École polytechnique de Louvain

Is the Real Driving Emission methodology really representative of the people's driving behaviour? Study of the driving patterns in Brussels.

Author: **Camille Lambert**
Supervisors: **Hervé Jeanmart, Francesco Contino**
Readers: **François Boveroux, Paul Fisette, Benoît Herman**
Academic year 2018-2019
Master [120] in Mechanical Engineering

Abstract

More and more cars are circulating on the roads and this significantly impacts the health of the population as well as the air quality. In that sense, norms on the emitted pollutants of vehicles have been set by the European Union. These norms must be respected by all new vehicles which cannot be launched on the market if they do not pass the tests inspecting their emissions. For years, the emissions of the vehicles have been tested exclusively on a dynamometer in a laboratory. However, most of these testing do not reflect the reality since external factors such as traffic, weather, driving dynamics and shifting behaviour are not taken into account. Therefore, real driving cycles are needed because they are more dynamic. To that extent, the Real Driving Emission (RDE) test has been developed over the years and implemented in 2017. The goal of this test is to measure emissions in real time on public roads. Basically, the vehicle is instrumented with portable emissions measurement systems and the main parameters are the average speed, the distance, the maximal speed and the lower and upper dynamic conditions. This work aims at analysing if these conditions are relevant and representative of Brussels drivers behaviour.

Therefore, data loggers recording twelve parameters have been developed and installed in the car of 51 Brussels drivers from one to two weeks. Then, the data were sorted and compared to the conditions of the RDE. The results show that Brussels people's driving behaviour does not match with the requirements of the RDE for the distance and the stop proportion but were acceptable for the dynamic conditions, the maximal speed and the average speed. The influence of external factors which are the fuel type, whether the vehicle is a company car, the category of displacement, of power and of year of construction, the type of gearbox, whether the engine was cold when it started, the time of the week and the time of the day has also been analysed to understand what impacts the most the RDE conditions. Moreover, a preliminary study was conducted on the data outside of Brussels in one vehicle. They have been compared to data from Brussels and they seem to indicate that the location has a strong impact on the parameters studied.

Acknowledgements

First, I would like to express my special thanks to my supervisors, Professor Hervé Jeanmart and Professor Francesco Contino for their guidance, support, advises and availability throughout this thesis.

Then, I would also extend my gratefulness to François Boveroux for his patience, help and kindness throughout the year but also for the development of the devices required for the study and his involvement in the management of many drivers for the research.

I also thank Professor Paul Fisette and Professor Benoît Herman for accepting the role of readers and therefore for adding more work on their stack of theses to examine.

The data loggers would not have been built without Lily and her meticulousness appreciated during the research and development of these devices.

The study could not have been conducted without the participation of the 51 drivers which I sincerely thank for their help and involvement.

Furthermore, I would like to acknowledge Catherine Rasse for her precious advises in statistics, Arnaud Roisin for helping me when Matlab was driving me crazy, Christophe Wilms for helping me to find drivers and to handle so many dataloggers, and Marie-Charlotte Sparenberg as well as my father, Didier Lambert for their time spent in the proofreading of my thesis.

Finally, I could not finish these thanks without submitting my heartiest gratitude to my family, friends and flatmates for their unconditional support, continuous encouragement and for listening to my complaints. This thesis completes five years of study that I could not have passed without them.

Contents

Glossary

The next list describes several acronyms and abbreviations that will be later used within the body of the document. These are explained within the body of the text where appropriate.

CAN	Controller Area Network
CH_4	Methane
DU	Dobson Unit
ECU	Engine control unit
EU	European Union
HC	Hydrocarbons
NEDC	New European Driving Cycle
NMHC	Non-methane hydrocarbons
NO_x	Nitrogen oxides
OBD	On-board diagnostics
PEMS	Portable emissions measurement systems
PKE	Positive Kinetic Energy
PN	particle numbers
Q-Q plot	Quantile-quantile plot
RDE	Real Driving Emissions
RPA	Relative positive acceleration
RPM	Rotations Per Minute
THC	Total Hydrocarbons
VOCs	Volatile organic compounds
WLTP	Worldwide harmonised Light vehicles Test Procedure

CHAPTER 1

Introduction

The first prototype of a car appeared in the late 18th century. Only one hundred years later, the design of combustion engines as well as the launch of manufactured cars allowed the growing success of this mode of transportation.

The amount of vehicles worldwide (of all types, including busses and heavy commercial vehicles) reached 1.28 billion in 2015 [1]. Within Europe, almost 400 million vehicles were recorded in 2016 [2] whereas in Belgium, in 2017, around 8 million vehicles were counted, including 5.7 million cars. More recently, according to an article in the newspaper *Le Soir*, the amount of private cars reached 5.85 million on August, 1st 2018. This represents a rise of 1.2% compared to the previous year [3]. FIGURE 1.1 shows the evolution of the number of cars over the years in Belgium, based on the data from FEBIAC, the Belgian federation of the automobile and cycles [4]. All of these facts unequivocally show the growing use of cars in Belgium.

Figure 1.1: Increase of the amount of cars in Belgium over the years 2002 - 2017 [4].

Car traffic is one of the factors affecting both quality of air and climate changes. Car users are now sensitised to these matters and recently a shift of consumer choice was observed. Diesel motors sales are now in decline mainly due to regulations regarding low emission zones in cities (they decreased by 4.4% from 2017 to 2018 [3]) whereas motors using alternative fuel such as electric, compressed natural gas and even hydrogen engines are in progress.

This introduction further describes three main topics: 1) the effects of vehicles on air quality and health, 2) the different norms implemented to decrease the emission levels and 3) the testing and cycles created to verify that the new cars fulfil the requirements.

On October 28, 2018, Greenpeace Belgium has released an interactive map from the data given by the ESA (European Space Agency), showing the world in terms of NO_2 concentration and highlighting the 50 biggest hotspots. The average NO_2 concentration of the world is 0.07 DU (Dobson Unit[1]). The map, zoomed in on Belgium, is shown in FIGURE 1.2. Belgium is one of the most polluted country of Europe, with Antwerp being one of the fifty hotspots in the world.

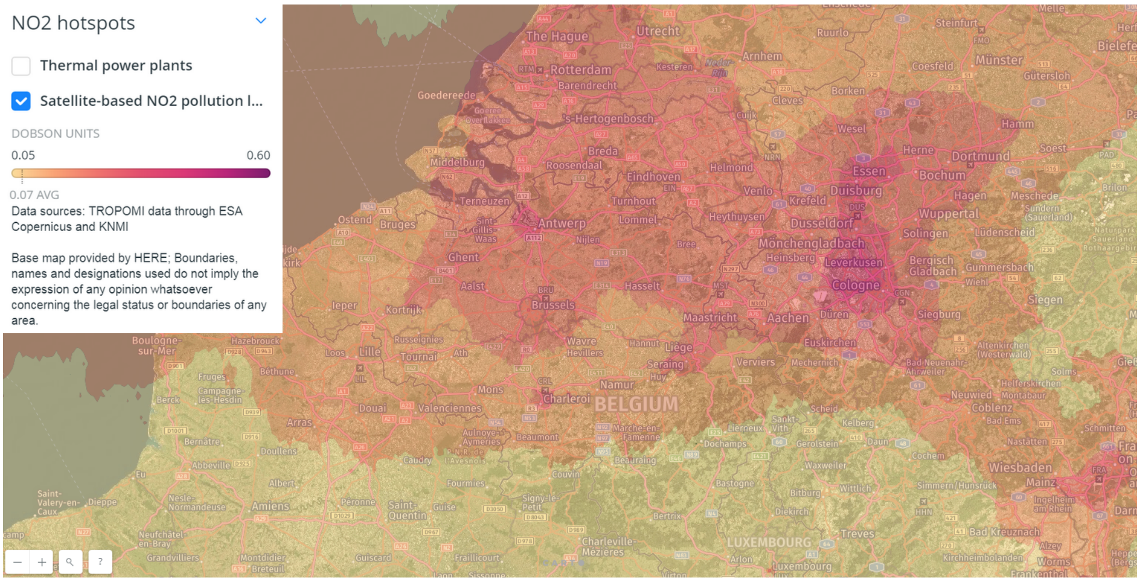

Figure 1.2: Satellite-based NO_2 pollution levels in Belgium, from Greenpeace, 2018 [6].

In Antwerp, the port industry has a huge impact on air pollution. However, according to Joeri Thijs, "*the traffic is the major cause of the impact of the pollution in Antwerp on our health and especially on our children's. We are more exposed to the tailpipes from the cars than the air of the chimneys from the port in Antwerp*". The comparison in FIGURE 1.3 proves that statement, since the NO_2 level increases during the workdays whereas most of port activities keep on at the weekend. Indeed, the NO_2 level increases up to 0.3 DU, in Antwerp and to 0.25 DU in Brussels during the workdays whereas it is around

[1] *The Dobson Unit is used to measure the amount of any gas in the atmosphere based on the thickness of the layer that gas would form if all other gases were removed and if the gas was subjected to sea-level pressure*, from the Dictionary of Environment and Conservation [5].

0.25 DU and less than 0.2 DU for Antwerp and Brussels respectively at the weekend. In Brussels, the reduction reaches -31.6% at the weekends [6].

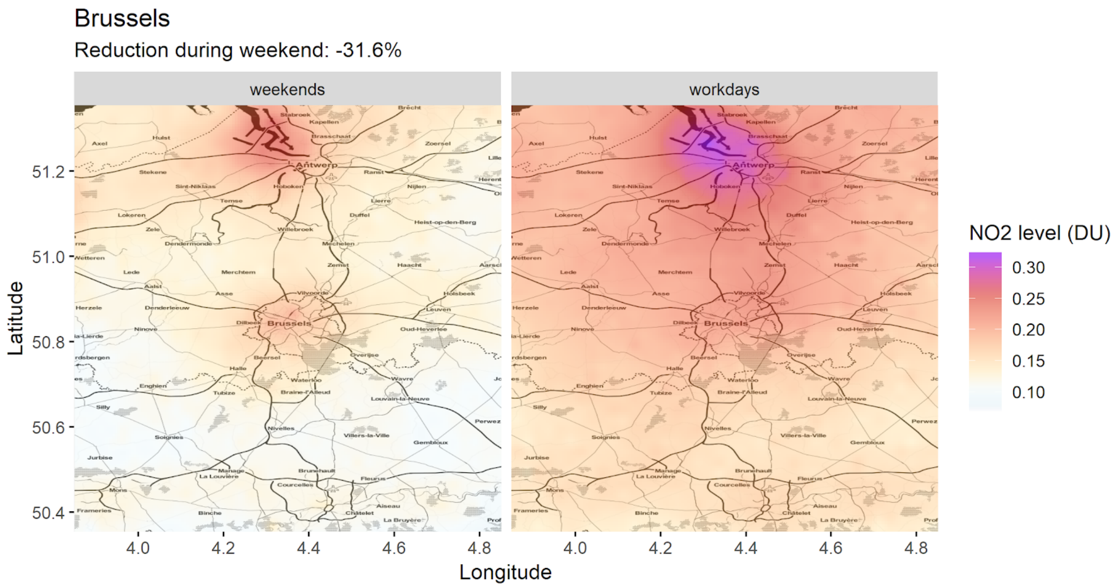

Figure 1.3: Increase of NO_2 levels in Brussels and Antwerp during the workdays compared to the weekends, from Greenpeace, 2018 [6].

The consequences of the atmospheric pollution on health are serious. Every year, 3.3 million premature deaths and many more associated cases of illness around the world are recorded due to poor air quality [7].
According to the European Environment Agency's annual report from 2017, air pollution has a significant impact on health (especially in urban areas) but also on economy since it increases medical costs and reduces productivity during the working days. This report also shows that the most serious pollutants in Europe in terms of harming citizens are particulate matter (PM), Nitrogen dioxide (NO_2) and ground-level ozone (O_3). In Belgium more precisely, in 2014, 8340 premature deaths were attributed to the exposure of PM2.5 (Particulate matter with a diameter of 2.5 μm or less), 1870 from NO_2 and 190 from O_3 [8]. However, PM2.5 does not only come from vehicles but also from natural sources such as seismic or geothermic activities or from other human activities which are mainly the energy production and the agricultural sector [9].

These effects on health impact everybody. Cyclists and bus passengers are even more exposed to traffic-related air pollutants than car drivers if we consider closed car windows [10]. However, these conditions may vary with the street design of the town under investigation, or with the study design such as the sample number, weather and season [11]. Luckily, a lot of policies and norms are developed nowadays to reduce traffic-related air pollution. Thanks to that, the emissions and population exposure are decreasing and this has significant positive effects on health [12].

For example, the European Union introduced in 1992 the first European emission standards (which had effect on cars launched in 1993), called Euro 1, for the light duty vehicles (lighter than 3500 kg). There are subcategories of each norm depending on the type of light duty vehicles (for example: passenger cars or light commercial vehicles heavier than 1760 kg). There are corresponding standards for heavy-duty vehicles (busses and lorries). Since then, five more severe Euro norms on emissions emerged to improve the air quality as shown in FIGURE 1.4 for the case of the norms on Diesel passenger cars. Each colour represents a different norm and the year indicated is the year of commissioning of the cars that must respect the new standards. This representation demonstrates that each norm became increasingly limited and strict in terms of particulate matter emissions (from 140 mg/km to 4.5 mg/km) and in terms of NO_x and Hydrocarbons (HC) together (from 970 mg/km to 170 mg/km) and that illustrates the impact of awareness of the effects of emissions.

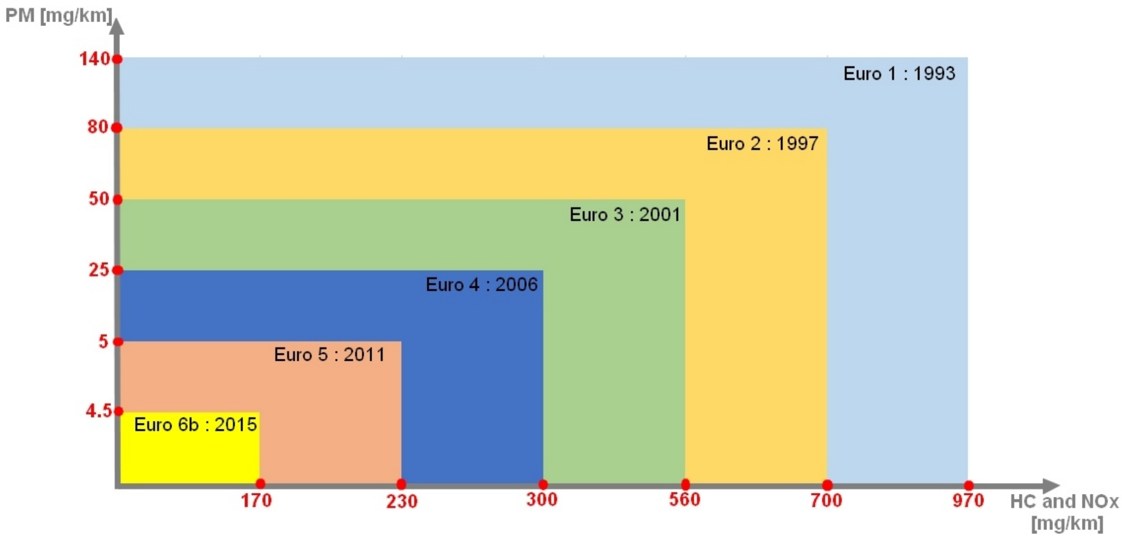

Figure 1.4: Reduction of the EU emissions standards for Diesel passenger cars in terms of particle and NO_x + HC emissions.

The goal was to fix maximal emissions limits, for new vehicles sold in the European Union, with respect to the four following main emissions: Carbon monoxide (CO), Hydrocarbons, Nitrogen oxides (NO_x) and Particulate matter (PM) [13]. These norms helped to reduce CO emissions by 82% for diesel-engine cars. For gasoline cars, the CO, the hydrocarbon and the NO_x emissions have decreased by 63%, 50% and 84% respectively. Particulate matter emissions have been down by 96%. New cars that do not comply with these norms cannot be sold in the EU.

