Characterization of drivers' energetic efficiency
Identification and evaluation of driving parameters related to energy efficiency

Master's thesis in Applied Mechanics

MATHIEU MAISONNEUVE

Department of Applied Mechanics
Division of Vehicle Safety and the Accident Prevention Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2013

Master's thesis 2013:17

MASTER'S THESIS IN APPLIED MECHANICS

Characterization of drivers' energetic efficiency

Identification and evaluation of driving parameters related to energy efficiency

MATHIEU MAISONNEUVE

Department of Applied Mechanics
Division of Vehicle Safety and the Accident Prevention Group
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2013

Characterization of drivers' energetic efficiency
Identification and evaluation of driving parameters related to energy efficiency

MATHIEU MAISONNEUVE

Master's thesis 2013:17
ISSN 1652-8557
Department of Applied Mechanics
Division of Vehicle Safety and the Accident Prevention Group
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
Volvo FH16: reducing fuel consumption has been one of the main objectives when designing the new Volvo range. However, to show its full potential, an efficient truck needs an efficient driver.

Chalmers Reproservice
Göteborg, Sweden 2013

Characterization of drivers' energetic efficiency
Identification and evaluation of driving parameters related to energy efficiency

Master's thesis in Applied Mechanics
MATHIEU MAISONNEUVE
Department of Applied Mechanics
Division of Vehicle Safety and the Accident Prevention Group
Chalmers University of Technology

Abstract

Fuel efficiency has become a very important objective in the automotive industry, and especially in the truck industry. While the price of fuel is increasing, the need to ensure timely deliveries of goods is still high. While engineers are working hard to improve the global powertrain efficiency, a lot can be done by changing drivers' behaviour. This approach is called eco-driving and refers to the driver's possibility of using the truck's kinetic energy in an efficient way, e.g. by accelerating softer or anticipating traffic to plan the driving

A state-of-the-art review in terms of current eco-driving system is included, covering systems developed for different hardware platforms and using different technologies. Then, this study identifies driving parameters relevant for eco-driving, based on previous research and using a dataset collected on-road in dedicated experiments. The aim for this part has been to establish correlations between efficient driving and drivers' behaviour.

Further, drivers' behaviour with respect to different road events, such as different curve radii, roundabouts or road altitude variation, is assessed in relation to driving-efficiency. Finally, drivers are ranked with respect to driving-efficiency using a grading system based on relevant driving parameters.

The results show that the ability to limit the speed variations was the most important for driving-efficiency, as expected, but also the variations of angle on both throttle and brake pedals were identified as relevant. This work can be used as a platform for application of similar methods to larger sets of data and preferably using naturalistically collected driving.

Keywords: Fuel efficiency, Truck industry, Drivers' behaviour, Eco driving, Correlation, Relevant driving parameters, Road events, Grading system

Acknowledgements

My first and major thanks go to my tutor Gilles Scouarnec, as well as to Claire d'Agostino, for their invaluable help and their permanent good mood during this thesis. It has been a real pleasure to work on this project with them.

I do really thanks Jonas Bärgman, for supervising my work despite the distance, seriously and with interest for the topic. It was hard to follow me in those particular conditions but he did it perfectly and I couldn't have finished the report without his help.

I would also like to thanks my manager Pierre Lavorini, as well as all the Driving Performance team. The so positive environment within this team really made me enjoy the thesis.

I also thanks all the people who have helped me all along those 6 months at Volvo, when I was facing problems or when I needed data.

Abbreviations

ADAS: Advanced Driver Assistance Systems. Refers to devices assisting the driver in his perception of road environment. Although this term comes from active safety field, it is also widely used to qualify devices aiming at changing drivers' behavior to improve fuel efficiency.

CAN: Controller Area Network. Vehicle network protocol, created by Bosch in 1986, which allows devices to communicate with each other without the need of a central bus master.

EEM: Energy Efficiency Management. This criterion represents the capacity of a driver to use truck energy efficiently. It considers fuel consumption to browse a given distance, but also the corresponding time. EEM can be split into sub-categories: indeed it would be different for traction and non-traction phases. For traction phases (throttle 0), it means using engine energy as efficiently as possible, whereas in non-traction phases (throttle = 0) it means how the driver uses the truck kinetic energy.

GPS: Global Positioning System. Satellite navigation system giving position (longitute, latitude, altitude) anywhere on the Earth.

HMI: Human Machine Interface. Refers to the way a system communicates with user: this usually done through visual signal (e.g. screen), audio signal (e.g. voice) or haptic signal (e.g. vibration).

GTT: Group Truck Technology. It is the branch of Volvo group that deals with the entire chain from research to final delivery. Volvo Group Trucks Technology is operational since January 1, 2012, when activities within Volvo 3P, Volvo Powertrain, Volvo Parts, Volvo Technology and Non-Automotive Purchasing were merged together [51].

PKE: Postive Kinetic Energy. Refers to the driver's ability to keep the vehicle kinetic energy as low as possible on speed variation phases[17]. Further details are given in subsection 2.3.2.

Summary

During this study, a first phase consisted in reading different documents about the topic. This literature review included mainly different scientific articles, but also of Volvo internal engineering reports. Updating the existing benchmark study was another way to get a clearer understanding of what is currently developed by different companies, competitors or suppliers, in terms of ADAS for fuel efficiency optimization based on driver behaviour change.

The dataset to be used in the study, was when then explored. There was already an existing code for processing driving data in this context, taking route data as input and providing driving parameters related to fuel efficiency as output. However this code had to be updated a lot, adapted to a new data format, and large sections of new code was needed to address the objectives specific to this study.

Driving parameters previously identified as relevant for driving-efficiency were investigated. This was a mandatory step from a methodology point of view, but also to check the validity of the previous study. Furthermore, new customers data related to different types of use were then analysed, to identify the different uses with respect to driving-efficiency through driving parameters.

Three types of road layout were then analysed with respect to drivers behaviour and parameters influencing driving-efficiency, namely roundabouts, road altitude variations and curves.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

1.1.1 Fuel saving concern in the truck industry

For many years now, fuel saving has been one of the main focus areas in the automotive industry. . As fuel producers have reached their maximum capacity, the rise of countries such as China and Brazil has made the demand for more efficient cars and driving grow quickly. A higher demand combined to a constant production has led to a dramatic increase of fuel prices: the truck industry has been very affected by this evolution, since both annual mileage and average fuel consumption on a truck is much higher than on a car.

In 2012, Volvo released the new FH serie (for high range), with Euro 6 engines[52], and claim that fuel consumption has dropped by 10% compared to the former FH (Euro 5 engines). The *Euro* regulations have been implemented since 1988 by the European Union mainly to reduce emissions. The evolution, in terms of NOx (nitrous oxides) and PM (particulate matter), the 2 main pollutants, is shown in Figure 1.1.

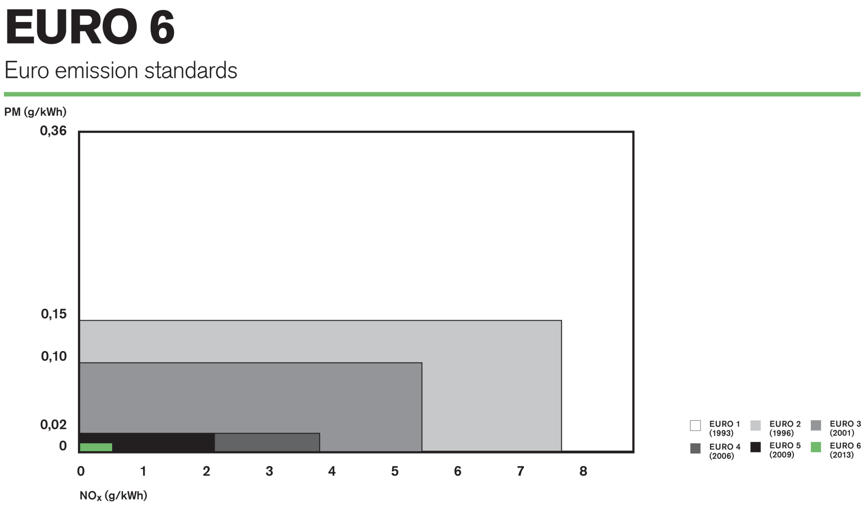

Figure 1.1: *Evolution of Euro norms*

Furthermore, between *Euro 1* and *Euro 6* the maximum CO (carbon monoxide) emission level was divided by 3, and the maximum HC emission level was divided by 10. Although the requirements on reduced emissions need to be met, customers of trucks are also requiring lower fuel consumption.

As engineers have strived to increase engine efficiency, new features have also been developed, such as *I-See*, a slope prediction and management tool whose functioning is explained on Figure 1.2.

However, several studies have shown that drivers' influence on fuel consumption is huge: considering the effects of real-world driving conditions, efficient driving behaviours could reduce fuel use by 20% on aggressively driven cycles and by 5-10% on more moderately driven trips[31].

The driver can be considered the most influential parameter in a truck with respect to fuel consumption, that is the reason why *Rational driving* is becoming more and more important. *Rational driving* is the optimization of the driving in terms of multiple parameters i.e. energy consumption, safety, travel time, comfort and other parameters[42]. Cummins, an American truck engine manufacturer, made a comparison of different parameters that affect fuel consumption, showing the percentage of fuel one can save. Figure 1.3 shows the summary of their findings[48], and especially the fact that driving behaviour can have a bigger impact than parameters such as tire condition or external temperature. There is much to do in terms of *Rational driving* as truck manufacturers start to figure out the relationship between driving style and fuel consumption. By understanding this relationship, they would be able to adapt their trucks' environment and technology to

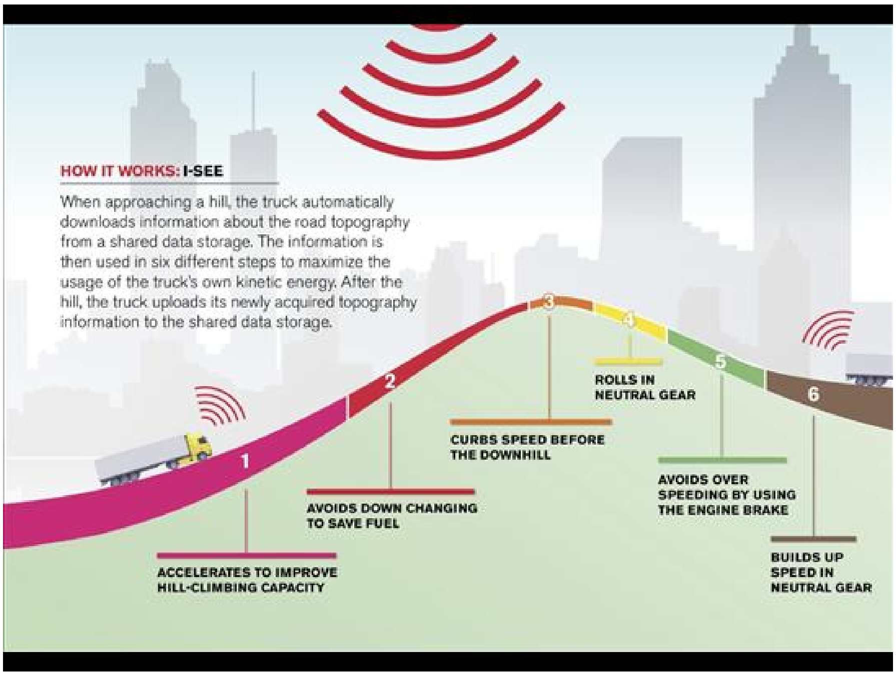

Figure 1.2: *Explaining of Volvo I-see system*

Impact Cause	Special Notes	% Effect
Engine and drive line "break-in" effect	After 10K miles the MPG improves approx 2% to 5%	2% to 5%
Tire tread depth effect	MPG improve by approx 6% from 100% tread depth tire (new), to a 50% tread depth tire	0% to 6%
Running one gear down effect	One gear down you decrease mpg by approx 3% due to gear mesh effect. The engine cruising rpm is approx 400+ rpm higher which decreases mpg by approx 4%	6% to 8%
Engine speed (proper gearing) effect	An engine geared to run 1450 rpm at 65 mph gets approx 4% better mpg than one geared to run 1600 rpm at 65 mph. If the truck is geared too high it will force a driver to run one gear down. (See the effect on one gear down in this matrix)	4%
Transmission gear mesh effect	Transmissions lose approx .75% for each gear mesh. Therefore an OD trans, running in OD, has 4 gear meshes and would lose about 2% to 3% mpg compared to a direct drive trans.	2% to 3%
Aerodynamics effect	Full aero aids can improve mpg by approx 15% above 50 mph. Trailer gap must be minimized, < 30 inches. Double trailers decrease MPG by 5%.	0% to 15%
Winter effect	Higher density air, wind (cross and head), more idle time, blended fuel (lower BTU), snow, more driveline drag, etc. 8% - 15% decreased MPG compared to Summer	8% to 15%
Cooling fan on time while driving	Fan HP increases with rpm (X HP cubed). When cooling fans are running they use between 1.5 and 3.5 gallons of fuel per hour. Increasing fan run distance from 30% to 50% will decrease MPG by between 3% to 5%.	3% to 5%
Speed effect	-.1 mpg / 1mph > 55 mph. This is a rule of thumb that is hard to beat and is based on aerodynamic drag.	
Idle time (%) effect	Engines use .5 gal/hr at 650 rpm and 1.0 gal/hr at 1,000 rpm. Reducing idle time from 50% to 25% can improve mpg 2% to 4%	0% to 10%
Driver variability effect	Up to 30% difference between a good mpg driver and a poor mpg driver (in each fleet). Recommend using LBSC.	0% to 30%

Figure 1.3: *Influence of different parameters on fuel consumption*

support the driving in reaching lower fuel consumption while not reducing on-time delivery etc.

1.1.2 Rational driving in the literature

One of the major objectives of *Rational driving* is to manage to change drivers' behaviour. Indeed, even if relevant driving parameters are clearly identified, drivers may not be eager to change their habits. Yet, they have a high influence on fuel consumption. The influence of human behaviour on commercial vehicle driving performance was developed by Maincent [37] with a definition of *Rational driving*. She makes the difference with *Eco-driving*, and shows that it tends to be more adapted to trucks as it takes into account the productivity, related to average speed, while the second one only considers fuel consumption.

However, Maincent [37] also makes a distinction between 2 types of scenarii: long haul and delivery. In the first case, the route is usually simple and the driver is not as affected by time requirements or pressure from the customer. The main stress cause is the risk of theft (during the night for instance). In the second case, the route is much more complex and drivers usually have to perform several deliveries per day. There are much more interactions with the other road users and the customers in the delivery scenario. Hence, the efficiency is not affected by the same parameters: that is the reason why she states the type of use as the main parameter when it comes to discussing about drivers' efficiency.

Ericsson[17], through the study of a larger amount of data, manages to point out a list of relevant driving parameters relevant across all different cycles. Those parameters are driver dependent, i.e. they allow a comparison between drivers regarding their efficiency. She especially highlights the influence of parameters related to acceleration and power demand, gear changing behaviour, and speed level.

Maincent[37] faced difficulties in finding parameters influencing long haul truck drivers' energetic efficiency. However, in complex situations, many parameters were found to be relevant. Some of them directly depend on the driver, such as the braking behaviour, while some of them are driver independent factors. By studying the bus drivers' efficiency on a same route, but at different times of the day, Liimatainen [35] confirms that a difference shall be made between factors related to the driver and to its external environment. According to him, drivers can be ranked provided that the influence of driver independent factors is known and quantified. He mentions the influence of those factors, such as traffic jam, on the fuel consumption: the time of the day, for instance, is shown to influence the fuel consumption by up to 46%. This influence of external parameters such as traffic lights or other vehicles, has led some universities to develop ADAS system that help the driver improve her/his efficiency by the use of horizon prediction system. Based on GPS data, they allow the driver to anticipate the preceding vehicle's action and status of traffic signals. In a model that he developed, Kamal et al.[32] determines the appropriate vehicle control input by using model predictive control. He estimates that some adjustments remain to be done, as the travel time is increased by 17% when the system is used, but he manages to lower the fuel consumption by more than 10%

Wada et al.[54] suggests that ADAS systems should also take into account the fact that drivers can get annoyed with it. Based on experiments, he shows that ADAS systems can lead to low driver motivation or boredom. This phenomenon results in higher fuel consumption, which is the opposite of what the system is developed for. Maincent[37] considers that drivers can adopt an efficient behaviour if it doesn't affect their driving pleasure, while J. Gonder[31] pretends that drivers are discouraged when the system is not easy enough to use.

The same observation led Syed et al.[49] to develop a haptic throttle pedal, i.e. a vibration is generated when the driver pressure is too high. If they highlight the difficulty of achieving driveability criteria with such a system, they conclude that fuel consumption can be dropped by 10% on a mixed driving cycle.

In order to quantify the influence of drivers' behaviour on fuel consumption, Andrieu and Pierre[4], have developed a global indicator based on statistical models. One of these models is a complex model based on 4 basic rules of eco-driving (i.e. *shift as soon as possible*, *maintain a steady speed*, *anticipate traffic flow*, *decelerate smoothly*) that provides efficient feedback to the driver. This model manages to correctly identify an average of 83% of the drivers regarding their driving parameters The system also uses GPS data for horizon prevision purpose. With a system simply based on PKE (a parameter defined in Subsection 2.3.2), they managed to find satisfying results, very close to what was found with the model based on the 4 basic rules.

1.2 Meta comparison of driver coaching systems

With the spread of new technologies and the rise of fuel prices, many driver coaching systems have been developed in the last 10 years. Those systems aim at helping drivers improve their efficiency by providing them with advice. One can distinguish 4 different types:

- Systems developed by manufacturers
- Systems developed by suppliers
- Systems developed for smart-phones
- Systems developed by universities

Using those system brings to an average fuel consumption which can be reduced by 5 to 20%. If there are many existing systems, most of them use the same input parameters and the same type of HMI. They are mainly based on the 4 parameters shown in Figure 1.4. GPS signal can be used, for most advanced systems, for horizon prediction (i.e. upcoming road). Road altitude variations or traffic jam, for instance, can indeed be predicted before being visible by the driver. Among the other parameters which are used in those systems, one can note the retarder use frequency and amplitude (retarder is a device used for slowing down the vehicle, while brakes are used to stop the vehicle) in the case of commercial vehicles. Most of ADAS devices display their messages through a HMI, usually a screen, such as the one developed by TomTom and shown on Figure 1.5. The previously mentioned data are weighted and computed, before being compared to an ideal behaviour calculated by the system. Drivers are given a grade, regarding their driving ability in different fields. Table 1.1 represents a list of the different systems which have been identified during this project.

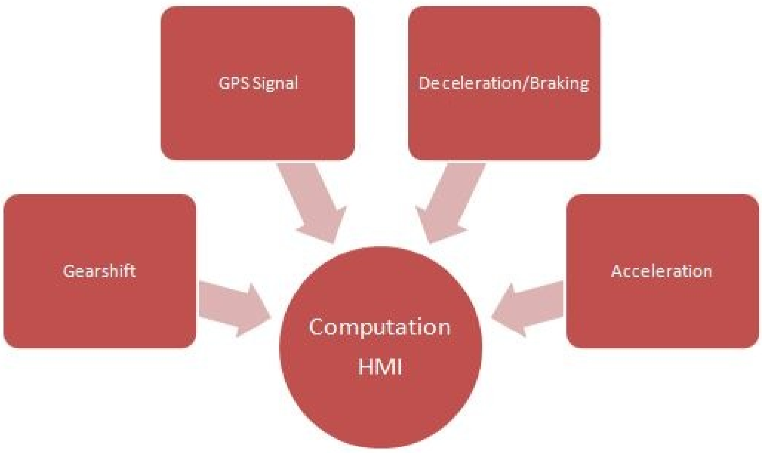

Figure 1.4: *Driver coaching system functioning*

Table 1.1: ADAS Benchmark list

Name	HMI	Type	Fuel economy	Vehicle	Real time /Post treatment	GPS
AA EcoDrive[5]	Smart-phone	Developer	10 to 20%	Car	Real time/Post treatment	x
BMW EcoPro[6]	Vehicle	Manufacturer	20%	Car	Real time	x
BMW MINIMALISM Analyser	Vehicle	Manufacturer	Car	-	Real time/Post treatment	x
Bosch EcoLogic[23]	Vehicle	Supplier	5%	Trucks	Real time	-
Bully Dog Heavy Duty WatchDog[14]	Specific	Developer	-	Trucks	Real time/Post treatment	-

Name	HMI	Type	Fuel economy	Vehicle	Real time /Post treatment	GPS
Crambo Labs Econav[3]	Smart-phone/GPS	Developer	16%	Car	Real time	x
Daf ATe[20]	Vehicle	Manufacturer	-	Trucks	Real time	-
European union Eco-Driver[10]	Vehicle	Public/Private project	Up to 20%	Car/Trucks	Real time	-
FIAT eco:Drive[19]	Vehicle	Manufacturer	6 to 16%	Car	Real time/Post treatment	-
Ford Wayne State Uni.[49]	Pedal (haptic)	Manufacturer/University	10%	Car	Real time	-
Fuso Eco-Fleet Pro[21]	Vehicle	Manufacturer	10%	Trucks	Real time/Post treatment	-
Garmin ecoRoute HD[22]	GPS	Supplier	-	Car	Real time	x
HDM Solutions ECO-MOBIL[16]	Computer	Developer	-	Car/Trucks	Real time/Post treatment	x
Hino Drive Master PRO[24]	Vehicle	Manufacturer	-	Trucks	Real time/Post treatment	-
Honda EDAS[25]	Vehicle	Manufacturer	-	Hybrid car	Real time/Post treatment	-
Hunter R&T green-Meter[47]	Smart-phone	Developer	-	Car	Real time	-
Isuzu Mimamori-kun Online Service[27]	Vehicle	Manufacturer	-	Trucks	Real time/ Post treatment	x
Isuzu Vehicle Health Reports[28]	Vehicle	Manufacturer	-	Trucks	Post treatment	-
Iveco DSE[30]	Vehicle	Manufacturer	5 to 12%	Trucks	Real time/Post treatment	x
Iveco Blue&Me[29]	Vehicle	Manufacturer	-	Trucks	Real time/Post treatment	x
Kagawa Uni.[54]	Vehicle	University	-	Car	Real time	-
KeyDriving KDTS[15]	Computer	Supplier	-	Car	Post treatment	-
Kia ECO driving system[11]	Vehicle	Manufacturer	10 to 20%	Car	Real time	-
Kyushu Uni.[32]	Vehicle	University	13%	Car	Real time	-
Mercedes FleetBoard EcoSupport[36]	Vehicle	Manufacturer	5 to 15%	Trucks	Real time/Post treatment	-
Nissan ECO Pedal[12]	Vehicle	Manufacturer	5 to 10%	Car	Real time	-
Nomadic Solutions EcoGyzer[7]	Smart-phone	Developer	-	Car	Real time/Post treatment	x
PLX Kiwi Trip Computer[41]	Specific	Developer	-	Car	Real time	-
Qualcomm ADBR[40]	Vehicle	Supplier	5 to 15%	Trucks	Real time/Post treatment	-
Renault DrivingEco2[44]	Tablet	Manufacturer	10%	Car	Real time/Post treatment	-
Scania Active Prediction[2]	Vehicle	Manufacturer	3%	Trucks	Real time	x
Scania Driver Support[39]	Vehicle	Manufacturer	11%	Trucks	Real time	-
TomTom EcoDriving[1]	GPS	Supplier	-	Car	Real time/Post treatment	x
UD Nenpioh[9]	Vehicle	Manufacturer	12%	Trucks	Real time/Post treatment	-

Name	HMI	Type	Fuel economy	Vehicle	Real time /Post treatment	GPS
Uni. Carlos III Artemisa[13]	Smart-phone	University	-	Car/Trucks	Real time	x
Veolia EcoDriving assistant	Vehicle	Public transport	8%	Bus	Real time	-
Volvo Eco-Roll[26]	Vehicle	Manufacturer	-	Trucks	Real time	-
Volvo Driving Coaching[53]	Vehicle	Manufacturer	-	Trucks	Real time	-

It is also interesting to see that some universities have been working on the topic: they have often developed a different approach to drivers' efficiency. For instance, Kagawa Uni.[54] consider that drivers quickly get annoyed by driver coaching, and their effort tend to decrease as time goes by. Around this human reaction, they developed a system allowing long term fuel consumption by presenting eco driving as a "challenge", whose level gets adapted to the driver level. The problematic of city driving is also broached, with a study carried out by Kyushu Uni.[32], who developed a system for crowded environment, with a huge focus on surrounding vehicles.

1.3 Objectives of the study

This study has been carried out within the department *Complete Vehicle - Feature Verification & Validation* at Volvo facilities in Lyon. It aims at quantifying and modelling how driver behaviour can influence fuel consumption. This especially includes driver reactions when facing an event. *Event* refers to any section of the road where the driver needs to take action, e.g. a roundabout or a traffic signal. Driver behaviour is initially divided into 5 categories, described in subsection 2.3.2.

The study is based on 2 types of driving data:

- Internal data: this refers to sessions organised by AB Volvo. They gather drivers from different backgrounds, driving on a same road and with the same truck. Those sessions allow a comparison between the drivers as they are assessed in the same conditions.
- Customer data: this refers to records gathered at customers' facilities. They represent real-life cycles, and give information about how trucks are actually used by customers.

Those 2 types of data have specificities which are interesting for the different objectives to achieve within this thesis, which are:

- Finding out which driving parameters affect fuel consumption, and generally speaking, driving efficiency.
- Evaluating the influence on driving efficiency of the way drivers react to different road events.
- Being able to classify different types of truck use into different categories (e.g. long haul, delivery) by observing CAN signals only, regardless the driver.
- Implementing a system that grades and ranks drivers in terms of efficiency based on driving parameters.

1.4 Building on previous research

1.4.1 PhD

This master thesis is part of an Advanced Engineering project started in 2012 called CVLY1106, intending to develop a driver model generator for dynamo-meter, powertrain rigs and simulation platform. The CVLY1106 project's backbone is a PhD students work, whose dissertation is to be delivered by the end of 2014. This

Figure 1.5: *Example of a driver coaching system interface*

PhD covers the different events one can distinguish during a driving cycle, and investigates drivers' behaviour by splitting it into different classes, from efficient behaviour to non efficient one.

