapitre VIII que la régression linéaire permettait d'étudier la relation entre une variable dépendante quantitative Y et une ou plusieurs variables explicatives, elles aussi quantitatives. Cependant, lorsque la variable Y n'est pas quantitative mais qualitative ou catégorielle, le modèle de régression linéaire n'est pas approprié. Un cas particulier est celui pour lequel la variable Y est binaire, c'est-à-dire qu'elle ne prend que deux valeurs (Y=1 si une caractéristique est possédée et Y=0 sinon). Y suit alors une loi de Bernoulli caractérisée par les probabilités suivantes : $p = P(Y=1)$ et $1-p = P(Y=0)$.

L'objectif de la régression logistique est celui de toute régression : modéliser l'espérance conditionnelle $E(Y \mid X = x)$. On veut connaitre la valeur moyenne de Y pour toute valeur de X. Pour une valeur Y valant 0 ou 1, cette valeur moyenne est la probabilité que Y=1. On a donc : $E(Y \mid X = x) = P(Y = 1 \mid X = x)$.

Notons $\pi(x) = P(Y = 1 \mid X = x)$ la probabilité que Y prenne la valeur 1 pour X fixé à *x*. La variable Y suit alors une loi de Bernoulli de paramètre $\pi(x)$. En régression logistique binaire simple, c'est-à-dire dans le cas d'une seule variable explicative X, on suppose que la probabilité $\pi(x)$ s'écrit sous la forme suivante :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1)}.$$

De manière équivalente, on suppose que le rapport des côtes (ou odds ratio), aussi appelé logit, respecte la relation linéaire suivante :

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

Notons que la régression logistique binaire est en fait un cas particulier du modèle linéaire généralisé. Pour une variable Y suivant une loi de Bernoulli de paramètre $\pi(x)$, ce modèle s'écrit : $g(\pi(x)) = \beta_0 + \beta_1 x$ où la fonction g est appelée fonction de lien. En régression logistique, g est la fonction logit définie par $g(x) = \frac{x}{1-x}$ qui est l'inverse de la fonction logistique $F(x) = \frac{e^x}{1+e^x}$. La fonction logistique est bien adaptée à la modélisation de probabilités car elle prend ses valeurs entre 0 et 1 selon une courbe en S.

Le modèle peut être étendu à l'analyse d'une variable réponse binaire Y en fonction de plusieurs variables explicatives $X_1, ..., X_k$ qui peuvent être quantitatives ou qualitatives. Le modèle logistique a alors pour expression :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}$$

ou de manière équivalente : $\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$

L'estimation du modèle de régression logistique consiste alors à estimer les coefficients β_i de ce modèle. Pour cela, on utilise la méthode du maximum de vraisemblance. En supposant que les observations sont indépendantes, cette vraisemblance, qui correspond à la probabilité d'observer les *n* données (x_i, y_i) calculée en fonction des paramètres inconnus β_i, s'écrit :

$$L(\beta_0, \beta_1, ..., \beta_k) = \prod_{i=1}^{n} \pi(x_i)^{y_i} (1 - \pi(x_i))^{y_i}$$

La maximisation de cette vraisemblance n'a pas de solution analytique, on utilisera donc des méthodes numériques afin d'obtenir les estimations $(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ des paramètres $(\beta_0, \beta_1, ..., \beta_k)$.

2.2 Modèle logistique avec cinq variables explicatives

A la section 1, nous avons obtenu un modèle de régression linéaire constitué de cinq variables explicatives permettant de prédire la consommation moyenne. Nous avons décidé de reprendre ce modèle et de l'appliquer à la régression logistique. Notre variable à expliquer est ici la variable Trajet, qui prend deux valeurs : « normale » (codé 0) et « économique » (codé 1). Les variables explicatives sont les mêmes que précédemment : MeanPosAcc, TempsArrete, TempsFreinMoteur, PKE et SumFreinage. Les résultats obtenus par la procédure LOGISTIC du logiciel SAS sont donnés aux tableaux IX.4, IX.5, IX.6 et IX.7.

2.2.1 Construction du modèle et critères de qualité

Afin de tester l'influence globale des cinq variables explicatives, on utilise le test du rapport de vraisemblance qui teste l'hypothèse $H_0 : \beta_1 = ... = \beta_k = 0$ contre l'hypothèse H_1 : au moins un des $\beta_j \neq 0$. La statistique utilisée est la suivante :

$G = -2 \log \dfrac{L(constante)}{L(toutes les \text{ var } iables)}$ où L désigne la vraisemblance.

Sous l'hypothèse H_0, G suit une loi du khi-deux à k degrés de liberté (ici k=5).
Dans notre cas, G=39.36 (cf. Likelihood Ratio Tab. IX.4), ce qui correspond à un niveau de signification inférieur à 0.0001. On rejette donc l'hypothèse H_0, et on en déduit qu'au moins une des cinq variables explicatives est significative.

Les estimations des paramètres sont données au tableau IX.5. Le modèle est donc le suivant :

Logit[P [Trajet=Eco | MeanPosAcc, TempsArrete, TempsFreinMoteur, PKE, SumFreinage]]
= 38.579 – 1.432*MeanPosAcc – 0.855*TempsArrete + 0.138*TempsfreinMoteur
- 44.699*PKE – 0.201*SumFreinage.

Pour obtenir la probabilité estimée qu'un trajet soit « économique », il suffit alors d'utiliser la fonction inverse de la fonction logit : $F(x) = \dfrac{e^x}{1+e^x}$.

Le tableau IX.5 donne aussi la statistique de Wald qui permet de tester la significativité de chaque coefficient. On teste l'hypothèse $H_0 : \beta_j = 0$ contre $H_1 : \beta_j \neq 0$ à l'aide de la statistique : $W = \dfrac{\hat{\beta}_j}{S\hat{E}(\hat{\beta}_j)}$. Sous H_0, cette statistique suit une loi du khi-deux à un degré de liberté. Dans notre cas, seule la variable TempsFreinMoteur ne semble clairement pas significative (p-value=0.4992).

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	60.224	30.868
SC	61.962	41.294
-2 Log L	58.224	18.868

R-Square	0.6082	Max-rescaled R-Square	0.8110

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	39.3559	5	<.0001
Score	22.9698	5	0.0003
Wald	6.5314	5	0.2579

Tab. IX.4 – Résultats de la procédure LOGISTIC du logiciel SAS : critères de qualité du modèle à cinq variables explicatives.