The New European Driving Cycle (NEDC) has been introduced in 1997 in the EU to assess the emissions of cars. Driving Cycles are graphs of the speed in function of time which represent driving patterns in a region or city. They are used to simulate driving conditions on a laboratory chassis dynamometer for evaluation of fuel consumption and exhaust emissions [14]. The NEDC has been used in the EU since then. Basically, every car is tested, before its launching, on a dynamometer that simulates the resistance of the drag force and the resistance for the vehicle's mass. The goal of this test is to see if the car respects the emissions imposed and if they do, they are permitted to be put on the market. The test takes place in a laboratory under well-defined ambient conditions of temperature and humidity. The motor must be cold (between 20°C and 30°C). The test lasts for 19 minutes and 40 seconds and accelerations are performed with the car to reach speeds with respect to FIGURE 1.5. The first 800 seconds of the cycle represent the urban driving and the last part simulates an extra-urban driving behaviour. The

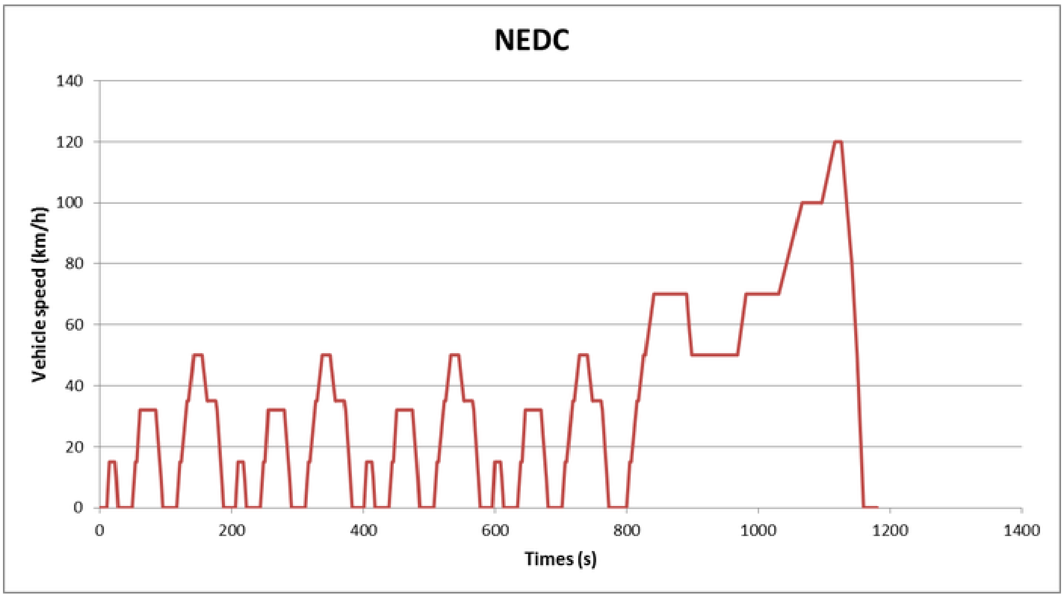

Figure 1.5: Speed of the vehicle in function of time for the NEDC, from [15].

emissions of the car under investigation are then measured and compared to the norms. However, it has been noticed over the years that the NEDC was only covering a small part of the engine operating range and is not representative enough of the reality [16]. In fact, this test underestimates emissions by up to 30% and 50% for cold and hot emissions respectively [17]. This is due to the too soft accelerations in the cycle and to the too high numbers speed cruises.

The Worldwide harmonised Light vehicles Test Procedures (WLTP) have been developed by the European Union, Japan and India to replace the NEDC [18]. It is also conducted on a dynamometer. This laboratory procedure measures fuel consumption and CO_2 emissions for cars and vans. Three classes of WLTP exist based on the power to mass ratio (PMR) of the car. The first class includes vehicles with a PMR lower than 22 W/kg; which corresponds to low power vehicles. The second class groups the cars with a PMR from 23 W/kg to 33 W/kg. The last class covers automobiles with a PMR higher than 34 W/kg. The values computed by the WLTP are more representative of the reality than the ones computed with the NEDC [19] and this method uses a more dynamic cycle [20] that the car must follow. FIGURE 1.6 illustrates the WLTP cycle for the third class of vehicles.

However, most of these testing do not entirely reflect the reality since the external

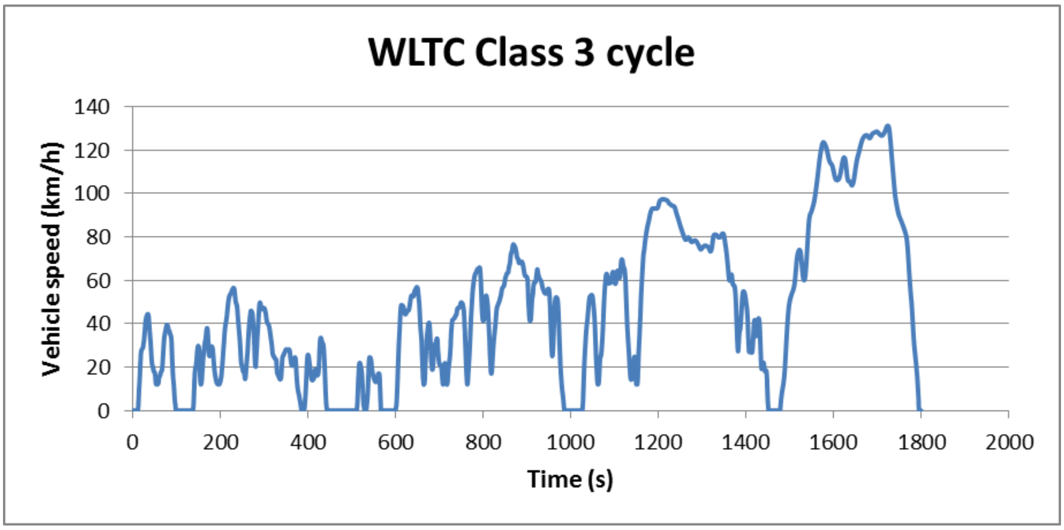

Figure 1.6: Speed of the vehicle in function of time for the third class of WLTP, from [15].

factors such as traffic, weather, driving dynamics and shifting behaviour are not taken into account. Moreover, the real conditions lead to higher pollutant emissions [14].

To compensate the gap between the laboratory testing and the reality, the European Commission has introduced the Real Driving Emissions test procedure (RDE) on the first of September 2017 [21]. In other words, the EU vehicle emissions legislation imposed that the new cars comply with real driving emissions requirements. It permits making cars less polluting over a better and extended set of operating conditions compared to the former cycles and tests [22]. This test of cars emissions uses portable emissions measurement systems and the data are recorded on-board while driving on public roads. It is important to keep in mind that RDE does not replace laboratory tests on a dynamometer but it complements it. This way of evaluating emissions could prevent car manufacturers from cheating such as the Volkswagen emissions scandal. Indeed, Volkswagen rigged millions of engines by activating their emissions controls only during the testing in the laboratory. A small spy program could detect when the car was undergoing an emission test and could interlock an intern mechanism which reduced the emissions. For that reason, their vehicles passed the emissions testing but in reality their exhaust emissions were significantly higher [23].

The subject of this thesis arose in this context of pollution and mobility. The goal of this work is to study the driving patterns of drivers, to determine if the new Real Driving Emission methodology is really representative of people's driving behaviour, by analysing the data collected from 32 cars and comparing it with the data considered in the RDE test procedure. This question is approached by considering drivers in Brussels.

The next chapter concerns the RDE procedure. In this test procedure, a car equipped with measurement systems drives on-road and many parameters are measured, especially driving and engine parameters (speed of the vehicle, GPS coordinates, temperatures of different parts of the car,...) and emission measurements. Conditions have been set on multiple factors to frame the test such as the maximal distance driven, the altitude difference, the ambient temperature, the stop percentage of the trips, the maximum speed and many more. All these variables and the computation of some of them are studied in this chapter. If the parameter under study of the driving car does not respect one of the conditions, the trip is invalid regarding the RDE methodology and is not taken into account in the assessment of the emissions of the vehicle under investigation. Thus, the data of the whole trip are set apart from the analysis and are not considered, even if the journey went well. The target of this work is to determine whether these conditions are coherent according to Brussels drivers. For instance, the driving distance on urban roads in the RDE procedure must be of at least 16 km but do Brussels drivers generally ride that distance?

The third chapter explains the methodology of this study by discussing the recording of the parameters, the choice of the subjects for the study and the sorting and computations of the data.

The fourth chapter is focused on one subject. This permits analysing more clearly why some of these conditions are not respected. The recording frequency, the handling of inaccuracies in the recording of the parameters and the deep analysis of the conditions compared to the expected values regarding the RDE are analysed.

Finally, the fifth chapter statistically analyses all drivers and trips together. The goal is first to describe the conditions regarding the validity of the trip and then to determine the links between the variables under study for the RDE methodology and other external factors such as the fuel type for instance. The impact of the parameters on the validity of the trip regarding the RDE procedure are also analysed.

CHAPTER 2

Real driving emissions test procedure

This chapter introduces the real driving emissions test procedure. First, the origin of this test is presented and then the measurement system is explained along with the recorded parameters imposed by the RDE methodology. As evoked in the introduction, scientists from the EU had to set some boundaries and guidelines to frame the RDE test. To understand the choice of these boundary parameters, some literature is reviewed. The boundaries for the RDE test and their computations end this chapter.

2.1 Introduction to a new test

The real driving emissions test procedure has been created in 2015 by the European Union. The RDE project is born after a publication from the Joint Research Centre (which is the European Commission's scientific and technical research laboratory) claiming that the NEDC did not accurately capture the NO_x emissions for normal use of diesel cars. Therefore, a team of researchers published a new legislative text split in four packages to introduce the new complementary test procedure. The first step procedure was voted by the Member States of the European Commission in May 2015 and the test has been implemented for a monitoring phase in 2016. Indeed, automakers needed to participate in this monitoring period, where they had to validate the RDE test and publish the emissions results. The RDE measurements for NO_x emissions became mandatory for new cars models from September 2017. Afterwards, this test has been extended for other pollutants such as particles numbers (PN) which are emitted from direct injection engines. In this package, the cold starts are taken into account, since they have a significant impact on small trips. All the new cars will be affected by the four packages of the legislation in 2021 and the new vans in 2022. The project is still under development and new policy packages are currently in evolution. This test complements the analyses of the WLTP (introduced in the previous chapter) on a dynamometer in the laboratory in order to reflect more accurately the real conditions [19]. The WLTP has been improved over the years to ensure more representative estimations of fuel consumption and CO_2 emissions. A standardised fuel consumption and energy consumption monitoring will be mandatory on-board of every vehicle in 2021 to provide valuable data to the consumer but also to compare the laboratory results with the reality. The car manufacturers must indicate the real-world emission performances in the certificate of conformity of the vehicle, to ensure a better transparency for every stakeholder [24].

According to the RDE methodology, the vehicle is driven on public roads and many conditions are tested and recorded. The conditions are the following: low and high altitudes, year-round temperatures, additional vehicle payload, up- and down-hill driving, urban roads (low speed), rural roads (medium speed) and motorways (high speed). Data are thus collected with specific on-board equipment to verify the compliance with the existing norms [25].

2.2 The portable emissions measurement system (PEMS)

The portable emissions measurement system records the data on-board. During the experiment, the PEMS records a lot of information and the vehicle passes the test or not in function of its performances compared to the enforced limits. The PEMS is installed at the back of the vehicle and is linked to the tailpipe. Many technologies exist and one example is shown in FIGURE 2.1. PEMS measurements technologies are improving and miniaturised. The actual system weights roughly 70 kg compared to 200 kg originally.

Figure 2.1: Example of a PEMS, from Motortalk, 2017 [26].

This system reports many parameters such as the exhaust mass flow rate, the emissions, the temperature, the humidity and the component concentrations for instance. A particle analyser is connected to its tailpipe via sampling probes. The recorded parameters are listed in TABLE 2.1. The recording frequencies must reach at least 1 Hz.

Table 2.1: Recorded parameters for a RDE test, from the European Commission regulations, 2016 [27].

Parameter	Unit	Source
Emissions		
Methane (CH_4) concentration	ppm	Gas analyser
Non-methane hydrocarbons (NMHC) concentration	ppm	Gas analyser
CO concentration	ppm	Gas analyser
CO_2 concentration	ppm	Gas analyser
NO concentration	ppm	Gas analyser
NO_2 concentration	ppm	Gas analyser
Total Hydrocarbons (THC) concentration	ppm	Gas analyser
PN concentration	$\#/m^3$	Gas analyser
Exhaust mass flow rate	kg/s	Exhaust flow meter
External conditions		
Ambient humidity	%	Sensor
Ambient temperature	K	Sensor
Ambient pressure	kPa	Sensor
Speed and GPS data		
Vehicle speed	km/h	Sensor (e.g. optical or micro-wave sensor), GPS or ECU
Vehicle latitude	Degree	GPS
Vehicle longitude	Degree	GPS
Vehicle altitude*	m	GPS or Sensor
Engine conditions		
Exhaust gas temperature*	K	Sensor
Engine coolant temperature*	K	Sensor or ECU
Engine speed*	rpm	Sensor or ECU
Engine torque*	Nm	Sensor or ECU
Torque at driven axle*	Nm	Rim torque meter
Pedal position*	%	Sensor or ECU
Engine fuel flow	g/s	Sensor or ECU
Engine intake air flow	g/s	Sensor or ECU
Fault status*	—	ECU

Parameter	Unit	Source
Intake air flow temperature	K	Sensor or ECU
Regeneration status*	—	ECU
Engine oil temperature*	K	Sensor or ECU
Actual gear*	#	ECU
Desired gear*	#	ECU

The parameters marked with an asterisk (*) need to be determined only if the vehicle status and operating conditions must be verified. The data are reported from various sources in the PEMS: from a gas analyser, from an exhaust flow meter (EFM), from diverse sensors, from a GPS antenna, from the Engine control unit (ECU) by the OBDII socket and from a rim torque meter. The ECU is the brain of the car which controls all actuators in the engine and monitors all information. It ensures the optimal performance of the engine. All engines parameters can be extracted from the ECU, except for the exhaust gas temperature and the torque at driven axle. The GPS antenna must be as high as possible to capture a maximum of information. The data must be collected, measured and recorded all along the test even if the car is turned off during the experiment.

The goal of the emissions measurements is to compute the emissions by kilometre and compare them to the EU norms. The CH_4, NMHC, CO, CO_2, NO, NO_2 and THC concentrations must be measured on a wet basis meaning that the moisture content in the flue stream is not removed. The emissions sampling is conducted downstream of the exhaust flow meter if possible where the influence of ambient air downstream of the sampling point is minimal.

If the altitude is measured by the GPS, it must be validated by comparing the latitude, longitude and the altitude from the GPS with topographic maps. If the difference between the two ways of computation are more than 40 meters, it must be manually corrected and marked. Correspondingly, the obvious errors regarding the speed recordings must be corrected by installing a dead reckoning sensor for instance whose goal is to deduct the position of the vehicle based on its speed and its last known position. The corrected information cannot exceed an uninterrupted time period of 120 seconds or a total of 300 seconds. Then, the consistency of the vehicle speed from the GPS must be studied and corrected by computing the total trip distance with the reference data from a sensor, from topographic maps or from the validated ECU. The distance computed from the GPS cannot diverge from more than 4% of the reference. If these constraints are not respected, the results of the test are invalid and therefore not taken into account in the assessment of emissions of the car under study [27].

The PEMS system must always be validated in the laboratory on a dynamometer [28]. This is included in the WLTP. The goal of this check is to ensure that the system is well installed and not to compare the data from the PEMS with the data recorded on the dynamometer. The PEMS validation must respect the tolerances listed in Table 2.2, about the distance and the emissions measurements. If the requirements are unmet, the system must be corrected and the validation must be reiterated until the system passes the test. Efficiency, leaks and calibration checks are also mandatory.

Parameter [unit]	**Permissible absolute tolerance**
Distance [km]	250 m of the laboratory reference.
THC [mg/km]	15 mg/km or 15% of the laboratory reference, whichever is larger.
CH_4 [mg/km]	15 mg/km or 15% of the laboratory reference, whichever is larger.
NMHC [mg/km]	20 mg/km or 20% of the laboratory reference, whichever is larger.
PN [#/km]	10^{11} p/km or 50 % of the laboratory reference, whichever is larger.
CO [mg/km]	150 mg/km or 15% of the laboratory reference, whichever is larger.
CO_2 [g/km]	10 g/km or 10% of the laboratory reference, whichever is larger.
NO_x [mg/km]	15 mg/km or 15% of the laboratory reference, whichever is larger.

Table 2.2: Permissible tolerances for the PEMS system, from B. Giechaskiel, 2018 [28].

The trips are evaluated and then validated if they are driven not too smoothly or not too aggressively. During the monitoring phase, pollutant emissions have to be documented. Not-to-exceed limits (NTE) were introduced in the RDE legislation to control the engine emissions and their mathematical expression is

$$\mathrm{NTE}_{\mathrm{pollutant}} = \mathrm{CF}_{\mathrm{pollutant}} \cdot \mathrm{TF}(p_1, ...p_n) \cdot \text{EURO-6}$$

with

- $\mathrm{NTE}_{\mathrm{pollutant}}$: the not-to-exceed emissions factor [g/km].
- $\mathrm{CF}_{\mathrm{pollutant}}$: a conformity factor for the pollutants. This factor is annually reviewed, since the exhaust treatment and engine the management are continuously improving. It has already been reduced once for the NO_x emissions, in the fourth act of the legislation, from a factor of 2.1 (for all new vehicles from September 2017) to a factor of 1.43 (for all new vehicles from January 2020). The conformity factor for the PN is 1.5 for all new models from September 2017 [29].
- $\mathrm{TF}(p_1, ...p_n)$: the transfer function to normalise the driving conditions, with the parameters p_i (i = 1,...,n). However, the regulatory text of the second RDE package set this function to 1 in 2016 [30].
- EURO-6: the Euro 6 emissions limits.