After identifying the connection between behaviours during different events and fuel consumption, the profiles are to be implemented within a virtual driver model in order to be used to control vehicle simulations. However, within this thesis, no study or development of the virtual driver is planned.

1.4.2 Previous master thesis - carried out in 2012

Generalities

A previous Master thesis (internal report), carried out between January and June 2012, delivered a set of first conclusions on drivers' influence on fuel efficiency. This study was divided into 2 parts:

- Benchmark study: existing ADAS were listed and explained, with a focus on passive systems (i.e. systems which only inform the driver, but which don't act on the vehicle). They were assessed and grouped regarding parameters used and level of development. This concerned systems developed by suppliers, but also by competitors or even mobile phone systems available on applications stores.
- Analysis of correlation between fuel economy and driving parameters: this part was the main part, and focused on finding out what makes a driver fuel efficient or not. Both simple criterion and multicriteria were carried out, and average speed was also taken into account.

This Master thesis has drawn first conclusions about the influence of different driving parameters on the fuel consumption. The set of data used for the analysis came from a PhD carried out until 2010 by Annick Maincent, within Renault Trucks, and called *Comportements humains, Activités finalisées et Conception de systèmes d'assistance àla conduite de Vehicules Industriels ("Human behaviours, finalized activities and conception of driver coaching systems for commercial vehicles")*[37].

The set of data used consisted of 33 drivers who travelled back and forth a 170 km route between Lyon and Voiron (south east of Lyon), gathering a wide range of situations (urban, highways, etc.). This field test was recorded between February 2003 and April 2004. The 33 drivers were divided into the followings categories:

- 23 "experts" drivers:
  - Professional drivers sent by external companies
  - 14 of them had received special training in rational driving
  - Average of 14 years with driving license E (super heavy)
  - Average of 500 km travelled per working day

- 7 ”non experts” drivers:
  - Employees of RT driving less than 5000 km per year
  - 3 of them had received special training in rational driving
  - Average of 18 years with driving license E (super heavy)
  - No professional experience
  - Less than 1000 km travelled per year
- 3 “reference” drivers: specialized drivers in rational driving within Renault Trucks

All drivers drove the same vehicle: a Renault Premium long-haul tractor, with 420 horse power engine, coupled to a semi-trailer, with a gross vehicle weight of 38 tons. The complete set was similar Figure 1.6. Up to 25% of difference between the best and the worst drivers' fuel consumption was observed, for an equivalent driving time.

Figure 1.6: *Truck used for driving records*

Results and conclusions

This study determined relevant correlations for the set of data detailed above. Those results are shown in Table 1.3, and refer to driving parameters detailed in Subsection 2.3.2. Due to the fact that most trucks are equipped

Table 1.2: Results from the previous study

Criteria	Correlation
Respect of Diesel engine green band	0.75
PKE	0,58
Even driving (average duration of phases)	0.57
Anticipation (average duration of phases)	0.49
Coasting (rate)	0.49
Braking (number of events)	0.41

with automatic gearboxes, the engine green band (i.e. the engine speed range for which the engine efficiency is the best) doesn't make a difference between drivers: that is the reason why it is not taken into account in this study any more. In addition, in a truck environment, the delivery time schedule is usually extremely tight: unfortunately, only *Fuel consumption* is used as a reference for the results given above. However, it is important to consider average speed as well. When taking this parameters into account, trends are found and listed in Table 1.3. Those trends remain global, i.e. showing up criteria only.

Table 1.3: Relevant parameters when taking the speed into account

Fuel Consumption (%)	Average Speed (%)	Relevant Parameters
100	0	Coasting, Anticipation, Even driving, Speed in roundabouts, PKE
80	20	Braking, Even driving
70	30	Coasting, Even driving

Driver categorization (i.e. efficient, mean and non-efficient drivers) were calculated from the percentages of *Fuel consumption* and *Average speed*, the sum of both being 100: *% Fuel consumption + % Average speed = 100.* The best results, shown on Table 1.4, were found with the second combination (*80/20*).

Table 1.4: Classification rate of driving parameters

Parameters	Classification rate
Even driving / Average duration	0,994
Braking / Number of events	0,984
Coasting rate	0,592
Anticipation / Average duration	0,592
Braking rate	0,592
PKE	0,544
Coasting / Average speed variation	0,455

The current thesis aims at complementing this previous thesis, and was carried out with different driving cycles, different drivers, and a different truck. Data used for this study are detailed in subsection 2.1.1. Particularly, it will be important to compare relevant parameters found for both studies, to figure out if they are the same whatever the type of trucks.

2 Methodology

2.1 Presentation of the data

2.1.1 Generalities about the 5 customers

During this study, a total of 5 customers' driving data has been used. The term *Customer* here refers to both internal (Volvo drivers) and external customer (delivery companies for instance). In the following customer is used both to describe a single driver and groups of drivers, depending on the context.

Customer 4 only was used for most of the study, except for Section 3.3, where all 5 customers were used. Customer 4 and 5 are Volvo internal drivers recorded during experiments. This means that they drove the same truck, on the same route: hence, for each of these drivers, it is possible to make a comparison between the drivers. For the drivers in these customers data, the influence of external environment is limited and can be compared more easily than for the other customers. For both customer 4 and 5, there was no delivery to insure, and the drivers knew they were recorded for fuel consumption purposes. They were just to follow the route that had been given to them, going from point A to point B. These drivers did this driving as part of their normal working time and was not compensated for their time in any other way. The fact that they knew the reason for the study and that they were under no time pressure, they may have driven more efficient than if making a real delivery.

The previous master thesis, mentioned in Subsection 1.4.2, already used Customer 5's data. At that time, customer 4's data hadn't been recorded yet.

The three other customers are external companies, located in the Lyon's area. They were recorded in their actual working environment. The first and third customers are driving on the same type of cycle, however the third customer is considered as very non efficient (according his employer's feedback) while the first customer has 3 normal drivers. Those customers drive mainly in a city environment. Customer 2 represents yet another type of use: it is a refuse truck company, i.e. company which collects garbage. This use is very specific, as it is mainly made of harsh acceleration/braking phases.

Table 2.1 sums up some details about the different data used.

Table 2.1: Detail of the 5 records used for the study

	Customer 1	Customer 2	Customer 3	Customer 4	Customer 5
Engine power (hp)	180	280	310	240	420
Norm	Euro 5	Euro 4	Euro 5	Euro 5	Euro 3
Type	Urban / Regional	Refuse	Urban	Regional	Regional
Number of drivers	3	1	1	11	23
Fuel consumption (L/100km)	21	N/A	N/A	26	56
Average speed without stop (km/h)	51	20	35	66	55
Specific Energy Consumption (kwh/km)	0.78	3.67	1.29	0.95	2.17
Mass (t)	6.7 to 8.5	14 to 26	11 to 15.3	13.3	40

All those data are quite diverse: they concern different trucks and different loads. The number of drivers is also very different with very few drivers data for customer one through three. The very low number of drivers for these customers makes it difficult give more than indications of results. However, data from all customers is used together only for a single part of the study: the identification of the types of use (i.e. refuse, regional, etc.). So those differences are on purpose and don't affect the results. In order to get only one value per customer, all the drivers' values for each customer are averaged: for instance, for customer 4, the values used are the average of the 11 drivers.

2.1.2 Customer 4's records used for the driving behaviour analysis

Generalities

A set of 11 drivers was recorded between May and July 2012, on a 150 km route around Lyon. In this study, 5 sections representing 64 km of driving are considered: each of them illustrates a certain type of driving conditions (e.g. highway) where all drivers can be compared. The 11 drivers used the same truck, a Volvo FL whose characteristics are given in Table 2.2. In this report, the drivers are named from A to K. Among these drivers, 2 (D and E) are part of the Driving Performance team within Volvo GTT: they are professional test drivers, and are used as a reference. The other drivers are temporary drivers. They have different level of experience, from inexperienced (K) to very experienced (B). This truck, shown on Figure 2.1 is usually

Characteristic	Value
Customer	Internal vehicle
Chassis type	FL R4x2
Engine type	DXi7 Euro 5
Power	240 hp
Real load	14 tons (43% on the front / 57% on the rear)

Table 2.2: Characteristics of the truck used for the driving session

Figure 2.1: *Truck used for the driving session*

dedicated to urban distribution and is fairly light: this is the main difference with the truck used in the previous study, which was much heavier. Mass remained the same for all drivers and all driving, and an automated manual gearbox was used: drivers were told to keep automatic mode, which means that the shifting strategy is the same for all drivers in this study.

Driving conditions

The 5 road sections which are driven by the customer 4's 11 drivers represent typical driving conditions. The sections have different length and speed profiles, and present different particularities. Especially, they recreate some of the types of use mentioned in the previous part: urban (i.e city centres), regional (suburbs) or highways. A full cycle is also made out of the concatenation of the 5 sections. In the following the different sections are described. All the maps shown up in this report were obtained from Google Maps.

- Section 1B: this section is quite short, with a straight line and 3 events (roundabouts). This is located in the suburbs of Lyon where the speed limitation for trucks is 80 km/h.
  - Length: 5870 m.

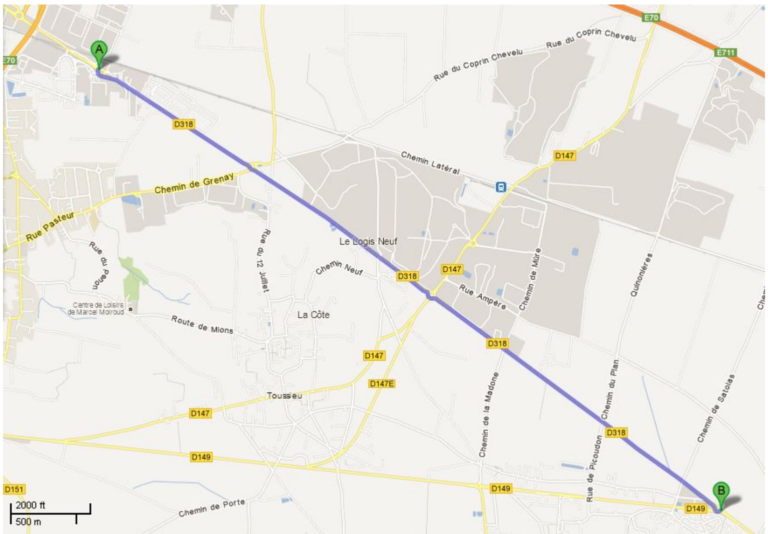

Figure 2.2: *Map of section 1B*

- Average speed across all drivers: 53 km/h.

* Section 2B: this section is short as well, and is considered because of its hilly aspect. The speed limitation is 80 km/h, and there is only traffic light.

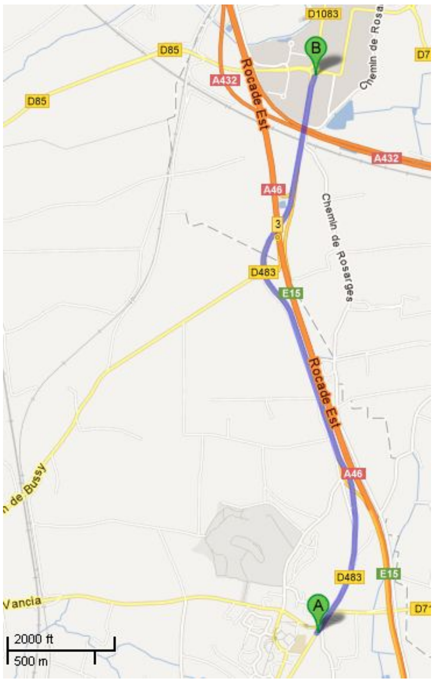

Figure 2.3: *Map of section 2B*

- Length: 3375 m.

- Average speed across all drivers: 62.4 km/h.

* Section 4B: this section represents a long haul usage on a highway where the speed limit is 90 km/h This section is by far the longest. There are a few occasions where the vehicle has to slow down: mainly tolls

or lane changes.

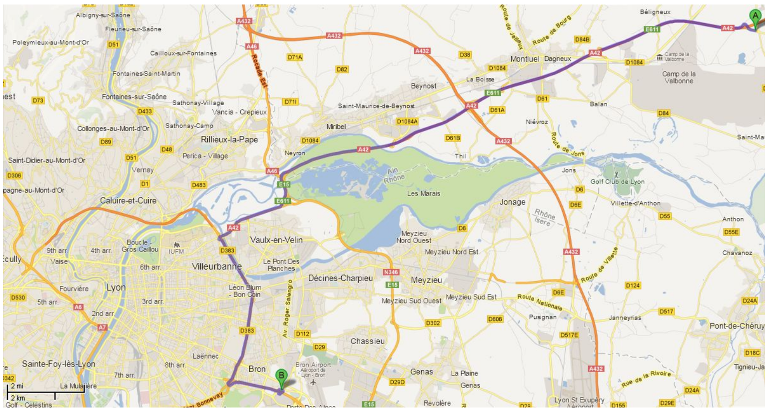

Figure 2.4: *Map of section 4B*

– Length: 46890 m.

– Average speed across all drivers: 79.1 km/h.

- Section 111: this section is located in an urban environment, behind Volvo GTT Lyon head-quarter, with a speed limit of 50 km/h. This short section includes few turns, until a traffic light.

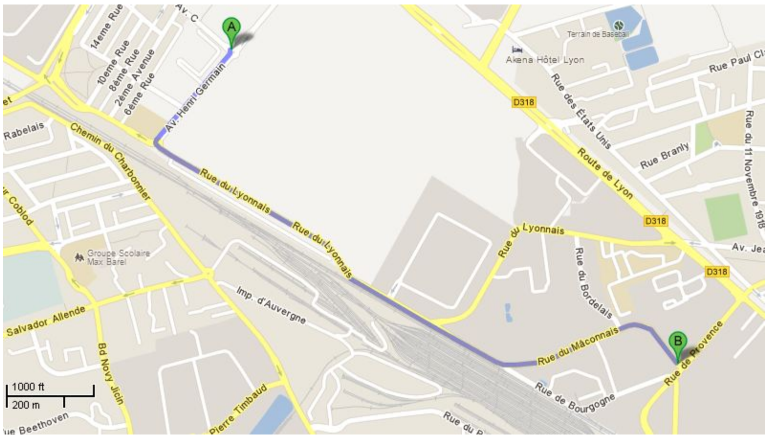

Figure 2.5: *Map of section 111*

– Length: 2560 m.

– Average speed across all drivers: 42.1 km/h.

- Section 112: this section is located right after section 111, with the same driving conditions. The section is a little bit longer, also has a few corners and roundabouts, explaining why the average speed is much lower than the speed limit.

– Length: 5180 m.

– Average speed across all drivers: 34.2 km/h.

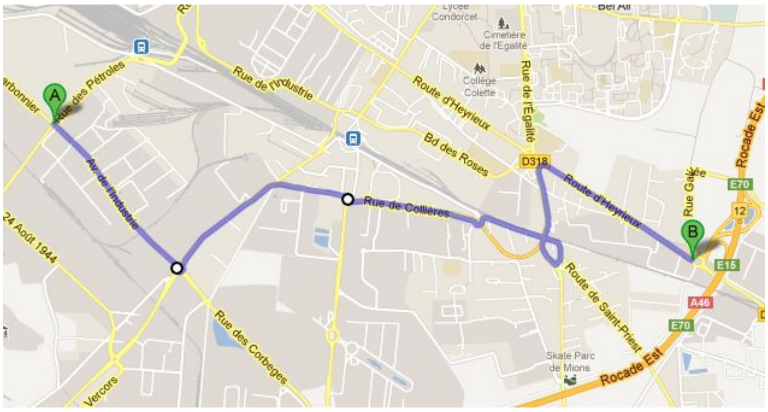

Figure 2.6: *Map of section 112*

2.2 Background about correlation tools

2.2.1 Correlation tools used in this study

Within this study, two approaches are used: the single criterion analysis, and the multicriteria analysis. The different approaches are combined in order to strengthen the results. Both methods have their advantages (respectively disadvantages): the first approach is easy to comprehend, but has issues with reliability of results when having few data points and in not considering multiple parameters. The second allows consideration of several parameters at the same time, while at least in this study the underlying algorithms providing the results are not as transparent. Further details are given in the next subsections.

Single criterion analysis

This study is mostly based on regression calculation: most of it was completed using Microsoft Excel to post-process data obtained from Matlab code. Many approaches exist but this study considers only two of them:

- Polynomial regression: the relationship between two parameters X and Y is assumed to be a polynomial function, given by Equation [8].

$$Y = a + bX + ... + nX^n \tag{2.1}$$

  Where $a, b, ...n$ are constant and found as part of the regression.

  This approach is simple and widely used, even on a complex system such as a truck. Only 1^{st} and 2^{nd} order regressions were used in this study. However, in such an environment, it is very likely that a linear regression can't be found: that is the reason why the report focuses on the 2^{nd} order trendline, which fits better to non-ideal environment. When only a low number of input data is available, as it is the case in this study, it is more reasonable to keep a low order regression, in order to stay true to reality.

  For visualization the regressions are in this report often drawn on scatter graphs. Those graph usually have 11 dots, each dot represent a driver. In Section 3.3, there are only 5 dots: each dot corresponding to one customer.

- Pearson's correlation coefficient: measure of the correlation (linear dependence) between two parameters X and Y. For a population, the Pearson's coefficient ρcan be calculated as the ratio of the covariance of the two parameters and their standard deviations[45]:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2.2}$$

  This function is defined between -1 and 1 and the better the correlation, the higher the value: a value of 0 means that there is no correlation at all between the parameters.

In this study, the squared value of Pearson's coefficient is used: as the initial coefficient is defined between -1 and 1, it's squared value is defined between 0 and 1. This coefficient is called *Coefficient of determination*, and is denoted R^2. R^2 is widely used in statistics to evaluate the dependency of two parameters.

When driving data (CAN signals) were post processed in Matlab, the results from this post processing were stored in an Excel file. Microsoft Excel was also used for the calculation of all the correlations.

However, the previously described methods have a main drawback: they don't allow comparison of a combination of independent parameters to a reference (i.e. dependent parameter), but only one independent parameter at a time. To also include multi-parameter interactions, another method has been used: the multicriteria analysis.

Multi criteria analysis

Following the previous phase, a multi criteria analysis was carried out in order to find the relationship between criteria that could have dependencies and also to set up a criteria hierarchy. To do so, a data mining software called Weka (*Waikato Environment for Knowledge Analysis*) and developed by the University of Waikato (Hamilton, New Zeland), was used. This freeware contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization[46]. It also contains a set of algorithms such as Naive Bayes and decisions trees. Two of these algorithms were mainly used for this study:

- **J48**: this is an open source Java implementation of C4.5, a statistical classification algorithm used to generate a decision tree. This algorithm produces a tree providing the hierarchy between the different driving parameters investigated[43]. The goal of tree induction is to find a subdivision that is fine enough to capture the structure[38]. It is interesting because it clearly states the most critical parameters: through the provided tree, one can easily see which parameters are more relevant. The complete set of parameters can be evaluated, and appear in the nodes (see Figure 2.7). From one input occurrence, and from the use of those parameters, the algorithm is then able to determine which class this occurrence belongs to.

- **Simple Logistics**: this uses logistic regression to model how the probability p of an event may be affected by one or more independent parameters [38]. This algorithms classifies occurrences (input data) into different classes (e.g. efficient drivers), and each of these classes is defined by an equation depending on the parameters assessed.

  $$C = a_1x_1 + a_2x_2 + ... + a_nx_n \tag{2.3}$$

  Where C is the class where the instance is classified, $x...x_n$ the different parameters (parameters), and $a...a_n$ are coefficients.
  In the case of the Simple Logistics algorithm, this equation is linear. An equation is generated for each class, as a linear function of driving parameters.

  For the whole multicriteria study, all the driving parameters are normalized: the objective is to give the same weight to all the parameters in order to compare them in a reliable way. The point is to get, for each parameter, a highest value equal to 1 and a lowest equal to 0.

  Simple Logistics equations work through a system of relative ranking between the different classes. There are as many equations as classes: each equation corresponds to a class. In the current case, there are three equations, related to efficient, mean and non-efficient drivers. Actual efficient (respectively mean, non-efficient) drivers, when the equation related to efficient (respectively mean, non-efficient) drivers is computed, are ranked above other drivers.

In this thesis, for each road section a set of three equations is created, one for each of the categories (i.e. efficient, mean, non-efficient). Each individual equation has its own relevant parameters for efficiency related to the section and the class. Since there was a very large set of parameters to start with, different sets were used for the different sections. Several iterations were run in order to determine which parameters to include in the logistic regression. This is close to the method used by the *Random Forest* explained below. This was used because the total number of parameters evaluated during this study ended up to be very large.

As it was very difficult to compare all those parameters in a single iteration, it was decided to run different iterations with different sets of data each time. This allowed a better evaluation of the relevant parameters: for each iteration, the parameters which were not important (i.e. considered as non relevant by *Weka*) were eliminated. In the end, iterations were run with sets of relevant parameters only, and those parameters were classified.

In Appendix D, a few more correlation tools are described.

2.2.2 Utilisation of *Weka*

Weka uses a system of classification. It means that it checks how many drivers are correctly classified according to the solution generated (tree for *J48*) or set of equations for *Simple Logistics*). If the solution works perfectly, *Weka* manages to find which class each driver belongs to. In other words, the software checks its own work (i.e. the efficiency of the solution generated). However, it can happen that some drivers are classified in a wrong class (e.g., a driver classified as a *mean*, while he is actually *non efficient*).

Explanation of the two multicriteria methods used

The two previously described multicriteria methods are very interesting but one needs to understand exactly how to interpret their results. That is the reason why an example is given for each of them.

- **Explanation of a decision tree**
  An example of decision tree is shown in Figure 2.7[33]. This example illustrates the decision process of a

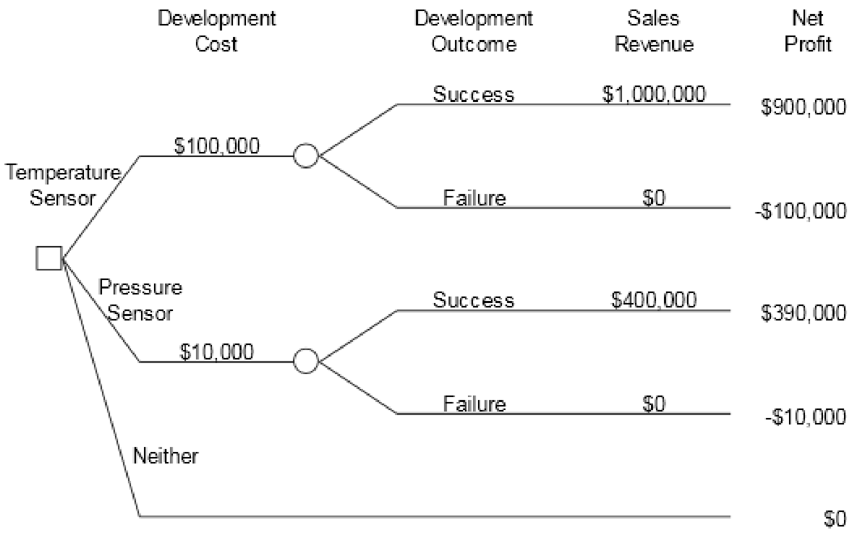

Figure 2.7: *Example of a Decision tree*

  company concerning the development of a product. Three solutions are available: a temperature sensor, a pressure sensor or nothing.

  If they choose to set up a sensor, they can either success or fail. The cost is higher for the temperature sensor, but the possible profit is higher. If they choose the last solution - nothing - there is no change, so neither profit nor loss.

  But this can be read the other way around, from the end to the beginning. Thereby, if the net profit is neutral, it means that none of the sensors was chosen. If the company makes 900 000$, it means that the temperature sensor was chosen and the development was successful.

- **Explanation of *Simple Logistics* output equations**
  This example explains how to interpret the equations which are given by the *Simple Logistics* algorithm that Weka produces. Consider the following set of equations:

$$Good = 2.7 + 3.4 \times A - 5.7 \times B + 0.3 \times C \tag{2.4}$$

$$Mean = 4.2 + 1.5 \times A + 3.1 \times B - 1.7 \times C \tag{2.5}$$

$$Bad = 3.1 - 2.5 \times A + 1.8 \times B - 0.7 \times C \tag{2.6}$$

  Here A, B and C can be, for instance, *Braking rate*, *Average brake pedal angle* and *PKE*. and they are predicting if the driver is a good, mean or bad driver with respect to efficiency. As input to Weka's creation of equations e.g. the mean value over a section, for each of a larger set of parameters and for each driver used. The input set of parameters are likely larger than the ones Weka uses in the output equation. The created equations can be interpreted as :

  - The parameters A, B and C that Weka kept out of a larger set of parameters are the most relevant for the dependent parameter (reference).
  - Providing that all the parameters are normalized: the higher the value is of a constant related to the parameter, the more influential the parameter is.

  In this study, the normalization was simply done by setting all the values between 0 and 1. E.g., considering the following vector: 1, 2, 3, 4, 5. The lowest value (1) is set to 0, while the highest value (5) is set to 1. The normalized vector is then: 0, 0.25, 0.5, 0.75, 1. This doesn't follow the mathematical definition of *normalization*, however this change is used by *Weka* and works fine as it allows the comparison of parameters which have different magnitude.

  When using the equations to classify a new driver, the same type of data is extracted for the same profile the set of equations is aimed at describing. That is, the mean of the parameters over the same type of section is used. When the new drivers data has been run in the equations the result is interpreted as: a bad driver has the highest value in the bad equation, while a good driver has the highest value in the "good" equation.

  Note that since this study had so little data no validation dataset was separated to validate the findings. This means that the equations are only indicative and may be sub-optimized to the dataset.

Weka software configuration

After assessing different driving parameters independently from each others, they were computed together in *Weka*. With this software, one can choose which driving parameters to compute and the classes (dependent parameter the driving parameters are predicting) is discretized. In this case, it was discretized into bad, mean and good drivers.