Analyse des estimations de la vraisemblance maximum					
Paramètre	DF	Estimation	Erreur std	Khi 2 de Wald	Pr > Khi 2
Intercept	1	38.5787	17.9639	4.6121	0.0317
MeanPosAcc	1	-1.4317	0.7675	3.4800	0.0621
TempsArrete	1	-0.8550	0.4507	3.5995	0.0578
TempsFreinMoteur	1	0.1384	0.2048	0.4566	0.4992
PKE	1	-44.6990	23.2515	3.6957	0.0546
SumFreinage	1	-0.2011	0.0889	5.1137	0.0237

Tab. IX.5 – Résultats de la procédure LOGISTIC du logiciel SAS : estimation des paramètres et test de Wald pour le modèle à cinq variables explicatives.

2.2.2 Interprétation des coefficients : odds-ratio

L'odds-ratio, appelé aussi rapport de cotes, d'une variable explicative numérique mesure l'évolution du rapport des probabilités d'apparition de l'événement Y=1 contre Y=0, lorsque la variable explicative X_i passe de x à x+1. Ceci s'écrit :

$$OR = \frac{\pi(x+1)/[1-\pi(x+1)]}{\pi(x)/[1-\pi(x)]} = e^{\beta_i}$$

C'est le facteur par lequel la cote est multipliée lorsque la variable explicative X_i passe de x à x+1. L'odds-ratio d'une variable explicative est égal à l'exponentielle du coefficient associé.
Les odds-ratio de notre modèle sont donnés au tableau IX.6. Prenons par exemple, la variable MeanPosAcc. L'odds-ratio correspondant est égal à 0.239, ce qui signifie que si on augmente la variable MeanPosAcc de 1 unité (c'est-à-dire de 1%), alors la chance que le trajet soit « économique » baisse de 76.1%. Au contraire, si on augmente la variable TempsFreinMoteur de 1%, la chance que le trajet soit « économique » augmente de 14.8% (OR=1.148). Le tableau IX.6 montre que les chances d'être en éco-conduite baissent fortement dès que la variable PKE augmente.

Estimations des rapports de cotes			
Effet	**Point Estimate**	**95% Limites de confiance de Wald**	
MeanPosAcc	0.239	0.053	1.075
TempsArrete	0.425	0.176	1.029
TempsFreinMoteur	1.148	0.769	1.716
PKE	<0.001	<0.001	2.394
SumFreinage	0.818	0.687	0.974

Tab. IX.6 – Résultats de la procédure LOGISTIC du logiciel SAS : odds-ratios et intervalles de confiance à 95% pour le modèle à cinq variables explicatives.

2.2.3 Capacité prédictive du modèle

Afin d'évaluer la capacité prédictive du modèle, on compare la valeur observée Y_i avec la valeur prévue par le modèle $\hat{Y}_i$. On utilise le modèle logistique pour affecter un individu à l'une des classes définies par la variable réponse Y (« normale » ou « économique »). Si la probabilité calculée $\hat{\pi}_i$ est supérieure ou égale à un seuil s, on décide d'affecter l'individu i à la classe [Y=1] (« économique »), sinon il est affecté à la classe [Y=0] (« normale »). Le tableau IX.7 montre que, pour un seuil s=0.6, seulement 4 trajets ont été mal classés parmi les 42 trajets observés (1 trajets « normal » a été affecté à la classe « économique », et 3 trajets « économique » ont été affectés à la classe « normal »). Le tableau IX.8 est une table de classification pour différents niveaux de probabilité allant de 0.1 à 0.9 par pas de 0.1. Si nous choisissons 0.5 comme seuil, nous obtenons que 85.7% des individus sont classés correctement. En particulier, parmi ceux qui ont une probabilité supérieure à 0.5 d'être

« économique », 18 le sont et 3 ne le sont pas (Faux Positifs). Il en est de même, parmi ceux qui ont une probabilité inférieure à 0.5 d'être « économique » : 18 ne sont effectivement pas « économique » mais 3 le sont (Faux Négatifs).
Le tableau IX.8 contient deux indices importants : la sensibilité et la spécificité. La sensibilité est la proportion de vraies valeurs 1, qui sont prédites valoir 1. Lorsque le seuil s augmente, alors la sensibilité diminue puisque la règle devient plus exigeante pour l'affectation dans la classe [Y=1]. La spécificité est la proportion de vraies valeurs 0, qui sont prédites valoir 0. Lorsque le seuil s augmente, alors la spécificité augmente puisque la règle devient moins exigeante pour l'affectation dans la classe [Y=0]. Ces deux indices permettent de construire la courbe ROC obtenue en représentant la sensibilité en fonction de 1 moins la spécificité pour chaque valeur du seuil s. Une mesure du pouvoir discriminant de la régression logistique est obtenue à l'aide de l'aire sous la courbe ROC. Elle représente une estimation de la probabilité qu'une paire (i,i') avec $y_i = 1$ et $y_i = 0$, tirée au hasard, soit concordante, c'est-à-dire $\hat{\pi}_i > \hat{\pi}_{i'}$. Plus l'aire sous la courbe ROC est proche de 1, meilleure est la discrimination. La courbe ROC associée à notre modèle est représentée figure IX.2. L'aire sous la courbe ROC est égale à 0.968, ce qui montre que la discrimination est très bonne.

Table de Trajet par predy			
Trajet	**predy**		
Fréquence Pourcentage	**Normale**	**Economique**	**Total**
Normale	20 47.62	1 2.38	21 50.00
Economique	3 7.14	18 42.86	21 50.00
Total	23 54.76	19 45.24	42 100.00

Tab. IX.7 – Proportion de bien classés et de mal classés parmi les 42 trajets pour un seuil s=0.6.

Table de classification									
	Correct		Incorrect		Pourcentages				
Niveau de prob.	Événement	Non-événement	Événement	Non-événement	Correct	Sensi-bilité	Spéci-ficité	POS fausse	NEG fausse
0.100	19	13	8	2	76.2	90.5	61.9	29.6	13.3
0.200	19	14	7	2	78.6	90.5	66.7	26.9	12.5
0.300	18	14	7	3	76.2	85.7	66.7	28.0	17.6
0.400	18	16	5	3	81.0	85.7	76.2	21.7	15.8
0.500	18	18	3	3	85.7	85.7	85.7	14.3	14.3
0.600	16	18	3	5	81.0	76.2	85.7	15.8	21.7
0.700	15	18	3	6	78.6	71.4	85.7	16.7	25.0
0.800	14	19	2	7	78.6	66.7	90.5	12.5	26.9
0.900	13	20	1	8	78.6	61.9	95.2	7.1	28.6

Tab. IX.8 – Résultats de la procédure LOGISTIC du logiciel SAS : Table de classification pour différents niveaux de probabilité.