In order to reduce the amount of tests, the RDE legislation divided the vehicles into different families based on their powertrain and engine-block type or number and disposition of cylinders. In each family, at least one vehicle must be tested with the PEMS. Within the same category, engine sizes can differ from 22% for car engines with a displacement greater than 1500 cc (cubic centimetre) up to 32% for engines smaller than 1500 cc [31].

2.3 Choice of parameters: literature review

The RDE methodology imposes conditions for the trip on the distance, on the average speed, on the stop proportion, on the altitude, on the dynamic conditions and on the maximal speed as explained in SECTION 2.4. The goal of this section is to understand, by looking at the literature review and spotting the interesting factors, why those parameters were chosen. This section is divided in multiple parts. First, the assessment parameters usually used in these kinds of studies are presented. Then the effects of speed and acceleration on pollution are evoked. The effects of temperature and altitude on emission are also covered, followed by the way of dividing the data into categories. The importance of these factors are presented through these previous studies. Finally, the interest of the conditions chosen for the RDE based on the literature review ends this section. All of these topics are important for the choice of the conditions of the RDE which are presented in the next section.

In the context of this work, driving patterns are analysed. A driving pattern is defined as the way for a driver to handle his or her vehicle and it is described by the speed and acceleration changes. They are studied based on real-world driving data and include the driving range and driving conditions [32]. Indeed, they are influenced not only by drivers but also by the localisation and the road conditions [33] and they affect fuel consumption and exhaust emissions [34].

2.3.1 Assessment parameters

Many parameters influence the driving patterns, such as street types, number of lanes, traffic conditions, type of cars, weather conditions and travel behaviours [34]. Eva Ericsson illustrated the cause effect model of variability in driving patterns as shown in FIGURE 2.2. In this scheme, the factors influencing the driving patterns are listed and clustered in six categories: travel behaviour, driver, vehicles, weather, traffic and street environment factors.
Twenty-six parameters evaluate the effect of street environment, traffic conditions and driver conditions. To conduct this analysis, a Volvo has been driven on specific roads by randomly chosen subjects in Lund and 439 driving patterns were collected. Each parameter could be stored in one of the three following categories of measurements: level, oscillation or distribution measurements. The level measurements are the standard deviation and the mean values of speed and acceleration (positive and negative); the oscillation measurements explain the jerkiness of the speed profile; while the distribution ones describe the percentages of speed, acceleration and deceleration in a certain range. The result of the analysis of variance (ANOVA) conducted is shown in TABLE 2.3.

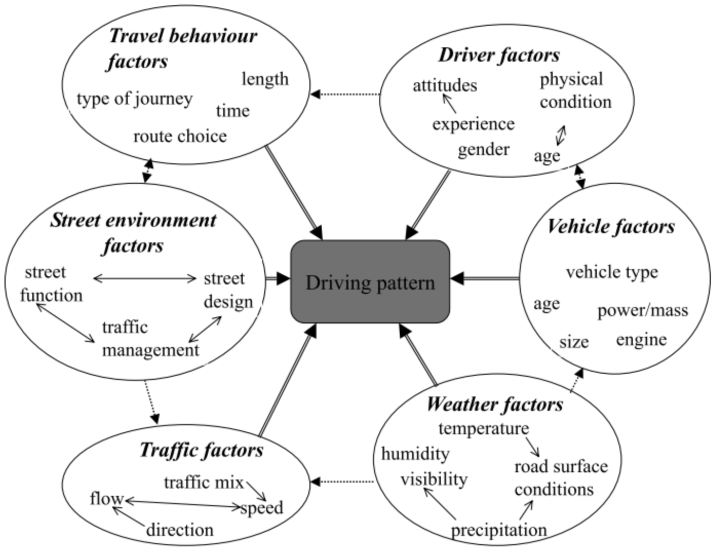

Figure 2.2: Cause effect model of variability in driving patterns, from E. Ericsson, 2000 [34].

In the second column, the type of measurement is indicated (L = level, O = oscillation and D = distribution). The p-value (represented by dots) describes how much the parameter of the first column varies with each effect studied. If the p-value is greater than 5%, it is not possible to state that the effect has an impact. This is illustrated with blank spaces in TABLE 2.3. Moreover, the smaller the p-value, the greater is the impact. Therefore, FIGURE 2.3 demonstrates that parameters are significantly influenced firstly by the street type and secondly by the driver. This justifies the need for real driving emission tests.

All researchers use various vehicles and driving parameters for their analyses. Lei Feng et *al.* considered fifteen parameters with specific weights to build a driving cycle pattern recognition as shown in TABLE 2.4 [35]. The need of weighting factors is justified by classification issues. Indeed, in this study, thirty-nine particular parameters were chosen based on Ericsson's work [36]. The concept of pattern recognition is to classify objects into different categories and classes. The assessment parameters are described by features and pattern recognition is made by extracting and classifying feature vectors. The problem is that the high dimension of these vectors affects the recognition of the driving cycle pattern in real-time. The weighting factors, computed by running simulation tests, are thus useful to reduce the number of dimensions. FIGURE 2.4 reveals that the most important factors according to Lei Feng et *al.*'s research [35] are the extreme deceleration time, the mid deceleration time, the mid high deceleration time, the high acceleration time and the extreme acceleration time.

Eva Ericsson [36] also studied, in 2001, the influence of driving pattern factors on fuel-use and exhaust emission factors. The goal of this research was first to find independent measurements describing the driving patterns and then compute their impact on fuel consumption, CO_2, HC and NO_x. After the analysis of 18 945 km of data from five

Table 2.3: Results of analysis of variance on different parameters, from E. Ericsson, 2000 [34].

Results of the analysis of variance[a]

Driving pattern measure	Measure type	Significant main effects				Significant interactions		
		Street type	Peak/ off-peak	Gender	Driver	Street type & Peak/ off-peak	Street type & Gender	Peak/ off-peak & Gender
v: average	L	•••	•••		•••		•	
v: standard deviation	L	•••			•••			
a: average	L	•••		•	•••		•••	
a: standard deviation	L	•••		••	•••		•••	
r: average	L	•••	•		•••	•	••	
r: standard deviation	L	•••	••		••	•••		
Integral a^2	O	•••	•	•	•••		•••	
RPA	O	•••			••		••	
Max/min > 2 km/h per 100 m	O	•••	•••		•••	••		
Max/min > 10 km/h per 100 m	O	•••			•••		•••	
Max/min > 2 km/h per 100 s	O	•••			•••			
Max/min > 10 km/h per 100 s	O	•••			••		•••	
% of time $0 < v < 15$ km/h	D	•••	••		•••			
% of time $15 < v < 30$ km/h	D	•••			•••	•		
% of time $30 < v < 50$ km/h	D	•••			•••	•		
% of time $50 < v < 70$ km/h	D	•••			•••			
% of time $r < -2.5$ m/s^2	D	••			•			
% of time $-2.5 < r < -1.5$ m/s^2	D	•••	•		•••	•	•••	
% of time $-1.5 < r < -1.0$ m/s^2	D	•••			•••			
% of time $-1.0 < r < -0.5$ m/s^2	D	•••			•••			•
% of time $-0.5 < r < 0$ m/s^2 s^{-2}	D	•••			•	•	•	
% of time $0 < a < 0.5$ m/s^2	D	•••		••	•			
% of time $0.5 < a < 1.0$ m/s^2	D	•••			•••			
% of time $1.0 < a < 1.5$ m/s^2	D	•••		••	••		•••	
% of time $1.5 < a < 2.5$ m/s^2	D	•••		•	•••		•••	
% of time $a < 2.5$ m/s^2 m	D	•		•	•••		•	

[a] v = speed, a = acceleration, r = deceleration, L = level measures, O = oscillation measure, D = distribution measures.
•$P < 0.05$.
••$P < 0.01$.
•••$P < 0.001$.

Table 2.4: Driving cycle feature members and corresponding weighting factors, from L. Feng et *al.*, 2012 [35].

Index Number (i)	Feature Parameters (a)	Weighting Factor (k)
1	Average Cycle Speed (m/s)	10
2	Positive Average Acceleration (a>0.1m/s^2)	1
3	Low Speed Time (15-30 Km/h)/Total Time (%)	10
4	Mid High Speed Time (70-90 Km/h)/Total Time (%)	100
5	High Speed Time (>90 Km/h)/Total Time (%)	10
6	Extreme Deceleration Time (a<-2.5 m/s^2)/Total Time (%)	1000
7	High Deceleration Time (a<-2 & a>-2.5 m/s^2)/Total Time (%)	1
8	Maximum Cycle Acceleration (m/s^2)	100
9	Maximum Cycle Speed (Km/h)	6
10	Standard Deviation of Cycle Speed (Km/h)	1
11	Mid Deceleration Time (a<-1 & a>-1.5 m/s^2)/Total Time (%)	1000
12	Mid High Deceleration Time (a>-2 & a<-1.5 m/s^2)/Total Time (%)	1000
13	Mid Acceleration Time (a>1.5 & a<2 m/s^2)/Total Time (%)	1
14	High Acceleration Time (a>2 & a<2.5 m/s^2)/Total Time (%)	1000
15	Extreme Acceleration Time (a>2.5 m/s^2)/Total Time (%)	1000

passenger cars, sixty-two factors arose and a factorial analysis shortened this list to sixteen independent driving pattern factors. The effect of these factors has been analysed according to the VETO model (an emission model developed by the Swedish National Road and Transport Research Institute). Moreover, a regression based on the measured emissions has been made and has been used to validate the results. The emissions have been measured with the equipment developed by Rototest AB, a Swedish consultant company. K. Brundell-Freij and Eva Ericsson [37] summarised the results of Ericsson's analysis reported in Table 2.5. This table describes the sixteen factors having a significant effect on emissions and fuel consumption. The plus and minus signs represent the magnitude of the effect.

Table 2.5: Driving pattern factors with significant overall effect on emissions and fuel consumption, from K. Brundell-Freij and Eva Ericsson, 2005 [37].

Factor	**Fuel**	**CO_2**	**HC**	**NO_x**
Deceleration factor *Typical parameter*: Average deceleration	-	-		
Factor for acceleration requiring high power *Typical parameter*: RPA	++++	++++	+++	++++
Stop factor *Typical parameter*: % of speed < 2 [km/h]	+++++	+++++		
Speed oscillation factor *Typical parameter*: Frequency of local max/min values of speed curve per 100 s	++	++		
Factor for acceleration requiring moderate power *Typical parameter*: % of time when v^*a_{pos} is 3-6 [m^2/s^3]	++	++		
Extreme acceleration factor *Typical parameter*: % of time at acceleration over 1.5 [m/s^2]	++	++	+++++	++++
Factor for speeds from 15 to 30 [km/h] *Typical parameter*: % of time at speeds from 15 to 30 [km/h]				-
Factor for speeds from 90 to 110 [km/h] *Typical parameter*: % of time at speeds from 90 to 110 [km/h]				
Factor for speeds from 70 to 90 [km/h] *Typical parameter*: % of time at speeds from 70 to 90 [km/h]	-	-		
Factor for speeds from 50 to 70 [km/h] *Typical parameter*: % of time at speeds from 50 to 70 [km/h]	- -	- -		

Factor	Fuel	CO_2	HC	NO_x
Factor for late gear changing from 2nd and 3rd gear *Typical parameter*: % of time engine speed 2500–3500 rpm in 3rd gear	+	+	++	+++
Factor for engine speeds > 3500 rpm *Typical parameter*: % of time engine speed is > 3500 rpm			++	++
Factor for speeds > 110 [km/h] *Typical parameter*: % of time speed is > 110 [km/h]				
Factor for moderate engine speeds in 2nd and 3rd gear. *Typical parameter*: % of time engine speed 1500–2500 rpm in 2nd gear	- -	- -		- -
Factor for low-engine speeds in 4th gear *Typical parameter*: % of time engine speed is < 1500 rpm in 4th gear	-	-		-
Factor for low-engine speeds in 5th gear *Typical parameter*: % of time engine speed is < 1500 rpm in 5th gear	-	-		-

The factor for acceleration requiring high power and the extreme acceleration factor are highlighted to be the most influential factors on the fuel consumption, the CO_2, the HC and the NO_x emissions. The stop factor has a significant effect on the fuel consumption and the CO_2 emissions.

2.3.2 Effect of speed and acceleration on pollution

P. Höglund and J. Niittymäki analysed the effects of speed changes and humps on the environment, especially on air pollution [38]. The computer program EMCA has been used to compute the emissions based on the continuous emission matrices recorded in the vehicles on the studied roads. The data recorded on-board were completed by traffic simulation with the HUTSIM simulation model and micro-modelling of traffic behaviour, fuel consumption and exhaust pollution. Fuel consumption, exhaust pollution and traffic efficiency were then compared at various time of the day (following peak hour traffic) and for two streets with a permitted speed of 30 [km/h] and 50 [km/h]. The conclusion is that the changes of speed in the traffic flow have a strong influence on the amount of air pollutants emitted.
The results of driving analyses are also strongly dependent on sites. Therefore, each study must be conducted at different times of the day and on different roads. Indeed, the vehicle

emissions depend on the driving patterns.

Concerning the pollution, P. Shridhar Bokare et *al.* [39] came to the same conclusion as P. Höglund and J. Niittymäki. In their paper about the effect of speed, acceleration and deceleration of a small gasoline car on its tail pipe emission, various acceleration levels were achieved and emission of Carbon Monoxide (CO), Carbon dioxide (CO_2), Hydrocarbon (HC) and Oxides of Nitrogen (NO_x) were analysed with a *t*-test. The Student's *t*-test, was used to assess if the emissions were the same for different acceleration levels. This test is also used in this work and explained in SECTION 5.2. The research concluded that the speed had an effect on emissions for a constant acceleration. Indeed, emissions rates decrease with high speed, for the same level of acceleration. Moreover, the acceleration rate of a car significantly impacts the tailpipes emissions. However, the deceleration does not have an impact on emissions for small cars.

2.3.3 Effect of temperature and altitude on emissions

A. Nagpure et *al.* [40] analysed the impact of altitude on emission rates of CO, NO_x and VOCs (Volatile organic compounds). The research has been made on gasoline-driven light-duty commercial vehicles in three Indian cities: Delhi, Dehradum and Mussoorie. The conclusion is that CO and VOCs emissions increase with the altitude. However, the NO_x emissions decrease with the altitude. The results are illustrated in FIGURE 2.3. The temperatures are lower at higher altitudes [40]. Therefore, for higher external

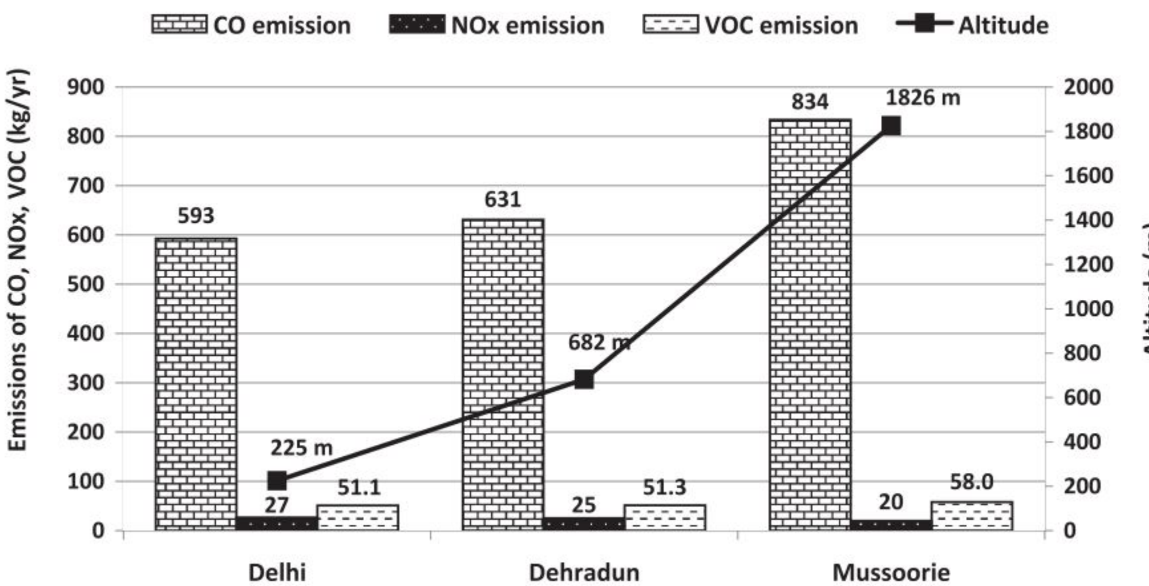

Figure 2.3: Annual average emission rates (kg/yr) of CO, NO_x and VOCs for three Indian cities, from A. Nagpure et *al.*, 2011 [40].

temperatures, the emissions of CO and VOCs are lower whereas the emissions of NO_x are higher.