Indeed, on all the sections, there are 3 clear groups: the 2 internal expert drivers (D and E), the 2 bad drivers (C and K), and the 7 remaining drivers. Based on Volvo internal feedbacks, an estimation of the drivers distribution on the European market had already been made. This study tries to remain close to this distribution:

- Non efficient drivers: non efficient drivers represent around 10 to 15% of the total population.
- Mean drivers: normal drivers represent around 80% of the total population.
- Efficient drivers: efficient drivers represent around 5 to 10% of the total population.

Note however that it is a very rough approximation to generalize to the population from two drivers in each of the two groups . In this study this only is indicative for the drivers used.

Before the iteration process, *Weka* is aware of the number of drivers that each class contain. But it doesn't know who those drivers are: that is why it uses the driving parameters to classify each driver in the suitable class. The classification rate is used to see the percentage of drivers that are correctly classified.

However, classification was defined regarding the reference class (*Eco coefficient* in this case): hence, the choice was mainly restricted by its actual values. On all the driving sections, and regarding their *Eco coefficient*, the drivers were split into 3 classes: indeed, there were clear gaps between the three groups, making two groups less relevant. The configuration for the 5 sections is given in Table 2.3.

The drivers were split in those 3 classes based on their *Eco coefficient* values calculated from the following

Table 2.3: Configuration of classes for the 5 section

Sections	Number of bad drivers	Number of mean drivers	Number of good drivers
1B	2	6	3
2B	3	5	3
4B	1	7	3
111	2	6	3
112	4	5	2
Full cycle	3	5	3

equation:

$$EcoCoefficient = \frac{AverageSpeed}{FuelConsumption} \tag{2.7}$$

Appendix A shows the values taken by this coefficient for all the sections: the choice was quite logical as 3 groups clearly appeared for each section. Unlike Section 2.3.1, *PKE* is not taken into account here: the objective is just to get a first approximation of each driver's efficiency based on their speed and consumption.

2.3 Driving criteria and parameters

2.3.1 Parameter used as a reference

The reference parameter (dependent parameter) is very important as it is a basic way to evaluate drivers regarding their efficiency. For instance, the reference parameter was used to determine the classes given in Subsection 2.2.2: this was, however, possible only because those drivers were driving the same truck, on the same road, making the comparison consistent. Indeed, in terms of efficiency, a driver who drives a light truck on a highway can't be compared to another driver driving a refuse truck, as the driving conditions are too different.

The study was initially based on *Fuel consumption* only. However, the results were hard to synthesize as the correlations found were too low and too sensitive. In addition, this was not representative enough of the driving efficiency since it was not taking at least speed into account. That is the reason why it was decided to create a new parameter out of both fuel consumption and average speed.

All along this study, this parameter called *Eco coefficient* was used as a reference parameter (dependent). This coefficient is particularly interesting as it takes into account the *Fuel consumption*, but also the *Average speed*, for the following reasons:

- Fuel consumption: due to the rising cost of fuel and its high role in the fleet expenses, fuel consumption has become the most important parameter to optimize.
- Average speed: truck drivers are ruled by their delivery schedule. Even if they have to focus on lowering their fuel consumption, they need to keep on being quick enough to deliver goods in time. In addition to that, it is worth mentioning that driving and working time is regulated.

A first version of this coefficient was taking into account those two parameters only. However, a second version was introduced, where the parameter PKE was also taken into account.

PKE is defined in Subsection 2.3.2. This parameter shows the drivers' ability to use their truck's kinetic energy as part of their driving. As a consequence, this parameter is low for good drivers. This means that they accelerate less often and with less amplitude than non efficient drivers. That is the reason why the whole *Eco coefficient* is to be maximized by drivers to be overall good rational drivers.

Finally, *Eco coefficient* is defined by the following equation:

$$EcoCoefficient = \frac{AverageSpeed}{PKE \times FuelConsumption} \tag{2.8}$$

As a consequence, fuel consumption needs to be reduced without affecting average speed: thus, best drivers are the ones who have a high *Eco coefficient*. It means that they combine both a good fuel consumption, with a good average speed. Indeed, a driver with a very low fuel consumption can't be considered as a good driver if her/his average speed is too low: this would cause her/him miss to delay deliveries, which is extremely costly for a company.

2.3.2 Definition of the types of criteria

All along this study, the driving behaviour is split into 5 different criteria. Hence, each criterion refers to an aspect of driving behaviour. Five main types of criteria can be identified. They are defined regarding conditions that don't depend on a specific truck but on a vehicle state:

- Braking: a braking phase occurs each time the drivers press the brake pedal. In this study, only relevant phases are considered (defined as braking longer than 1 sec and with a pedal angle higher than 5%). In Appendix B, a curve shows the reasons for this value. Indeed, a value of 20% was initially chosen (as a feedback from the previous master thesis mentioned in section 1.4), but the actual braking curves happened to oscillate around this 20% threshold. The same problem was noticed for lower values, and 5% finally happened to be a good compromise. Indeed, it avoids oscillations while making sure that the driver actually intends to brake. That is the reason why it gives a more accurate estimation of the actual braking events.

- Anticipation: planning, predicting and taking decisions based on the surrounding environment when an event is about to happen. It is calculated by taking the time between the end of a traction phase and the beginning of a braking phase.
  Traction phase:

$$EngineTorque \geq 0 \tag{2.9}$$
$$VehicleSpeed \geq 0 \tag{2.10}$$
$$FuelConsumption \neq 0 \tag{2.11}$$
$$BrakePedal = 0 \tag{2.12}$$

  Braking phase:

$$BrakePedal \geq 5\% \tag{2.13}$$
$$VehicleSpeed \geq 0 \tag{2.14}$$

  The *Anticipation* phenomenon is illustrated in Figure 2.8 with a timeline.

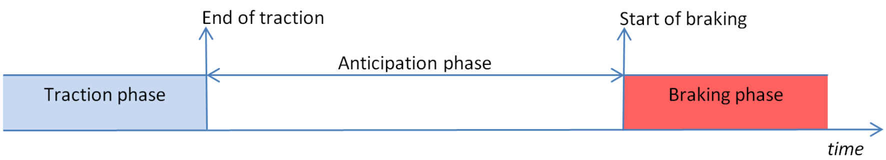

Figure 2.8: *Anticipation phenomenon*

- Coasting: refers to driving phases where nothing but kinetic energy contributes to move the vehicle forward. It is calculated by taking the time spent between two traction phases, when driving downhill for instance.

$$BrakePedal = 0 \tag{2.15}$$
$$FuelConsumption = 0 \tag{2.16}$$
$$VehicleSpeed \geq 0 \tag{2.17}$$
$$EngineTorque = 0 \tag{2.18}$$
$$RetarderTorque = 0 \tag{2.19}$$

- Even driving: refers to the fact of keeping the vehicle speed as constant as possible. Parameters related to sudden speed change are also used in this report: they refer to the phases outside of the range below. The threshold comes from the study mentioned in Section 1.4.

$$|\nabla VehicleSpeed| \leq 1km/h/sec \tag{2.20}$$
$$VehicleSpeed \geq 2km/h \tag{2.21}$$

- PKE: this refers to the way drivers use their truck's kinetic energy. This criterion is special as it doesn't correspond to any kind of driver reaction, but it was decided before the thesis to consider it as an independent criterion. PKE is defined as *the difference between final and initial speeds when the acceleration is positive, divided by total travelled distance*[17]:

$$\frac{\sum \left[(V_f^2 - V_i^2)_{dV/dt>0} \right]}{TotalDistance} \tag{2.22}$$

2.3.3 Driving parameters evaluated

A code was written on Matlab during a previous study. This algorithm computes driving session records as an input, and calculates driving parameters as an output for each criterion. The identification of the relevant driving parameters was made following the process given in Figure 2.9.

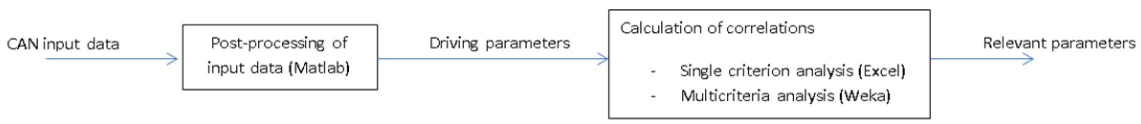

Figure 2.9: *Relevant parameter identification algorithm detailed*

Table 2.4 lists all the parameters calculated during the study, for the identification of relevant driving parameters. Further tables are given in Appendix E: those tables refer to the parameters calculated in Section 2.4, so related to road features. In this study:

- *Rate* (Coasting rate for instance) refers to the ratio between event total duration and the total cycle duration.

- *Number of events* (Braking for instance) refers to the total number of events observed on the cycle.

- *Delta* refers to the difference between initial and final speed, for an anticipation phase for instance.

- *Average* refers to the average value taken by a parameter for all the occurrences of a same cycle.

It is also worth noting that *Sudden speed change* is classified into the *Even driving* criterion even if it refers to speed change phases that are beyond a given range (+/- 1 km/h per second for instance).

Table 2.4: Output from Matlab code

Criteria	Driving Parameters
Anticipation	Duration of events (Average) - sec
	Number of events (Average)
	Initial speed of event (Average) - km/h
	Final speed of event (Average) - km/h
	Kinetic energy spent during event (Average) - J/kg
	Duration of event * Mean speed during event (Average) - s*km/h
	Duration of event * Initial speed of event (Average) - s*km/h
Braking	Rate for Pedal Angle > 5% - %
	Rate for Pedal Angle > 10% - %
	Rate for Pedal Angle > 20% - %
	Number of events for Pedal Angle > 5%
	Number of events for Pedal Angle > 10%
	Number of events for Pedal Angle > 20%
	Duration of events (Average) - sec
	Duration of event * Mean speed during the event (Average) - s*km/h
	Duration of event * Initial speed of the event (Average) - s*km/h
	Initial speed of event (Average) - km/h
	Final speed of event (Average) - km/h
	Delta speed of event (Average) - km/h
	Deceleration during event (Average) - $(m/sec)^2$
	Kinetic energy spent during event (Average) - J/kg
	Rate for Retarder Use - %
	Number of events for Retarder
Coasting	Number of events
	Rate - %
	Duration of events (Average) - sec
	Final speed of event (Average) * Brake pedal angle (Average) - deg*km/h
	Initial speed of event (Average) - km/h
	Final speed of event (Average) - km/h
	Delta speed of event (Average) - km/h
	Kinetic energy spent during event (Average) - J/kg
Even Driving	Duration of events (Average) - sec
	Number of events (Average)
	Rate - %
	Duration of events for Speed variations (Average) - sec
	Number of events for Speed variations (Average)
	Rate for Speed variations (Average) - %
Others	Positive Kinetic Energy (PKE) - $(m/sec)^2$
	Engine torque (Average) - Nm
	Retarder torque (Average) - Nm
	Throttle pedal angle (Average) - deg
	Brake pedal angle (Average) - deg
	Throttle pedal angle / Brake pedal angle (Average)
	Throttle pedal variations (Sum) - deg/sec
	Brake pedal variations (Sum) - deg/sec
	Throttle pedal variations (Sum) * Brake pedal variations (Sum) - $(deg/sec)^2$
	Braking force (Average) - N

2.4 Study of road features

2.4.1 Road sections considered

In Sections 3.2 and 3.3, only parameters related to drivers behaviour are taken into account. However, it is also important to consider their environment. This part aims to find out what driving parameters are relevant when facing different types of event, more specifically roundabouts, curves and large changes in altitude. As the code written for the previous parts couldn't be used for those specific cases, a new code was written, and many parameters were calculated. The 11 drivers used previously are used again for this part: the objectives are the same, as one aims to determine relevant parameters that identifies good, mean and bad drivers behaviour. In this part, 3 of these events are investigated:

- **Roundabouts**: in order to assess this criterion, it was decided to focus on 1B. Indeed, this section is short and simple, and 3 roundabouts can easily be extracted out of it. Figure 2.10 shows a detail of this section. Section 1B is 5.8 km long, and is made of a straight line interspersed of 3 roundabouts: 2 of

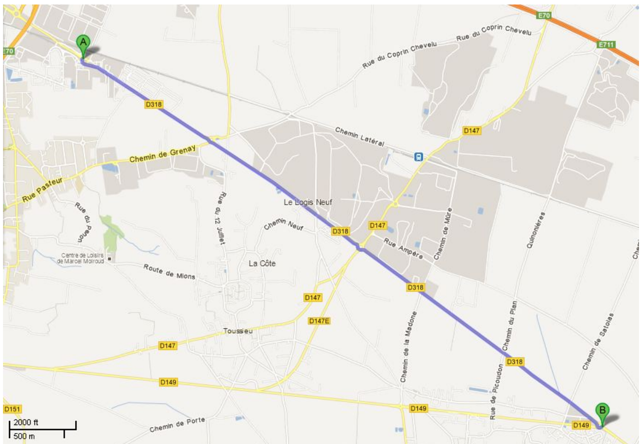

Figure 2.10: *Section 1B used for evaluating drivers behaviour face to roundabouts*

  them are crossed (180°), and drivers make a U-turn on the last one. In order to get the most reliable data, and thanks to video recording of driving sessions, it was decided not to take into account events polluted by disturbances: most of them were caused by traffic jam. Unfortunately, out of 33 events (3 roundabouts and 11 drivers), only 16 were clean enough to be used. However, to deepen the study, the study was also carried out on Sections 111 and 112: indeed, each of them contains one 270°roundabout.

- **Curve radius**: in order to assess this criterion, it was decided to focus on sections 111 and 112 (urban environment). These sections are short and simple as well, as shown on Figures 2.11 and 2.12. Section

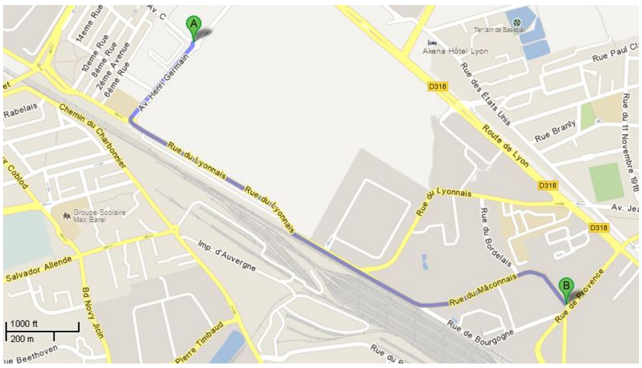

Figure 2.11: *Section 111 used for evaluating drivers behaviour face to curves*

  111 is 2.6 km long, and is made of 1 roundabout and 3 corners. As it is in an urban environment, the speed limitation is 50 km/h. After the roundabout, there is one slight baffle whose first corner is about

30°. The second event is a long curve of about 60°. The last corner is a 70°corner followed by a traffic light, and traffic jam remains very limited. Figure 2.12 shows Section 112: this section is 5.2 km long,

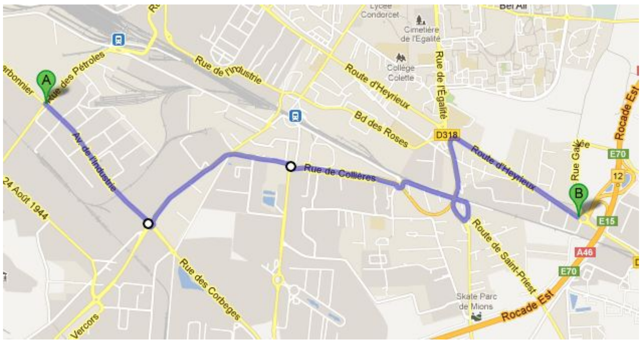

Figure 2.12: *Section 112 used for evaluating drivers behaviour face to curves*

ans is made as well of 1 roundabout and 3 corners. The speed limitation is still 50 km/h. It starts with a 270°roundabout followed by a slight lane change on the right, of about 50°. Drivers then have to take a 430°loop crossing an intersection. The final corner is a 130°curve laying around a roundabout.

- **Road altitude**: in order to assess this criterion, it was decided to focus on 2B. This section was designed on purpose as it consists in a straight line with high altitude variations. Once again, it is a short section of 3.4 km, detailed on Figure 2.13. For such a study, it is important to use a reliable slope signal,

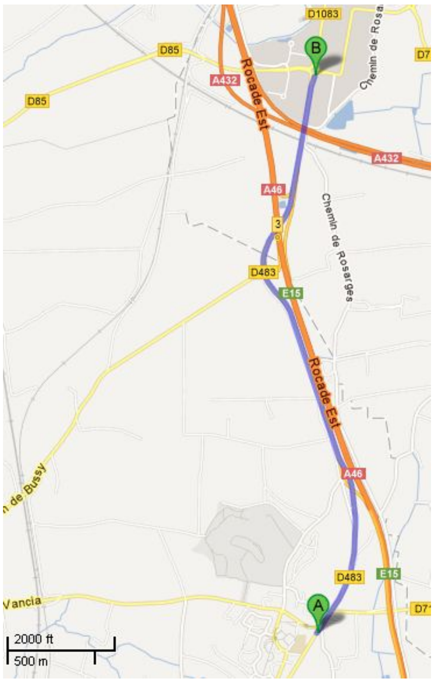

Figure 2.13: *Section 2B used for evaluating drivers behaviour face to altitude variation*

which is not the case of GPS output: if longitude and latitude are very accurate, altitude signal is approximative and there are some big differences between drivers on a same cycle. That is the reason why it has been decided to use an input signal generated by an internal software. Figure 2.14 shows the difference between GPS data (right) and internal software signal (left). All slopes weren't considered: an algorithm, function of vehicle speed, was used to determine whether a slope was relevant or not. This determination is based on a very simple linear model. It depends on the speed and is shown in Appendix C. For uphill phases, the higher the speed is, the lower the threshold at which a slope is considered as important is. For downhill phases, it is the opposite, due to inertia. For most of the drivers, four relevant uphill phases and five downhill were considered on this section.

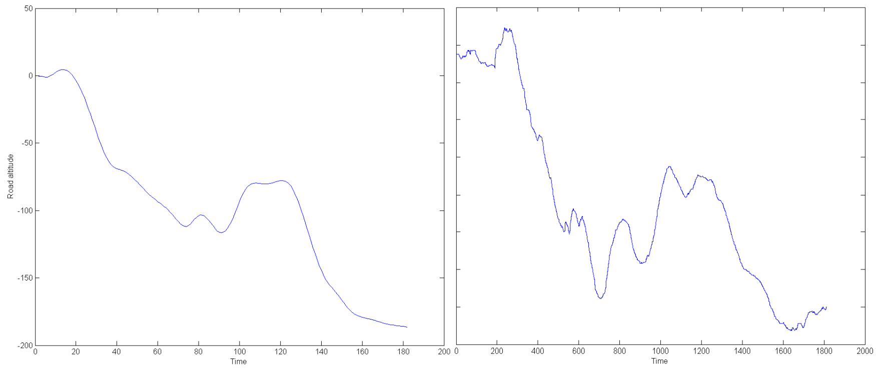

Figure 2.14: *Comparision of road altitude using 2 methods*

In this part, *Eco coefficient* was used as a reference, for two reasons: first, to keep the same reference as section 3.2, and then because the correlations were higher than with *Fuel consumption*. An exception was made for *Section 111*: this section has the particularity of having a strong correlation between *Fuel consumption* and *Average speed*. As *Eco coefficient* is the ratio of those two quantities, the effect is cancelled out and the results become senseless. That is the reason why *Fuel consumption* was used for this section. Whatever the reference, it is calculated on the complete section for each driver. This means that each driver has 5 eco coefficient or fuel consumption values, related to the 5 sections.

2.4.2 Limitations and assumptions

The whole parameter list which was investigated during this study are given in Appendix E. However, some assumptions were made and may need to be explained for a better understanding.

During this study, the exact place where the event takes place is defined as:

- For a roundabout, the GPS coordinates of the centre of the roundabout. Those coordinates were found thanks to *Google Map*.
- For a turn, the GPS coordinates were measured in the middle of the turn. Once again, *Google Map* was used to get those coordinates.

Study of road altitude variations

In this part, road altitude variations are split into two: uphill and downhill phases, and those phases are split into two as well. Indeed, one distinguish 2 parts: a first one where curve percentage is increasing, and a second one where it is decreasing. Those 2 parts are shown on Figure 2.15. It was decided to make the difference between behaviour on the whole downhill/uphill section (*on downhill phases*) and on peaks (*once uphill*) as they affect the results a lot in spite of their limited dependence.

Study of roundabouts

Section 1B used for this analysis is made of 3 roundabouts homogeneously spread enough to split the section into 3 parts. Those parts are much longer than the event, and it actually allows to target easily braking and throttle events related to each roundabout, knowing its GPS coordinates. The main difficulty is that only 16 events out of 33 can be considered: the remaining events are troubled by disturbances such as traffic jam. As a results, for each event, there is only a limited of values. Hence, it was decided to analyse average vectors of events, in order to maximize the reliability and the number of values. Another consequence is that 10 drivers only are considered as one of them faced disturbances in all 3 roundabouts.

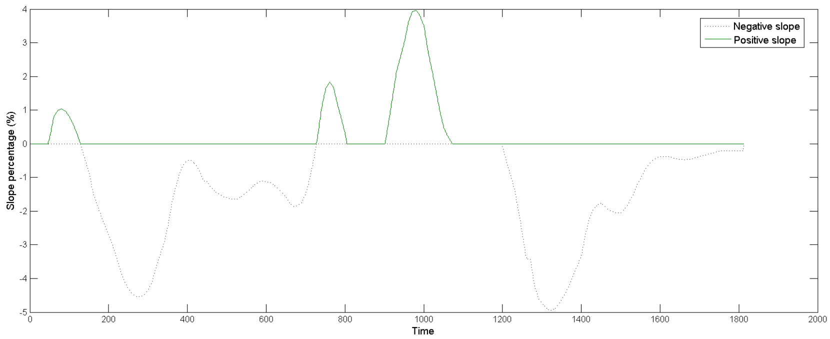

Figure 2.15: *Negative and positive slope curves*

Study of curves

For this section, it was decided to start considering data 200 meters before the event and to stop 100 m after: it is a good compromise as it happened to be enough to represent driver behaviour on events. As the 2 sections used for this analysis have some similarities, it was decided to group events of both of them so that analyses can be run in parallel and results can be compared:

- Roundabouts of both sections are grouped: indeed, they are similar and speed profiles are also alike, as shown on Figure 2.16.

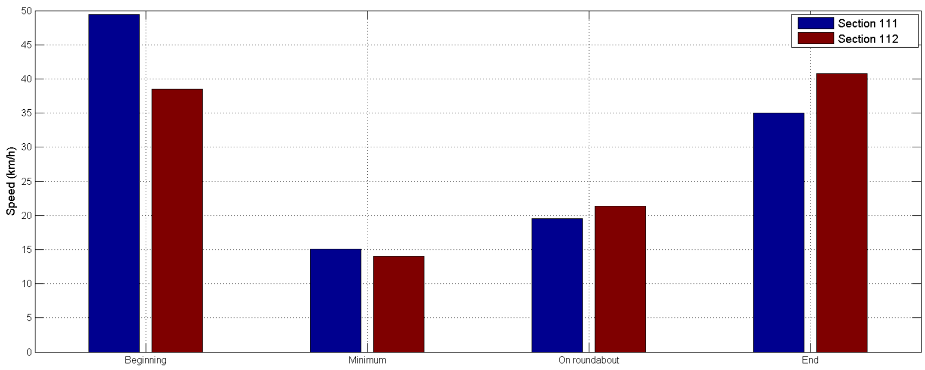

Figure 2.16: *Speed profiles for roundabouts on sections 111 and 112*

- Curves 2 for section 112 and curve 3 for section 111 are also grouped: curve shape is very close, as well as angle. But speed profiles and variations along the corner are also similar, as shown on Figure 2.17.

The vehicle speed profile plays an important role in this study. Many drivig parameters are investigated but the ones related to the speed are particularly important. The are three important speeds to measure: before, during and after the event. Those three speeds were finally combined in a single parameter: $Speedduringevent^2/(Initialspeed \times Finalspeed)$. The speed on the event is more important, that is the reason why it is squared. That is also a way to get an adimensional coefficient.

Efficient drivers usually arrive slowly on an event, keep their speed on the event, and then accelerate in a smooth way. On the opposite, non efficient drivers arrive faster, brake in order to reach a low speed on the event, and then accelerate strongly again. That is to differentiate those two behaviours that the coefficient is defined as a ratio.

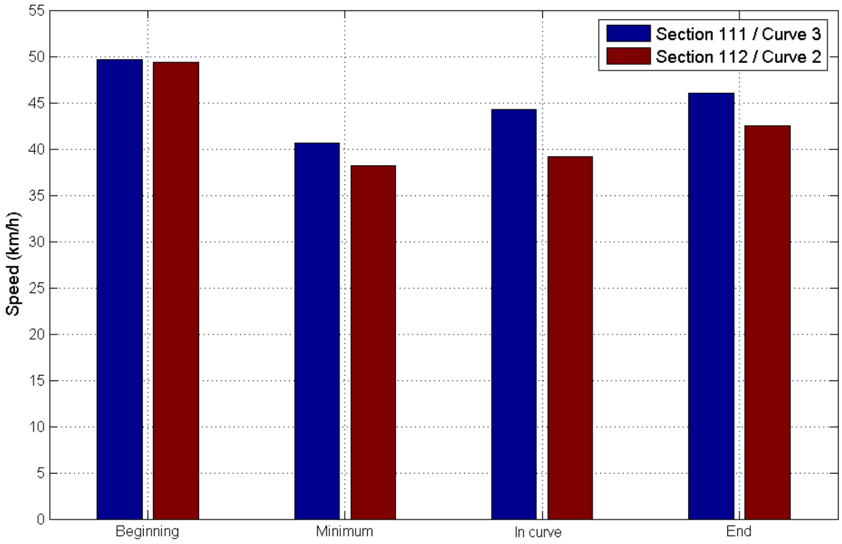

Figure 2.17: *Speed profiles for curves 2 on section 112 and 3 on section 111*

2.5 Creation of a grading system for truck drivers efficiency ranking

The objective of this section is to set up a grading system which would give drivers a grade based on their driving parameters. Doing so is possible thanks to the fact that relevant parameters are identified in the previous parts. Hence, the advantage of such a system is that drivers are ranked with parameters that are predictive of a combination of fuel consumption, speed and PKE. That is, external factors i.e. wind are not affecting the result, only the driver behaviour.