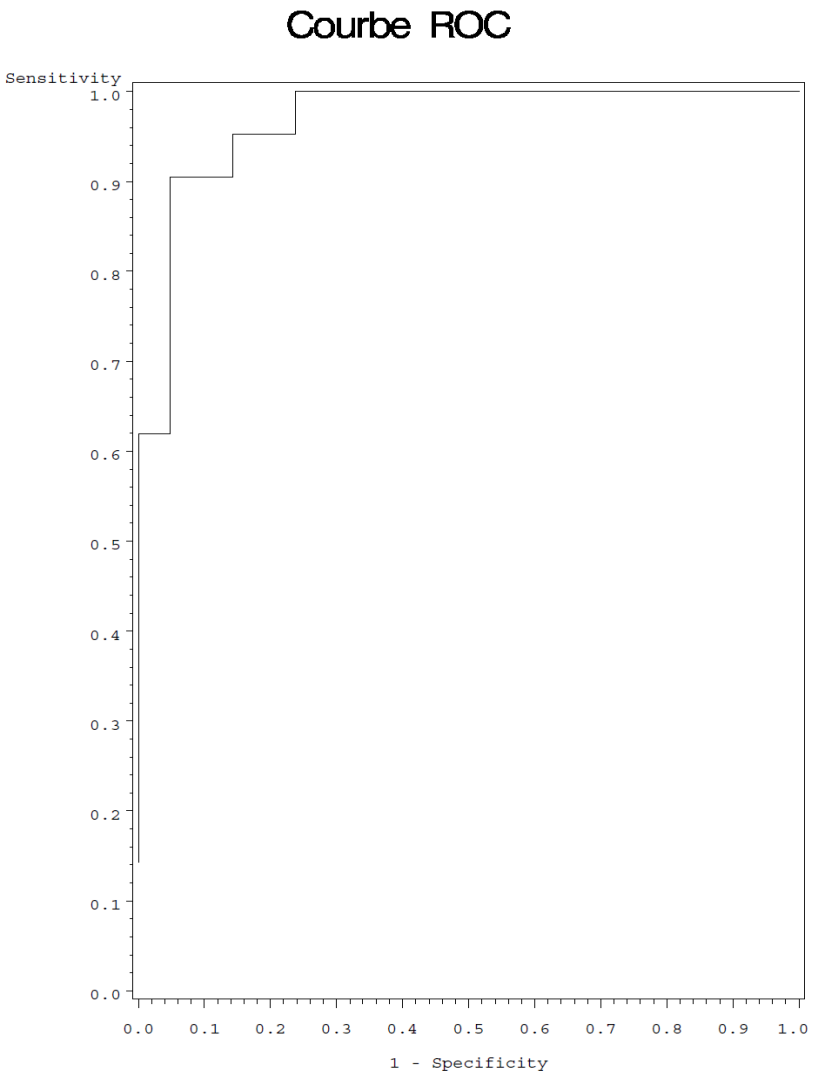


Fig. IX.2 – Courbe ROC pour le modèle avec cinq variables explicatives.

2.3 Modèle logistique avec une variable explicative

Afin de simplifier le modèle, nous avons effectué une régression logistique par la méthode stepwise avec les dix variables sélectionnées par la classification ascendante hiérarchique (cf. Tab. VIII.2). Les différentes étapes de la sélection séquentielle sont résumées dans le tableau IX.9. On observe que seule la variable PKE a été conservée dans le modèle.

Récapitulatif sur la sélection séquentielle								
	Effet							
Étape	Saisi	Supprimé	DF	Nombre dans	Khi 2 du score	Khi 2 de Wald	Pr > Khi 2	Libellé de variable
1	PKE		1	1	19.4701		<.0001	PKE
2	SumFreinage		1	2	4.3812		0.0363	Nombre de freinages
3		SumFreinage	1	1		3.5618	0.0591	Nombre de freinages

Tab. IX.9 – Récapitulatif des différentes étapes de la régression logistique avec la méthode stepwise.

Les différents critères de qualité du modèle sont donnés Tab. IX.10. La statistique associée au test du rapport de vraisemblance est G=26.27, ce qui correspond à un niveau de signification inférieur à 0.0001. On en déduit donc que la variable PKE est significative.

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	60.224	35.953
SC	61.962	39.428
-2 Log L	58.224	31.953

R-Square	0.4650	Max-rescaled R-Square	0.6200

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	26.2712	1	<.0001
Score	19.4701	1	<.0001
Wald	10.8287	1	0.0010

Tab. IX.10 – Résultats de la procédure LOGISTIC du logiciel SAS : critères de qualité du modèle à une variable explicative.

Les estimations des paramètres sont données au tableau IX.11. Le modèle est le suivant :

Logit[P [Trajet=Eco | PKE]] = 9.91 – 34.558*PKE.

Le test de Wald confirme la significativité de la variable PKE (p-value = 0.001).

Analyse des estimations de la vraisemblance maximum					
Paramètre	**DF**	**Estimation**	**Erreur std**	**Khi 2 de Wald**	**Pr > Khi 2**
Intercept	1	9.9097	2.9862	11.0123	0.0009
PKE	1	-34.5579	10.5017	10.8287	0.0010

Tab. IX.11 – Résultats de la procédure LOGISTIC du logiciel SAS : estimation des paramètres et test de Wald pour le modèle à une variable explicative.

L'odds-ratio de la variable PKE étant inférieur à 0.001 (cf. tab. IX.12), on en déduit un très fort impact de la variable PKE sur la probabilité d'être en situation d'éco-conduite : dès que la variable PKE augmente, les chances d'être en éco-conduite baissent fortement.

Estimations des rapports de cotes			
Effet	**Point Estimate**	**95% Limites de confiance de Wald**	
PKE	<0.001	<0.001	<0.001

Tab. IX.12 – Résultats de la procédure LOGISTIC du logiciel SAS : odds-ratios et intervalles de confiance à 95% pour le modèle à une variable explicative.

La figure IX.3 montre l'évolution de la probabilité estimée d'être en éco-conduite en fonction de la variable PKE. On observe que la probabilité P(Y=1 | X=*x*) diminue lorsque *x* augmente selon une courbe en S. Plus précisément, la probabilité d'être en éco-conduite est inférieure à 0.5 dès que PKE est supérieure à 0.3. La figure IX.3 représente aussi les valeurs observées (on rappelle que 1= « économique » et 0= « normal »), ainsi que les valeurs prédites pour un seuil s=0.6. Le tableau IX.13 montre que, pour ce seuil, seulement 7 trajets sont mal classés parmi les 42 trajets observés (3 trajets « normal » ont été affectés à la classe « économique », et 4 trajets « économique » ont été affectés à la classe « normal »).