2.3.4 Division in categories

M. André represented the actual driving of European cars in real driving conditions with the ARTEMIS European driving cycle [17]. This cycle is useful to evaluate and measure the pollutant emissions from passenger cars. The method followed is shown in Figure 2.4. The first step was to observe the vehicles uses by instrumenting them. A relatively

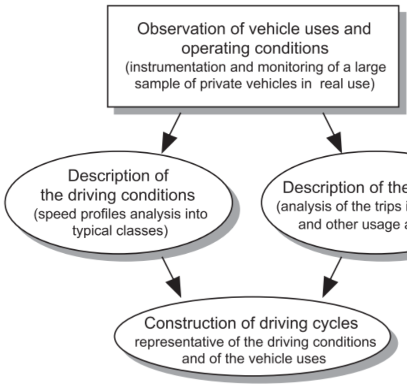

Figure 2.4: Main steps for developing a real-driving cycle, from M. André, 2004 [17].

large sample of private cars had to be considered. The two next steps concerned the analysis of the driving conditions and the vehicles uses. For the first one, the description was made by recording speed and by computing the speed times acceleration distribution with these profiles. The goal was to classify the driving conditions in different classes. The second description method clustered the data in different trip classes. The last step was the development of the driving cycle.
The database was constructed by using the data from two European research projects, the DRIVE-MODEM one, that collected on-board data from France, the UK and Germany and from the HYZEM research project in Greece that used the same recording methodology. Other data from different programs were used to validate the driving cycles afterwards.
By identifying and describing the data, twelve sequences emerged from the typical driving conditions: urban, congested urban, road and motorway. These twelve classes are characterised by features listed in Table 2.6. A pertinent partition of the data has thus been made based on the speed profiles. To characterise the vehicle uses, the same methodology has been applied and three categories stood out: motorway, rural roads and urban. The trip classes were thus described by their speed and driving conditions.

Table 2.6: A description of European driving conditions for twelve typical classes, obtained by automatic clustering of speed profiles recorded on-board, from M. André, 2004 [17].

Classes of driving conditions and their description			Percentage of total mileage	Running speed (km/h)	Average speed (km/h)	Stop duration (in %)	Stop rate (stop/km)	Average positive acceleration (m/s^2)
1	congested urban	high stop duration	3.7	25.9	10.2	60.8	3.9	0.87
2			5.9	23.6	15.9	32.7	3.0	0.81
3		low steady speeds	2.4	16.5	13.2	19.5	3.4	0.67
4	free-flow urban		5.1	28.0	26.1	6.7	0.97	0.65
5		unsteady speeds	12.2	35.6	32.3	9.1	0.98	0.81
6	secondary roads	unsteady speeds	10.8	52.2	48.8	6.6	0.41	0.75
7			8.8	45.5	43.8	3.7	0.39	0.63
8		steady speeds	7.2	65.0	64.0	1.5	0.15	0.55
9	main roads	unsteady speeds	11.8	75.0	72.5	3.3	0.15	0.67
11			6.2	86.1	85.7	0.4	0.04	0.48
10	motorways	unsteady speeds	10.4	115.6	114.9	0.7	0.03	0.53
12			15.6	123.8	123.7	0.1	0.01	0.40

2.3.5 Interest of the conditions chosen for the RDE based on the literature review

A lot of interesting conclusions emerged from the literature review. The main parameters adopted in the RDE test procedure and presented in the scope of this work are the average and maximal speed, the stop percentage, the altitude, the temperature and the distance. Two dynamic factors -the relative positive acceleration and the 95th percentile of the product of the speed times the positive acceleration- are also selected for the RDE methodology and are explained in this section.

The researches show the need of recording and analysing the speed, the acceleration and the deceleration in the study, since they significantly impact the pollution. In fact, acceleration and deceleration times are considered as the most important factors in driving patterns analysis following Lei Feng et *al.* as seen in Section 2.3.1. This need is fulfilled by the dynamic conditions in the RDE which are the relative positive acceleration and the 95th percentile of the product of the speed times the positive acceleration.

The relative positive acceleration (RPA) is a value that characterises the load of the trip [16, 21]. It is representing the power needed to accelerate the vehicle all along the cycle. If the trip requires many low power-demand accelerations, the RPA is low [41]. This parameter describes the softness of the driving patterns and it is the lower dynamic condition of the RDE test. It is computed from the acceleration and the speed as it is explained later in Section 2.4.5. The interest of computing the RPA lies in its huge influence on the emissions of fuel, CO_2, HC and NO_x as explained in Table 2.5 from Section 2.3.1. Moreover, this criterion is strongly affected by the street type and the driver. These two latter are taken into account in the RDE test procedure, therefore this parameter is optimal as condition.

The upper dynamic condition is the 95th percentile of the product of the speed times the positive acceleration. Taking the 95th percentile means that the value of $v \cdot a_{pos}$ is 95% of the time under this amount, while the remaining 5% of the time are above that number.

In this expression, v is the speed in km/h and a_{pos} is the vector of the accelerations greater than 0.1 m/s^2. All values above this condition show an aggressive driving behaviour. From TABLE 2.5 of SECTION 2.3.1, the extreme accelerations encountered in aggressive driving behaviour have a strong impact on emissions. The importance of the extreme accelerations is also underlined by Lei Feng et *al.* in SECTION 2.3.1. Therefore, an upper limit was necessary for the RDE conditions.

SECTION 2.3.3 has proven that the altitude and temperature have an impact on CO VOC and NO_x emissions. The fuel consumption and CO_2 emissions are affected by the stop proportion factor. In Ericsson's study, this factor is the percentage of speed strictly inferior to 2 km/h whereas in the RDE methodology, it is the percentage of speed strictly inferior to 1 km/h. To summarise, all these parameters have been chosen as conditions by the makers of the RDE procedure because they have a strong impact on emissions.

2.4 Boundaries and parameter computation

This section covers the definition of a RDE trip and the boundaries set for the RDE whose conditions are detailed and their computations described. These boundaries were put to frame the RDE procedure test, to ensure a certain structure and framework. If the boundary is not respected during the test, the parameters computed are not taken into account in the assessment of emissions and the trip is said to be "invalid". It is really important to understand that this definition of validity is only related to the fulfilment of the conditions of the RDE.

2.4.1 Definition of a trip and boundaries

A RDE trip consists in a sequence, from the start of the engine to the stop of it, of three driving categories based on speed as established by M. André (explained in SECTION 2.3.4):

1. Urban: for cars with a speed up to 60 [km/h] included.
2. Rural: for cars with a speed from 60 to 90 [km/h] ($>$60 [km/h] - $\leqslant$90 [km/h]).
3. Motorway: for cars with a speed strictly above 90 [km/h].

The shares of urban, rural and motorway driving are classified by instantaneous speed. The urban, rural and motorway shares must be selected on a topographic map. The urban roads must have a speed limit of 60 km/h. Even if the speed limit is higher than this value on the urban part, the driver cannot exceed 60 km/h.

Each trip must be evenly distributed between three categories with a 10% tolerance, except for the lowest urban share where the tolerance is 5% to ensure at least 29% of valid data in the urban conditions. Each part of the trip for each category must happen continuously. However, the rural part can be briefly interrupted by small period of urban ride and the motorway part can be briefly interrupted by urban or rural periods, for cases such as tollgates or for urban crossings.

The RDE test must be performed during workdays on road pavements [27]. The test cannot be conducted at the weekend as defined in the Council Regulation from 1971 about periods, dates and time limits [42]. Table 2.7 summarises the conditions.

Table 2.7: Specifications for each category of the RDE test from the International Council on Clean Transportation (ICCT), 2017 [31].

Trip specifications		**Boundaries of the RDE test**
Total trip duration		Between 90 and 120 minutes
Average speeds	Urban	15 - 40 [km/h]
	Rural	60 - 90 [km/h]
	Motorway	>90 [km/h] and at least 5 minutes >100 [km/h]
Trip composition	Urban	29% to 44% of distance
	Rural	23% to 43% of distance
	Motorway	23% to 43% of distance
Distance	Urban	>16 km
	Rural	>16 km
	Motorway	>16 km

The average speed in the urban part is expected to be within 15 and 40 [km/h], including stops. The stop periods are considered when the speed of the car is strictly smaller than 1 [km/h] and they must cover at least 6% of the duration of the urban trip [27]. The RDE test methodology shares the concern of the effect of speed and acceleration on pollution by implementing the dynamic conditions. These and other factors such as the temperature or the altitude have been added. There are indexed in Table 2.8.

The maximal speed is expected to be 145 [km/h]. However a tolerance of 15 [km/h] is applicable if this speed is maintained at maximum 3% of the duration of the trip on the motorway. Before the RDE test, the vehicle must have been driven for at least 30 min and then parked for a period of 6 to 56 hours. After the first engine ignition, the idling time is limited to fifteen seconds [43]. For every parameter, if the data are outside the boundaries, the whole trip must be deleted. For the altitude and the ambient temperature, the moderate and extended cases appear. If the data are included in the range for the extended case, the emissions measured must be divided by a factor of 1.6.

Table 2.8: conditions for RDE tests, from the International Council on Clean Transportation (ICCT), 2017 [31].

Parameters		Conditions of the legislation
Payload		Less than 90% of maximum vehicle weight
Altitude	Moderate	0 to 700 [m]
	Extended	700 to 1300 [m]
Altitude difference		<100 m altitude difference between start and finish
Cumulative altitude gain		1200 [m] for every 100 [km] driven
Ambient temperature	Moderate	0[°C] to 30[°C]
	Extended	From -7[°C] to 0[°C] and 30[°C] to 35[°C]
Stop percentage		Between 6% and 30% of urban time
Maximum speed		145 [km/h] 160 [km/h] for 3% of motorway driving time
Dynamic boundary conditions	Maximum metric	$(v \cdot a_{\mathrm{pos}})_k[95]$
	Minimum metric	RPA
Use of auxiliary systems		Free to use as in real life (operation not recorded)

2.4.2 Computation of the mean velocity

The mean velocity must be computed as

$$\boxed{\bar{v_k} = \frac{\sum_i v_{i,k}}{N_k}, \quad [\mathrm{km/h}] \qquad i = 1 \text{ to } N_k}$$

with

- k: urban, rural or motorway.
- $v_{i,k}$: the instantaneous speed for each category [km/h].
- N_k: the total amount of samples for the urban, rural and motorway classes.

2.4.3 Computation of the acceleration

In the RDE test, the acceleration is computed over the whole time-based speed trace (1 Hz resolution) from second 1 to second t_t (last second), following the equation

$$\boxed{a_i = \frac{(v_{i+1} - v_{i-1})}{2 \cdot 3.6}, \qquad [\mathrm{m/s^2}] \qquad i = 1 \text{ to } N_t}$$

with

- a_i: the acceleration at the instant i $[\mathrm{m/s^2}]$.
- v_i: the speed at the instant i [km/h].
- N_t: the total sampling.
- for i = 1: $v_{i-1} = 0$.
- for i = N_t: $v_{i+1} = 0$.

2.4.4 Computation of $(v \cdot a)_i$ and $(v \cdot a_{\mathbf{pos}})_k[95]$

From the acceleration, the product $(v \cdot a)_i$ is

$$\boxed{(v \cdot a)_i = \frac{v_i \cdot a_i}{3.6} \quad [m^2/s^3] \qquad i = 1 \text{ to } N_t}$$

with

- a_i: the acceleration at the instant i $[m/s^2]$.
- v_i: the speed at the instant i [km/h].

After the computation of the acceleration and $(v \cdot a)_i$, the values of v_i, $(v \cdot a)_i$ and a_i are classified into ascending speed.
For each trip, all instantaneous speeds are looked at and the data are classified in one of the three categories regarding their instantaneous speed.

- All data from each trip with $v_i \leqslant 60$[km/h] belong to the urban category.
- All data from each trip with $v_i > 60$[km/h] and $v_i \leqslant 90$[km/h] belong to the rural category.
- All data from each trip with $v_i > 90$[km/h] belong to the motorway category.

For the upper condition, $\mathrm{a_{pos}}$ is needed which is a vector made of the instantaneous accelerations larger than $0.1[m/s^2]$. The amount of data with $a_i > 0.1[m/s^2]$ must be at least of 150 for every class of speed. The 95th percentile of the product of vehicle speed per positive acceleration greater than 0.1 $\mathrm{m/s^2}$ for urban, rural and motorway shares can then be computed.

Firstly, the values of $(v \cdot a)_{i,k}$ for each category are sorted in ascendant order for each dataset with a_{pos} (defined as $a_{i,k} \geqslant 0.1[m/s^2]$, with k the urban, rural or motorway part) and M_t, the total number of data points, is determined. Secondly, the percentile values are assigned to $(v \cdot a)_{i,k}$ with $a_{i,k} \geqslant 0.1[m/s^2]$ as follows:

- the lowest value of $v \cdot a_{\text{pos}}$ gets the percentile $1/M_k$, the second lowest value gets $2/M_k$, the third one gets $3/M_k$ and the last one gets the percentile $M_k/M_k = 100\%$.
- $(\mathrm{v} \cdot \mathrm{a}_{\text{pos}})_k[95]$ equals $(v \cdot a_{\text{pos}})_{j,k}$ with $\mathrm{j}/M_k = 95\%$.
- If $\mathrm{j}/M_k = 95\%$ cannot be met, $(\mathrm{v} \cdot \mathrm{a}_{\text{pos}})_k[95]$ is calculated by linear interpolation between consecutive samples j and j+1 with $\mathrm{j}/M_k < 95\ \%$ and $(\mathrm{j}+1)/M_k > 95\%$.

2.4.5 Computation of the relative positive acceleration (RPA)

The relative positive acceleration for each category is defined as

$$\boxed{\mathrm{RPA}_k = \frac{\sum_j (\Delta t \cdot (v \cdot a_{\text{pos}})_{j,k}))}{(\sum_i d_{i,k})} \qquad [m/s^2]}$$

with

- $d_{i,k}$: the distance driven in each category for a single point.
- $\mathrm{j} = 1$ to M_k, with M_k the amount of data points in urban, rural and motorway categories for positive accelerations only.
- $\mathrm{i} = 1$ to N_k with N_k the amount of data points in urban, rural and motorway categories.
- Δt: the time difference which is equal to 1 second in the case of the RDE.

2.4.6 Dynamic conditions

The dynamic conditions from Table 2.8 go along with a graph shown in Figure 2.5.

To validate the trip, the product of the speed times the positive acceleration must be, for each category, below the line of the left graph (the high dynamic condition) whereas the RPA values must be above the low dynamic condition showed in the right graph. Three cases are illustrated on the graphs. Only the trip represented by the red dot is valid regarding the dynamic conditions imposed by the RDE since it respects these conditions for the three classes (urban, rural and motorway). The encircled dots are outside the limits.

These conditions are mathematically defined as follows.

1. For the higher dynamic condition, the trip is invalid if

$$\begin{cases} \bar{v}_k \leqslant 74.6 \text{ [km/h]} \quad \text{and} \quad (v \cdot a_{\text{pos}})_k[95] > (0.136 \cdot \bar{v}_k + 14.44) \\ \bar{v}_k > 74.6 \text{ [km/h]} \quad \text{and} \quad (v \cdot a_{\text{pos}})_k[95] > (0.0742 \cdot \bar{v}_k + 18.966) \end{cases}$$

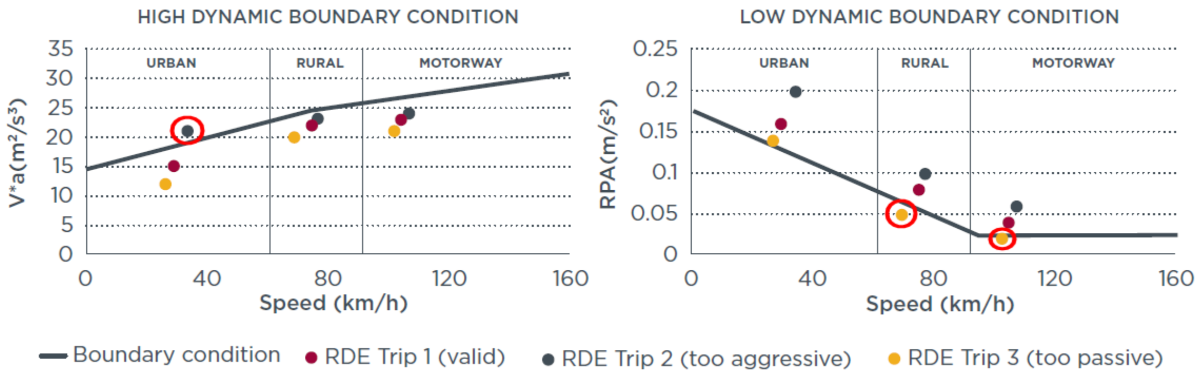

Figure 2.5: Dynamic conditions with three illustrative RDE trips, from the International council on Clean Transportation, 2017 [31].

2. For the lower dynamic condition, the trip is invalid if

$$\begin{cases} \bar{v}_k \leqslant 94.05 \text{ [km/h]} & \text{and} \quad RPA_k < (-0.0016 \cdot \bar{v}_k + 0.1755) \\ \bar{v}_k > 94.05 \text{ [km/h]} & \text{and} \quad RPA_k < 0.025 \end{cases}$$

with $\bar{v}_k$ the mean velocity of the trip for the category k (urban, rural or motorway).