Indeed, the easiest way would be to rank drivers regarding their *Eco coefficient*. But this depends a lot on the actual driving conditions, which can vary a lot (not in the current case, as those conditions are the same for all the drivers). That is the reason why a grade, independent from those conditions, and based on driving performance, is a better representation of drivers' energetic efficiency.

2.5.1 System based on Simple Logistics Equation

Simple Logistics algorithm generates a set of output linear equations depending on driving parameters. Those equations are found for each of the 5 sections, but also for the full cycle. However, they differ a lot, as each section represents a specific driving environment. That is the reason why the driver ranking cannot be made with only one equation.

As explained in Section 2.2.2, *Simple Logistics* generates one equation for each class, i.e. 3 equations for the current study. As there are 5 sections, it means that the total system is made of 15 equations. The equations related to each class are weighted regarding the *section length* and summed:

$$Equation = \sum_{i=1}^{N=5} W_i \times \sum_{j=1}^{M} x_{i,j} \times C_{i,j} \tag{2.23}$$

Where W_i is the weight of the section i, $C_{i,j}$ the coefficient multiplying the driving parameter, and $x_{i,j}$ is the driving parameter j for the section i (for instance, *Braking rate* on Section *4B*).
The weight of each section is calculated as follows:

$$W_i = \frac{Length_i}{\sum_{i=1}^{N=5} Length_i} \tag{2.24}$$

The final system consists of 3 equations, and the global parameters, i.e. for the full cycle, are taken in input. If the system works, bad drivers (respectively mean drivers, good drivers) have the highest resulting value for the equation related to *Bad drivers* (respectively *Mean drivers*, *Good drivers*).

Each driver is given a grade but this grade is based on 1 of 3 equations (the equation related to the driver's class: an efficient driver gets his grade from the equation related to efficient drivers for instance). This method is interesting when it comes to classifying drivers into classes, but not to get an absolute grade. Especially as the classification is based on the higher grade for each equation.

2.5.2 System based on single criterion analysis

The previous method is interesting because it is very simple, however it may not be powerful enough. That is the reason for this second method to set up a driver ranking. Instead of using *Simple Logistics* equations, this method is based on single criterion analysis.

The complete process is summed up in Figure 2.18, assuming that the relevant parameters for the global cycle are already known in input.

Figure 2.18: *Driver ranking algorithm detailed*

First of all, a set of parameters giving strong correlations on the full cycle was identified by evaluating all the driving parameters and keeping only the relevant ones. Those parameters were also chosen so that all the driving criteria were represented. They were then calculated for the 5 sections. There were few cases where the correlations were too low for being really relevant: those cases were not considered. For each parameter corresponding to each section, the maximum and minimum values were written down in a table.

It would have been preferable to calculate the full range of driving parameters on the 5 sections, however for computation time reasons, it was not possible. The computation of the whole section took a while, but had to be launched only one time. However, even though this computation would have been shorter on the sections (as they are shorter than the full cycle), it would have had to be run 5 times (as there are 5 sections). It was decided to use only those parameters that with single criterion analysis were relevant when applied to the full cycle. That is, a list of parameters were extracted, this list was then used for the grading equation creation for the individual sections. This assumption may not be completely right, but accurate enough to be made, and interesting when considering the gain in terms of computation time.

The parameters were then normalized. The whole set of parameters was then classified with *Weka* and each parameter was given a coefficient of classification to signify its relative importance compared to the other parameters (for instance a parameter with a coefficient of 1.5 is considered as more important than another parameter with a coefficient of 0.7). For each section:

$$Pm_{i,j,k} = \frac{X_{i,j,worst} - x_{i,j,k}}{X_{i,j,worst} - X_{i,j,best}} \tag{2.25}$$

The global score on the section is then calculated by summing all parameters:

$$Score_{j,k} = \frac{\sum_{j=1}^{M} C_{i,j} \times Pm_{i,j,k}}{\sum_{j=1}^{M} C_{i,j}} \tag{2.26}$$

Where $Pn_{i,j,k}$ is the value taken on the section i by the parameter j for driver k after its normalization, $X_{i,j,worst}$ and $X_{i,j,best}$ are the lowest (respectively the best) value among the 11 drivers, $x_{i,j,k}$ is the value of the parameter for the driver k, and $C_{i,j}$ is the coefficient of classification.

The score obtained is between 0 and 10 . A global score can then be calculated out of the five scores, where W_i is described in Equation 2.24:

$$Score_{gobal,k} = \sum_{i=1}^{N=5} Score_{i,k} \times W_i \tag{2.27}$$

This global score is between 0 and 10 as well, which easily allows a comparison between drivers. Those limits were chosen arbitrarily, in order to provide an easy format. The grade is based on a complete set of driving parameters and not only on fuel consumption or mean speed, which makes it more consistent to the global driving behaviour, and less dependent on external parameters such as the wind for instance.

3 Results

3.1 Drivers distribution within Customer 4's records

3.1.1 Generalities

In Figure 3.1, one can see that the 11 drivers have homogeneous average speed, but one driver is much higher in terms of fuel consumption:

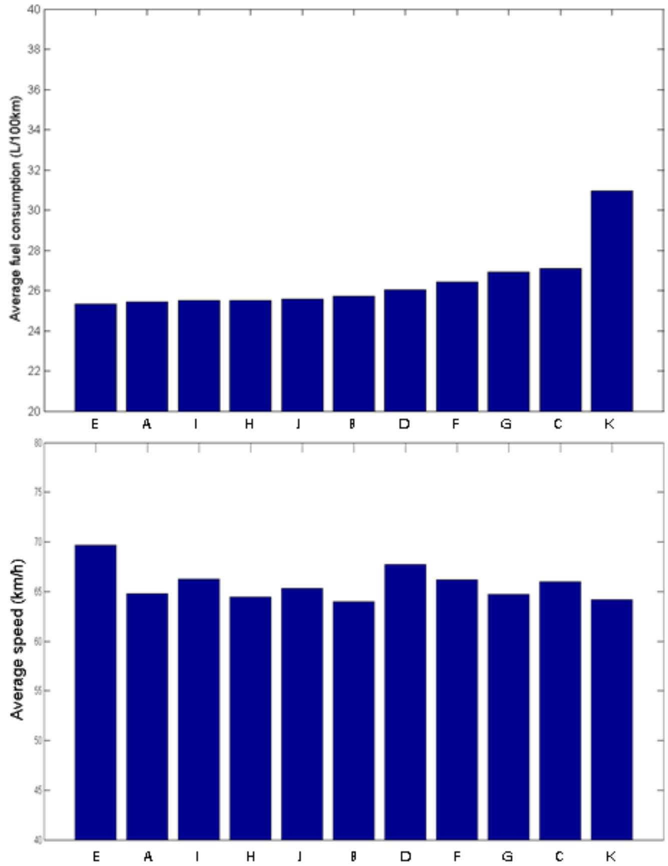

Figure 3.1: *Drivers' average fuel consumption and average speed on the cycle*

3.1.2 Specificities

In the previous study, fuel consumption values were more constantly spread. In the current one, these values are very close to each others, except for K whose consumption is much higher. K's high consumption is remarkable and creates a sort of non-homogeneity among the values, which can make the analysis much more complicated. Indeed, having one single driver with such a fuel consumption while the 10 other drivers have close results can affect the results.

The major problem with this customer is that it contains very few values. Indeed, with only 11 drivers, it is hard to generalize the conclusions. Statistical studies are usually carried out with several hundreds of values at least, so that 1 single driver doesn't influence the results. However, when only few drivers are investigated, a big sensitivity is expected, and it is hard to find. In such a case, it is much harder to detect non-normal behaviours.

In Figure 3.2, variations of fuel consumption for each driver compared to the average value, for both the previous and the current set of data, are shown, making the difference more obvious. On each graph, the 2 horizontal lines represent the highest and lowest fuel consumption values. K has an extreme non efficient

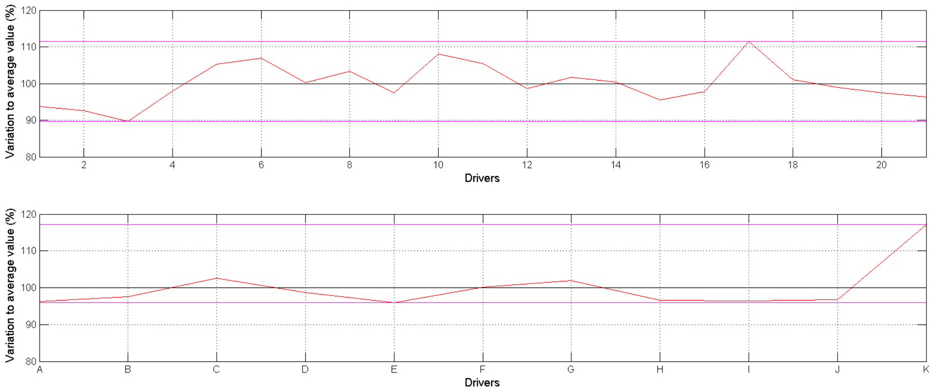

Figure 3.2: *Variations of fuel consumption for each driver compared to the average value - Top: previous, bottom: current*

behaviour: the next parts show that this problem is sometimes hard to deal with as the influence of such a behaviour is big. However, *K* wasn't the only driver causing troubles. *C* had also a strange behaviour, more difficult to apprehend. Indeed, especially on highway cycles, *C* was acting as a good driver, coasting a lot for instance. But his behaviour was too extreme and his coasting value were dramatically high compared to other drivers (he was coasting 3 times more than the average). If *K* was amplifying the results through his stereotyped behaviour, *C* was lowering them as he had high fuel consumption values but was apparently acting efficiently.

As the complete sample only contains 11 drivers, one cannot reasonably get rid of a record: that is the reason why *K* was kept. However, for in some tests, *C* was not taken into account. This is because his behaviour and his performance are in contradiction. Especially, some investigations are to be done to figure out if *C* actually faced disturbances (wind) during his driving session, which could explain the problem.

3.2 Identification of the relevant driving parameters

3.2.1 Evaluation of the basic driving parameters

The 11 drivers' *Eco coefficient* values, the referent parameter introduced in Subsection 2.3.1, are shown in Figure 3.3. A range of simulations was run over the different driving parameters, with a huge impact of sensitivity. This problem is particularly true for *K* or C. As written in Subsection 2.1.2, *K* behaves like a non-efficient driver, and is easily predictable. But *C* behaves well, has high coasting and low braking rates for instance, while having a high fuel consumption: this is much harder to understand, and makes this driver very specific. That is the reason why it was somehow expected to find it apart on the clouds.

The connection between fuel consumption and average speed is not obvious, as can be seen in Figure 3.4. Indeed, the 2 expert drivers manage to be fast and efficient. However, if good drivers are fast, it does not mean that fast drivers are good. As a consequence, the correlation between these 2 parameters is quite low. The results of the new range of iterations are then compared, in Figure 3.5 to the ones obtained from the correlation with fuel consumption, including K. Appendix G goes over how the study based on *Fuel consumption* was done. Linear regressions were calculated for the complete set of driving parameters: *K* was causing the regression to change sign for 26% of the parameters (instead of 70% when *Fuel consumption* was used as a reference). The sensitivity is much lower by using *Eco coefficient*. The rate remains high, but it is now easier to draw conclusions. Conclusions can be written for each criterion:

- **Braking**: correlation values are a little bit lower than what was found with *Fuel consumption*. However, the main concern, discussed previously, is whether results are sensitive to *K* or not: as said above, this

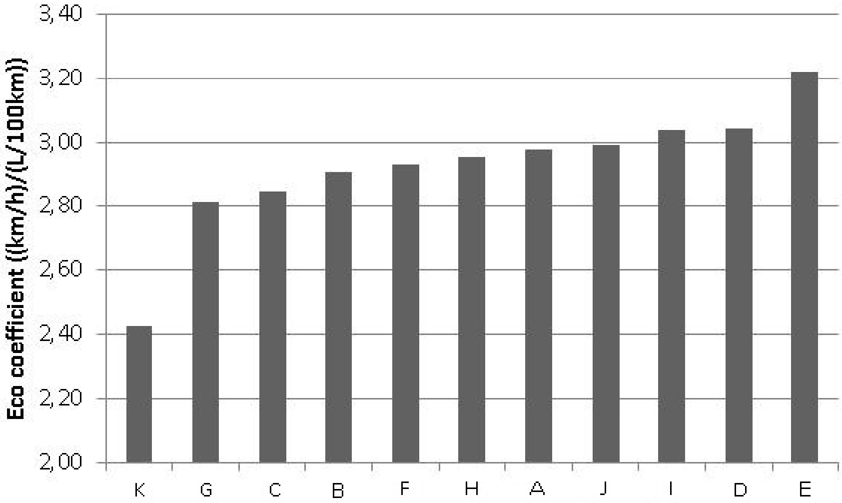

Figure 3.3: *Eco coefficient for the 11 drivers*

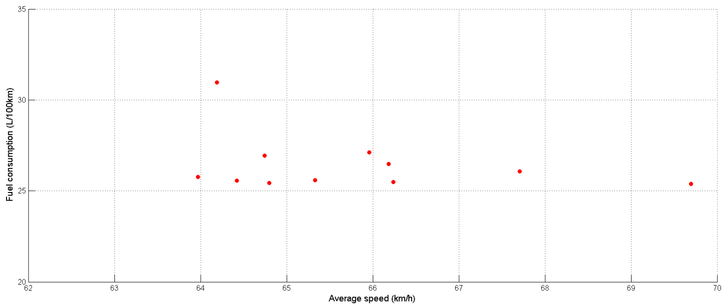

Figure 3.4: *Average speed and fuel consumption of the set of drivers*

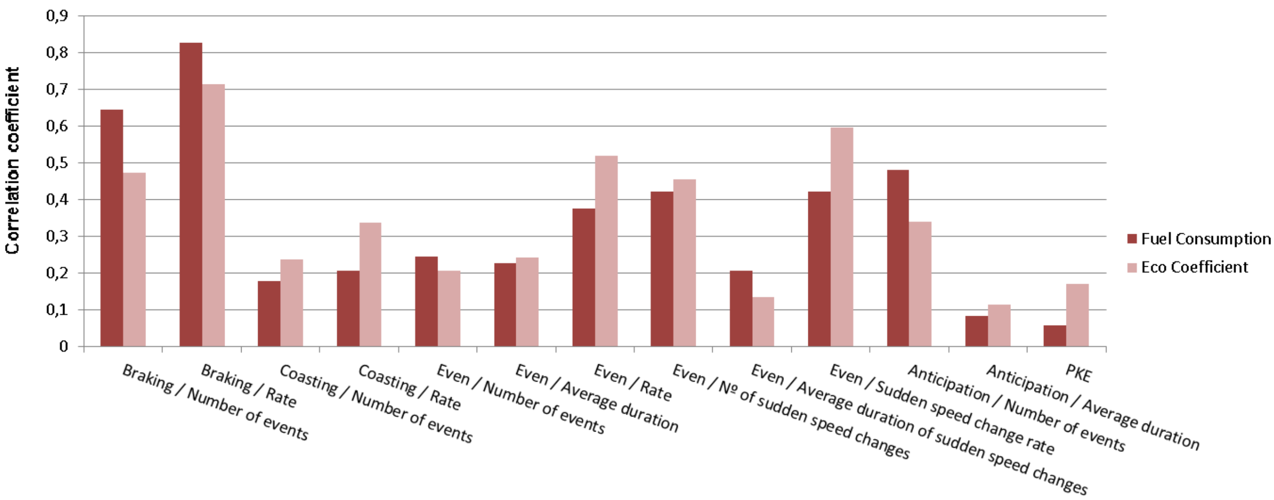

Figure 3.5: *Results found from correlation with fuel consumption and eco coefficient*

problem is partly solved for most of the criteria but remains remarkable for the *Braking* one. Finally, further parameters will be investigated in the next parts in order to find relevant results.

- **Coasting**: results are better when *Eco coefficient* is chosen as a reference. The difference is slight, but the sensitivity to K is reduced. Without C, the trend is almost linear, and a value of 0.36 is found for *Coasting rate*.
- **Even driving**: values found previously were already good, and they are even better now. The 2 most relevant parameters are *Even driving rate* and *Sudden speed change rate*, however *Number of sudden speed change phases* is also interesting.
- **Anticipation**: as for *Braking*, the correlations are a bit lower than what was found with *Fuel consumption*. There is no parameter giving strong results, and it is hard to draw conclusions. For the moment, *Number of events* is kept. However, a first thought would tend to think that a high number of events is better, but in fact good drivers have a low number of events (and poor drivers, a high number), but higher average duration. As for *Coasting*, there is a high influence of C, and the results are better without him.
- **PKE**: the correlations are still lower than expected, considering what they were in the previous study. In spite of these disappointing results, there is a clear trend showing that best drivers manage to keep a very low PKE, i.e. to lower their acceleration amplitudes. Indeed, for this criterion, the influence of K is a little bit lower but remains high.

Finally, if using *Eco coefficient* instead of *Fuel consumption* as a reference doesn't dramatically improve the results, the sensitivity is lowered. In addition to that, this reference is more relevant as it also takes into account the average speed.

However, some correlations remain to find, especially for *Braking*: that is the reason why further parameters are investigated in the following part.

3.2.2 Evaluation of further driving parameters

In the previous part, some interesting results were found. Nevertheless, another parameters are yet to be investigated: indeed, the correlations remain too low for some criteria, such as *Braking* for instance. These parameters, for most of them, are new: hence, they can't be compared with previous study. However, the previous parts have shown that there is a very limited connection between the results of both studies.

Mean Braking Force In order to finally obtain a reliable driving parameter for *Braking*, it was decided to determine the mean braking force applied by each driver during the cycle: indeed, on today's truck, there is only a low connection between brake pedal angle and actual braking force. Driving force can't be obtained from the CAN bus, however it can be calculated from the parameters recorded and computed with the equations given in Appendix F.

To each driver corresponds one mean braking value, as shown in Figure 3.6. This graph is interesting as

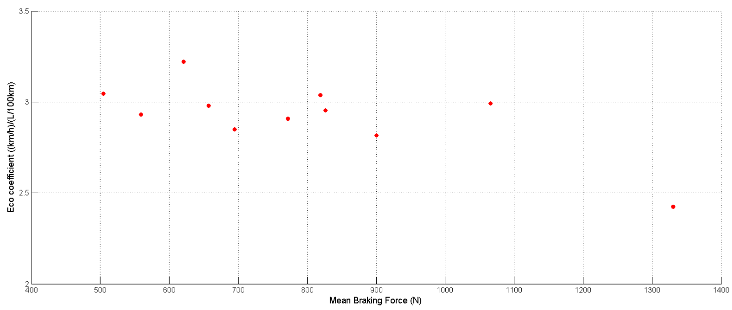

Figure 3.6: *Drivers' mean braking force on cycle*

it shows a good trend for a braking parameter: indeed, as one could guess, the higher the braking force, the lower the *Eco coefficient*. The trend isn't bad but once again, there is a very high dependence on K (bottom right dot), that makes this parameter too sensitive.

Braking and Anticipation speeds It was decided to take a look at initial and final speeds, for both *Braking* and *Anticipation* criteria, as shown in Figure 3.7. Those parameters don't give strong enough trends:

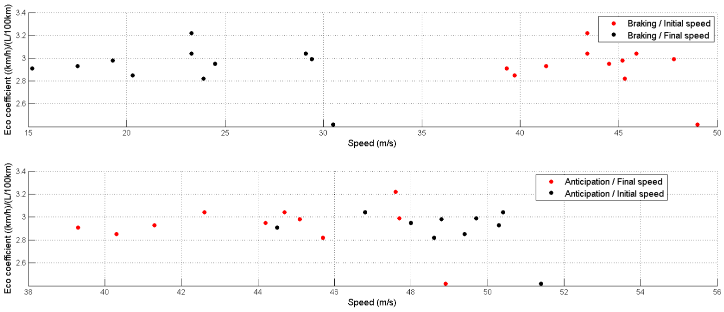

Figure 3.7: *Initial and final speeds for Braking and Anticipation criteria*

for this reason, those parameters are not kept. It is also important to mention that many iterations were carried out on various parameters related to those 2 criteria (durations, speeds, accelerations, etc.), as well as combinations: however, none of them were relevant. The correlations were as low and sensitive as those for the speeds just above. One combination was finally found to give strong connections: it is detailed in the next paragraph.

Driver behaviour related parameters One problem, especially true for *Coasting*, was the binary behaviour of C in terms of torque demand. Several attempts to take into account this particularity without affecting the results were made, concerning pedal solicitations. One way to characterize drivers' behaviour can be to figure out how often and how strong they press throttle and brake pedals: that is the reason why these 2 pedal variations were calculated all along the cycle and summed in absolute value. They somehow show the brutal behaviour that some drivers can have, whereas good drivers are very smooth. The following parameters were calculated:

- Average throttle percentage
- Average brake pedal percentage
- Sum of throttle gradient
- Sum of brake pedal gradient
- Sum and product of the 2 quantities above

These parameters, especially the average brake pedal angle and the product of gradients, gave satisfying results, as shown in Figure 3.8 (the product of gradients is divided by 10^6 as actual values are huge).

On these 2 graphs, a clear trend shows that a driver who has a high pedal activity has a lower *Eco coefficient*: the nervousness can, in this case, be measured through the gradient quantity, and the results are much better when considering the brake pedal.

Combination of parameters As it turned out to be tough to find relevant parameters for *Braking* and *Anticipation* criteria, a combination of parameters was made: *Anticipation final speed* $\times$*Average brake pedal angle*. Indeed, those 2 parameters are minimized by efficient drivers: they manage to end up their anticipation phases at a low enough speed, so that they don't have to brake too strong. Results are shown in Figure 3.9, they are globally satisfying for most drivers.

For *Coasting*, *Initial* and *Final speeds* were squared and combined into a kinetic energy parameter: however, the weight is not taken into account as it is constant on the whole cycle for all the drivers. Results are shown

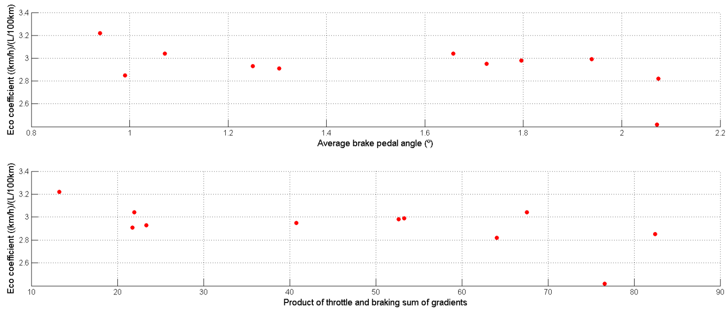

Figure 3.8: *Driver behaviour related parameters*

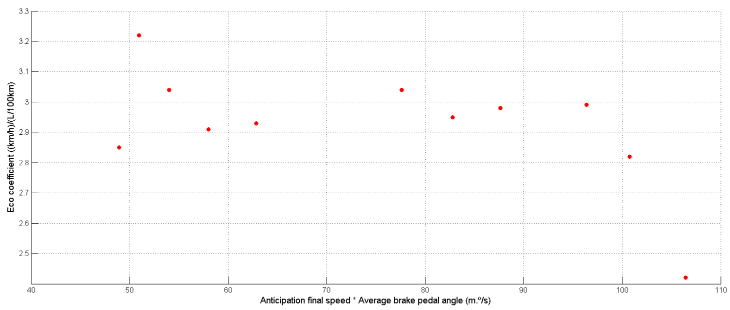

Figure 3.9: *Anticipation final speed * Average brake pedal angle*

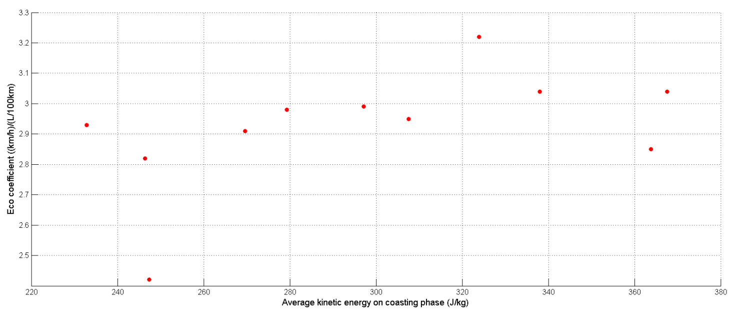

Figure 3.10: *Average kinetic energy on coasting phases*

in Figure 3.10. This parameter shows with success how much energy drivers manage to spend in coasting phases: good drivers have it much higher than non-efficient drivers.

3.2.3 Multicriteria analysis

In the following part, the results of the multicriteria analysis carried out with *Weka* are shortly presented. They are based on 2 methods: a decision tree (*J48* algorithm) and a set of equations (*Simple logistics* algorithm). Those methods are described in Section 2.2.1. The results are commented results are given for each of the 5 sections, and also for the global cycle. Those equations and decision trees are shown in Appendix E.

In this part, not all the parameters were used. Indeed, for computation time reason, the additional parameters were calculated for the full cycle only, but not on the 5 sections separately. That is the reason why Subsection 3.2.3 contains only basic parameters while Subsection 3.2.3 contains all the parameters.

Analysis per section

Section 1B In this section, only 1 driver is classified in the wrong class. Apart from that, *Speed change rate* plays an important role, as a high rate is enough to determine non-efficient drivers. To determine mean and efficient drivers, the algorithm used the number of events of *Braking pedal angle* higher than 5%, as well as *Average duration of anticipation*. Simple Logistics equations gave a perfect score of 11 drivers correctly classified, and especially shows that non-efficient driver have a very high *Speed change rate*.

The *Time* parameter isn't really relevant as it just shows that efficient drivers are fast, what doesn't mean that fast drivers are efficient. Again, the low number of driver can account for this.