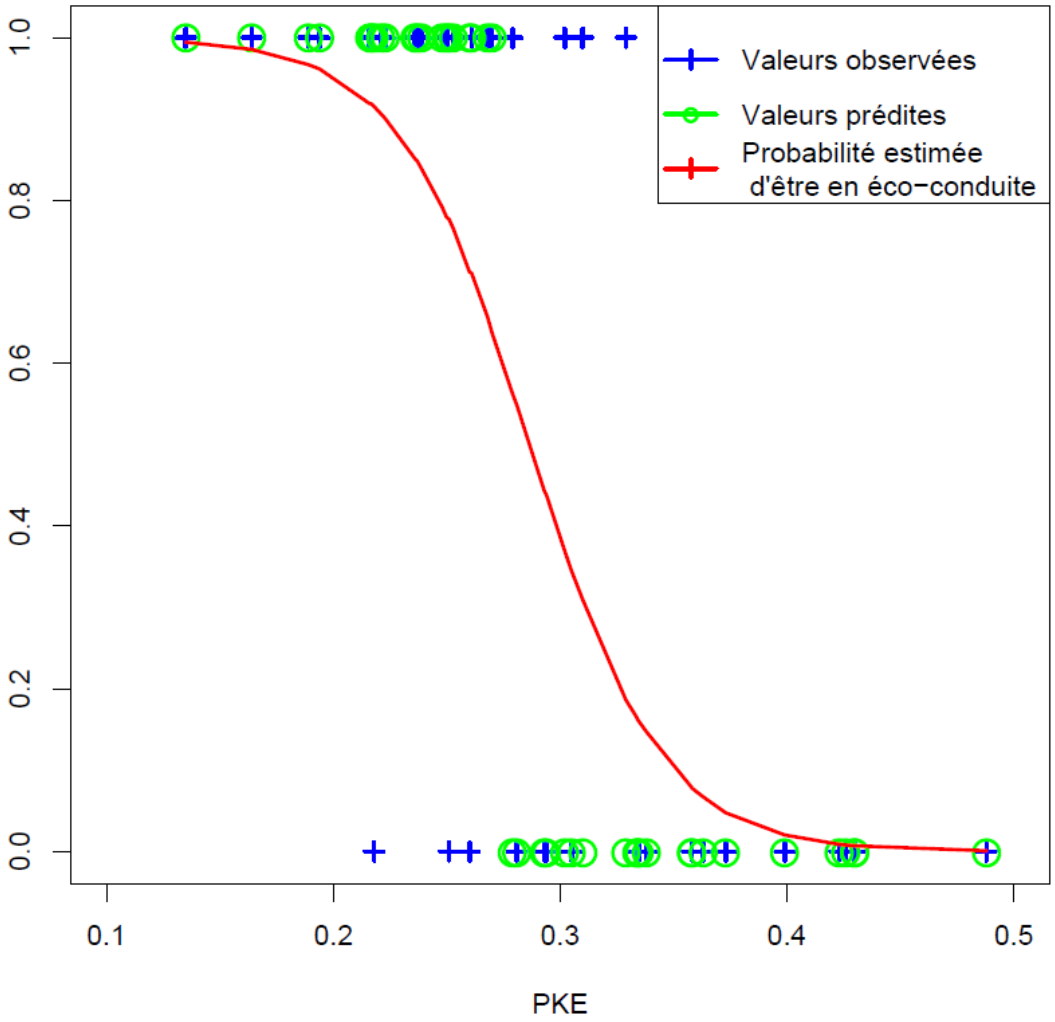


Fig. IX.3 – Graphique représentant les valeurs observées, les valeurs prédites pour un seuil s=0.6, ainsi que la probabilité estimée d'être en éco-conduite en fonction de la variable PKE.

Table de Trajet par predy			
Trajet	**predy**		
Fréquence Pourcentage	**Normale**	**Economique**	**Total**
Normale	18 42.86	3 7.14	21 50.00
Economique	4 9.52	17 40.48	21 50.00
Total	22 52.38	20 47.62	42 100.0 0

Tab. IX.13 – Proportion de bien classés et de mal classés parmi les 42 trajets pour un seuil s=0.6.

Le tableau IX.14 donne la table de classification pour différents niveaux de probabilité allant de 0.1 à 0.9 par pas de 0.1. Si nous choisissons 0.5 comme seuil, nous obtenons que 83.3% des individus sont classés correctement. En particulier, parmi ceux qui ont une probabilité supérieure à 0.5 d'être « économique », 18 le sont et 4 ne le sont pas (Faux Positifs). Parmi ceux qui ont une probabilité inférieure à 0.5 d'être « économique », 17 ne sont effectivement pas « économique » mais 3 le sont (Faux Négatifs). La courbe ROC est représentée figure IX.4. L'aire sous la courbe ROC est égale 0.907, ce qui montre que la discrimination reste excellente même si elle est un peu moins bonne qu'avec le modèle à cinq variables explicatives présenté à la section 2.2.

Table de classification									
Niveau de prob.	Correct		Incorrect		Pourcentages				
	Événement	Non-événement	Événement	Non-événement	Correct	Sensi-bilité	Spéci-ficité	POS fausse	NEG fausse
0.100	21	10	11	0	73.8	100.0	47.6	34.4	0.0
0.200	20	14	7	1	81.0	95.2	66.7	25.9	6.7
0.300	19	14	7	2	78.6	90.5	66.7	26.9	12.5
0.400	18	15	6	3	78.6	85.7	71.4	25.0	16.7
0.500	18	17	4	3	83.3	85.7	81.0	18.2	15.0
0.600	17	18	3	4	83.3	81.0	85.7	15.0	18.2
0.700	14	18	3	7	76.2	66.7	85.7	17.6	28.0
0.800	11	19	2	10	71.4	52.4	90.5	15.4	34.5
0.900	7	20	1	14	64.3	33.3	95.2	12.5	41.2

Tab. IX.14 – Résultats de la procédure LOGISTIC du logiciel SAS : Table de classification pour différents niveaux de probabilité.