2.4.7 Cold start

In the second package of the policy, the EU analyses the cold starts. The cold start period encloses the first five minutes after initial start of the combustion engine. If the temperature of the cooling system is known and accurate, the period of cold start ends when the cooling temperature is equal to 343 K or after five minutes at the latest.
The cold starts are responsible for poorer air quality, especially in the cities, where the trips are shorter. Some 27 ± 5% of trips may contain a cold start, since they are driven in Europe after vehicle parking of at least 3 to 8 h. Based on all collected data, the mean and median driving distances between two consecutive cold starts reaches 36 ± 16 km and 30 ± 13 km, respectively. However, if only urban trips are taken into account, the distances between two consecutive cold starts reach 25 ± 16 km and 27 ± 8 km [44].

The requirements for the cold start period are that the maximum idling time after the ignition of the engine must be 30 seconds and all the stops must last less than 90 seconds.

CHAPTER 3

Methodology

3.1 Presentation of the study

To analyse the driving patterns of drivers, devices measuring driving parameters were installed in the cars of 51 volunteer people for one to two weeks (as covered in SECTION 3.3). Indeed, the data are recorded on-board, with an Arduino assembled by Yan Liu and François Boveroux. Arduino boards read inputs (in this case: data from the computer of the car, data from a GPS and data from the accelerometers) and convert it into an output. The design and the manufacturing of the devices is not covered in this thesis. These devices must be plugged in the OBDII socket and require a power input (e.g. cigarette lighter socket) and twelve parameters are measured as explained in SECTION 3.2.

Each participant of the study had to plug in one of these devices in their car for one to two weeks. Then, the data were accessible via WIFI or with the SD card on which the files were created. The appliance was then moved from one subject to another. Once installed in the car, the device recorded every trip of the car. A trip begins when the driver starts its engine until the moment he turns it off. The system of start and stop does not create a new trip.

Data have been recorded from mid February until April. Among the 51 drivers selected, only 32 could be exploited. Indeed, troubles occurred with the SD card, with the battery and compatibility problems arose. In total, 899 trips were recorded which accounts for 4690 km, with a total driving time of 414 hours 4 minutes and 22 seconds.

This chapter presents the methodology by explaining how the data were recorded, the choice and management of the subjects, the sorting of the data, the parameters assessed and the division in categories.

3.2 Recording of the data

The first challenge was to decide what to measure. The data loggers installed in the vehicles could read and store dozens information of vehicle sensors available on the vehicle's CAN (Control Area Network) bus via the OBD (On-board diagnostics) socket, such as the fuel tank level input, the throttle angle, the engine exhaust flow rate or the fuel rate. OBD systems give diagnostic information on the state of the vehicle from the on-board computer. However, because of recording frequencies issues, the quantity of parameters to save from the car system through the OBD socket has been limited to three recorded at high frequency and a fourth one less frequently. Indeed, most of researchers studying driving patterns and driving cycles logged data at a frequency of 1 Hz. The main articles concerned can be found at references [17, 21, 35, 39, 41, 45, 46, 47]. In all those articles, this choice has never been justified. According to M. Quddus [48], any recording frequency greater than 1 Hz is considered as high frequency. Therefore, it could be thought that this frequency has been chosen because it is the highest of the "low frequencies". Moreover, a higher frequency induces more noise. It could thus be the good compromise between accuracy and high amount of data recorded. In fact, depending on the device, returning many parameters disallows to work with high frequencies.

For this study it was decided to use a recording frequency between 2 Hz and 5 Hz. The number of parameters has been limited because adding more parameters significantly slows the whole device down. The analysis of this choice is made in SECTION 4.2.

Therefore, only 12 main parameters were measured by three different components: a GPS chip, an accelerometer and the ECU from the OBD socket. They are listed below.

1. Date and time [UTC[1]]
2. Latitude [Decimal form]
3. Longitude [Decimal form]
4. Vehicle speed from the GPS [milliknots[2]]
5. Altitude [cm]
6. Acceleration x [Counts[3]]
7. Acceleration y [Counts]
8. Acceleration z [Counts]
9. Load [%] (it indicates a percentage of peak available torque [49].)
10. Engine speed [RPM]

[1] Coordinated Universal Time

[2] 1 m/s = 1943,84617179 mkn

[3] Counts are post-filtered accelerometer values. They vary based on the frequency and intensity of the raw acceleration. The filtering process by which counts are produced is proprietary to ActiGraph.

11. Vehicle speed from the OBD [km/h]
12. Engine coolant temperature [°C]

The first five parameters are measured by the GPS chip, the accelerations are measured by a digital accelerometer and the last four parameters are recorded from the OBD socket. The acceleration is measured in three dimensions. However, in this work, data from the accelerometer are not analysed due to its complexity.

For frequency issues mentioned above, it has been decided to only register the temperature (as a fourth driving parameter) every hundred recording points. The temperature of the engine is mostly used to count the number of cold starts. Moreover, the temperature does not change as quickly as the other factors. Therefore, this parameter can be recorded less frequently than the others.

However, after the analysis, it has been noticed that the load and the engine speed were not considered as the main parameters concerned about the RDE conditions and requirements. That is why these two parameters have not been used. However, these parameters are very useful for further work on emissions.

The GPS chip is continually powered by a battery, automatically reloaded with the connection to the cigarette lighter socket. However, the Arduino needs power from the cigarette lighter socket to turn on. Consequently, as soon as the driver starts the engine, the whole device powers up and it is ready to record after approximately 45 seconds. Every time the car is started, the data logger generates a file in text format for each trip with the 12 parameters listed above.

3.3 Choice and management of the subjects

51 drivers have been chosen for this study. They were selected among the acquaintances of the research team. Therefore, the representativeness of the subjects to the overall population is not respected. This car fleet biases the study since it does not represent the whole population of Brussels. To have a representative fleet, other aspects should have been selected according to decisive selection criteria with respect to the Brussels population such as the proportions of male/female, the age of the drivers, their social class, the type of cars (age, displacement, power, type of fuel) and whether the vehicle is a company car.

However, two requirements had to be fulfilled to take part in the study. Firstly, the drivers needed to drive in Brussels mainly. Secondly, their car had to be compatible with the Arduino, by using a CAN bus (11 or 29 bits) as data transmitter from the car to the reader. Actually, almost every car younger than 10 years respects the second constraint. The goal of the thesis is to characterise the main trends and not to compare the drivers. In other words, the driving patterns are of interest and not the drivers. That is why the car needed to be driven in the same conditions as usual and as if there was no device recording data. For instance, having several people riding the same vehicle is not a problem.

Each participant needed to fill in a form with information about himself/herself (name, age, gender) and about the car (brand and model, year of construction, displacement, power, type of gearbox and type of fuel). Moreover, they were asked if the vehicle was a company car since it influences a lot its use. The start and end date of recordings and the information about their car are listed in Appendix A.1 for the 32 datasets that could be exploited and in Appendix A.2 for the 19 failed recordings.

3.4 Sorting of the data

The data needed to be sorted due to recording inaccuracies and the trips outside of Brussels were set apart. As a reminder, a trip is considered as the journey travelled on the road from the start of the engine until the complete stop of it.

3.4.1 Failed data

Sometimes the recordings failed. Some consequent troubles were reported such as a broken cable, a system that ran out of battery in the middle of the experiment of one driver or incompatible cars. Among the 51 participants, the 19 failed measurements were due to three main causes: undetected incompatibility with the car, damaged or unreadable files (therefore no access to the SD card) or less than 30% of distance inside of Brussels as it is mentioned later in Section 3.4.2. The reason of failure for each car is reported in Appendix A.2.

Yet, for the 32 valid drivers, some smaller failures could happen within a trip. For example, when the vehicle runs through a tunnel, the GPS signal is lost and consequently, the data in the file is zero for the GPS coordinates. Therefore, the data were sorted by applying conditions on each parameter of each trip, to only keep the valid measurements. These are listed in Table 3.1. This table summarises the conditions of failure for each parameter and their explanation.

As soon as an incoherent measurement of speed, altitude or temperature is detected, it is registered in a different vector for each parameter. Later on, erroneous data are not considered in the computations. It permits using the well-recorded data anyway, without having to delete the whole trip, even if the data logger failed at some point. However, even if the GPS measurement fails, the line of data is still considered for the computations, since the GPS is only useful to see whether the trip is in Brussels. Therefore, with the information from the OBD, the computations are not affected by the signal loss of the GPS. Most of the time, the GPS is the component that fails, especially for the altitude. The GPS chip was the most unstable component because it captured signals from satellites and interferences could occur. The altitude came directly from the GPS chip but was less accurate than the GPS coordinates (latitude and longitude). This is justified by the fact that the latitude and longitude can be located with three satellites whereas the detection of the altitude must require the mobilisation of at least four satellites [50]. These errors significantly impact the reliability of source data.

Table 3.1: Conditions of recording failures.

Parameters	**Condition(s) of potential failure**	**Remark**
Time, latitude, longitude	= 0	If the GPS fails, it returns zero.
Speed from GPS, speed from OBD	1) <0 km/h 2) >240 km/h 3) dv/dt >5.5 m/s^2	Sometimes, the speed goes up or down too quickly with respect to the difference of time. Therefore, the acceleration must never be greater than 5.5 m/s^2 between two recorded points which corresponds to get from 0 km/h to 100 km/h in 5 seconds.
Altitude	<-10m 2) >200m	The data are recorded in Brussels and its suburbs. The altitude never exceeds 200 m.
Engine coolant temperature	<-20°C 2) >120°C	The temperature should be between -20°C and 120°C.

The velocity data supplied by the CAN bus was more accurate since it came directly from the car by a cable, so less interference occurred. The velocity data from the GPS is compared to the velocity data supplied by the CAN bus of a random car in FIGURE 3.1.

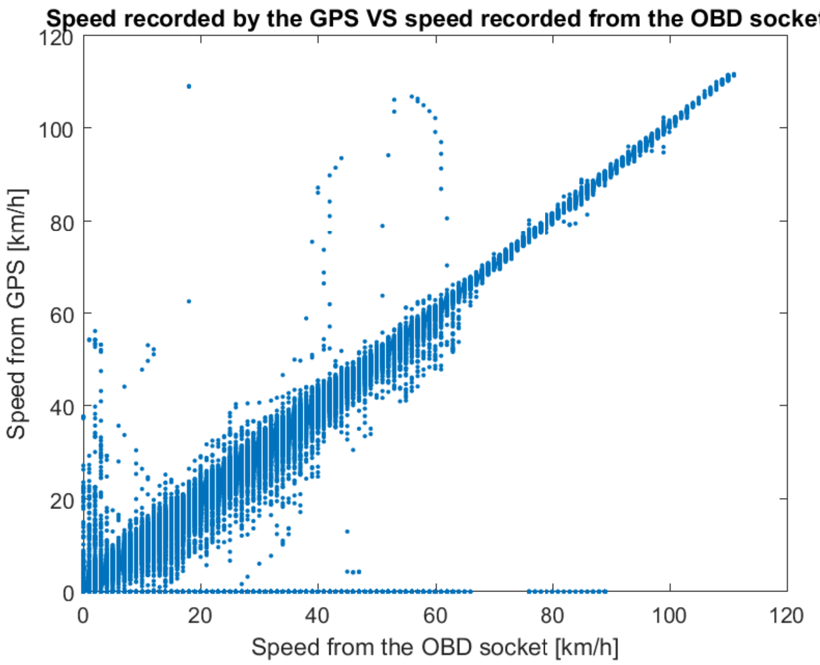

Figure 3.1: Speed recorded with the GPS in function of the speed from the OBD socket, for a random trip.

Since the GPS data are more often missing and the data from the OBD have the same trends as the speed from the GPS, the speed from the OBD is used in the further computations. However, if the data of the GPS were always available as in FIGURE 3.2, it would be more interesting to use that one since its velocity accuracy is 0.05 m/s in the data sheet whereas the errors of the speed from the OBD can go up to ±10% compared to the real speed. This figure shows that the speeds are pretty coherent. The speed from

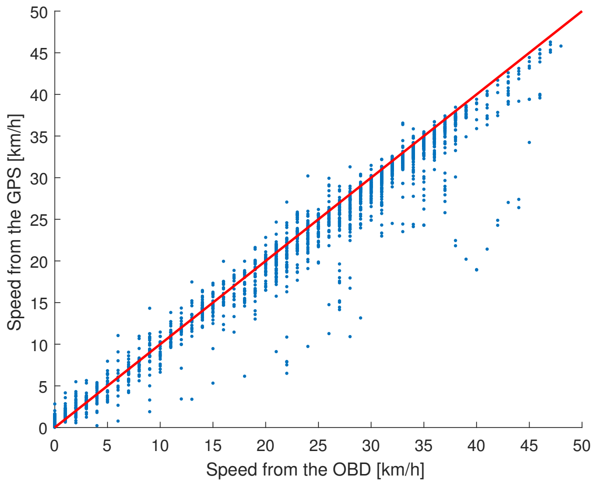

Figure 3.2: Speed from the GPS in function of the speed from the OBD for a trip without loss of GPS signal.

the GPS tend to be slightly smaller than the one from the OBD.

3.4.2 Data outside of Brussels

For this work, it has been decided to analyse trips in Brussels. In the context of this work, Brussels has been considered as a circle as shown in FIGURE 3.3. The extreme latitudes and longitudes of the Brussels-capital Region were found via Google Maps, then the centre was computed by making the differences between those extreme coordinates. The radius has been calculated by looking at the distance between the centre and the most extreme point. Therefore, the circle is an overestimation of the real region.

The routes were not chosen in advance, so the vehicles could go anywhere on different types of roads. Yet, each trip completely outside of this zone has not been taken into account for the calculations. However, if at least 30% of the distance of the trip was inside the zone, the trip was kept. All trips with a driving distance smaller than one kilometre were automatically removed.

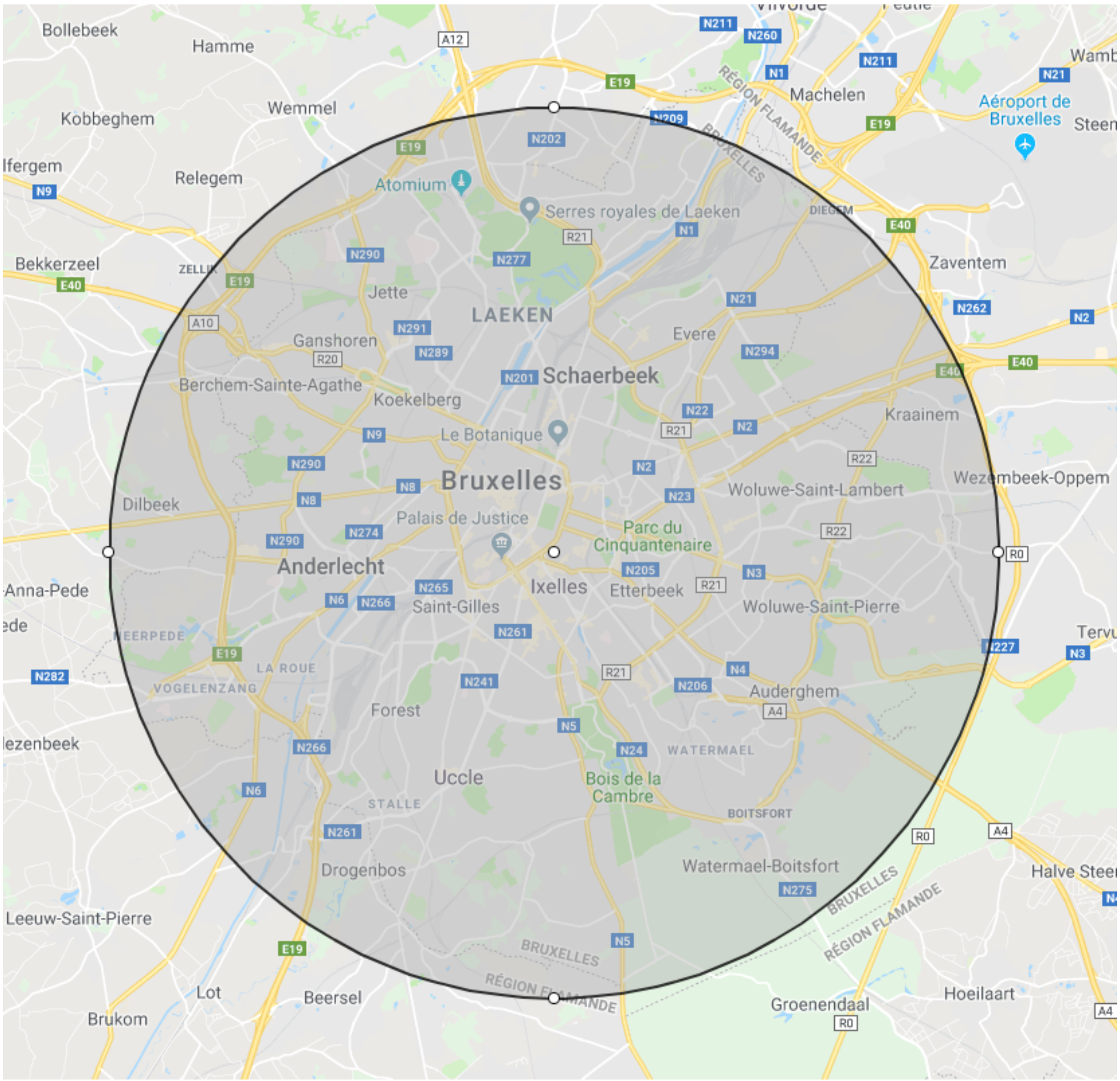

Figure 3.3: Approximation of Brussels, from Google Maps.