Section 2B Once again, only 1 driver is classified in the wrong class. However, in this case, *Coasting rate* is the most relevant parameter: indeed, efficient driver have a high coasting rate. Then, *Even driving duration* is used to distinguish non-efficient and mean drivers. Simple Logistics equations manage to classify the 11 drivers, and highlights the low *Coasting rate* of non-efficient drivers, contrary to efficient drivers.

Section 4B Out of 11 drivers, only 1 driver was classified in *Non-efficient* instead of *Mean*. In this section, efficient drivers are distinguished by both a low *Braking rate* and a high *Even driving duration*, while the non-efficient driver has this parameter low. Simple Logistics equations manage to classify the 11 drivers, and especially shows that non-efficient drivers tend to have a higher *Number of braking events* compared to other drivers.

Section 111 The J48 algorithm classifies correctly 10 drivers, regarding *PKE* mainly, and *Coasting rate* then. Simple Logistics equations manage to classify only 8 drivers, with *Even driving rate* and *Number of coasting events* especially.

Section 112 Even if 10 drivers are classified, the tree is odd as the same parameter is used 2 times: indeed, *Even duration* is high for efficient drivers, low for non-efficient drivers, and normal in between. Simple Logistics equations manage to classify 10 drivers as well, highlighting the low *PKE* of efficient drivers as well as the high *Number of braking events* of non-efficient drivers.

Analysis on the full cycle

The same analysis was run for the global cycle. For this cycle, the set of parameters was much bigger: indeed, due to calculation time, all the additional parameters detailed in the previous part were calculated for the global cycle only (running them on each of the 5 sections, for the 11 drivers, can take a while).

Basic parameters First of all, basic parameters, such as those assessed above, were taken into account. As seen in several sections before, the two most critical parameters are *Even driving duration*, distinguishing non-efficient drivers, and *Coasting rate*, distinguishing efficient and mean drivers. Once again in this case, 10 out of 11 drivers were correctly classified: this is not perfect, but remains a good score. Simple Logistics algorithm manages to classify all the drivers. It shows the difficulty for non-efficient drivers to keep a constant speed, and shows that efficient drivers brake more times but the global rate is lower.

Additional parameters Again, the additional parameters were added after the basic parameters. Indeed, those parameters did not show good enough results in the analysis described in Section 3.2.

The main parameter is *Average coasting kinetic energy*, and it shows that efficient drivers manage to spend more energy in coasting phases. The second parameter is the one that combines *Anticipation final speed* and *Braking angle*: non efficient drivers have this parameter quite high.

When classifying parameters, it appears that *Product of gradients* and *Coasting average kinetic energy* are both considered as having a stronger influence on the results, which confirms what was written in the previous part.

3.2.4 Summary of the results

The results in this section are as before only based on 11 drivers. The results are therefore sensitive to individual drivers behaviour. The results shown are only indicative of which parameters may be relevant, even though the parameters for which the correlation was good but the sensitivity too high weren't retained.

Single criterion analysis

In the end, Table 3.1 summarizes the results of both the previous study (mentioned in Subsection 1.4.2) and the current one. They are based on the basic parameters only, i.e. Subsection 3.2.1, as the additional parameters were not calculated in the previous study.

The correlations were higher in the previous study, and the hierarchy is also very different. *Braking* is relevant in both cases, but this criterion was too sensitive. The correlations are low for *Anticiaption* as well and show that, as well as for *Coasting*. However, for *Coasting*, the hierarchy is conserved (even if the influence of Driver *C* affected the results a lot). The difference between the driving parameters which belong to a same criterion makes the comparison difficult. Indeed, one can't say that one criterion is above the others.

Table 3.1: Synthesis of coefficient of determination R^2 values for the basic parameters

Driving Parameters	Polynomial Regression Previous Study	Polynomial Regression Current Study
Braking		
Number of events	0.54	0.47
Braking rate	0.42	0.71
Coasting		
Number of events	0.58	0.24
Coasting rate	0.62	0.34
Even driving		
Number of events	0.32	0.21
Average duration	0.57	0.24
Even driving rate	0.46	0.52
Sudden speed change		
Number of events	0.17	0.45
Average duration	0.60	0.13
Sudden speed changes rate	0.50	0.59
Anticipation		
Number of events	0.54	0.34
Average duration	0.64	0.11
PKE	0.58	0.17

The additional parameters calculated in Subsection 3.2.2 which give relevant results, i.e. high enough correlations, are summed up into Table 3.2.

Table 3.2: Synthesis of coefficient of determination R^2 values for the additional parameters

Driving Parameters	Polynomial Regression Current Study
Average braking value	0.64
Braking / Initial speed	0.57
Braking / Final speed	0.26
Anticipation / Initial speed	0.10
Anticipation / Final speed	0.27
Product of gradients	0.49
Average Brake Pedal Angle	0.37
Coasting / Average kinetic energy	0.43
Anticipation final speed * Average brake pedal angle	0.54

Multi criteria analysis

The multi criteria analysis confirms most of the conclusions which were written in the single criterion one. The basic parameters which are the highest influence are *Coasting rate* and parameters related to *Even driving*. *Braking rate* also appears several times, but it was shown in Section 3.2 that this parameter was very sensitive. The additional parameters which are the most relevant are the same as those which were observed previously.

3.3 Identification of the type of use with driving parameters

Although this report mostly focuses on splitting drivers regarding their fuel efficiency, it is also interesting to be able to differentiate various truck uses by evaluating driving parameters. The main objective is to be able to figure out the type of cycle to run on a dynamo-meter to simulate driving for different truck uses, in order to determine, for instance, driver efficiency. Indeed, one cannot evaluate truck driver efficiency if the cycle is unknown, assuming that efficiency is affected by different parameters regarding the type of cycle.

In addition to that, knowing the type of truck and the type of cycle, the fuel consumption can be roughly estimated: from 10 L/100km for a light truck, it can reach 50 L/100km for a refuse or a long haul vehicle. Driver efficiency then affects the results, but the impact on fuel consumption is limited to 10 to 20%. This is not so relevant, as the different between two types of use is much bigger than that.

Given that for three of the five customers analysed here, the number of drivers is very low (1 or 2), it is not possible to generalize the results. That is, the results are only indicative.

Multi criteria analysis wasn't used for this part. Indeed, time didn't allow it, and it was judge as non necessary. First because the number of customers was very low (5), and also because the single criterion analysis alone gave satisfying enough results.

3.3.1 Reference used for comparison

In section 3.2, *Eco coefficient (energy)* was used for comparison: it was particularly suitable because it gives a reliable information on drivers efficiency. The basic idea, in this part, was to keep the same coefficient to make the distinction between the different types of uses. However, one can see in Subsection 2.1.1 that fuel consumption is not available on 2 of these trucks: that is the reason why it was decided to calculate an *Eco coefficient (energy)* depending on energy consumption instead: thereafter, this coefficient is called *Eco coefficient (energy)*. To get this coefficient, the total energy consumption is calculated with the engine torque and speed, and is then divided by the total distance driven by the truck. It is relevant to compare these 2 quantities as all trucks share the same technology (Diesel engine), with very similar efficiency. Figure 3.11 shows the value of this coefficient for the 5 records. Finally, this coefficient is chosen as a reference for comparing the 5 customers. The exact values are given in Table 3.3, and are especially important as they are used in all the graphs later on in this section.

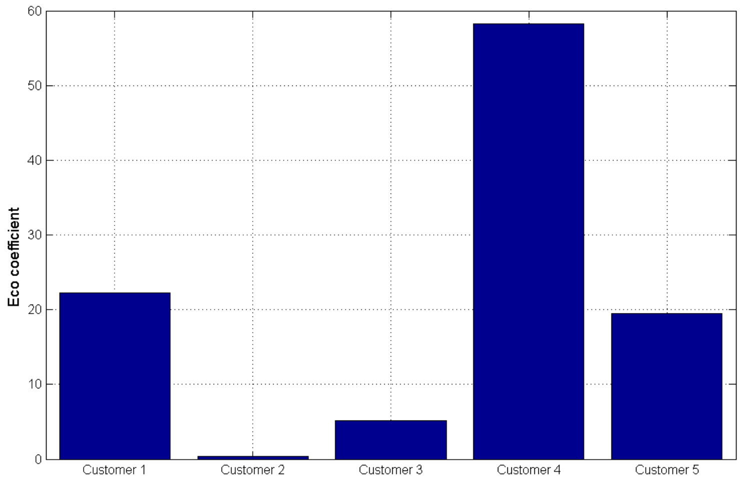

Figure 3.11: *Eco coefficient (energy) for the 5 customers*

Table 3.3: Eco coefficient (energy) values for the 5 customers

	Customer 1	Customer 2	Customer 3	Customer 4	Customer 5
Coefficient value	2.22	0.04	0.52	5.82	1.95

3.3.2 Relevant parameters

Following the same methodology as for driver classification, the objective is to find at least one relevant driving parameter per type of criteria, which makes possible the distinction between types of use. While the previous correlations were made with a set of 11 values, there are only 5 values to compare in the current one: that is the reason why a 2nd order polynomial regression isn't used any more. A 1^{st} order polynomial regression is used instead (further details are given in Section 2.2.1), in order to avoid having too high correlations, what would be artificial due to the small set of data. For customers with more than 1 driver, an average value is taken. Alternative methods was considered, but due to the very limited dataset this simple approach was chosen.

The same parameters as those investigated in Section 3.2 are used: this made the task much faster. Unfortunately, in this case, some parameters couldn't be calculated with customer 5, while trends with the four other customers were promising, due to data not available. In addition, it is also much easier to differentiate truck uses than drivers themselves on a same road. Hence, there was no need to look for new parameters as available ones gave accurate enough results. Obviously, in order to be compared, some of these parameters had to be divided by the distance of the section, as it was different for the 5 trucks.

Anticipation and Braking For those two criteria, *Average initial speed* and *Number of events per km* happened to give the strongest results. The number of events is almost the same in both cases as anticipation phases are followed by braking phases. Figure 3.12 shows that the 5 in the following order: Customer 2 (*Refuse*), Customer 3 (*Urban*), Customer 1 (*Urban*), Customer 5 (*Regional*) and Customer 4 (*Distribution*) on the 2 first graph.

For anticipation and average braking speed (i.e. vehicle speed when the driver starts anticipating or braking) the trend for this order is a positive slope, while for number of events for anticipation and braking (i.e. the braking or anticipation frequency per kilometre) the trend is a negative slope.

Coasting Once again, the only valuable parameter is *Coasting rate*, as shown in Figure 3.13.

Two values appear to be extreme: *Refuse* (bottom left), which is a very tough use, and *Distribution* (top right), which is much smoother. In between, the difference is not obvious between the 2 urban trucks and the regional one, even if they are quite different. The 2 urban trucks have a similar *Coasting rate*, which is a good

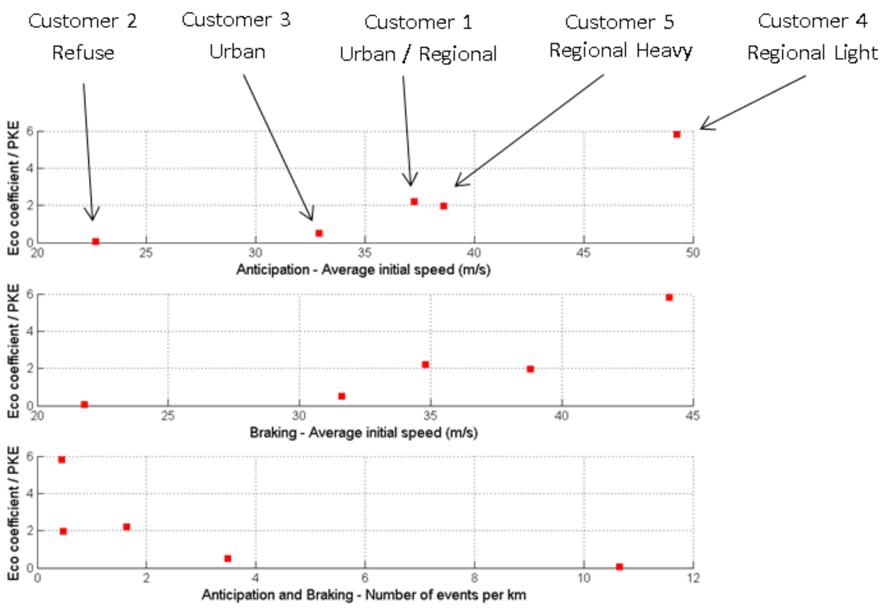

Figure 3.12: *Anticipation and braking - Initial speed and number of events*

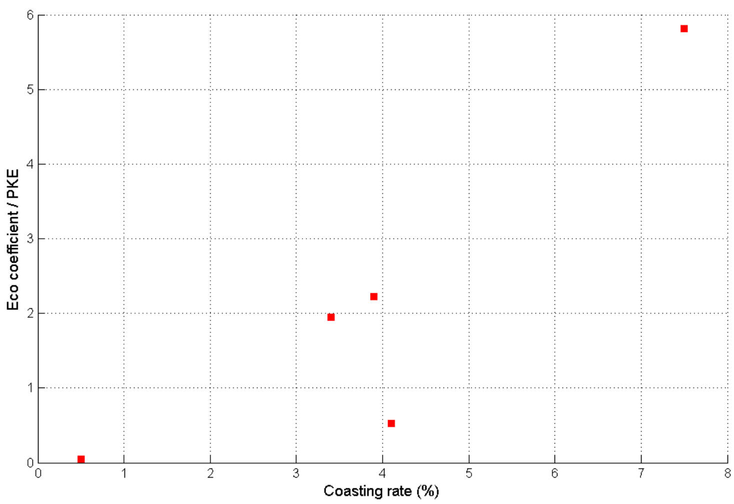

Figure 3.13: *Coasting rate*

point as they are used on the same type of route: hence, it is possible to recognize this type of use in spite of the fact that the drivers are quite different.

Even driving It was decided to keep a single parameter for this criterion, *Even driving rate*, without taking into account the sudden speed change values. Indeed, as for *Coasting*, it is interesting to see that customers 1 and 3, despite their different in terms of *Eco coefficient (energy)*, can be recognized on the graph, among other dots. The long haul truck, customer 5, is also easy to recognize as its rate is especially high, as well as customer 4. Those 2 cycles have much more highway and extra urban phases, and it looks much clearer in Figure 3.14.

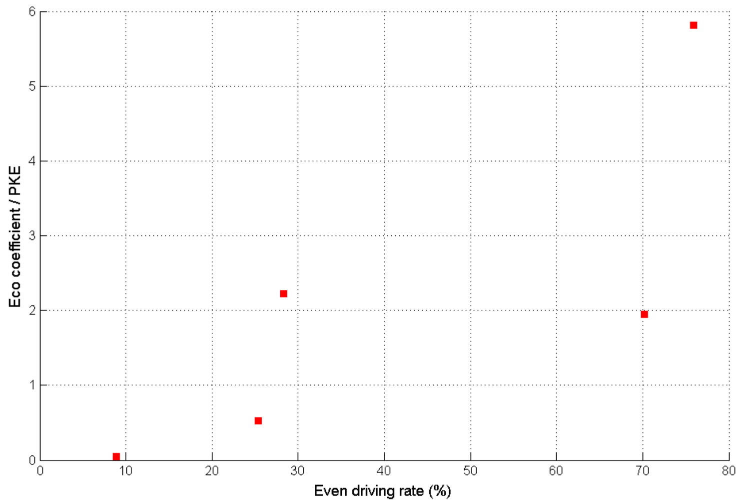

Figure 3.14: *Even driving - Number of events and Sudden speed change rate*

Miscellaneous *PKE* is also considered as relevant: of course, having it as both reference and attribute explains why the correlation is good. However, it really makes a difference between uses, as shown in Figure 3.15. Through its very specific driving cycle, mainly made of stop and start phases, a refuse truck leads to a

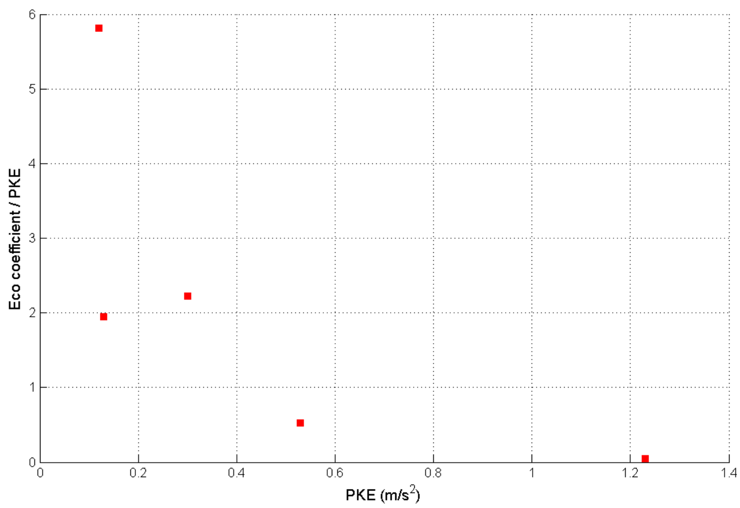

Figure 3.15: *PKE values for 5 customers*

high PKE. On the opposite, for *Distribution* and *Regional* cycles, PKE is very close to 0. In between, one can easily distinguish the 2 urban cycles.

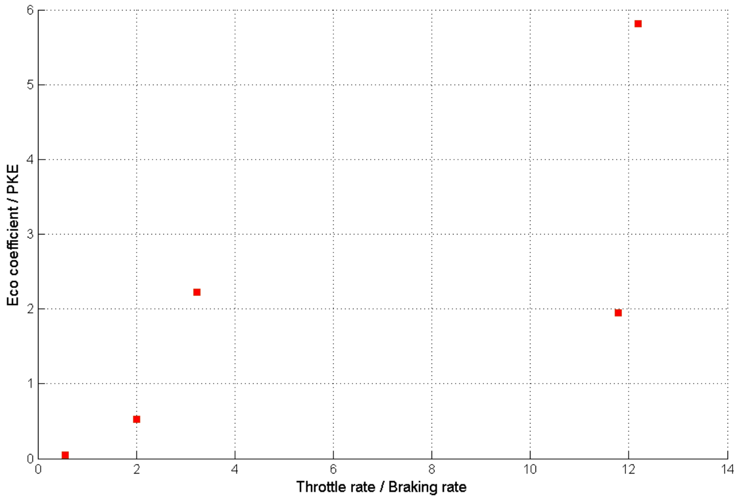

Figure 3.16: *Throttle rate / Braking rate for 5 customers*

Finally, the last parameter to be considered is the ratio *Throttle rate / Braking rate*. This parameter, like the previous one, makes a clear distinction between 3 behaviours: *Customer 2*, refuse, brakes a lot, so the ratio is low. Then *Customer 4* and *Customer 5* spend most of their time on extra urban sections, where their braking frequency is low. In between, on urban section, one can see *Customer 1* and *Customer 3*: once again, in spite of the fact that those 2 customers drive on the same kind of cycle but have a very different behaviour, it is easy to recognize them.

3.4 Influence of driver behaviour regarding road layout on fuel consumption

This section makes a direct connection between the drivers and the road. Indeed, most of the parameters investigated in the previous parts are general and are present in different kinds on situation, i.e. they are not related to any road event. In order to establish the relationship between the the road layout and the fuel efficiency, the following part focuses on 3 types of events related to road layout:

- Road altitude: this focuses on slopes, and how drivers deal with them. Especially, it is interesting to observe the way they coast for instance, or how strong they accelerate on uphill phases.
- Roundabouts: this focuses on roundabouts only, but it plays an important role on fuel efficiency as average speed in such an event is low. Drivers' speed profile is investigated, as well as their capacity to anticipate.
- Curve radius: this covers a large range of events, that is the reason why it is split into different types of curve (large and small curves). As for roundabouts, there is a special focus on the speed profile, as well as the braking behaviour.

This part only focuses on relevant parameters, i.e. parameters whose correlations are too low are not mentioned.

3.4.1 Road altitude

Relevant parameters

One noticeable parameter is *Average torque in downhill phases* and highlights the fact that drivers with a higher *Eco coefficient* seem to use those phases to increase their speed before an uphill phase, as shown in

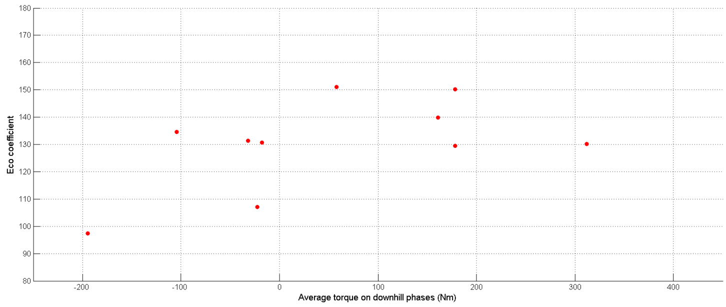

Figure 3.17: *Average torque on downhill phases on section 2B*

Figure 3.17. This trend is confirmed by *Average speed on downhill phases*, where non efficient drivers have lower speed than efficient drivers, as shown in Figure 3.18. However, those trends remain extremely small, even if they confirmed what was said about *Volvo I-see* feature in Figure 1.2. When it comes to speed uphill,

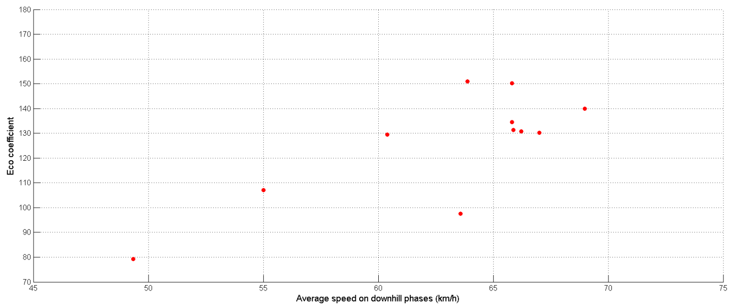

Figure 3.18: *Average speed on downhill phases*

it appears that best drivers reach top of the hill with a higher vehicle speed, which could have been thought to be the opposite. However, Figure 3.19 shows that this speed is much lower than speed limit for all the drivers. However, they don't have higher throttle pedal angle values on uphill phases: their values are even lower. So

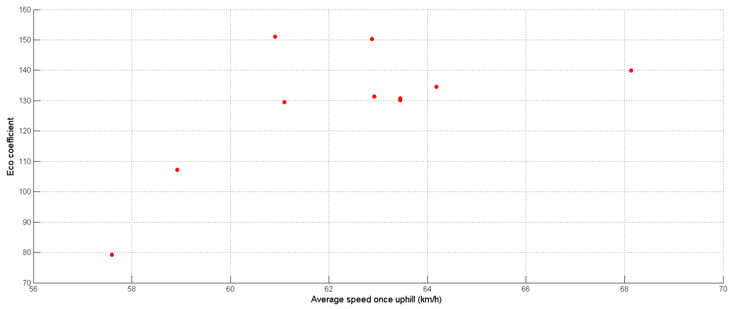

Figure 3.19: *Average speed once uphill*

those drivers happen to reach top of hills at a high speed, while their throttle pedal values remains low: this gain is due to their high speed on downhill phases observed above. Stronger acceleration before starting an uphill phase is also advised on *I-see* system.

Multicriteria analysis

A multicriteria analysis was run with 3 classes: efficient (3), mean (5) and non-efficient (3) drivers. The reason for such a subdivision are given in Subsection 2.2.2

If the decision tree manages to classify only 9 drivers out of 11, 2 of the previously noticed parameters are also important, as shown in Figure 3.20.

According to this tree, *Average speed on downhill phases* is enough to identify non-efficient drivers, while

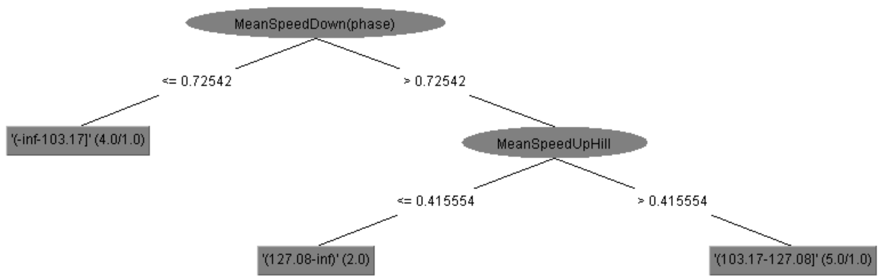

Figure 3.20: *Decision tree for section 2B*

another parameter is taken into account to make the distinction between mean and efficient drivers: *Average speed once uphill.* When classifying parameters, it appears again that the first parameter remains the most important one. Simple Logistics oddly classifies 8 drivers only when this parameters is taken into account, while 9 are classified without it:

$$BadDrivers = 4.74 + [MeanTorqueDown] * -4.49 + [Averagethrottleangle/UpNeg] * -5.46 \quad (3.1)$$
$$MeanDrivers = -0.84 + [MeanSpeedUpHill] * -3.03 + [Averagethrottleangle/UpNeg] * 4.44 \quad (3.2)$$
$$GoodDrivers = -3.04 + [MeanTorqueDown] * 2.7 + [Averagethrottleangle/UpPos] * 2.79 \quad (3.3)$$

Removing *Average speed on downhill phases* makes other parameters emerge: especially, one can see that *Average torque on downhill phases* is of high importance. Those equations also signify that non-efficient drivers tend to lower their throttle pedal angle on decreasing part of uphill phases, i.e. when slope percentage is going from its maximum to zero. In the same time, efficient drivers accelerate more in the first part, when slope percentage is going from zero to its maximum: once again, this is the behaviour adopted by *I-see*.