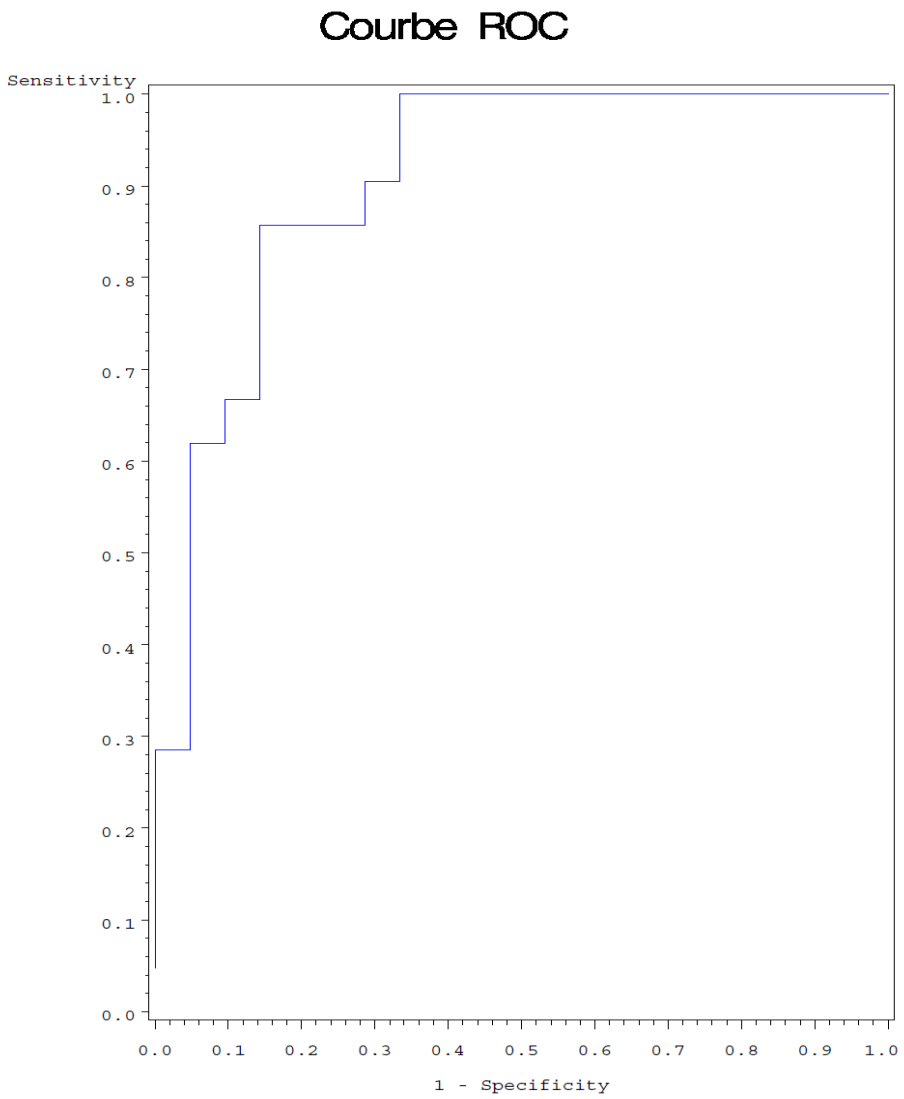


Fig. IX.4 – Courbe ROC pour le modèle avec une variable explicative.

Conclusion

Les résultats de la régression linéaire sur les 42 trajets de notre base de données nous ont permis de construire un bon modèle de prédiction de la consommation moyenne dépendant de cinq variables explicatives : la moyenne du pourcentage d'enfoncement de la pédale d'accélérateur, le pourcentage de temps passé à l'arrêt, le pourcentage de temps passé en frein moteur, le nombre de freinages, et la variable PKE. Nous avons ensuite obtenu, par régression logistique, deux modèles de prédiction de la probabilité d'être en situation d'éco-conduite : l'un comportant les cinq variables explicatives citées précédemment, et l'autre constitué uniquement de la variable PKE. Ces deux modèles présentent un fort pouvoir prédictif. Le premier modèle est meilleur, mais le deuxième présente l'avantage d'être beaucoup plus simple puisque ne dépendant que d'une seule variable. Notons enfin que ces résultats permettent de montrer que PKE est un indicateur très pertinent pour caractériser l'éco-conduite. De plus, cet indicateur est très facile à mesurer puisqu'il ne dépend que de la vitesse.

Chapitre X

Conclusion générale

Ce travail nous a permis de confirmer les avantages de l'éco-conduite en termes de réduction de consommation, et d'analyser la mise en pratique de principes simples destinés à économiser du carburant. Cette expérience a montré qu'une simple formation théorique permettait d'améliorer significativement les performances des conducteurs en termes de gain de consommation, mais il semble que ces résultats pourraient être encore améliorés par une formation pratique car certains principes comme l'utilisation du frein moteur sont encore peu familiers des conducteurs. Les techniques d'éco-conduite pourraient ainsi être intégrées dans la formation et l'examen des conducteurs, comme c'est déjà le cas dans certains « pays experts » [1], afin d'inculquer ces techniques aux apprentis-conducteurs.

Nous avons ensuite déterminé dix indicateurs permettant de discriminer les trajets effectués en conduite normale, et ceux effectués avec une conduite dite économique. Puis nous avons construit un modèle de prédiction de la consommation moyenne basé sur cinq indicateurs très caractéristiques du style de conduite. Nous avons aussi construit deux modèles de prédiction de la probabilité d'être en situation d'éco-conduite, dont l'un ne dépendant que d'une seule variable : Positive Kinetic Energy (PKE). Cet indicateur est très pertinent pour caractériser l'éco-conduite et présente l'avantage de ne dépendre que de la vitesse. La mise au point de ces deux modèles de régression logistique permettrait ainsi de créer un indice d'éco-conduite, calculé à partir des informations relatives aux parcours d'un conducteur, et que celui-ci pourrait chercher à améliorer afin d'économiser sa consommation de carburant. Fiat a d'ailleurs mis en place un système Eco Drive sur la Grande Punto et sur la 500, permettant à ses clients de calculer un « éco index » traduisant leur comportement de conduite [18].

Les utilisations potentielles de la mise au point de ces modèles de prédiction sont diverses. On pourrait envisager la réalisation d'un système d'aide à l'éco-conduite : le conducteur pourrait choisir la consommation moyenne souhaitée en l/100km pour son parcours, et plus la consommation choisie serait faible, plus la marge de manœuvre du véhicule serait limitée (pas d'accélérations brusques sauf en cas d'urgence : système de kickdown). Un système permettant de choisir son mode de conduite (confort ou sportif) et agissant pour modifier le point de changement pour les rapports de boîte a d'ailleurs été mis en place par Audi.
Une autre perspective serait d'analyser différentes portions d'un trajet afin de déterminer les zones non favorables à l'éco-conduite, et ainsi d'adapter l'infrastructure de la route à ce nouveau style de conduite. En effet, depuis une dizaine d'années les véhicules ont été modifiés afin de réduire la consommation de carburant, mais on peut se demander si l'infrastructure est réellement adaptée à l'éco-conduite. Notre étude a ainsi montré que le pourcentage de temps passé au-delà de la vitesse de consigne reste élevé en éco-conduite

(25% en moyenne), bien que la conduite soit plus souple. Les consignes de limitation de vitesse ne semblent donc pas adaptées à une conduite basée sur une anticipation des freinages et le maintien d'une vitesse constante.