3.5 Performance assessment

Once the data are sorted, the assessment parameters are computed. At the end of the computations, a loop on all the trips for every driver enables to return the values for each parameter of all the trips for one driver.

The RDE test focuses on emissions and fewer driving conditions are considered. Therefore, in this work, the idea is to compute parameters from the RDE test procedure, for every trip in Brussels. TABLE 3.2 recaps the measurements for every components and the ensuing parameters. These factors have been chosen based on the RDE computations, more precisely from the SECTION 2.4. All the computations followed by a number under brackets are explained in SECTION 3.5.1 at the corresponding item.

Table 3.2: Variables measured and computed by each component of the device.

Components of the device	**Measurements**	**Purpose / variables computed**
GPS Chip	- Time - Latitude - Longitude - Altitude - Speed of vehicle	- Verification of the locations (1) - Determination of time of the week (trip at the weekend or not) - Peak or off-peak hours - Distance [m] (2) - Duration of the trip [s] - Altitude difference [m]
OBDII information	- Load (not used) - Engine speed (not used) - Speed of vehicle - Engine coolant temperature	- Average speed [km/h] (3) - Stop proportion [%] (4) - Distance by integrating the speed over time [m] (5) - Acceleration - $(v \cdot a_{pos})_k[95]$ [m^2/s^3] - Real Positive Acceleration [m/s^2] (7) - Cold start
Accelerometer	- Accelerations x, y, z	Not used

3.5.1 Details and explanations of the variables computed

In this section, some calculations are explained and detailed with respect to Table 3.2.

1. **Verification of the locations**: See Section 3.4.2.

2. **Distance**: The distance d is found from the radius of the earth R and the central angle Θ by

$$\boxed{d = \Theta \cdot R}$$

In order to compute the angle Θ between two points based on their latitude and longitude, the Haversine formula is useful. It is given by:

$$\boxed{\mathrm{hav}\,(\Theta) = \mathrm{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\,\mathrm{hav}(\lambda_2 - \lambda_1)}$$

with

- $\mathrm{hav}(\Theta) = \sin^2(\frac{\Theta}{2})$.
- φ_i = Latitude of point i.
- λ_i = Longitude of point i.

Therefore,

$$d = 2R\arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \text{ [m]}$$

The instantaneous distances are found with this formula, and the total distance of each trip is thus the sum of all these instantaneous distances.

3. **Average speed**: the speed from the OBD is used for the computations, since it is more reliable than the speed from the GPS. The average speed $\overline{v}$ is defined as

$$\boxed{\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i \qquad [\mathrm{km/h}]}$$

with n representing the number of recorded points in each trip and v_i the instantaneous speed.

4. **Stop proportion**: a stop is considered for speeds lower than 1 [km/h].

5. **Distance by integrating the speed over time**: Since the GPS is not always accurate, the distance has also been computed by integrating the speed from the OBD over time.

$$\boxed{d_i = \int_0^t \frac{v_i}{3.6} \cdot \mathrm{d}t \qquad [\mathrm{m}]}$$

with d_i and v_i the instantaneous cumulated distance ([m]) and velocity ([km/h]) respectively and t the time of the trip ([s]).

6. **Acceleration**: In this case, the car is accelerating if $a \geqslant 0.1$ [m/s^2]. Since the recording frequency is not the same as in the RDE testing, the acceleration definition in this case differs a bit from the definition in Section 2.4.3 and is simply computed as

$$\boxed{a_i = \frac{v_{i+1} - v_{i-1}}{3.6 \cdot (t_{i+1} - t_{i-1})} \qquad [\mathrm{m/s^2}] \qquad i = 1 \text{ to } N_t}$$

with

- v_i, the speed in km/h at time-step i.
- t_i, the time in seconds at time-step i.
- N_t, the total sampling.

The acceleration is smoothed out. This reduces the noise induced by the choice of recording the data at 3 Hz as covered in Section 4.2 where signals of 1 Hz as in the RDE test and signals of 3 Hz are compared.

7. **Real positive acceleration**: See Section 2.4.5.

8. **Cold start**: The cold start period ends when the engine coolant temperature reaches 70°C.

3.6 Division in three categories

The three categories directly come from the RDE definition in Section 2.4:

1. $v_{\mathrm{i}} \leqslant 60$ [km/h]
2. $v_{\mathrm{i}} > 60$ [km/h] and $v_{\mathrm{i}} \leqslant 90$ [km/h]
3. $v_{\mathrm{i}} > 90$ [km/h]

with v_{i} the instantaneous speed. To sort by category, every point of data has been classified regarding its instantaneous speed and has been organised in the urban, rural or motorway class. Then, the computations of the parameters are made three times by trip: one for each category. Once all the data are sorted by categories, the parameters for each trip of every driver are compared to the conditions.

As explained in Section 3.4.2, data come mainly from Brussels. Therefore, the rural and motorway parts are not represented nor entirely covered in the route of the trip and they do not happen continuously. However, artificial classes were made based on the instantaneous speeds of the the trip. This means that if the driver was on a road limited to 70 km/h, all points of data larger than 60 km/h were classified in the rural category, even if the driver was circulating on urban roads. Thereby, data are analysed with caution.

CHAPTER 4

Detailed analysis for one driver

In this chapter, data of the eighth driver are analysed in detail. First, the choice of the computation method for the distance is explained. Then, then the impact of the recording frequency is studied, followed by the establishment of inaccuracies in the recording of the parameters. Finally, all conditions are analysed in detail and an overall summary closes this chapter.

The model of the car is X1s Drive 18d from BWM, 2017. The displacement of this diesel engine is 1995 cm^3 and the power reaches 110 kW. This company car, equipped with a manual gearbox, has been driven by a woman.

The choice of the studied driver was made based on the distance travelled in Brussels. Indeed, almost 200 000 points of data were useful to the study which represents almost 324 km driven and it accounts for more than fifteen hours on the roads. All available GPS data from this driver are displayed on the map in FIGURE 4.1. Her job is to sell products to physicians therefore she drives a lot (mostly small trips) between medical offices. The device has been installed in her car from the 22/02/2019 at 18:30 until the 03/03/2019 at 19:10. Ten trips out of the seventy-one have been dismissed because more than 30% of their distance was out of the zone established in SECTION 3.4.2. TABLE 4.1 has been built considering all the data of all the trips together. All the values of this table are deepened in the present chapter.

This driver rides mainly in the centre of Brussels. Therefore, 94.10% of the distance represents the first category of speeds ($v_i \leqslant 60$ km/h) whereas the two others are lowly exemplified, with only 5.8% of distance and 0.1% of distance for the rural and motorway classes respectively. As evoked in SECTION 2.4.1, the trip composition had to be of at least 23% of distance in rural and motorway parts and of at most 44% of distance for the urban one which is impossible if staying in the city.

The average speeds computed on all the trips together for each category are 20.29 km/h for the urban one, 67.84 km/h for the rural one and 92.14 km/h for the motorway.

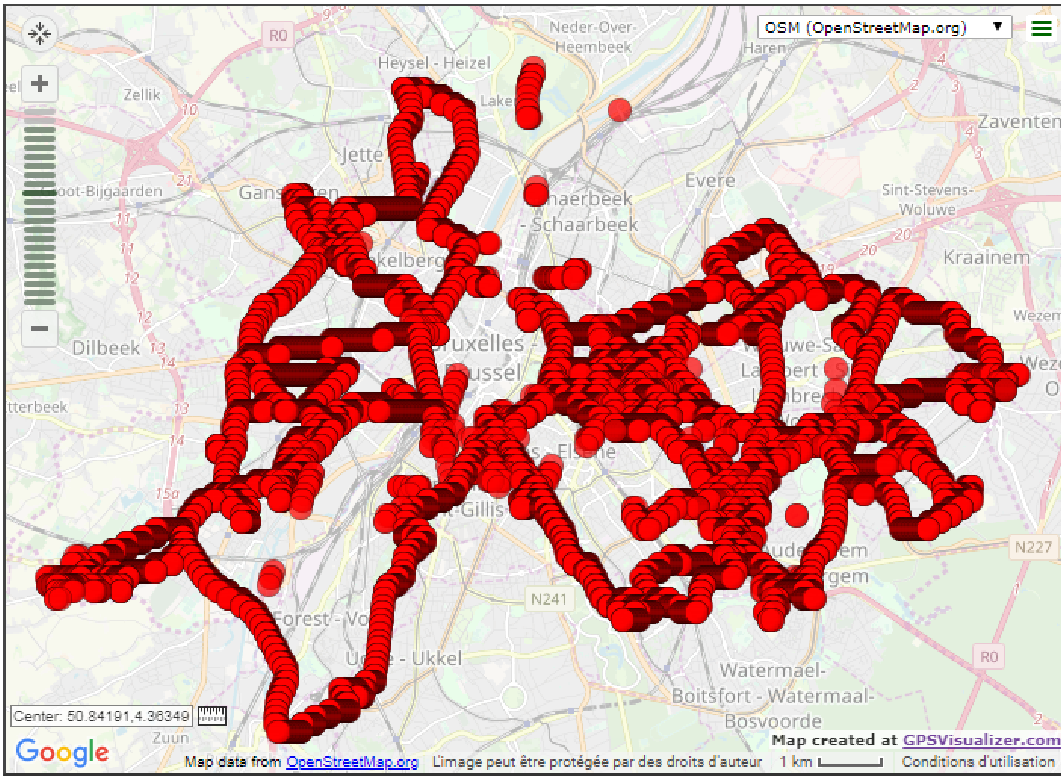

Figure 4.1: GPS data points of all the trips of the eighth driver.

4.1 Choice of the distance measurement

The total distance has been computed from two different sources with different methods as explained in SECTION 3.5. The distance has been calculated from the GPS coordinates but also by integrating the OBD speed over time. FIGURE 4.2 displays the comparison between those two computed distances, each dot representing one data point of all the trips jointly.

Each dot in FIGURE 4.2 represents an instantaneous distance from all trips of the eighth driver. It shows that the majority of dots are grouped around a 45 degree line which means that the distances are quite congruent. However, some points are gathered on the y-axis and some others are lost on the right of the graph. The first ones are due to the loss of GPS signal. This plays in the favour of the distance integrated from the speed and the time since the OBD data are always available. To understand the presence of odd points at the right of the graph, three of them were randomly analysed. Basically, the GPS coordinates of these points were entered in Google Maps to have a better vision of the details. For these three examples, the GPS coordinates have led to a place slightly off the road as shown in FIGURE 4.3. The coordinates entered in Google Maps are represented by white dots circled in black. The blue line is the portion of the road where the GPS data should have been since the yellow rectangle depicts the road. This proves that the GPS is not perfectly accurate even if the one used for the study is more precise than an average GPS since the horizontal position accuracy of the device used is 2.5 meters in the data sheets compared to 10 meters for an ordinary GPS. For these reasons, the distance integrated from the speed by the OBD and the time has been chosen for all computations.

Table 4.1: Summary of the main parameters for all the trips of the driver chosen.

Parameter	Value
Total distance useful to the study [km]	323.92
Total duration useful to the study [hh:mm:ss]	15:24:40
Number of trips useful to the study	61
Total number of trips recorded	71
Proportion of the trips in the urban category [%]	94.10
Proportion of the trips in the rural category [%]	5.80
Proportion of the trips in the motorway category [%]	0.10
Average velocity in the urban category [km/h]	20.29
Average velocity in the rural category [km/h]	67.84
Average velocity in the motorway category [km/h]	92.14
RPA in the urban category [m/s^2]	0.24
RPA in the rural category [m/s^2]	0.12
RPA in the motorway category [m/s^2]	0.11
$(v^*a_+)_1[95]$ [m^2/s^3]	5.20
$(v^*a_+)_2[95]$ [m^2/s^3]	10.53
$(v^*a_+)_3[95]$ [m^2/s^3]	11.83
Stop proportion in the urban category [%]	20.82

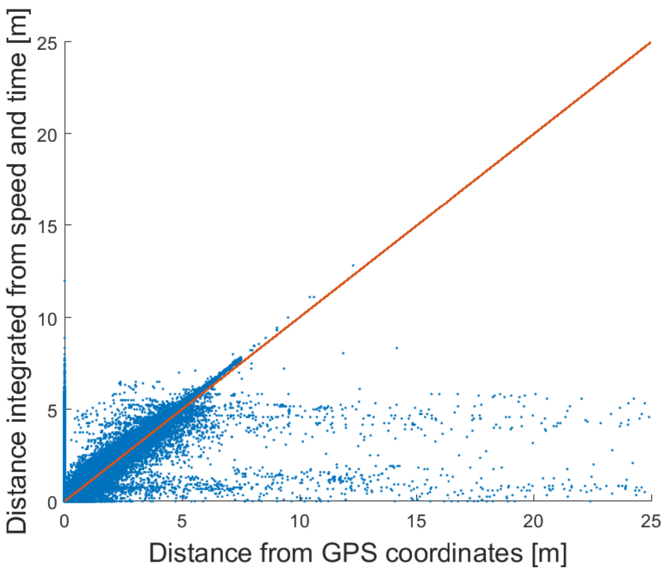

Figure 4.2: Distance integrated from speed and time in function of the distance computed from the GPS coordinates.

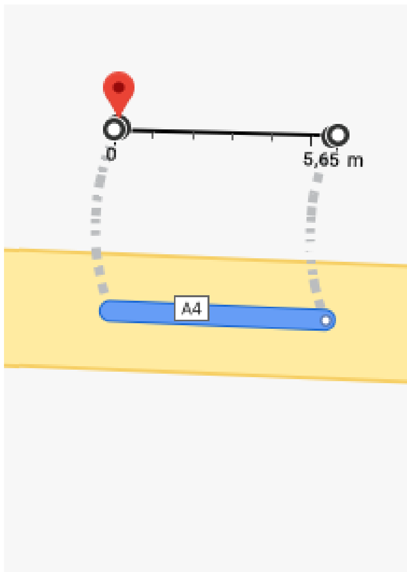

Figure 4.3: Example of GPS data for which the two distances computed are very different.

However, by doing the same computations on a larger interval, the distance computed from the speed and time and the distance computed from the GPS coordinates match better. Concerning the distance integrated, the same calculations were applied but on an interval of roughly 30 seconds ($dt = 30$) and instead of taking the instantaneous speed, the average speed on this interval is adopted. Concerning the distance from the GPS coordinates, the computations detailed in SECTION 3.5 are conducted but instead of computing the distances from two successive points, the distance is computed between two points spaced 30 seconds away.

FIGURE 4.4 represents the distance integrated from speed and time on an interval of 30 seconds and the distance computed from the GPS coordinates 30 seconds away from each other. This graph shows less abnormalities since all the singular data points are compensated with the others. Nevertheless, the instantaneous distances integrated from the speed and the time are used for all computations in this work.

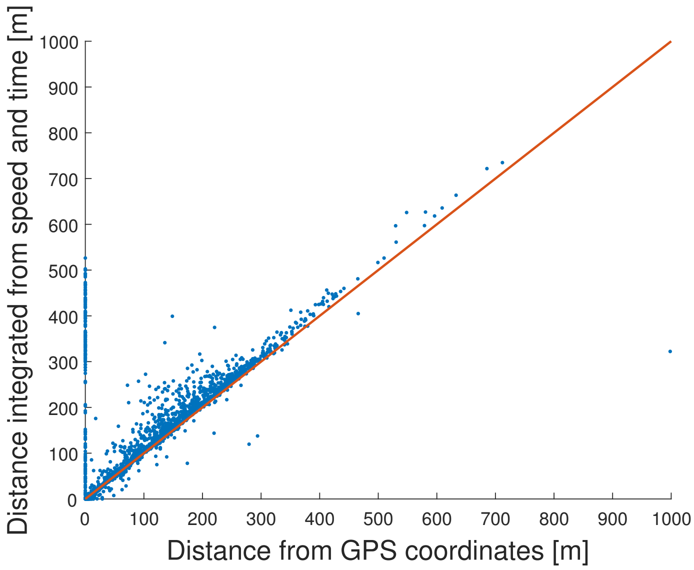

Figure 4.4: Comparison of the two ways of computing the distance on a larger interval.

4.2 Influence of frequency of measurements

The recording frequency has been a major topic of discussion during the conception of the data loggers. This has been tackled in SECTION 3.2. FIGURE 4.5 shows the proportion of recording frequencies for all the trips of the driver studied. The target of the frequency was 3 Hz. In FIGURE 4.5, around 94% were indeed 3 Hz. However, the system is not always accurate and some other frequencies were spotted. Practically five percent were recorded at 5 Hz and about 1% at 2 Hz. 0.05% of the frequencies were different from the three listed before. Among those exceptions, some were recorded at 1 Hz and the others were infinite frequencies which happens when the device reports two times the same line of data in the file.