3.4.2 Roundabouts

Relevant parameters

In this section, satisfying results were found for 5 parameters. The first is *Distance between lowest speed and event*, as shown in Figure 3.21. This graph shows that efficient drivers tend to anticipate events by having

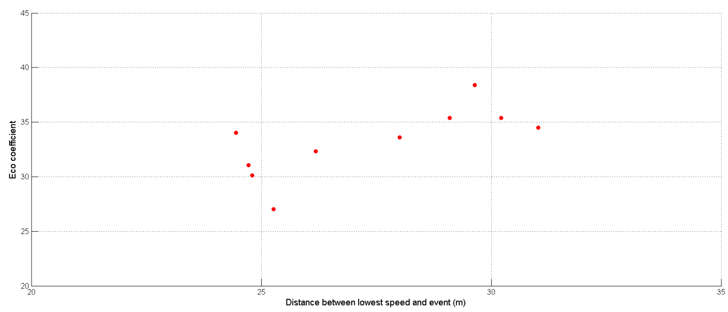

Figure 3.21: *Distance between lowest speed and event on section 1B*

their lowest speed before other drivers. However, surprising results are found about brake and throttle pedal angles. In the first case, it was found that efficient drivers were braking stronger, as shown in Figure 3.22. This is going against what is usually assumed: the same phenomenon was observed with *Throttle pedal angle*, which is awkward and remains to be clarified.

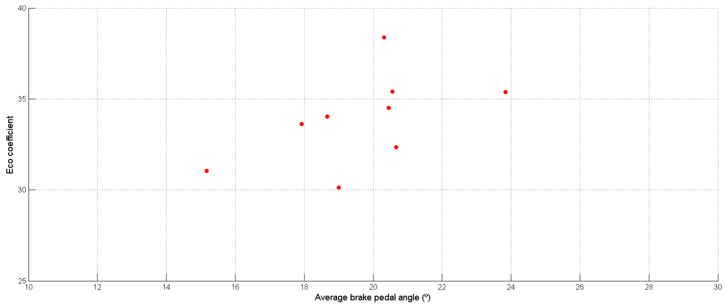

Figure 3.22: *Average brake pedal angle during braking phase on section 1B*

Multicriteria analysis

A multicriteria analysis was run with *Weka* in order to evaluate all parameters, with limited success. Drivers were split into 3 classes: efficient (3), mean (5), and non-efficient (2). This analysis was carried out with 10 drivers still, and non disturbed data. A decision tree was built, classifying only 8 drivers correctly, and shown in Figure 3.23. If it confirms that *Distance between lowest speed and event* distinguishes efficient drivers, it

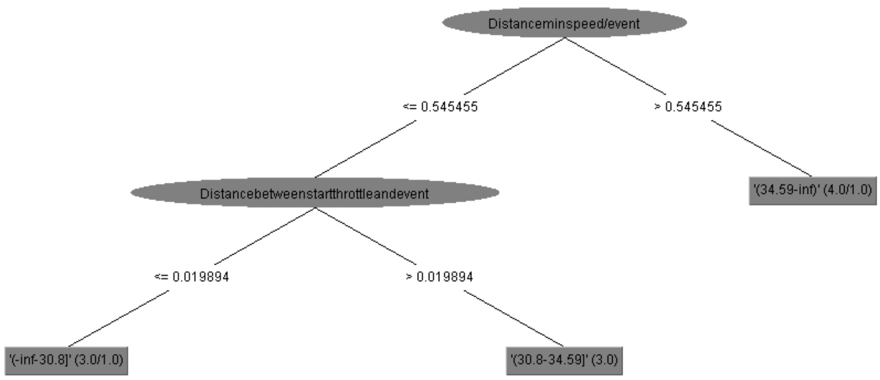

Figure 3.23: *Decision tree on section 1B*

also says that non-efficient drivers are identified with *Distance between beginning of throttle phase and event*: they have this parameter lower than other drivers, while other drivers usually start accelerating much earlier. This seems to be connected to their *Anticipation* capacity. Simple Logistics equations are also generated, with a classification rate of 9/10:

$$Baddrivers = 2.13 + [MeanBrakingAngle] * -2.95 + [SpeedAtStartThrottle] * -1.518 \quad (3.4)$$

$$Meandrivers = 2.02 + [MinspeedEvent] * -3.08 + [Distancebetweenstartthrottleandevent] * 1.52 \quad (3.5)$$

$$Gooddrivers = -2.57 + [Distanceminspeed/event] * 2.37 + [MinspeedEvent] * 2.36 \quad (3.6)$$

This set of equations confirms the results found with single criterion analysis, as non-efficient drivers are characterized by their low *Average brake pedal angle.* It also confirms that efficient drivers' lowest speed is high and happens early before event.

3.4.3 Curve radius

Relevant parameters per event

In this part, it was decided to focus only on events giving consistent results. Indeed, for some events, results were not relevant enough to be trusted, likely due to traffic flow issues such as traffic lights or queues. Especially, the 3^{rd} corner of section 112 contains disturbances due to the fact that there was a traffic light at the end of the corner. The same observation can be done for the fourth corner of section 111. Concerning the second corner of section 111, it was initially thought to give hints about slight angle corners: but it seems that this type of curve doesn't play a big role on global fuel efficiency.

Curves 111/3 and 112/2 Those 2 curves have the same speed profiles, and they share most of their relevant parameters: first of them concerns *Mean throttle pedal angle* after the event, as shown in Figure 3.24. The

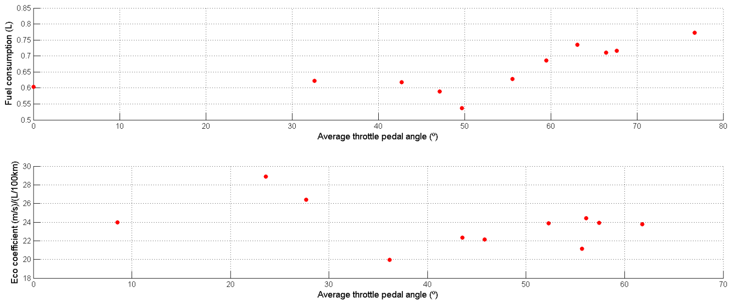

Figure 3.24: *Average throttle pedal angle after event - Top: section 111, bottom: section 112*

trends of both graphs show that too much throttle after event leads to much higher *Fuel consumption*. As a consequence, *Eco coefficient* drops: this means that the gain in terms of average speed is not relevant enough to be worth a high throttle angle. The same observation can be made with *Throttle angle* during the event: as the lowest speed usually occurs before the event, drivers start to accelerate during the event. Efficient drivers tend to turn at a high enough speed so that they don't have to accelerate too much afterwards.

Curve 112/4 The last curve is interesting as it is quite long, but with a high radius: total angle is about 130°but radius makes it progressive. The speed profile shape, and especially during event, plays an important role in this case. This profile is synthesized within a coefficient: $Speedduringevent^2/(Initialspeed \times Finalspeed)$. This coefficient is shown in Figure 3.25. If speed is to be kept high enough, it also means that

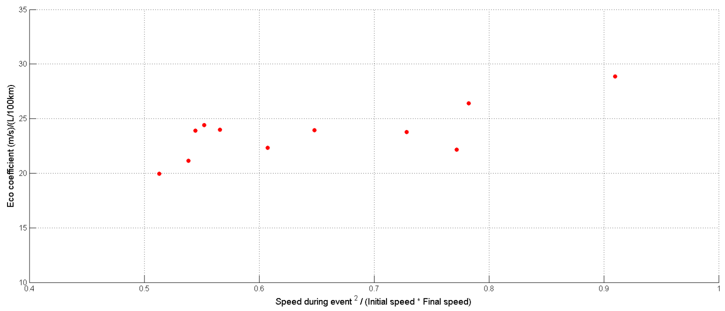

Figure 3.25: *Speed profile on curve 4, section 112*

drivers should avoid braking too strong in this kind of corner. One can figure it out by watching *Average brake pedal angle* during the event, where it is easy to see how *Eco coefficient* is affected by a too high brake pedal angle. This effect is shown in Figure 3.26. For this curve as for most of the previous ones, a too high throttle angle after event makes efficiency drop: this is strongly connected to previous observation as when braking too much, then drivers have to accelerate to compensate.

Multicriteria analysis

A global multicriteria analysis of both sections was run in order to put the previous method to the test. This aimed to verify if same parameters as above were giving same results, i.e. if relevant parameters were the same.

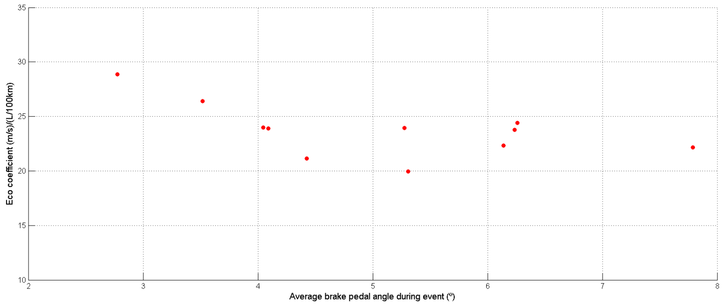

Figure 3.26: *Average brake pedal angle during event on curve 4, section 112*

Section 111 For this section, as *Fuel consumption* was used as a reference instead of *Eco coefficient* values, only two classes were implemented instead of three: efficient (6) and non-efficient (5) drivers. The reason for such a change is that, for this section, a strong relationship was found between *average speed* and *fuel consumption*. Since *Eco coefficient* is defined as the ratio of those two quantities, the effects were cancelled out. When looking at *fuel consumption*, the drivers can't really be split into three classes, and only two groups come out.

Figure 3.27 shows the results of J48 algorithm on this section, where 10 out of 11 drivers are classified correctly. *Average throttle pedal angle after event* is confirmed to play an important role: the algorithm takes

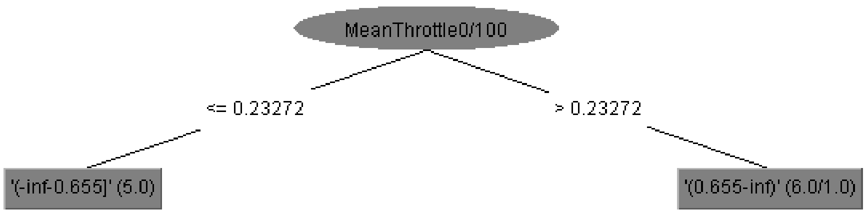

Figure 3.27: *Decision tree for Section 111*

only this parameter to make the difference, and highlights the fact that efficient drivers are moderate in their re-acceleration after an event. When classifying events, 3 parameters are clearly better: the previous parameter, of course, but also Average throttle pedal angle during event and also *Final speed on event*. All those parameters are connected to the drivers behaviour after event. Equations are also extracted using Simple Logistics algorithm, with a classification rate of 10/11:

$$Baddrivers = -2.09 + [MeanThrottle0/100] * 2.33 + [SpeedEnd] * 1.91 \quad (3.7)$$

$$Gooddrivers = 2.09 + [MeanThrottle0/100] * -2.33 + [SpeedEnd] * -1.91 \quad (3.8)$$

Those equations say the same thing as the decision tree: efficient drivers are those who don't accelerate too much after a curve. In the previous part, it was observed that they don't have to accelerate as their cornering speed is generally speaking higher.

Section 112 Considering *Eco coefficient*, 3 classes were used: efficient (2), mean (5) and non-efficient (4) drivers. Once again, 10 out of 11 were correctly classified in the decision tree, as shown in Figure 3.28 This tree shows once again that efficient drivers are remarkable through their a low *Average throttle pedal angle after event* but also through their lowest speed well ahead of the event: this phenomenon was already observed in the previous part. It also shows that non-efficient drivers are not especially noticeable through their throttle angle but rather through their lowest speed close to the event. Simple logistics gives the following set of

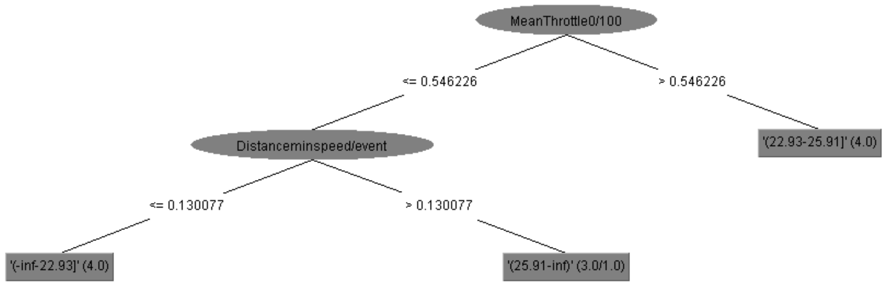

Figure 3.28: *Decision tree for Section 112*

equations, with a classification rate of 100%:

$$Baddrivers = 0.22 + [MeanThrottle] * 2.09 + [Speedwhile] * -2.43 + [While^2/Beg * End] * -2.56 \tag{3.9}$$

$$Meandrivers = -1.86 + [Distanceminspeed/event] * 1.73 + [MeanThrottle0/100] * 4.27 \tag{3.10}$$

$$Gooddrivers = 0.35 + [MeanThrottle] * -1.88 + [Where/IndEventInd2] * -2.18 + [While^2/Beg * End] * 2.41 \tag{3.11}$$

The ratio between *Speed on event* and *Initial speed * Final speed* is higher for efficient drivers and explain why they manage to keep their throttle value low. It is confirmed as *Speed on event* for non-efficient drivers is low.

3.5 Ranking of drivers regarding their efficiency

In Section 2.5, two different methods were described for the construction of this ranking system. However, the first gave results that showed little or no correlation between the studied parameters, making ranking impossible. Those results were put in Appendice I. They are not necessary for understanding of this section.

The 2^{nd} method only, i.e. based on single criterion analysis, is shown in this section. With this method, the single criterion analysis was first applied to the global cycle to extract the relevant parameters for the global cycle. These were then used as the starting list for applying the method to road sections 1 to 5. This means that only the parameters identified as relevant for the global cycle was used as possible parameters for the extraction of ranking equations for the 5 separate sections.

A huge sensitivity to Driver *C* was observed, and it was decided not to consider this driver as his behaviour is awkward and affects the results too much. This removal of an individual driver is in general not appropriate, but given the very low number of driver anyway one driver can have high influence of the results. A total of 78 parameters and parameter combinations were considered when running the single criterion analysis, resulting in 8 chosen relevant parameters, shown in Figure 3.29.

Results for each section

The parameters which gave relevant results are then calculated and normalized for the 5 sections (i.e. for each parameter, the lowest value is set to 0 and the highest is set to 1, see Equation 2.25). For each sections, iterations are run with *Weka* in order to determine the classification coefficient of those parameters. Those coefficients are shown for the 5 sections and for the global cycle (coefficients equal to 0 mean that the parameters is not taken into account).

Those coefficients are used to weight the driving parameters, to determine their level of importance. That is, as explained in Subsection 2.5.2, the global grade is made up of the sum of the weights of the sections (where the section weights are one over the distance for the section), and for each section the sum of the weighted parameters (where the parameter weights are from Table 3.4). Finally, a grade is obtained for each of the 5

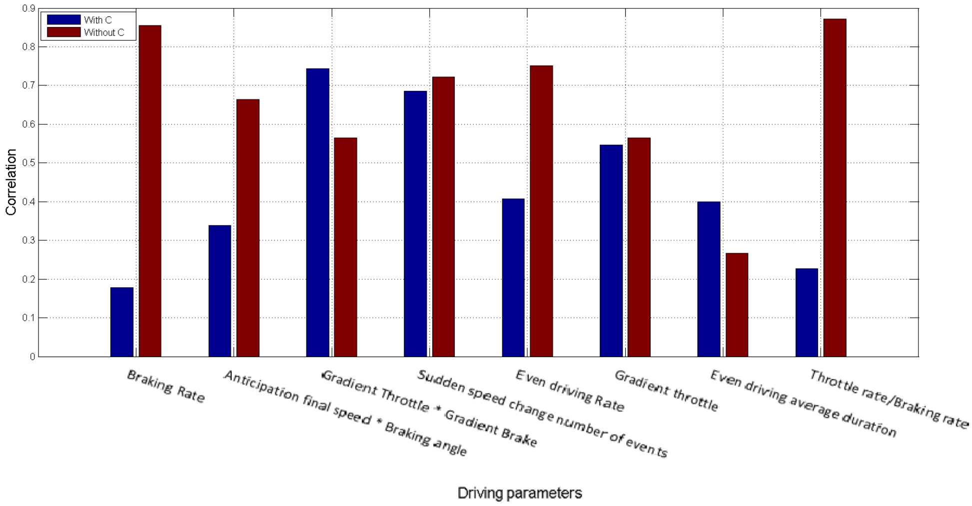

Figure 3.29: *Driver ranking following driving parameter method*

Table 3.4: Classification coefficient for the 5 sections and the global cycle

Sections	1B	2B	4B	111	112	Global
Braking Rate	0	0	0	0	0.722	1.573
Anticipation final speed * Braking angle	0	0	0	0	0.722	1.573
Gradient Throttle * Gradient Brake	0.722	0	0.88	0	0.722	0.845
Sudden speed change number of events	0	0	1.485	0	0.722	0.845
Even driving Rate	0.722	0.722	1	0.994	1.37	0.845
Gradient throttle	0.722	0	0	0	0.722	0.845
Even driving average duration	0	0	0	0	0	0.699
Throttle rate / Braking rate	0	0	0.88	0	0.722	0.651

Table 3.5: Drivers' grade for the 5 sections

Sections	1B	2B	4B	111	112
A	6.3	6.7	3.9	6.0	3.2
B	9.1	3.8	5.7	4.1	8.8
C	6.7	3.6	1.1	4.0	5.3
D	8.1	10.0	9.1	8.2	5.3
E	8.8	8.1	9.7	6.9	9.5
F	6.3	4.1	7.4	4.3	6.6
G	2.3	0.0	3.8	1.6	0.7
H	4.6	7.3	5.5	6.4	3.9
I	1.6	9.0	5.1	7.5	3.1
J	2.5	7.1	2.5	6.3	4.1
K	4.1	7.5	0.6	6.6	1.9

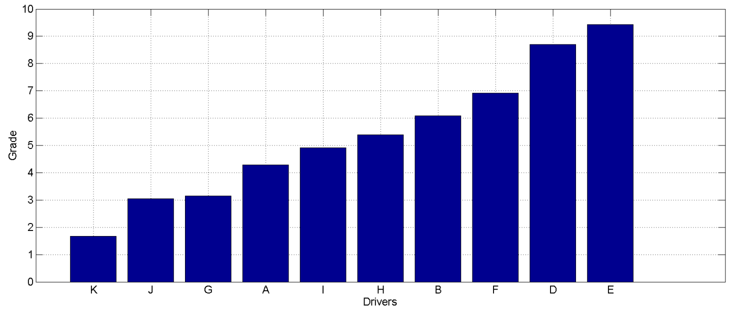

Figure 3.30: *Global ranking for the complete cycle*

sections, as shown in Table 3.5. The global grade is then calculated from those 5 grades, weigthed regarding the length of the section. Figure 3.30 shows this ranking, from the weakest to the best driver.

Comparison between the ranking and the physical measurements

Drivers are ranked depending on various references: *Grade*, $\frac{Ecocoefficient}{PKE}$, *Eco coefficient* and *Fuel consumption*.

Ranking	Grade	Eco coefficient / PKE	Eco coefficient	Fuel consumption
1	E	E	E	E
2	D	B	D	A
3	F	D	I	I
4	B	I	J	H
5	H	F	A	J
6	I	A	H	B
7	A	H	F	D
8	G	J	B	F
9	J	G	C	G
10	C	K	G	C
11	K	C	K	K

Table 3.6: Drivers ranking regarding different references

Table 3.6 sums up those results. *C* is given a grade, even if he was not considered in the process: this grade is determined with his driving parameters and with the classification coefficients given previously.

4 Discussion

4.1 Sensitivity of the set of data

All along this study, the data sensitivity has been a major issue. Indeed, with so few drivers, one single driver can have a huge influence on the correlations. Very high variations were observed, and even cases where there were inverted correlations when a driver was removed (which leads to a contradiction).

It is extremely hard to draw conclusions in such conditions: the solution that was found was to rely more on scatter graphs than on a single coefficient (R^2) in the interpretation of the results. This is not statistically rigorous and only provides indications, but given the low number of data point, alternatives were hard to find

Several times in this thesis, all the drivers were aligned on the graph, but one driver was really out of the cloud. In some of those cases, the driver was removed, when his behaviour was considered as *non-normal* (e.g. 4 times more coasting events than the other drivers). But overall, this solution was avoided (and when used, it was noted in the report): indeed, the number of drivers is too limited to get rid of one of them and consider them outliers, and it is hard to say what a *normal* behaviour is because the set of data is way too small to be representative of the whole population.

In addition to that, this study was carried out with a human bias due to the fact that drivers knew they were recorded: this is proven to influence the results, however this change is hard to quantify. In the future, an effort should be made to assess drivers in their real environment. This means with higher pace and without telling them that their fuel consumption will be recorded, or by telling them that they are recorded for other reasons.

When the study was set up, the intention was to use naturalistic driving data with more than 50 drivers having driven hundreds of hours each under different conditions. A database called *EuroFOT* was to be used, but unfortunately, problems with the data transformation and processing specific for this study occurred and those data finally couldn't be used.

EuroFOT is a European project launched in 2008, performing coordinated tests of Intelligent Vehicle Systems with ordinary drivers in real traffic[18].1000 vehicles from 9 European brands were used.

The naturalistic collection means that they are driving with the data collection daily under a long period of time, most of the time forgetting about being monitored. This data would have been very good for between and within driver evaluation of efficiency and the study could have produced more generalizable and statistically validated results.

4.2 Evaluation of driving criteria and comparison to previous results

4.2.1 Single criterion analysis

In this part, the fact that dealing with the current set of data was tricky became obvious, since the 11 drivers are not consistent in their behaviour with respect to the evaluated parameters, and since their low number creates high sensitivity. In addition to that, global parameters were targeted (i.e. valid for a representative driving cycle): then, a comparison could be made with what was found in the master thesis mentioned in Subsection 1.4.2. Even if the driving cycles were alike, the weigth of both trucks was very different and the results were finally not so comparable.

The first attempts, by combining parameters to *Fuel consumption*, gave poor correlations, and the results were very hard to interpret because of sensitivity problems. Changing to *Eco-coefficient* improved the correlations a lot. Overall, those results weren't completely convincing as some problems were still present (sensitivity), yet it was much better.

Out of these new parameters, some are too sensitive to individual drivers to be considered despite their good correlation values for the full set of drivers. For the complete set of parameters investigated, only the following ones are finally considered:

- **Braking and Anticipation**
For those 2 criteria, it was really hard to find satisfying results. Indeed, for most of the driving parameters which were investigated, the sensitivity was quite high, mainly because of Driver *K*. After many attempts, it was decided to combine *Anticipation* and *Braking*: this makes sense as the first one usually precedes the second one. It would have been preferable to find one parameter at least for each criterion, but still this combination works fine so it is conserved until a better solution is found. This parameter is the product of *Anticipation final speed* and *Average brake pedal angle*.

- **Coasting**
Coasting rate has revealed a slight dependence to Driver *C*. In spite of that, it remains the driving parameter which brings the best results about coasting behaviour. As mentioned several times in Part 3, Driver *C* is known for his specific behaviour and for the contradiction between this behaviour and his performance. The second interesting parameter in connection with coasting was *Average kinetic energy*: however, it was not considered for the ranking as some parameters are of greater importance. But the point with those two parameters is that they are quite different, and so represent different aspects of the coasting behaviour.

- **Even driving**
This criterion was the one for which the best results were found. It appears to be, in the case of Customer 4, very determining. When watching the results of the previous study mentioned in Subsection 1.4.2, it seems to be more relevant for a light truck than for a long haul truck, on comparable uses. There were several good parameters but the ones that are to be kept are *Even driving rate* and *Sudden speed change / Number of events*. It is interesting for several reasons: the correlations are very good and remain robust, and they represent different approaches (rate and number of events).

- **Miscellaneous**
New driver oriented parameters are satisfying: that is the reason why *Average brake pedal angle* and *Product of gradients* are kept. This last parameter is somehow difficult to visualize but it really shows, through the intensity of the action on the pedals, how nervous a driver can be, and this has a direct impact on fuel efficiency.

4.2.2 Multi criteria analysis

The multicriteria analysis was globally a success as the classification rate was good for all the sections. Subsection 3.2.3 highlights the difference between all of them, as they correspond to different driving situations.

In the end, parameters related to *Even driving* are very recurrent: this is especially true for long highway sections such as *4B*. The results show that *Braking* is not as important as one might think, even if it remains present on urban sections such as *112*, or sections with large speed reductions such as *1B*. This particularity may simple come from the fact that on highways, drivers almost never brake, so the impact is limited compared to other parameters such as their ability to keep a constant speed. Conversely, on urban sections, they brake much more often, or more intensely (e.g. when approaching a roundabout).

On those sections, *Braking* is dominating

Some of the new parameters which were implemented specifically in this study seem in general to be more relevant than the basic ones (implemented already in the previous study), especially because they include driver-oriented information. The way drivers use both throttle and brake pedals is quite interesting and gives a good glimpse of their efficiency. Using throttle and brake pedals works fine, as it really shows how often and how strong drivers use them. On today's vehicle, there is only a limited relationship between those pedals and the road (e.g. if the throttle pedal angle is 50%, that doesn't mean that the torque delivered by the engine is 50% of the maximal torque). In spite of this limited relationship, working on the pedal signals says a lot about how smooth the driving style is. The way they use their truck's kinetic energy was also investigated: taking the squared speed gave much better results than when the speed wasn't squared. As the mass is the same for all the drivers, this squared speed is equivalent to the kinetic energy, which is the actual important parameters in this case.

4.3 Identification of the type of use

The type of use analysis was able to relatively well separate the different customers and their individual types of uses. A reduced set of parameters was finally considered, representing all the driving criteria. The analysis method was primarily using scatter graph instead of statistics. This was due to the low number of customers (values) and that the interpretation is more qualitative than quantitative on terms of evaluation. The results range from an extremely harsh type of use (refuse) to a much smoother one (regional light). Those two types were opposites on the graphs, and, in between, the 3 other uses were spread as follows: urban, urban/regional, and regional heavy.

Through the 7 driving parameters which were conserved, it is now possible to identify different types of cycles in a reliable way. Each of them has its particularities, and the differences are described in the following.