Enfin un dernier point n'a pas pu être traité dans cette étude : le problème de la sécurité. En effet, si l'éco-conduite présente de nombreux avantages, certains aspects sécuritaires restent encore à étudier. Nous envisageons ainsi par la suite d'exploiter les données issues du radar et de la vidéo des pieds du conducteur afin de déterminer certains indicateurs de risque. En effet, le radar permettrait de calculer les distances inter-véhiculaires, et la vidéo des pieds du conducteur permettrait d'analyser les freinages à risque (pied non situé sur la pédale de frein juste avant le freinage).

Bibliographie

[1] CIECA. (2007). *Projet interne de la CIECA concernant l'intégration de l'éco-conduite dans la formation et l'examen des conducteurs de catégorie B.*

[2] Confais J., Le Guen M. (2006). *Premier pas en régression linéaire avec SAS.* Revue Modulad, Numéro 35, pages 220-363.

[3] Droesbeke J.-J., Lejeune M. et Saporta G. (2005). *Modèles statistiques pour données qualitatives.* Editions Technip.

[4] Ericsson E. (2001). *Independent driving pattern factors and their influence on fuel-use and exhaust emission factors.* Transportation research part D 6, 325-345.

[5] Johansson H., Färnlund J. et Engström C. (1999). *Impact of EcoDriving on emissions and fuel consumption.* Vägverket publication 1999:165E.

[6] Preud'homme E. (2006). *Cours de logiciel SAS, Tome 1, 2 et 3.* IUT 2 Grenoble, Département STID.

[7] Saint Pierre G. (2008). *Le système LAVIA et la consommation de carburant.* Rapport technique.

[8] Taffé P. (2004). *Cours de régression logistique apliquée.*

[9] Tenenhaus M. (2007). *Statistique : méthodes pour décrire, expliquer et prévoir.* Dunod.

[10] Tufféry S. (2007). *Data mining et statistique décisionnelle.* Editions Technip.

Sites internet :

[11] Projet Ecodrive de l'UE :
www.ecodrive.org

[12] Quality Alliance Eco-drive :
www.ecodrive.ch

[13] Projet Treatise de l'UE :
http://treatise.eu.com

[14] Maison de la sécurité routière des Yvelines :
www.securite-routiere78.fr

[15] Site consacré à l'éco-conduite :
www.ecoconduite.org

[16] Maîtrise des consommations de carburant dans le transport routier :
www.energeco.org

[17] Site de l'ADEME :
www.ademe.fr

[18] Système Eco Drive de Fiat :
http://www.fiat.fr/cgi-bin/pbrand.dll/FIAT_FRANCE/blueme/blueme.jsp

Annexe A

Tableau des indicateurs

Nom de la variable	Type	Description
MeanVitesse, StdVitesse, MaxVitesse	Réel	Moyenne, Ecart-type et Maximum de la vitesse instantanée
MeanAccel, StdAccel, MaxAccel, MinAccel	Réel	Moyenne, Ecart-type, Maximum et Minimum de l'accélération du véhicule
MeanRegime, StdRegime, MaxRegime	Réel	Moyenne, Ecart-type, Maximum et Minimum du régime moteur
MeanDepassement, StdDepassement	Réel	Moyenne et Ecart-type du dépassement de vitesse
MeanPosAcc, StdPosAcc, MaxPosAcc	Réel	Moyenne, Ecart-type et Maximum du pourcentage d'enfoncement de la pédale d'accélérateur
MeanRapBoite, StdRapBoite	Réel	Moyenne et Ecart-type du rapport de boîte
MeanAccelerationLaterale, StdAccelerationLaterale, MaxAccelerationLaterale, MinAccelerationLaterale	Réel	Moyenne, Ecart-type, Maximum et Minimum de l'accélération latérale
SumChangeVitesse	Entier	Nombre de changement de rapports de boîte
Distance	Réel	Distance parcourue (en km)
DureeMinutes	Horaire	Durée du trajet au format HH:MM:SS
ConsoBusCan	Réel	Consommation cumulée mesurée par le bus CAN en litres
ConsoDebitmetre	Réel	Consommation cumulée mesurée par le débitmètre en litres
ConsoCanL100	Réel	Consommation cumulée mesurée par le bus CAN en litres / 100 km
ConsoDebitL100	Réel	Consommation cumulée mesurée par le débitmètre en litres / 100 km
TempsArrete	Réel	Pourcentage de temps de trajet passé à l'arrêt
TempsHorsConsigne	Réel	Pourcentage de temps de trajet passé au-delà de la vitesse de consigne
TempsFreinMoteur	Réel	Pourcentage de temps de trajet passé en frein moteur
DistanceFreinMoteur	Réel	Pourcentage de distance passé en frein moteur
TempsPointMort	Réel	Pourcentage de temps de trajet passé au point mort
DistancePointMort	Réel	Pourcentage de distance passé au point mort

TempsVitesseNonOptimale	Réel	Pourcentage de temps de trajet passé avec une vitesse non-optimale (Régime moteur > 3000 tr/min et pression exercée sur la pédale d'accélérateur)
MeanRMChangeVitesse	Réel	Moyenne du régime moteur auquel se fait le changement de vitesse
SumArret	Entier	Nombre d'arrêts
ArretParKm	Réel	Nombre d'arrêts par km
MeanDureeArret	Horaire	Durée moyenne des arrêts
KmParLitre	Réel	Nombre de km réalisés par litre
SumDebraye	Entier	Nombre de débrayages
SumFreinage	Entier	Nombre de freinages
DistanceFreinageVit1	Réel	Distance parcourue en freinant
DistanceFreinagePercent	Réel	Pourcentage de distance parcourue en freinant
RPA	Réel	Relative Positive Acceleration $\frac{1}{x}\int va^{+}dt$ avec x=durée totale du trajet
PKE	Réel	Positive Kinetic Energy $\sum \left(v_f^2 - v_s^2\right) x$ when dv/dt>0, v_f =final speed, v_s =start speed, x=distance
IntegraleCarreAcc	Réel	Intégrale du carré de l'accélération $\frac{1}{n}\int a^2 dt$ avec n=nbre d'intervalles de temps
Indicateurs liés à chaque plage de régime moteur		
Regime_PCT_C1	Réel	Pourcentage de temps de trajet passé avec un régime moteur inférieur à 500 tr/min
Li_Total_C1	Réel	Volume total de carburant consommé (en l) avec un régime moteur inférieur à 500 tr/min
Li_Moyen_C1	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur inférieur à 500 tr/min
Li_PCT_C1	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur inférieur à 500 tr/min
Regime_PCT_C2	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 500 et 1000 tr/min
Li_Total_C2	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 500 et 1000 tr/min
Li_Moyen_C2	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 500 et 1000 tr/min
Li_PCT_C2	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur compris entre 500 et 1000 tr/min
Regime_PCT_C3	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 1000 et 1500 tr/min
Li_Total_C3	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 1000 et 1500 tr/min
Li_Moyen_C3	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 1000 et 1500 tr/min
Li_PCT_C3	Réel	Pourcentage de carburant consommé sur tout le trajet