Since most of researchers worked with a frequency of 1 Hz, a comparison between the signals has been made. Three signals have been computed. The first one, at 3 Hz, is the raw signal. The second one is the simulation of a signal at 1 Hz by taking one value out of three from the signal at 3Hz. The last one is a signal at 1 Hz computed by interpolation of the raw signal with a perfect vector of time at 1 Hz. The speeds of these three signals

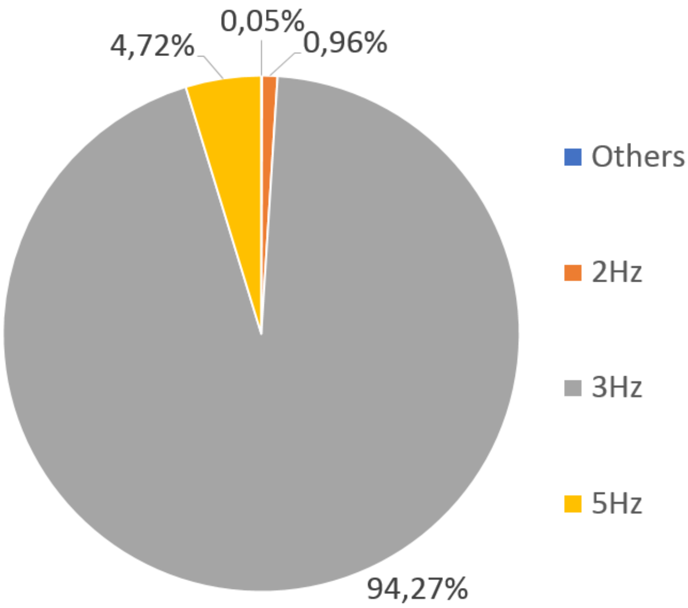

Figure 4.5: Proportion of the frequencies of all data for one subject.

for the first trip of the eighth driver are displayed in FIGURE 4.6. The shape is accurate for the three signals but by zooming at the curves in FIGURE 4.7, it can be seen that the raw signal at 3 Hz is more noisy.

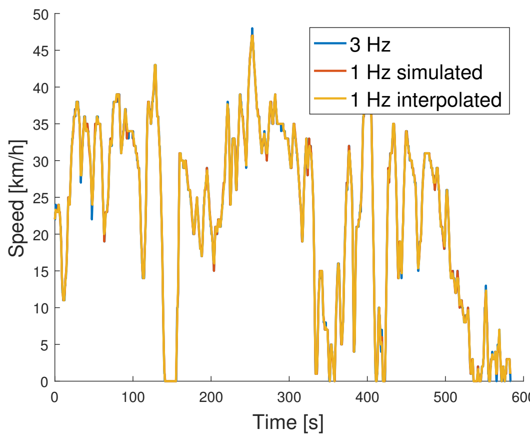

Figure 4.6: Speed in function of time for the three different frequencies.

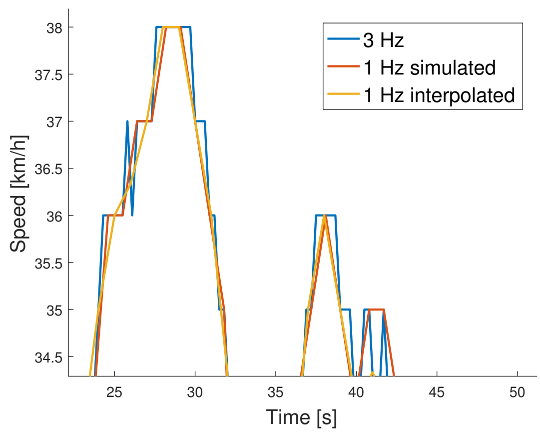

Figure 4.7: Zoom on the speed in function of time for the three different frequencies.

The acceleration has then been computed with respect to the formula in SECTION 3.5.1. The signals are displayed in FIGURE 4.8 and the zoom in FIGURE 4.9 illustrates that the signal at 3 Hz is indeed more noisy than the two others computed.

Therefore, it is interesting to see if the dynamic conditions are impacted by the recording frequency. By computing the relative positive acceleration and the 95th percentile of ($v \cdot a_{pos}$) for this trip, it appears that these two values are significantly higher for a frequency of 3 Hz. The values are reported in TABLE 4.2. In fact, the values of RPA are 17.85%

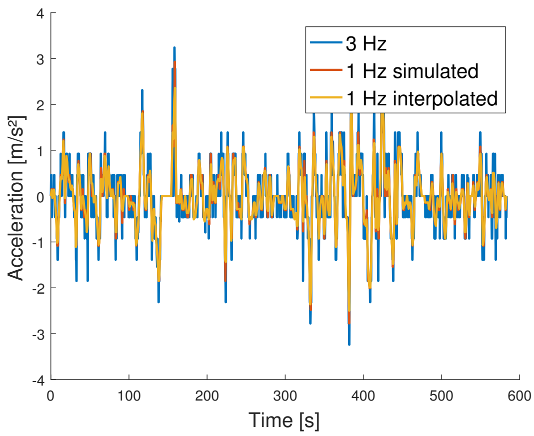

Figure 4.8: Acceleration in function of time for the three different frequencies.

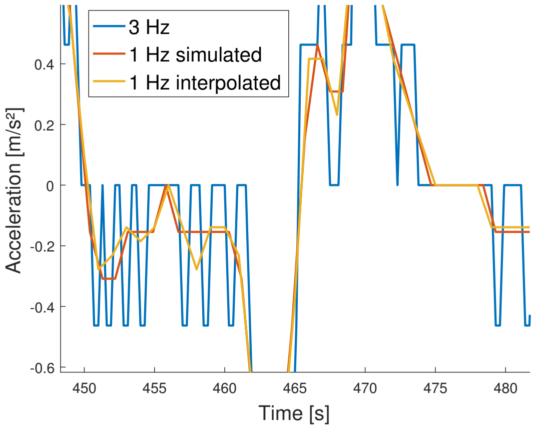

Figure 4.9: Zoom on the acceleration in function of time for the three different frequencies.

Table 4.2: Comparison of the dynamic conditions for the first trip of the eighth driver with the three different signals of acceleration.

Signal	RPA [m/s^2]	$(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$ [m^2/s^3]
3 Hz	0.79	10.80
1 Hz simulated	0.64	8.57
1 Hz interpolated	0.69	8.73

smaller for the simulated signal at 1 Hz and 11.69% smaller for the interpolated signal compared to the value computed with the acceleration at 3 Hz. Concerning the upper dynamic condition, it is 20.63% smaller for the simulated signal and 19.19% smaller for the interpolated one, compared to the value found with the raw acceleration. This conclusion could impact the results on the validity of some trips regarding those conditions. For this work, the raw frequency has been kept at 3 Hz. All computations were thus based on signals of 3 Hz. There is more noise with this method but less loss of information. However, for further studies, the choice of frequency of one hertz is sufficient and could even be better.

4.3 Inaccuracies in the recording of the parameters

As covered in Section 3.4.1, some recording issues arose. For the example studied in this chapter, the data from the OBD has never failed regarding the conditions of potential failure imposed in Section 3.4.1. However, as the case may be, the GPS data encountered some troubles. Figures 4.10 and 4.11 display the histograms of the proportions of time for which the GPS failed and the proportions of failure for the altitude.

The goal of a histogram is to picture the distribution of a quantitative dataset and to describe the pattern of the distribution. The histograms of this case represent the amount

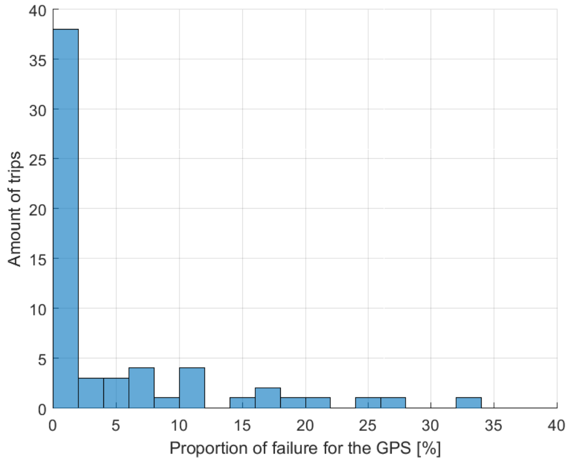

Figure 4.10: Histogram of the proportion of the inaccurate recordings for the GPS coordinates (latitude and longitude).

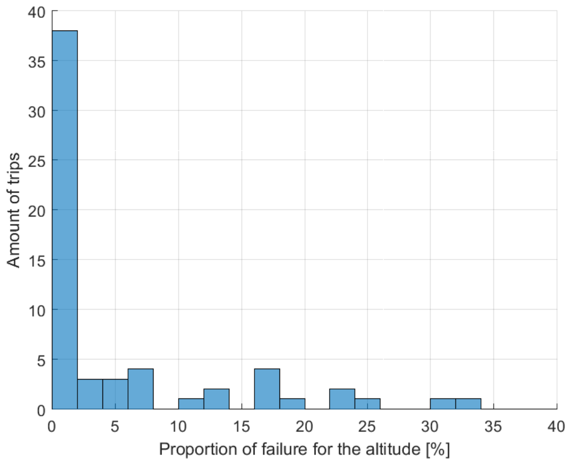

Figure 4.11: Histogram of the proportion of failure of the altitude records.

of trips within different categories of proportions of failure for the GPS coordinates and the altitude. For both cases, 38 out of 61 trips had failure proportion in the range [0% ; 2%] which is good. Yet, about twenty trips failed more than 2% of the time. The maximal values are 32.54 % for the GPS and 32.77% for the altitude. It means that approximately one third of the trip had no information about the localisation. The mean value for the GPS losses proportions is 4.77% and the median is 0%. For the altitude, the mean value of losses proportions is 5.23% of the time and the median is 0.06%. The median value is the number under which half of the information lies. For both cases, they are very small or zero. The 3-quantile, defined as the value under which 75% of the dataset lie, is 7.145% and 7.29% for the GPS coordinate and altitude inaccuracies respectively. This proves that trips losing their signal more than 7.2% of the time are not frequent.

4.4 Condition on the distance

The condition on the distance imposed by the RDE has not been respected at all. In fact, all the trips were below the sixteen kilometres required as displayed for the urban data in Figure 4.12. Concerning the proportions of the distance driven in each category, it has been detailed in Table 4.1. On that condition, the RDE requirements have been too severe. The maximal distance driven by the driver chosen was 13 km. This is obviously the hardest condition of the RDE procedure to fulfil while staying in the vicinity of Brussels. The influence of the cold start is smaller for bigger distances so this condition cannot be underestimated.

Moreover, the rural and motorway parts were under-represented. By summing the distances of all the trips for each category, the results are the following:

- 305 km in the urban part.
- 19 km in the rural part.

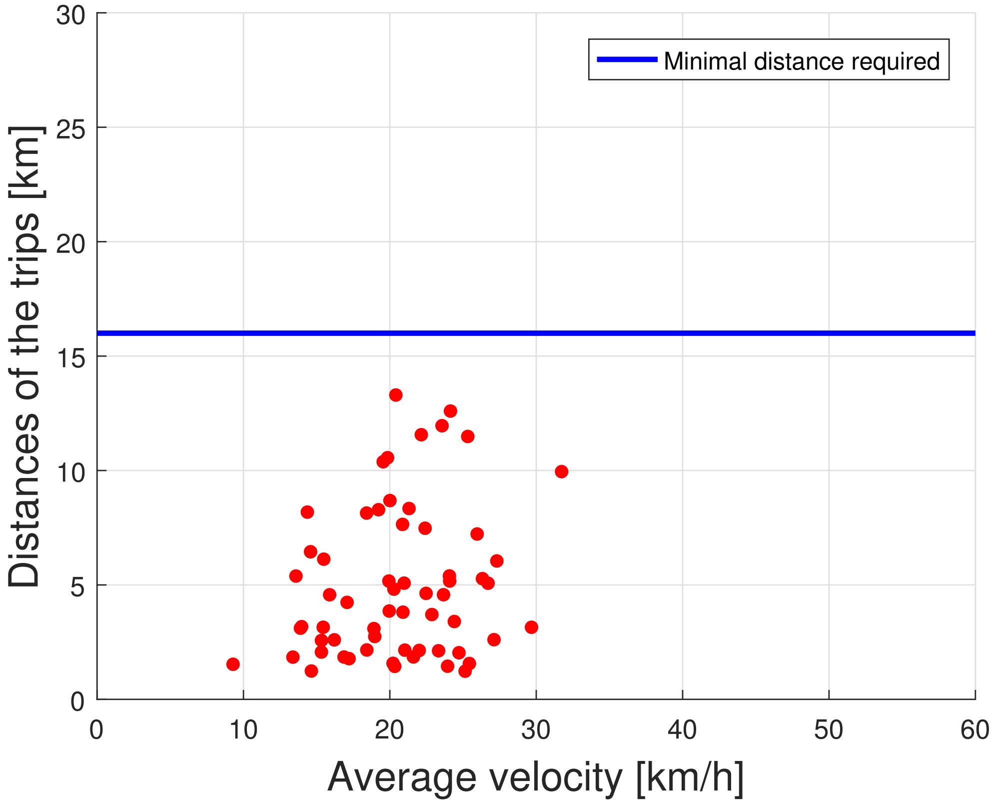

Figure 4.12: Distance of the trip on urban roads in function of the average velocity for each trip of the subject 8.

- 0.34 km in the motorway part.

Needless to say that both rural and motorway shares were far away of the condition on the distances. The maximal distance driven in one trip in the rural and motorway categories were 4.54 and 0.34 km respectively. This means that the motorway part has only been covered for 340 m in one trip. As a reminder, more than 16 km had to be ridden in these categories as well. The non respect of this condition leads theoretically to the invalidity of the RDE trip and therefore the emissions measured for this trip cannot be used to assess the car performance.

4.5 Lower condition: RPA

The graph of the relative positive acceleration for the subject under investigation is shown in FIGURE 4.13. For one trip, all data were classified in one of the three categories and the RPA has been computed for each of them. The blue line is the condition imposed by the RDE as explained in SECTION 2.4.6. Each blue dot represents one trip and the size of the marker is proportional to the distance of the trips. The short trips are thus depicted by smaller dots than the long ones. The pink points are the RPA computed from the matrix with the data from all the trips together. As reported in TABLE 4.1, their values are 0.24 m/s^2, 0.12 m/s^2 and 0.11 m/s^2 for each category.

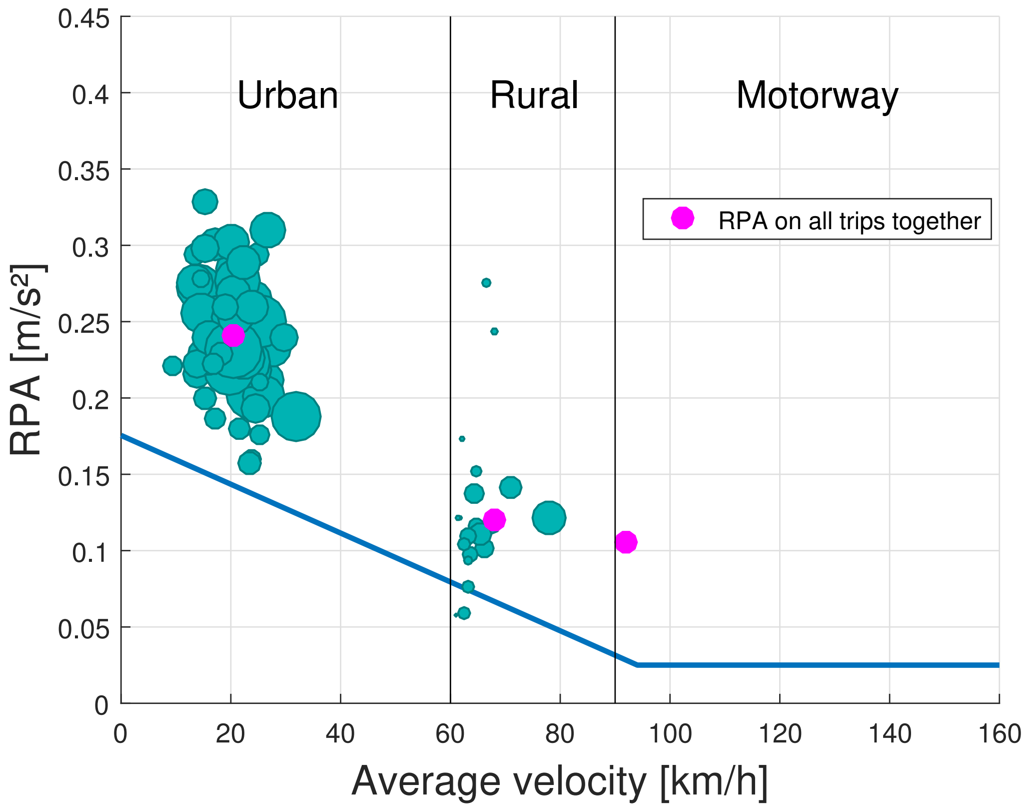

Figure 4.13: Relative positive acceleration in function of the average velocity for every trip of the subject 8.