Note that the very limited number of drivers for some customers makes it impossible to generalize also the results for *type of use*. However, the difference in the two extremes should be large enough that no driver should be able to have such bad driving in the smooth (e.g. regional light) that it would be interpreted as refuse use. The three other types of use (in the middle on the graphs) are likely to be more sensitive to the individual drivers for separation.

Refuse

This is probably the easiest type of use to identify as most of the values are quite extreme. Due to low average speeds, *Braking* and *Anticipation* speeds are low as well (lower than 25 km/h), but happen at a very high frequency compared to more standard uses (more than 10 times per km, which means 1 stop every 100 meters).

There is almost no place for *Coasting* with a global value of about 1%. The rare phases happened to be out of the working cycle, but rather before starting or after finishing it. A constant speed cannot be kept more than 10% of the time, with a very high PKE: this was predictable as such a use requires many acceleration and deceleration phases. On this cycle, the throttle/braking ratio is close to 1: it means that the driver is braking almost as much as she/he accelerates.

Urban

In the current study, *Urban* cycles were represented by two customers: one normal (customer 1), and one very non efficient (customer 3). It was interesting to see if those data were going to interact with the data around, as this cycle is in between *Refuse* and *Regional*. The good point is that both dots remain in the same order for each graph: *Anticipation* starts around 35 km/h while *Braking* starts shortly after, around 33 km/h, with lower speed for non efficient driver. The frequency is between 2 and 4 braking events per kilometre, with higher frequency for non efficient driver.

Both costumers have almost the same value of *Coasting rate*, 4%, and very close value of *Even driving rate*, 26%. However, the non efficient customer has a much higher value of *PKE*: this highlights the high number and the large amplitudes of accelerations. This point probably explains the difference between both customers in terms of efficiency. In spite of that, they have close values for *Throttle/Braking*. This can be explained by the fact that if the non efficient customer accelerates more, it also brakes more, balancing the ratio.

Regional

As detailed before, this customer has the heaviest load (40 tons). This makes a big difference compared to other customers, whose load is closer to 10 tons. Even though *Eco coefficient / PKE* is similar between customers 1 and 5, all the parameters are quite different. *Anticipation* and *Braking* happened at very similar speed, which is mainly due to the fact that with such a load, inertia is huge and deceleration is low. As the cycle is mainly out of the city centres, there are very few *Anticipation* phases and *Even driving rate* is much higher than in the previous cases. The same reason can explain a low *PKE*, as speed variation are limited on this kind of cycle. There are few *Braking* phases, with a high average speed on the cycle. As a consequence, the *Throttle / Braking* ratio is higher than 10 (which means that throttle pedal is pressed 10 times more than brake pedal).

Distribution

Customer 4's truck was initially made for distribution cycle. However, the cycle on which it was recorded is a regional cycle. The main difference with the previous regional truck (customer 5) is hence the weight.

This use has the highest *Eco coefficient*, due to a low load and an extra urban cycle. It is much closer to the urban cycle, and especially customer 1, than refuse. The *Anticipation* and *Braking* speeds are remarkable as they are quite high, unlike the number of events. This is normal as on a distribution cycle, there is no need to brake so often. But the average speed is higher than on a urban cycle (speed limitations are higher), hence the average braking speed is also higher. *Coasting rate* is almost twice as high as previous customer (customer 4), mainly due to the proportion of highway on the cycle. For the same reasons as regional truck, *Even driving rate* is large. With more than 75% even driving, it shows the smoothness of this kind of cycle, and how suitable it is when it comes to keeping a constant speed. Both cycles share some similarities, that is the reason why the same observation can be done for both *PKE* and *Throttle / Braking* where values are very close for regional and distribution trucks.

4.4 Influence of driver behaviour regarding road layout on fuel consumption

Despite the difficulty, parameters have been found for the three types of events. Further tests remain to be made, in order to improve those results: indeed, for some events (roundabouts in sections 111 and 112 for instance), there were some contradictions and results are not as reliable as expected.

As said earlier, only three weeks were dedicated to this work, which was too short: another couple of weeks would be needed to improve the code and make it more accurate. it would also be interesting to watch driving session video to find ad remove sections where drivers were disturbed (e.g. by traffic).

Road altitude variation

The analysis which was carried out on Section *2B* and it was hard to find relevant parameters. The fact that there were different phases, with different aspects, may have contributed to some differences between the phases. For instance, one parameter can be relevant on two phases, and not on two others.

As each slope is different, it could be interesting to classify them by length or percentage. For all of them, it appeared that the *Speed profile* was quite relevant. This means the average speed on downhill phases (so all along the slope) and the speed once uphill (so the speed at one specific point). The average torque on downhill phases, somehow related to the speed, is also relevant.

This makes sense, as this is the approach used by the Volvo *I-see* software for improving drivers' efficiency on slopes.

Roundabouts

The main problem in this analysis was that half of the data couldn't be used because of traffic jams. An average of all the roundabouts on Section *2B* had to be done with the remaining data, in order to have something substantial enough. Of course, there were still some problems with data sensitivity. As a consequence, there were very few parameters which were relevant. The most relevant aspect was related to *Anticipation*, and especially when drivers reach their lowest speed, and when they start accelerating again.

The speed profile wasn't so important in this case: previous studies had however shown that efficient drivers were faster in roundabouts. But this wasn't obvious along this study. Some odd results were found about the influence of braking intensity: indeed, according to those results, braking strongly was supposedly efficient. But it is really going against common sense, so it has to be checked again. Although the purpose of this study is to highlight the influence of some parameters which were unknown, common sense is still important to judge if a result is realistic or not.

Curve radius

Curve radii, like slopes, and unlike roundabouts, are always different. Here, they were simply split into two types, especially because the number of events was low.

On smaller curves (i.e. $< 70°$), the role of acceleration after event was highlighted. Efficient drivers manage to be quick without having to accelerate too strong after a curve. The assumption can be made that they don't have to accelerate because they keep a high speed during the event: however, it hasn't been possible to make sure this assumption was right.

On larger curves (i.e. $> 70°$), the speed profile was also proved to be important. The coefficient which was introduced for this purpose ($\frac{Speedduringevent^2}{Initialspeed \times Finalspeed}$) was very interesting for that. The correlations obtained with this coefficient were relevant. On those big curves, the braking behaviour during the event was also quite important: this is somehow related to what was said before: efficient drivers are able to anticipate, and they are usually faster in curve. That is the reason why they don't have to brake too strong.

4.5 Ranking of drivers regarding their efficiency

Among the 2 methods initially proposed, only one is chosen: the method based on *Single criterion* analysis. In this method, the critical point is that one driver is dismissed: however, as explained in Subsection 3.1.2, Driver C's behaviour is hard to explain. Indeed, despite his good behaviour on several criteria, he is among the *bad* drivers. One reason for that could be the wind, some investigations need to be carried out to find out whether it influenced the results or not.

Secondly, as for the classification by types of use, a new criterion is chosen as a reference. This criterion, ratio between *Eco coefficient* and *PKE*, increases dramatically the correlation values.

This ranking is based on driving parameters whose classification was drawn for the complete cycle: that is the reason why the classification coefficient is much higher for the *Global* cycle. However, the classification is much higher for some cycles. such as 1B, 4B and 112. For 2B and 111, the results are lower. All cycles are taken into account anyway: these low results don't affect the global grade, because the classification coefficients are equal to 0.

Some observations can be made concerning the influence of some parameters:

- **Braking rate** and $Anticipationfinalspeed \times Brakingangle$: surprisingly, this parameter is classified 1^{st} for the whole cycle, but this influence doesn't appear on the different sections (it only appears on section 112). One explanation for that it that, in spite of the fact that $Anticipationfinalspeed \times Brakingangle$ has been proved to be relevant (this was indeed shown in the previous parts)for each sections, there are some parameters which are even more relevant.
- **Even driving rate** and $GradientThrottle \times GradientBrake$: in Appendice I, one could be surprised not to see *Even driving rate* while the previous parts had shown how important this parameter was. This shows that the ranking based on *Simple Logistics* equation may not be fully reliable. On the other hand, the method considered during this study (i.e. based on driving parameters and not on *Simple Logistics* equations) appears much more reliable and brings results which are more consistent with what was found before, as the two parameters mentioned before are shown as relevant.
- **Even driving average duration**: this parameter appears on none of the section. This somehow confirms Figure 3.29, where the correlation is lower. So even if the classification coefficient is satisfying and the correlation, in absolute, is good, this parameter is probably not worth being kept.

Subsection 3.5 showed that drivers can have very different grade depending on the sections. This is merely explained by the specificities of each section (i.e. slope, roundabouts, etc.), but also by the driving conditions (traffic jam).

In only two cases, drivers are given an extreme grade (10 for D in *2B* and 0 for G in *2B*). This means that those drivers, for all the parameters which are taken into account for evaluating their score, are the best/worst: however, on *2B*, only 1 parameter is taken into account.

Table 3.6 sums up the drivers efficiency regarding different references. The most important is to see that the global order is maintained, especially between the *Eco coefficient* ranking and the current one (based on driving parameters). Figure 3.3 shows that three groups can easily be distinguished: this can also be observed on Figure 3.30. Indeed, E and D are on top on the ranking, while C and K are at the bottom. In between, small changes appear but they are minor. The gap between *good* and *mean* drivers, as well as between *bad* and *mean* drivers, is high: this is satisfying, as it is actually the case in the reality.

Nevertheless, some problems remain: J is given a low grade, being 9^{th} in the ranking, while his *Eco coefficient* is good. The opposite is observed for B, whose grade is much higher than his *Eco coefficient*.

Despite those differences, this ranking responds to what was targeted: a grade consistent with the drivers' *Eco coefficient* (one again, as the drivers were in similar conditions, their *Eco coefficients* can be compared) between 0 and 10 calculated from various driving parameters.

5 Conclusion

This study analysed drivers efficiency with respect to four main topics:

- Identification of parameters relevant for driving efficiency for different types of road,
- Detailed identification of parameters relevant for driving efficiency for three specific road features (altitude change, roundabouts and curves),
- Development of a ranking of drivers across all types of road types with respect to eco driving,
- Development of a system to identify the type of use of a truck based only on driving data.

All along this study, working with such a limited dataset has been a major issue, with some customers having data collected data only with one driver. In spite of this, it was important to focus on doing the best with the data available. The initial intentions were to have a large set of naturalistic data to work with, but for several reasons that was not later possible. Sub-optimization of results were considered inevitable with the limited dataset. Different types of regression analysis were used throughout the study, as well as classification and clustering software. Due to the very limited dataset much of the interpretation was qualitative inspection of graphs.

A set of parameters were identified as most relevant for driving efficiency. Driving efficiency was not defined only as fuel consumption but also included speed and kinetic energy. An additional set of parameters were identified for the three different road features altitude change, roundabouts and curves with respect to these parameters role in driving efficiency for these features. Further, a grading and ranking system was established based on driving parameters, making it possible to evaluate drivers without knowing the type of use and driving conditions. Finally, a system that fairly reliably could separate the different use of trucks based on driving parameters.

Future work should focus on using a larger set of data as well as more statistically rigorous analysis methods. Future data used should preferably be from naturalistic driving, where data is collected over such a long time and in a natural environment that the drivers forget that they are monitored.

References

[1] http://automotive.tomtom.com/software/eco.html. TomTom, 2012.
[2] *Active Prediction Presentation.* Scania, Dec. 2011.
[3] U. of Alcalá. *Report of Tests conducted on Vexia Econav.* Apr. 2010.
[4] C. Andrieu and G. Pierre. "Using statistical models to characterize eco-driving style with an aggregated indicator". *Intelligent Vehicles Symposium (IV), 2012 IEEE.* 2012, pp. 63–68. DOI: 10.1109/IVS.2012.6232197.
[5] T. A. Association. http://www.theaa.com/apps/eco-drive.html. 2012.
[6] BMW. http://www.youtube.com/watch?v=253tkqP_1y4. 2012.
[7] *Boîtier et application PC EcoGyzer.* Nomadic Solutions. Jan. 2012.
[8] M. Bremer. *Polynomial Regression Models.* San José State University, 2012.
[9] L. Brun. *Functional evaluation of NENPIOH PDA System on road.* UD Trucks, Jan. 2009.
[10] O. Carsten. *ecoDriver project overview.* Mar. 2012.
[11] H. J. Cho. *Eco Driving System.* http://kia-buzz.com/eco-driving-system/. Kia Motors Corporation, June 2008.
[12] N. M. Company. *ECO Pedal.* http://www.nissan-global.com/EN/TECHNOLOGY/OVERVIEW/eco_pedal.html.
[13] V. Corcoba Magana and M. Munoz-Organero. "Artemisa: An eco-driving assistant for Android Os". *Consumer Electronics - Berlin (ICCE-Berlin), 2011 IEEE International Conference on.* 2011, pp. 211–215. DOI: 10.1109/ICCE-Berlin.2011.6031794.
[14] B. Dog. *Heavy Duty WatchDog Flyer.* http://bigrig.bullydog.com/docs/Big Rig Flyer_WD.pdf. 2012.
[15] K. Driving. *KDTS.* http://www.keydriving.be/en/kdts_vrachtwagens_ma.html.
[16] *ECOMOBIL - Outil de formation a la conduite économique et écologique.* H.D.M Solutions Technologiques.
[17] E. Ericsson. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment* **6**.5 (2001), 325 –345. ISSN: 1361-9209. DOI: 10.1016/S1361-9209(01)00003-7. URL: http://www.sciencedirect.com/science/article/pii/S1361920901000037.
[18] A. Etemad. http://wiki.fot-net.eu/index.php?title=EuroFOT. Ford Research and Advanced Engineering Europe, 2008.
[19] Fiat. *Eco Driving Uncovered.* 2010.
[20] D. T. France. *DAF Advanced Transport Efficiency (ATe).* Sept. 2012.
[21] M. Fuso. *Eco Fleet Pro.* http://www.mitsubishi-fuso.com/jp/service/ecofleetpro/economy/main.html. 2011.
[22] Garmin. *Eco Route HD.* https://buy.garmin.com/en-GB/GB/prod38354.html. 2010.
[23] R. B. GmbH. *Driver assistance systems are becoming ever more capable.* Oct. 2012.
[24] Hino. *Drive Master Pro.* http://www.hino-sales.com/dmp/syousai.html. 2007.
[25] Honda. *Ecological Drive Assist System.* http://world.honda.com/news/2008/4081120Ecological-Drive-Assist-System/. 2008.
[26] *iShift - The new standard in fuel efficiency, safety and driver comfort.* Volvo Trucks, 2010.
[27] Isuzu. *Mimamori-kun Online Service.* http://www.japancorp.net/Article.Asp?Art_ID=14839. May 2007.
[28] Isuzu. *Vehicle Healt Reports.* http://www.isuzucv.com/images/healthreportbrochure.pdf. 2008.
[29] Iveco. *Blue&Me Fleet.* http://web.iveco.com/uk/Services/Pages/blue-me-fleet.aspx. 2008.
[30] Iveco. *Évaluation du style de conduite.* http://web.iveco.com/france/produits/pages/stralis-hi-way-evaluation-du-style-de-conduite.aspx. 2012.
[32] M. A. S. Kamal et al. "On board eco-driving system for varying road-traffic environments using model predictive control". *Control Applications (CCA), 2010 IEEE International Conference on.* 2010, pp. 1636–1641. DOI: 10.1109/CCA.2010.5611196.
[33] C. W. Kirkwood. *Decision Tree Primer.* Arizona State University. 2002.
[34] A. C. Leo Breiman. http://www.stat.berkeley.edu/ breiman/RandomForests/. University of California, Berkeley. 2001.
[35] H. Liimatainen. Utilization of Fuel Consumption Data in an Ecodriving Incentive System for Heavy-Duty Vehicle Drivers. *Intelligent Transportation Systems, IEEE Transactions on* **12**.4 (2011), 1087–1095. ISSN: 1524-9050. DOI: 10.1109/TITS.2011.2142182.
[36] C. Ludewig. *FleetBoard inside: telematics as standard on the new Actros.* Daimler, July 2011.

[37] A. Maincent. “Des sciences humaines aux sciences de l’ingénieur : Comportements humains, activités finalisées et conception de systèmes d’assistance à la conduite de Véhicules Industriels”. PhD thesis. Université Lumière Lyon 2, 2010.

[38] E. F. Niels Landwehr Mark Hall. “Logistic Model Trees”. *Proc 14th European Conference on Machine Learning*. Cavtat-Dubrovnik, Croatia. Springer, 2003, pp. 241–252.

[39] P.-E. Nordström. *Scania Driver Support – on-board economy and performance coaching*. Scania, Sept. 2009.

[40] *OmniExpress*. http://www.omnitracs.eu/files/4/OmniExpress%207-inch%20English.pdf. Qualcomm, 2013.

[41] T. Oravec. *PLX Presents the New Kiwi 2 Wifi*. PLX Devices Inc., 2009.

[42] O. ORFILA. *Effect of the percentage of drivers practicing ecodriving on fuel consumption, greenhouse gases emissions and traffic congestion*. Ifsttar. June 2011.

[43] J. R. Quinlan. *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN: 1-55860-238-0.

[44] *Renault Environment*. http://www.renault.co.uk/cars/environment/default.aspx. Renault, 2007.

[45] G. Rodríguez. *Generalized Linear Models: Linear Models for Continuous Data*. Princeton University, Sept. 2007.

[46] M. R.R. Bouckaert E.Frank. *WEKA Manual for Version 3-7-8*. The university of Waikato. 2013.

[47] H. R&T. *greenMeter 2.0 User Manual*. Aug. 2009.

[48] *Secrets of Better Fuel Economy: The Physics of MPG*. Cummins. 2009.

[49] F. Syed et al. “Adaptive real-time advisory system for fuel economy improvement in a hybrid electric vehicle”. *Fuzzy Information Processing Society, 2009. NAFIPS 2009. Annual Meeting of the North American*. 2009, pp. 1–7. DOI: 10.1109/NAFIPS.2009.5156404.

[50] J. F. Trevor Hastie Robert Tibshirani. *The Elements of Statistical Learning*. English. Vol. 2. Springer Series in Statistics, 2009. ISBN: 978-0-387-84857-0. DOI: 10.1007/978-0-387-84858-7.

[51] V. Trucks. http://www.volvogroup.com/group/global/en-gb/volvo%20group/our%20companies/gttechnology/pages/gtt.aspx. 2012.

[52] V. Trucks. http://fh.volvotrucks.com/en/index.html. 2012.

[53] M. Vassiliadis. *New Dynafleet app keeps track of the entire fleet*. Volvo Trucks, May 2012.

[54] T. Wada et al. “Proposal of an eco-driving assist system adaptive to driver’s skill”. *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. 2011, pp. 1880–1885. DOI: 10.1109/ITSC.2011.6083034.

[55] G. I. Webb, J. R. Boughton, and Z. Wang. Not So Naive Bayes: Aggregating One-Dependence Estimators. English. *Machine Learning* **58**.1 (2005), 5–24. ISSN: 0885-6125. DOI: 10.1007/s10994-005-4258-6. URL: http://dx.doi.org/10.1007/s10994-005-4258-6.

A *Eco coefficient* values for each section

The 5 following graphs present the *Eco coefficient* value for each driver, on each driving section. If, overall, the efficient and non efficient drivers are the same, there are some slight differences. They can be due to 2 reasons:

- External parameters: everything that has nothing to do with the truck or the driver. Most of the time, it is either the wind or the traffic jam.
- Driver ability: this is due to the fact that those sections are quite different. Some drivers may, for instance, feel more comfortable on highway sections than in city centres. This ability can affect their driving performance.

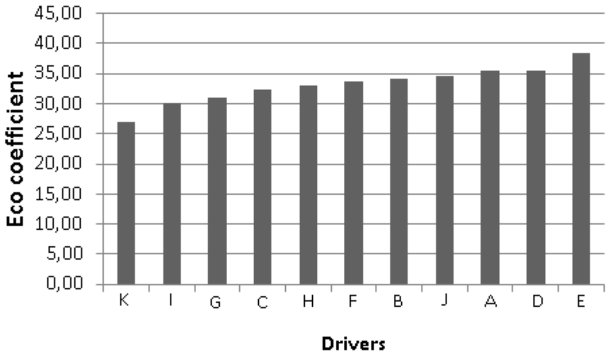

Figure A.1: *Eco coefficient for Section 1B*

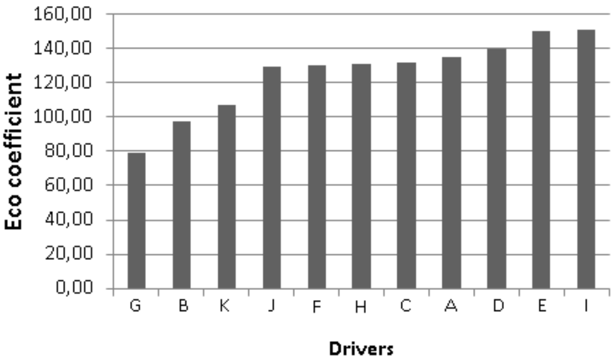

Figure A.2: *Eco coefficient for Section 2B*

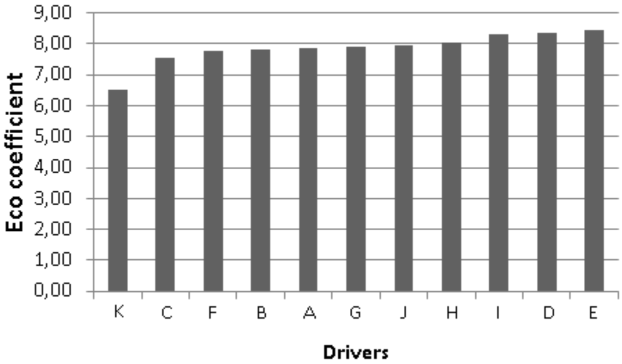

Figure A.3: *Eco coefficient for Section 4B*

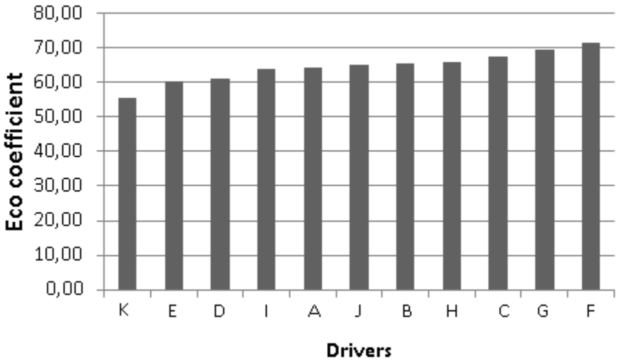

Figure A.4: *Eco coefficient for Section 111*

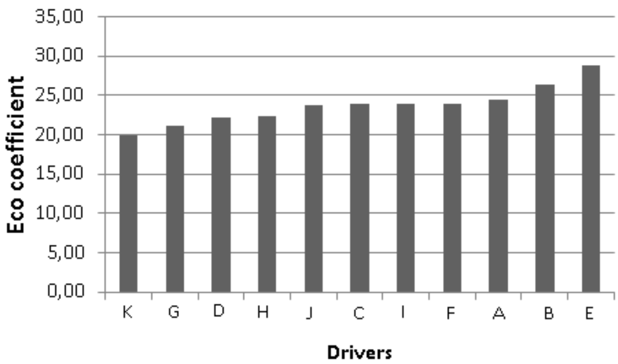

Figure A.5: *Eco coefficient for Section 112*

B Braking pedal curve

In Figure B.1, the brake pedal angle signal is not constant. That is the reason why the threshold to consider for brake phases identification is important. E.g., if a value of 20 %, is chosen, then the following happens: many additional phases pop up due to the curve oscillations around this limit. Finally, a lower threshold (5 %) is recommended. An alternative approach would be to include an hysteresis and thus, not producing multiple

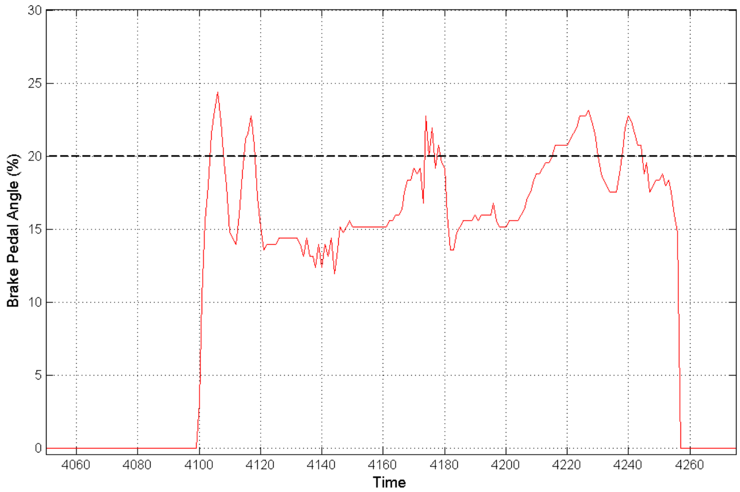

Figure B.1: *Braking pedal curve*

braking events even with erratic data around the threshold.

C Model used for the determination of relevant slopes

On the following curve, the curves represent the model which is used in order to determine whether a slope is relevant or not. This model is split into 2 cases:

- Downhill slopes: due to the vehicle inertia, too low slopes are not taken into account. The higher the speed, the higher the inertia: that is the reason why the threshold increases with the vehicle speed. This was determined based on previous internal tests.
- Uphill slopes: due to vehicle weight, the threshold decreases with the vehicle speed. Indeed, it is much more difficult for a vehicle to drive uphill at high speed.