		avec un régime moteur compris entre 1000 et 1500 tr/min
Regime_PCT_C4	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 1500 et 2000 tr/min
Li_Total_C4	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 1500 et 2000 tr/min
Li_Moyen_C4	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 1500 et 2000 tr/min
Li_PCT_C4	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur compris entre 1500 et 2000 tr/min
Regime_PCT_C5	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 2000 et 2500 tr/min
Li_Total_C5	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 2000 et 2500 tr/min
Li_Moyen_C5	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 2000 et 2500 tr/min
Li_PCT_C5	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur compris entre 2000 et 2500 tr/min
Regime_PCT_C6	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 2500 et 3000 tr/min
Li_Total_C6	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 2500 et 3000 tr/min
Li_Moyen_C6	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 2500 et 3000 tr/min
Li_PCT_C6	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur compris entre 2500 et 3000 tr/min
Regime_PCT_C7	Réel	Pourcentage de temps de trajet passé avec un régime moteur compris entre 3000 et 3500 tr/min
Li_Total_C7	Réel	Volume total de carburant consommé (en l) avec un régime moteur compris entre 3000 et 3500 tr/min
Li_Moyen_C7	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur compris entre 3000 et 3500 tr/min
Li_PCT_C7	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur compris entre 3000 et 3500 tr/min
Regime_PCT_C8	Réel	Pourcentage de temps de trajet passé avec un régime moteur supérieur à 3500 tr/min
Li_Total_C8	Réel	Volume total de carburant consommé (en l) avec un régime moteur supérieur à 3500 tr/min
Li_Moyen_C8	Réel	Volume moyen de carburant consommé (en l) avec un régime moteur supérieur à 3500 tr/min
Li_PCT_C8	Réel	Pourcentage de carburant consommé sur tout le trajet avec un régime moteur supérieur à 3500 tr/min

Pourcentage de temps passé dans chaque plage de vitesse		
TempsVInstantanee2	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 0 et 15 km/h
TempsVInstantanee3	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 15 et 30 km/h
TempsVInstantanee4	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 30 et 50 km/h
TempsVInstantanee5	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 50 et 70 km/h
TempsVInstantanee6	Réel	Pourcentage de temps de trajet passé avec une vitesse comprise entre 70 et 90 km/h
TempsVInstantanee7	Réel	Pourcentage de temps de trajet passé avec une vitesse supérieure à 90 km/h
Pourcentage de temps passé dans chaque plage d'accélération/décélération		
TempsAcceleration9	Réel	Pourcentage de temps de trajet passé avec décélération comprise entre -10 et -2.5 m/s²
TempsAcceleration10	Réel	Pourcentage de temps de trajet passé avec décélération comprise entre -2.5 et -1.5 m/s²
TempsAcceleration11	Réel	Pourcentage de temps de trajet passé avec décélération comprise entre -1.5 et -1 m/s²
TempsAcceleration12	Réel	Pourcentage de temps de trajet passé avec décélération comprise entre -1 et -0.5 m/s²
TempsAcceleration13	Réel	Pourcentage de temps de trajet passé avec décélération comprise entre -0.5 et 0 m/s²
TempsAcceleration14	Réel	Pourcentage de temps de trajet passé avec accélération comprise entre 0 et 0.5 m/s²
TempsAcceleration15	Réel	Pourcentage de temps de trajet passé avec accélération comprise entre 0.5 et 1 m/s²
TempsAcceleration16	Réel	Pourcentage de temps de trajet passé avec accélération comprise entre 1 et 1.5 m/s²
TempsAcceleration17	Réel	Pourcentage de temps de trajet passé avec accélération comprise entre 1.5 et 2.5 m/s²
TempsAcceleration18	Réel	Pourcentage de temps de trajet passé avec accélération comprise entre 2.5 et 10 m/s²
Pourcentage de temps passé dans chaque plage de la variable (Vitesse*Accélération)		
TempsVAcc19	Réel	Pourcentage de temps de trajet passé avec (va)<0
TempsVAcc20	Réel	Pourcentage de temps de trajet passé avec (va) compris entre 0 et 3
TempsVAcc21	Réel	Pourcentage de temps de trajet passé avec (va) compris entre 3 et 6
TempsVAcc22	Réel	Pourcentage de temps de trajet passé avec (va) compris entre 6 et 10
TempsVAcc23	Réel	Pourcentage de temps de trajet passé avec (va) compris entre 10 et 15
TempsVAcc24	Réel	Pourcentage de temps de trajet passé avec (va) supérieure à 15

Distance parcourue en freinant pour chaque plage de vitesse		
DistanceFreinageVit2	Réel	Distance parcourue en freinant avec une vitesse inférieure à 20 km/h
DistanceFreinageVit3	Réel	Distance parcourue en freinant avec une vitesse comprise entre 20 et 50 km/h
DistanceFreinageVit4	Réel	Distance parcourue en freinant avec une vitesse comprise entre 50 et 90 km/h
DistanceFreinageVit5	Réel	Distance parcourue en freinant avec une vitesse supérieure à 90 km/h
Pourcentage de temps passé dans chaque rapport de boîte		
TempsRapBoite26	Réel	Pourcentage de temps passé en vitesse 1
TempsRapBoite27	Réel	Pourcentage de temps passé en vitesse 2
TempsRapBoite28	Réel	Pourcentage de temps passé en vitesse 3
TempsRapBoite29	Réel	Pourcentage de temps passé en vitesse 4
TempsRapBoite30	Réel	Pourcentage de temps passé en vitesse 5

Annexe B

Tableau des corrélations pour les variables de chaque groupe obtenu par la classification hiérarchique ascendante

Table des corrélations pour les variables du premier groupe

	MeanVitesse	MeanPosAcc	stdPosAcc	TempsHorsConsigne
MeanVitesse	1.000	0.886	0.693	0.931
MeanPosAcc	0.886	1.000	0.864	0.910
stdPosAcc	0.693	0.864	1.000	0.731
TempsHorsConsigne	0.931	0.910	0.731	1.000
TempsVAcc21	-0.741	-0.811	-0.744	-0.793
MaxVitesse	0.718	0.768	0.681	0.814
MaxPosAcc	0.539	0.737	0.842	0.614
TempsVit100	0.819	0.859	0.735	0.932

	TempsVAcc21	MaxVitesse	MaxPosAcc	TempsVit100
MeanVitesse	-0.741	0.718	0.539	0.819
MeanPosAcc	-0.811	0.768	0.737	0.859
stdPosAcc	-0.744	0.681	0.842	0.735
TempsHorsConsigne	-0.793	0.814	0.614	0.932
TempsVAcc21	1.000	-0.668	-0.658	-0.739
MaxVitesse	-0.668	1.000	0.535	0.890
MaxPosAcc	-0.658	0.535	1.000	0.569
TempsVit100	-0.739	0.890	0.569	1.000

Table des corrélations pour les variables du troisième groupe

	MeanAccel	StdAccel	TempsVAcc22	TempsVAcc23	TempsVAcc24	PKE	RPA
MeanAccel	1.000	-0.422	0.212	0.409	-0.389	-0.375	-0.339
StdAccel	-0.422	1.000	-0.677	-0.886	0.921	0.951	0.967
TempsVAcc22	0.212	-0.677	1.000	0.807	-0.644	-0.789	-0.656
TempsVAcc23	0.409	-0.886	0.807	1.000	-0.817	-0.927	-0.870
TempsVAcc24	-0.389	0.921	-0.644	-0.817	1.000	0.878	0.930
PKE	-0.375	0.951	-0.789	-0.927	0.878	1.000	0.928
RPA	-0.339	0.967	-0.656	-0.870	0.930	0.928	1.000
IntegraleCarreAcc	-0.420	0.994	-0.659	-0.872	0.909	0.941	0.969
TempsVit70	0.331	-0.736	0.414	0.706	-0.584	-0.660	-0.684
TempsAccNeg	-0.465	0.950	-0.582	-0.813	0.900	0.893	0.902
TempsAccMed	0.424	-0.980	0.702	0.901	-0.895	-0.963	-0.949
TempsAccPos	-0.375	0.956	-0.757	-0.923	0.850	0.969	0.939

	IntegraleCarreAcc	TempsVit70	TempsAccNeg	TempsAccMed	TempsAccPos
MeanAccel	-0.420	0.331	-0.465	0.424	-0.375
StdAccel	0.994	-0.736	0.950	-0.980	0.956
TempsVAcc22	-0.659	0.414	-0.582	0.702	-0.757
TempsVAcc23	-0.872	0.706	-0.813	0.901	-0.923
TempsVAcc24	0.909	-0.584	0.900	-0.895	0.850
PKE	0.941	-0.660	0.893	-0.963	0.969
RPA	0.969	-0.684	0.902	-0.949	0.939
IntegraleCarreAcc	1.000	-0.750	0.956	-0.981	0.955
TempsVit70	-0.750	1.000	-0.741	0.730	-0.688
TempsAccNeg	0.956	-0.741	1.000	-0.963	0.892
TempsAccMed	-0.981	0.730	-0.963	1.000	-0.981
TempsAccPos	0.955	-0.688	0.892	-0.981	1.000

Table des corrélations pour les variables du quatrième groupe

	SumFreinage	DistanceFreinageVit1
SumFreinage	1.000	0.765
DistanceFreinageVit1	0.765	1.000

Table des corrélations pour les variables du cinquième et du sixième groupe

	Regime_PCT_C4	TempsRapBoite27	TempsRapBoite28	TempsRapBoite30
Regime_PCT_C4	1.000	-0.608	-0.700	0.772
TempsRapBoite27	-0.608	1.000	0.520	-0.683
TempsRapBoite28	-0.700	0.520	1.000	-0.841
TempsRapBoite30	0.772	-0.683	-0.841	1.000
Regime_PCT_C3	0.694	-0.347	-0.593	0.626
Regime_PCT_C5	-0.641	0.309	0.449	-0.529
TempsRapBoite29	-0.281	0.147	0.147	-0.555

	Regime_PCT_C3	Regime_PCT_C5	TempsRapBoite29
Regime_PCT_C4	0.694	-0.641	-0.281
TempsRapBoite27	-0.347	0.309	0.147
TempsRapBoite28	-0.593	0.449	0.147
TempsRapBoite30	0.626	-0.529	-0.555
Regime_PCT_C3	1.000	-0.805	-0.337
Regime_PCT_C5	-0.805	1.000	0.351
TempsRapBoite29	-0.337	0.351	1.000

Table des corrélations pour les variables du septième groupe

	MeanRegime	StdVitesse	stdRegime	Regime_PCT_C6	Regime_PCT_C7
MeanRegime	1.000	0.409	0.644	0.913	0.808
StdVitesse	0.409	1.000	0.698	0.413	0.512
stdRegime	0.644	0.698	1.000	0.642	0.672
Regime_PCT_C6	0.913	0.413	0.642	1.000	0.905
Regime_PCT_C7	0.808	0.512	0.672	0.905	1.000
Regime_PCT_C8	0.492	0.355	0.568	0.407	0.550
TempsVitesseNonOptimale	0.738	0.514	0.671	0.833	0.943
MeanRMChangeVitesse	0.856	0.448	0.596	0.725	0.686
MaxRegime	0.659	0.475	0.783	0.565	0.608

	Regime_PCT_C8	TempsVitesse NonOptimale	MeanRMChange Vitesse	MaxRegime
MeanRegime	0.492	0.738	0.856	0.659
StdVitesse	0.355	0.514	0.448	0.475
stdRegime	0.568	0.671	0.596	0.783
Regime_PCT_C6	0.407	0.833	0.725	0.565
Regime_PCT_C7	0.550	0.943	0.686	0.608
Regime_PCT_C8	1.000	0.568	0.620	0.648
TempsVitesseNonOptimale	0.568	1.000	0.577	0.518
MeanRMChangeVitesse	0.620	0.577	1.000	0.762
MaxRegime	0.648	0.518	0.762	1.000

Table des corrélations pour les variables du huitième groupe

	TempsArrete	Regime_PCT_C2	TempsVAcc20	TempsPointMort
TempsArrete	1.000	0.853	0.815	0.980
Regime_PCT_C2	0.853	1.000	0.748	0.867
TempsVAcc20	0.815	0.748	1.000	0.793
TempsPointMort	0.980	0.867	0.793	1.000
SumArret	0.641	0.577	0.513	0.696
MeanDureeArret	0.760	0.767	0.738	0.727
TempsVit30	0.686	0.821	0.589	0.739

	SumArret	MeanDureeArret	TempsVit30
TempsArrete	0.641	0.760	0.686
Regime_PCT_C2	0.577	0.767	0.821
TempsVAcc20	0.513	0.738	0.589
TempsPointMort	0.696	0.727	0.739
SumArret	1.000	0.272	0.437
MeanDureeArret	0.272	1.000	0.655
TempsVit30	0.437	0.655	1.000

Table des corrélations pour les variables du neuvième groupe

	TempsVAcc19	TempsFreinMoteur
TempsVAcc19	1.000	0.517
TempsFreinMoteur	0.517	1.000