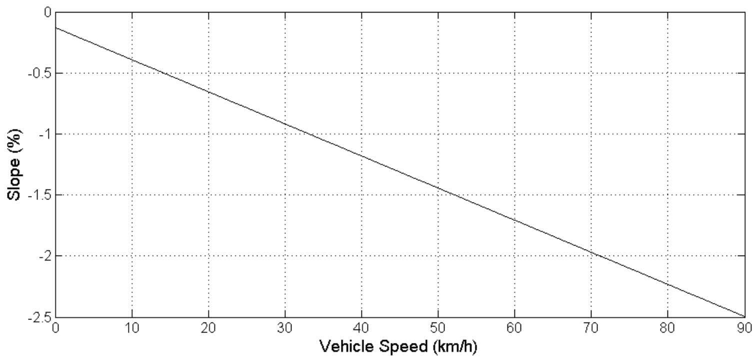

Figure C.1: *Relevant downhill phases*

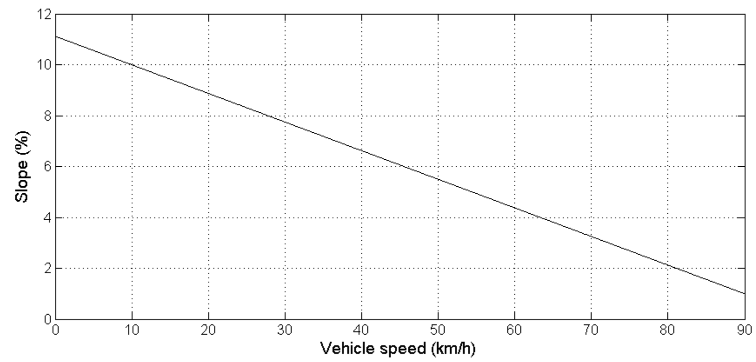

Figure C.2: *Relevant uphill phases*

D Description of other correlation tools

There are plenty of existing tools allowing a statistical analysis. Below are some of them that could also have been interesting for this study:

- **Random Forest**. This algorithm was introduced by Leo Breiman and Adele Cutler, and is based on decision trees[34]. Indeed, this algorithm generates many trees based on combinations of those parameters. If N is the number of cases and M the number of parameters, a number of cases $n \leq N$ and a number of parameters $m \ll M$ are chosen for the decision phase. The remaining parameters and cases are used for prediction, i.e. to estimate if the previously chosen values are accurate enough for the classification. This procedure is then iterated with different n and m. Each iteration products a decision tree. In the end, those trees are combined in order to identify the most relevant parameters.

This method is interesting for a large number of parameters as it allows to sort them out regarding their importance. However, it was decided to keep a simpler approach within this study, with the use of the J48 algorithm.

- **Naive Bayes**. This algorithm is inspired of Bayes' theorem, assuming that all the parameters are independent (which is not completely true, hence the name). This theorem is based on the following equation, considering 2 parameters A and B:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{D.1}$$

  Where $P(A)$ is the probability of A without taking B into account, $P(A|B)$ is the probability of A knowing B, $P(B|A)$ is the probability of B knowing A,and $P(B)$ is the probability of B without taking A into account.
  With the Naive Bayes algorithm, this method is applied to a much bigger number of parameters. Due to its simplicity, efficiency and efficacy, Naive Bayes is widely deployed for classification learning. It is limited, however, by the limitations of the attribute independence assumption[55], particularly in the current case, where the system (truck and driver) is very complex. A decision tree can also be produced out of this method[38].

Other methods exist, such as the *Nearest Neighbours* method or the *Kernel Smoothing* method for instance[50], however they are not detailed in this report.

E Driving parameters look-up table

Table E.1 gathers the parameters which are investigated in this study, for the determination of relevant driving parameters on steep road.

Table E.1: Output from Matlab code - Road altitude

Driving Parameters
Average throttle pedal angle on start of downhill phases
Average throttle pedal angle on end of downhill phases
Average throttle pedal angle on start of uphill phases
Average throttle pedal angle on end of uphill phases
Average speed on uphill phases
Average speed on downhill phases
Average speed once uphill
Average speed once downhill
Average torque on uphill phases
Average torque on downhill phases
Average brake pedal angle on start of downhill phases
Average brake pedal angle on end of downhill phases

Table E.2 gathers the parameters which are investigated for the determination of relevant driving parameters on roundabout.

Table E.2: Output from Matlab code - Roundabout

Driving Parameters
Distance between place of lowest speed and place where event happens
Length between lowest speed and event
Minimum speed on event
Distance between place where braking starts and event
Distance between place where braking ends and event
Distance between place where throttle starts and event
Length of braking phase
Length between end of braking phase and start of throttle phase
Average brake pedal angle during event
Average brake pedal angle before event
Speed when braking phase starts
Speed when braking phase ends
Speed when throttle starts

Table E.3 gathers the parameters which are investigated for the determination of relevant driving parameters in curve.

Table E.3: Output from Matlab code - Curve

Driving Parameters
Distance between place of lowest speed and place where event happens
Minimum speed on event
Average throttle pedal angle between 200 m and 100 m before event
Average throttle pedal angle between 100 m and 0 m before event
Average throttle pedal angle between 0 m and 100 m after event
Average throttle pedal angle during event
Average brake pedal angle between 200 m and 100 m before event
Average brake pedal angle between 100 m and 0 m before event
Average brake pedal angle between 0 m and 100 m after event
Average brake pedal angle during event
Speed during event
Speed before event
Speed after event
Speed during / (Speed before * Speed after)

F Calculation of *Mean Braking Force*

The following equations show how the *Mean Braking Force* was determined using only CAN signals.

- Acceleration

$$F_{acceleration} = Mass_{vehicle} \frac{\nabla Speed_{vehicle}}{\nabla Time} \tag{F.1}$$

- Aerodynamic

$$F_{aerodynamic} = \frac{1}{2} \times \rho \times SCx \times Speed^2_{vehicle} \tag{F.2}$$

- Rolling resistance

$$F_{rolling} = C_{rolling} \times Mass_{vehicle} \times g \tag{F.3}$$

- Slope

$$F_{slope} = Mass_{vehicle} \times g \times \sin(\alpha_{slope}) \tag{F.4}$$

- Traction

$$F_{traction} = \frac{Torque_{engine} \times Ratio_{gearratio}}{Radius_{wheel}} \tag{F.5}$$

According to Newton's second law:

$$\sum F_{net} = Mass_{vehicle} \times \frac{dV}{dt} \tag{F.6}$$

Thus, the braking force:

$$F_{braking} = F_{traction} - [F_{acceleration} + F_{aerodynamic} + F_{rolling} + F_{slope}] \tag{F.7}$$

G Relevant parameters identification based on *Fuel consumption*

The influence of Driver K on the results was very large: this kind of problem is hard to deal with when the complete sample is so limited. The calculations was run with and without Driver K, to find out about parameters sensitivity. Figure G.1 shows the results based on the R^2 (coefficient of correlation). It is extremely

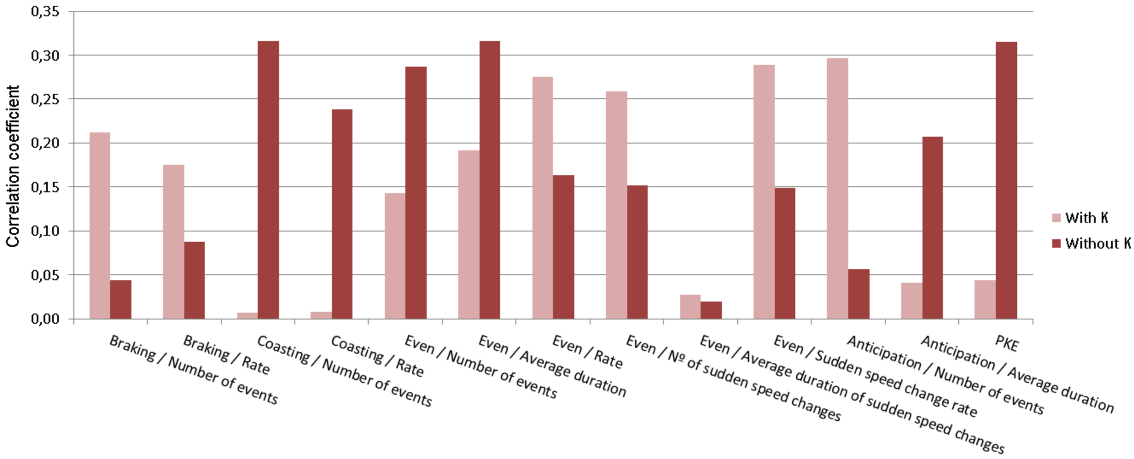

Figure G.1: *Comparison of the results found, with and without WI*

hard to draw conclusions based on these results. Indeed, the influence of K is huge, and there is a big gap between the different criteria. In addition, the correlation values are quite low: the best ones are just above 0.3. Despite this complexity, each of the 5 criteria is analysed.

- **Braking**: K is a beginner and it was noticed during the driving session, that he happened to brake a lot. As his fuel consumption is also the highest, which explains why the coefficients are much higher when he is taken into account. Figure G.2, where K is the top right dot, highlights this influence: For the

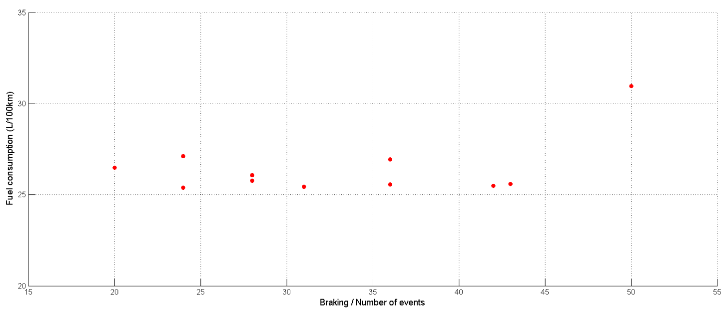

Figure G.2: *Scatter of braking events for the 11 drivers*

Braking criterion, no relevant conclusion can be drawn: indeed, due to this sensitivity to one individual driver, there is no parameter giving good results.

- **Coasting**: the 2 coasting criteria are even more sensitive. in Figure G.3, one can see that K is still out of the cloud (top left).

- **Even driving**: the results for this criterion are good and have a low sensitivity. One can already distinguish that even driving rate is an interesting parameter to consider, as well as the rate of sudden speed changes. For this criterion, K's behaviour is consistent with the other drivers. One can already see that the capacity of keeping a constant speed seems to play an important role when it comes to fuel efficiency.

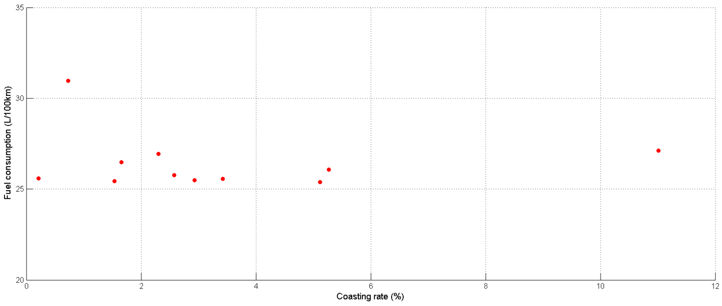

Figure G.3: *Scatter of coasting rate for the 11 drivers*

- **Anticipation**: it is hard to make sense of this criterion, as variations are different for the two driving parameters. In addition, the sensitivity is very high. Different parameters have to be investigated, such as initial and final speeds, in order to come up with something more reliable.
- **PKE**: the positive kinetic energy coefficient illustrates the acceleration amplitude. Efficient drivers should have a low PKE, i.e. low amplitudes. On the opposite, non-efficient drivers have a non efficient behaviour. Surprisingly, in spite of his high fuel consumption, K has a low PKE (as shown on the following scatter graph): that is the reason why the correlation is much lower when he is taken into account. However, the trend in Figure G.4 seems to indicate, considering the 10 remaining points, that a higher PKE leads to a higher fuel consumption.

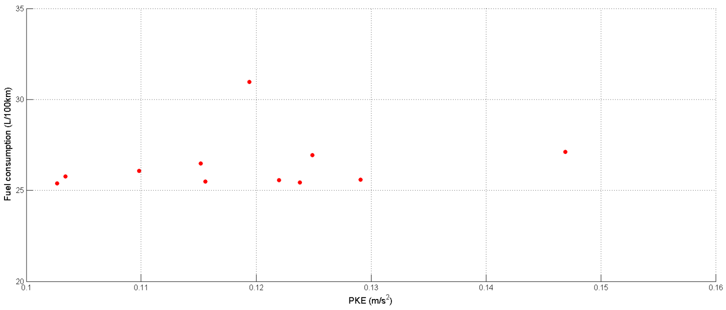

Figure G.4: *Scatter of PKE for the 11 drivers*

However, the correlation values found with this set of data are much lower than what the previous study highlighted. In Figure G.5, a comparison of the scatter graph for PKE illustrates this difference (top: current study, bottom: previous study). If *Even* and *Coasting* criteria can be considered as independent from K,

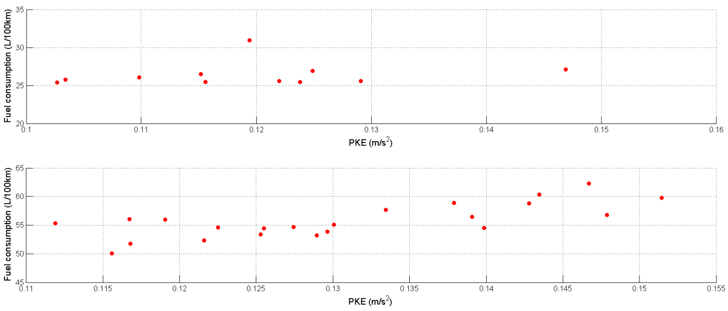

Figure G.5: *Comparison of the PKE values for current and previous studies*

Braking and *Anticipation* are much more complicated to handle.

H Results of the multicriteria analysis for relevant parameters determination

In the following part are shown the results of both *J48* (i.e. decision trees) and *Simple Logistics* (i.e. linear equations) algorithms. It consists of 2 parts: results for each section, and for the full cycle.

H.0.1 Analysis per section

Section 1B

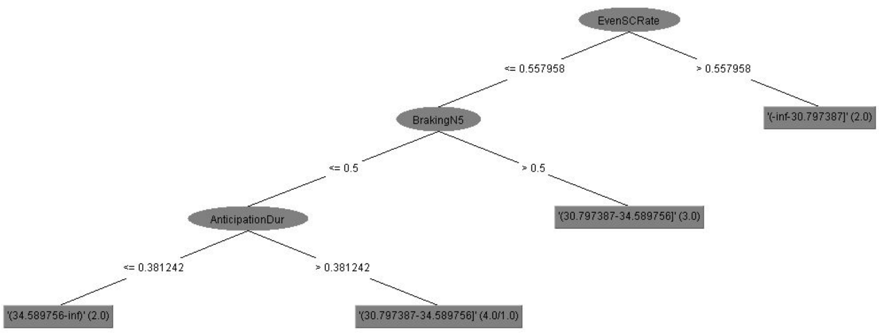

Figure H.1: *J48 Decision tree for Section 1B*

$$Baddrivers = -55.45 + [EvenSCRate] * 98.7 Meandrivers = 1.38 + [AnticipationDur] * 1.16 + [EvenDur] * -2.02 + [Eve \tag{H.1}$$

Section 2B

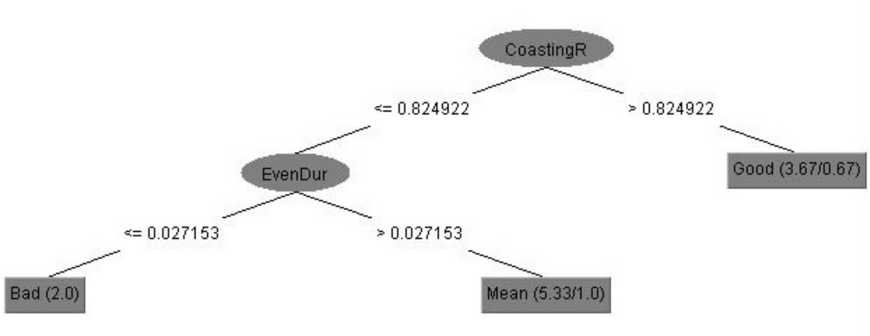

Figure H.2: *J48 Decision tree for Section 2B*

$$Baddrivers = 2.76 + [CoastingR] * -10.18 \tag{H.2}$$

$$Meandrivers = 1.85 + [AnticipationDur] * 0.6 + [BrakingRate] * -1.89 + [EvenDur] * -0.94 \tag{H.3}$$

$$Gooddrivers = -7.31 + [CoastingR] * 9.8 \tag{H.4}$$

Section 4B

$$Baddrivers = -16.01 + [BrakingN] * 20.86 \tag{H.5}$$

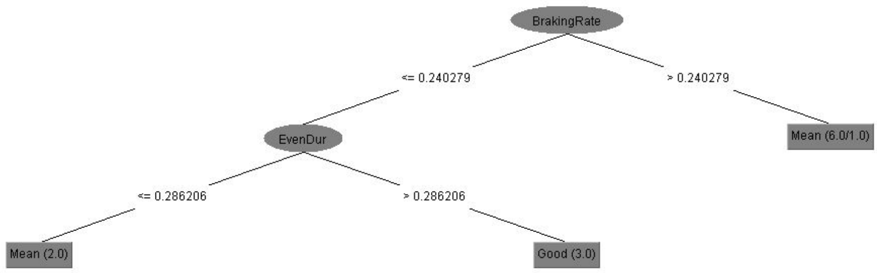

Figure H.3: *J48 Decision tree for Section 4B*

$$Meandrivers = -1.54 + [AnticipationDur] * 3.38 + [BrakingRate] * 4.26 + [BrakingN] * -1.54 + [EvenDur] * -1.32 \tag{H.6}$$

$$Gooddrivers = 4.16 + [AnticipationDur] * -3.45 + [BrakingRate] * -7.5 + [EvenDur] * 2.22 \tag{H.7}$$

Section 111

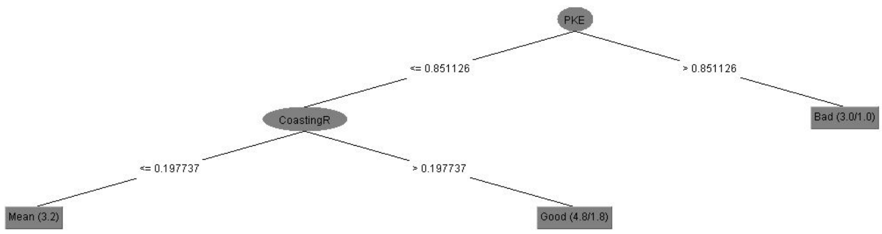

Figure H.4: *J48 Decision tree for Section 111*

$$Baddrivers = 0.37 + [EvenRate] * -2.06 \tag{H.8}$$
$$Meandrivers = 1.42 + [CoastingN] * -2.88 \tag{H.9}$$
$$Gooddrivers = -1 + [CoastingN] * 3 \tag{H.10}$$

Section 112

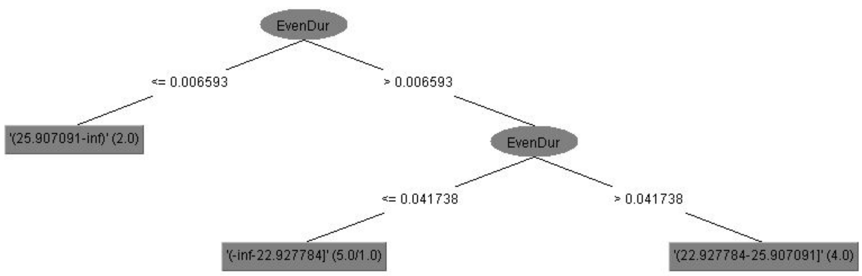

Figure H.5: *J48 Decision tree for Section 112*

$$Baddrivers = -1.25 + [BrakingN] * 2.85 \tag{H.11}$$
$$Meandrivers = -0.32 + [EvenDur] * 2.68 \tag{H.12}$$
$$Gooddrivers = 1.62 + [PKE] * -3.25 \tag{H.13}$$

H.0.2 Analysis on full cycle

Basic parameters

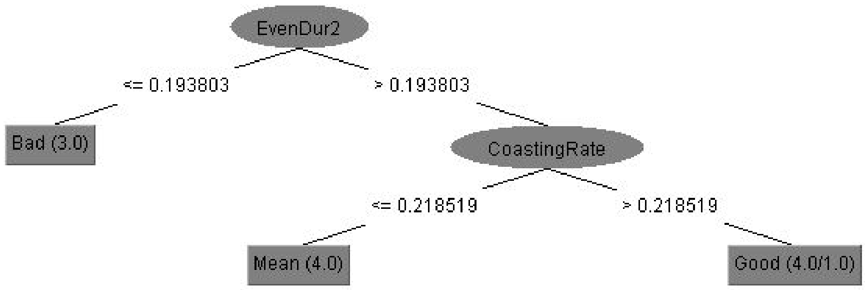

Figure H.6: *J48 Decision tree for full cycle - Basic parameters*

$$Baddrivers = -4.87 + [EvenDur2] * -30.01 + [EvenSCRate] * 15.3 \tag{H.14}$$
$$Meandrivers = 6.46 + [CoastingRate] * -17.54 + [BrakingN5] * -2.1 + [EvenSCRate] * -1.4 \tag{H.15}$$
$$Gooddrivers = 0.75 + [BrakingRate5] * -7.07 + [BrakingN5] * 5.72 \tag{H.16}$$

Additional parameters

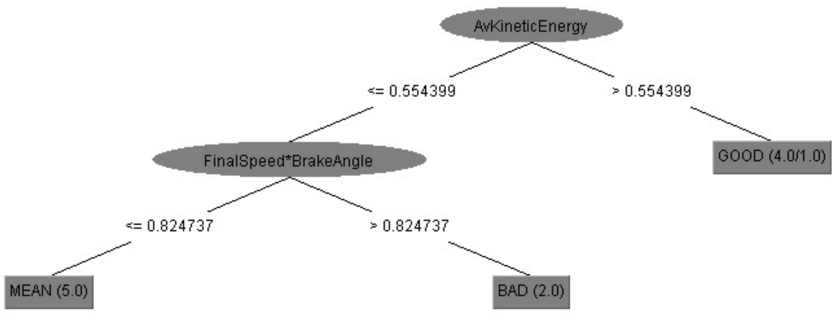

Figure H.7: *J48 Decision tree for full cycle - Additional parameters*

$$Baddrivers = -11.63 + [AvKineticEnergy] * -0.01 + [MeanGasPedal] * 0.1 + [Sumgrad(GasPedal)] * 0Meandrivers = \tag{H.17}$$

I Ranking based on Simple Logistics equations

Based on equations generated by *Simple logistics*, 3 coefficient matrices are calculated for the whole driving cycle. These 3 matrices refer to the 3 classes, bad, mean and good drivers, and their sizes are $[11 * 5]$. Indeed, 11 parameters are considered and coefficients are calculated for all 5 sections.

The results are shown on Tables I.1, I.2 and I.3.
The 3 matrices are as follows:

- Non-efficient drivers

Table I.1: Coefficients related to non-efficient drivers

Parameter	Section 1B	Section 2B	Section 4B	Section 111	Section 112
Coefficient	-55.45	-3.71	4.16	10.62	-1.25
Sudden speed change rate	98.7	0	0	0	0
Even driving average duration	0	0	2.22	0	0
Braking rate	0	0	-7.5	0	0
Braking number of events	0	0	0	0	2.85
PKE	0	0	0	0	0
Anticipation average duration	0	0	-3.45	0	0
Anticipation number of events	0	0	0	0	0
Coasting number of events	0	0	0	0	0
Coasting rate	0	-17.78	0	0	0
Even driving rate	0	0	0	-107.9	0

- Mean drivers

Table I.2: Coefficients related to mean drivers

Parameter Section	1B	Section 2B	Section 4B	Section 111	Section 112
Coefficient	1.38	-0.56	-1.54	3.84	-0.32
Sudden speed change rate	-17.35	2.72	0	0	0
Even driving average duration	-2.02	-0.53	-1.32	0	2.68
Braking rate	12.73	-0.78	4.26	0	0
Braking number of events	0	2.49	-1.54	10.21	0
PKE	0	0	0	2.71	0
Anticipation average duration	1.16	3.27	3.38	0	0
Anticipation number of events	0	0	0	-12.62	0
Coasting number of events	0	0	0	-4.62	0
Coasting rate	0	0	0	0	0
Even driving rate	0	0	0	0	0

- Good drivers

Driving parameters were normalized in *Weka* (cf. equation 2.25) for the calculation of the *Simple Logistics* equations, on each section. To fit these conditions, they were also normalized before being multiplied to these matrices. The result of this calculation is valid for the whole cycle, as it comes from the combination of the 5 sections. This matrix is then multiplied to each of the 3 matrices related to driver classes. Figure I.1 shows the results of this computation: In those equations, the class related to a given behaviour (*bad*, *mean* or *good*) is supposed to get the highest score on the corresponding equation. For instance, for the equation related to mean drivers, the mean driver class is supposed to have a grade higher than good and non-efficient driver classes.

Table I.3: Coefficients related to mean drivers

Parameter Section 1B	Section 2B	Section 4B	Section 111	Section 112	
Coefficient	21.26	-6.48	-16.01	-2.74	1.62
Sudden speed change rate	0	0	0	0	0
Even driving average duration	0	0	0	0	0
Braking rate	0	0	0	0	0
Braking number of events	0	0	20.86	-18.68	0
PKE	0	-20.89	0	0	-3.25
Anticipation average duration	0	0	0	0	0
Anticipation number of events	0	0	0	8.89	0
Coasting number of events	0	0	0	7.6	0
Coasting rate	0	16.39	0	0	0
Even driving rate	0	0	0	3	0

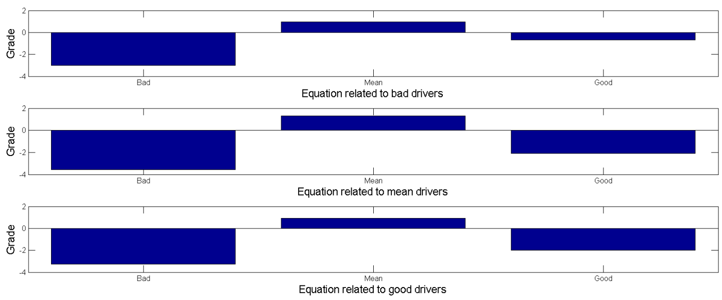

Figure I.1: *Driver ranking following Simple Logistics equation method*