For this driver, only two points were invalid regarding the RPA. Both came from the rural class. The information about those two invalid points is mentioned in TABLE 4.3.

Table 4.3: Information about the two invalid trips for the studied car regarding the RPA.

N° of the trip	$\mathbf{v_{avg}}$ **[km/h]**	**Distance [km]**	**RPA [m/s²]**	**Required value of RPA [m/s²]**
22	61.00	0.04	0.058	0.078
42	62.64	0.6	0.059	0.075

As stated before, the rural and motorway categories are not represented as they should be. Therefore, those two rejected values are not alarming. In fact, the distances covered in the rural part are 41 meters for the 22nd trip and 604 meters for the 42nd. The lack of data for these portions of trip (since the categories have been artificially made) proves that it cannot be stated that they are not representative of the RDE methodology. Moreover, the RPA of all the trips of the rural category is above the limit so it determines that those two outsiders are not very relevant. The most important part is the urban one and it can be seen that all the trips are above the limit. Since the data come from Brussels, the urban class is the point of interest. The analysis on all subjects in the next chapter is only based on the urban data.

4.6 Upper condition: $(v \cdot a_{\mathbf{pos}})_k[95]$

FIGURE 4.14 displays all values of $(\mathrm{v} \cdot \mathrm{a}_{pos})_k[95]$ for every trip. The blue line is the RDE condition as seen in SECTION 2.4.6. The pink dots represent the value of $(\mathrm{v} \cdot \mathrm{a}_{pos})_k[95]$ if all the trips were considered as one. These equal 5.2 m^2/s^3, 10.53 m^2/s^3 and 11.83 m^2/s^3 for the urban, rural and motorway parts respectively. For this driver, one trip is

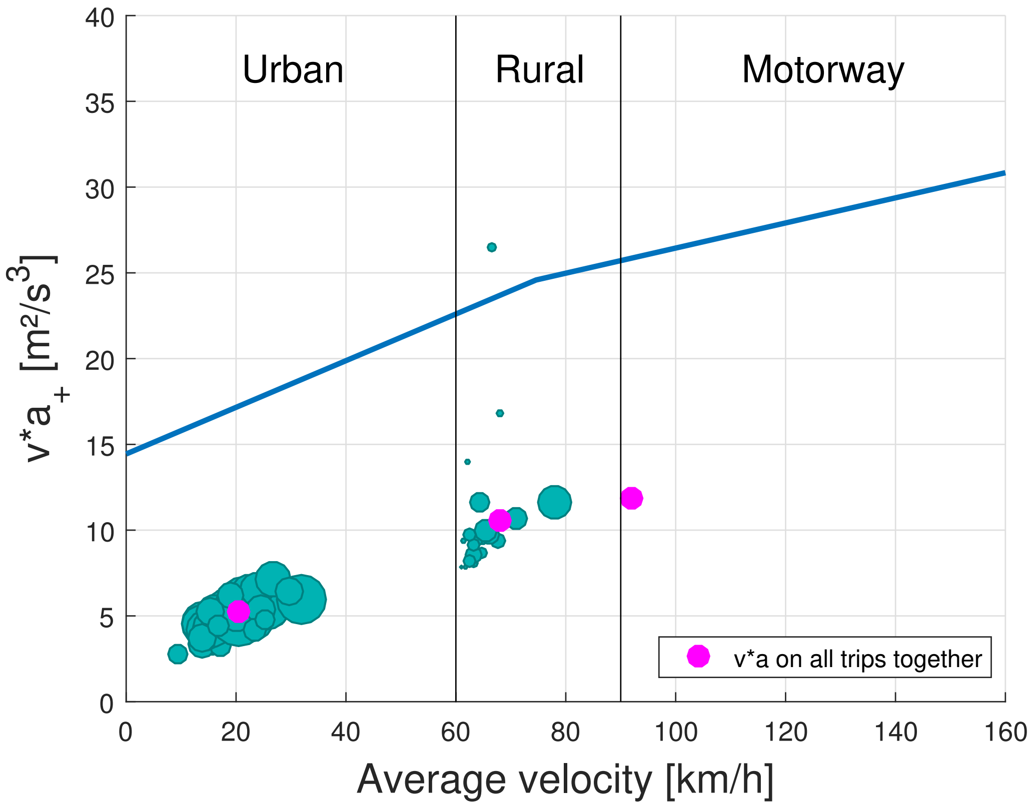

Figure 4.14: $(\mathrm{v} \cdot \mathrm{a}_{pos})_k[95]$ in function of the average speed for subject 8.

invalid regarding the upper condition. The information about that trip is indexed in TABLE 4.4. Again, the zone of interest is the urban one, where all the trips are below the upper limit. The invalid trip is 289 meters long which accentuates its insignificance. The validity of these trips would not be impacted by the computation of this value based on an acceleration recorded at 1 Hz.

Table 4.4: Information about the invalid trip for the studied car regarding $(\mathrm{v} \cdot \mathrm{a}_{pos})_k[95]$.

N° of the trip	$\mathbf{v_{avg}}$ **[km/h]**	**Distance [km]**	$\mathbf{v \cdot a_{pos}}$ $[\mathbf{m^2/s^3}]$	**Required value of** $\mathbf{v \cdot a_{pos}}$ $[\mathbf{m^2/s^3}]$
62	66.60	0.29	26.50	23.50

4.7 Stop proportion and average speed

FIGURE 4.15 represents two conditions for the urban part. First, the blue lines delimit

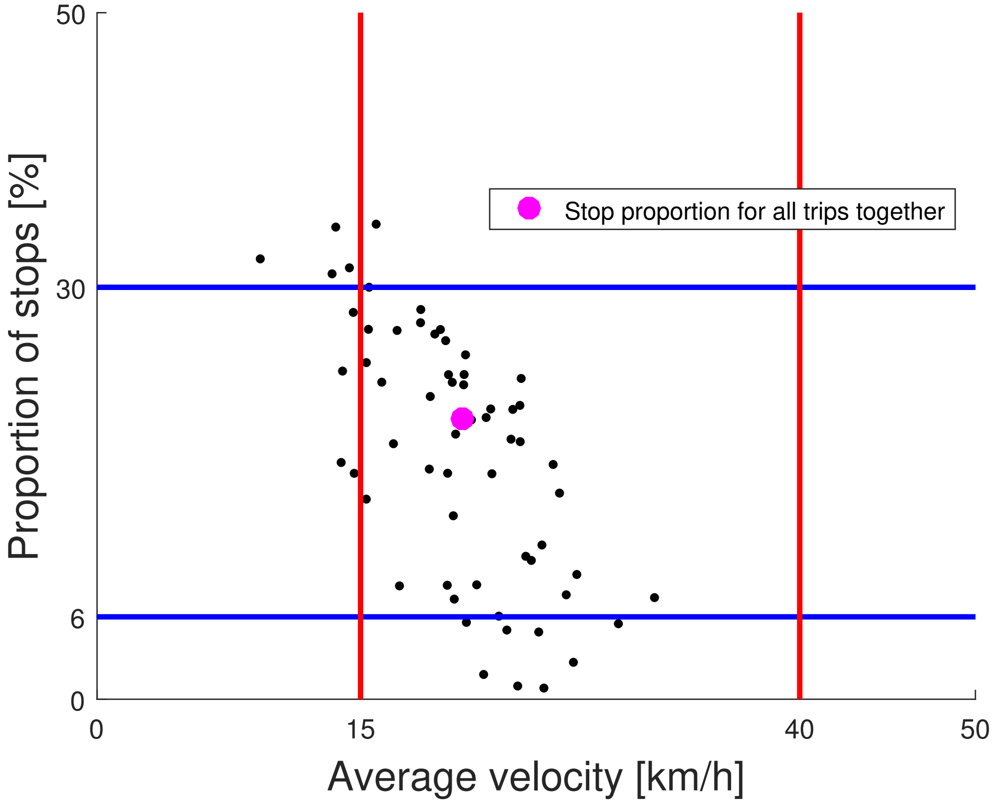

Figure 4.15: Proportion of stops in function of the average velocity for the urban category for the driver 8.

the upper and lower conditions on the stop proportion at 6% and 30% of urban time respectively. Secondly, the red lines show the conditions on the average speed which are equal to 15 km/h and 40 km/h. Again, the pink dot represents the value if all the trips were considered as one and it equals 20.82% of urban time.

The invalid trips according to the stop proportion are listed in TABLE 4.5.

Fourteen trips out of sixty-one are invalid with respect to the condition on the stop proportion. Eight of them are below the limit and six of them are beyond. For those who are beyond, it can be seen in FIGURE 4.15 that most of them have lower average velocities. In fact, four of them (number 5, 16, 32 and 45) are also invalid due to too low average speed. The trips 5, 16, 42 and 54 happened during peak hours as reported in TABLE 4.5. The peak hours were considered as 7 am - 10 am and 4 pm - 7 pm from Monday to Friday. These values have been arbitrarily chosen. This justifies thus the high value of stop proportion and the low average speed. However, it does not mean that all the trips during peak hours have a lower average speed and higher stop proportion as displayed in FIGURE 4.16, where many trips during peak hours stand within the conditions.

Table 4.5: Number of the trip, average speed, distance, stop proportion and time of the day of the invalid trips regarding the stop proportion condition.

N° of the trip	**Average speed of the trip [km/h]**	**Distance [km]**	**Stop proportion [%]**	**Time of the day**
2	21.02	2.15	5.60	Off-peak
5	13.38	1.85	30.98	Peak
7	27.12	2.61	2.68	Off-peak
16	14.37	8.18	31.42	Peak
29	23.94	1.45	0.96	Off-peak
32	13.59	5.39	34.38	Off-peak
42	15.48	6.12	30.00	Peak
45	9.29	1.53	32.07	Off-peak
48	22.01	2.14	1.80	Off-peak
54	15.89	4.57	34.60	Peak
57	25.14	1.23	4.89	Peak
68	23.33	2.12	5.04	Peak
69	25.44	1.57	0.81	Off-peak
71	29.68	3.15	5.50	Off-peak

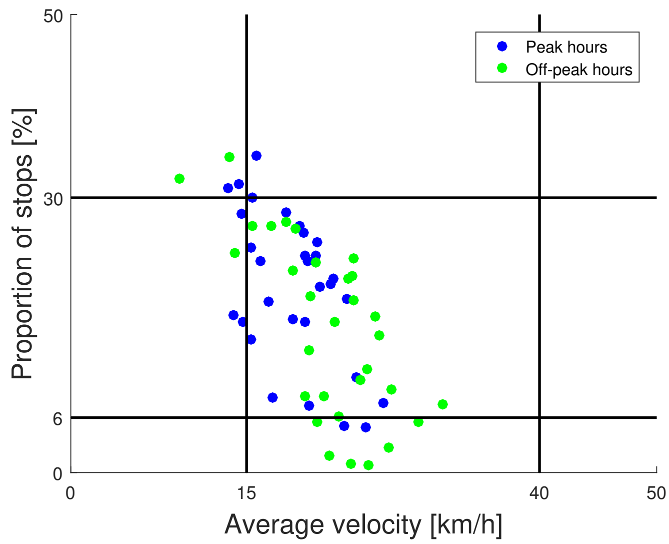

Figure 4.16: Proportion of stops in function of the average velocity for trips during peak hours and off-peak hours.

Concerning the two others, the 32 and 45, they have been recorded on a Wednesday and on a Thursday respectively, around 11:30 am. The abnormality of these trips could be explained by the fact that the driver could have been talking to somebody out of the car while having the engine on, or she followed an unexpected slow driver for instance. The 45th trip lasted for only 1.5 km and the 32nd for 5.4 km.

Concerning the trips with too low stop proportion, six out of eight happened during off-peak hours and all of them are valid regarding the average speed. The 57th and 68th trips happened during peak hours and their distance ridden was 1.5 km and 2.1 km and lasted for 175 seconds and 327 seconds respectively. These trips being so small, it is normal that their stop proportion is not in the norms and that the factor of peak hours did not affect this condition. Concerning the six trips during off-peak hours, they all lasted less than 7 minutes. Some of them did not fulfil the condition due to lack of data. Other cases could be that the driver rode during a particularly calm period, or on roads with no traffic lights.

The eight invalid trips regarding the average speed are listed in TABLE 4.6. All of them are below the lower limit. Four of them (5, 16, 32 and 45) were already analysed

Table 4.6: Number of the trip, average speed, distance and time of the day of the invalid trips regarding the average speed condition.

N° of the trip	**Mean speed of the trip [km/h]**	**Distance [km]**	**Time of the day**
5	13.38	1.85	Peak
16	14.36	8.18	Peak
31	13.90	3.11	Peak
32	13.59	5.39	Off-peak
38	14.64	1.24	Peak
41	14.58	6.45	Peak
44	13.97	3.18	Off-peak
45	9.29	1.53	Off-peak

since they were also invalid regarding the condition about the stop proportion. Three of the four others (31, 38, 41) happened during peak hours and lasted between 5 minutes (38th trip) and 26 minutes (41st trip). The 44th trip has been ridden during off-peak hours. The reason of their rejection might be because of the time of the day but could also be due to the fact that, for example, the roads were busy so there were slowdowns, or some perturbations of the road conditions occurred due to roadworks, or else a reduction of traffic lanes due to an accident or a car breakdown.

4.8 Altitude

In the RDE methodology, the altitude difference between the start and stop points must be less than 100 m. It should not be a problem in this case since Brussels is quite flat. However, FIGURE 4.17 shows that the 40th trip is out of limits.

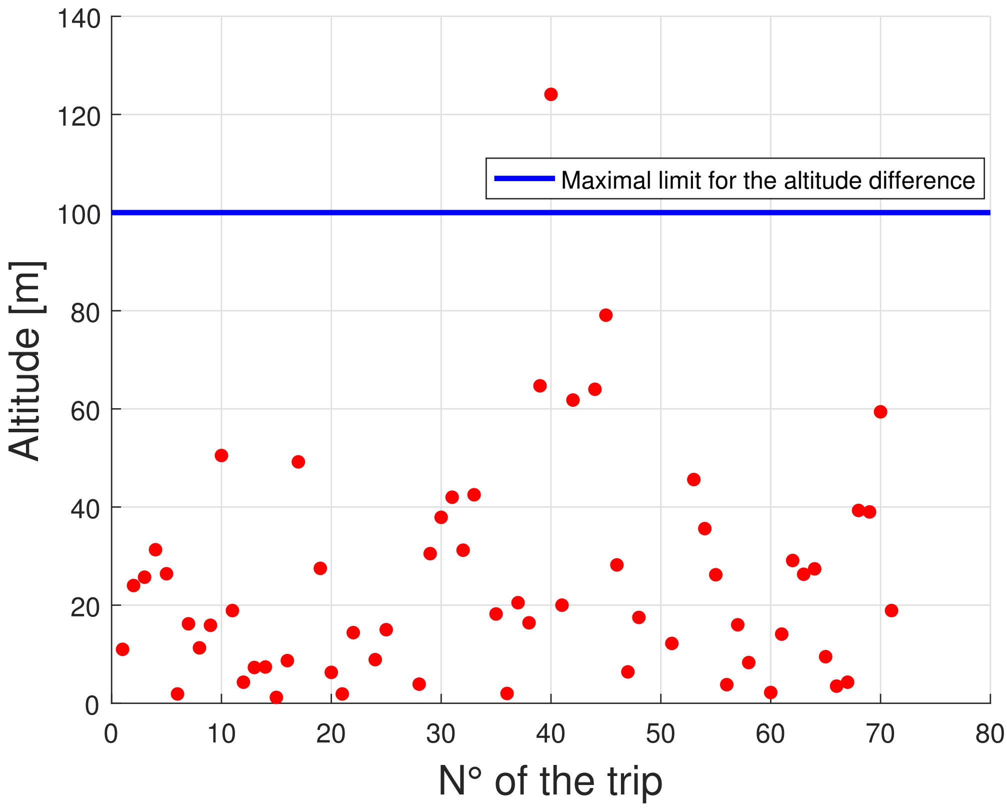

Figure 4.17: Condition on the altitude difference for every trip of driver 8.

Sometimes, the GPS is not connected in the first minutes of driving. In the case of the failed data, the first connection to satellite for the altitude happened after almost four minutes. This first value, taken to compute the altitude gain, was 191.40 m in the data retrieved. This was astonishing, given that the driver only rode in Brussels. By looking at the latitude and longitude, the exact place of this point of data was found and the real altitude, given by Google Maps, was 90 m. Therefore, this value is a pure example of incoherence from the GPS chip.

Another condition on the altitude is set in the RDE methodology and states that the cumulative altitude gain must be 1200 m at most for 100 km. Since the altitude is not accurate enough and often missing or wrong, this condition is not deeply examined in the scope of this work. However, five trips have been tested regarding this condition. The chosen trips did not suffer from loss of GPS signal. The following formula is used:

$$\begin{cases} \text{Cumulative altitude gain} & = \left(\sum_{i=1}^{n} d\mathrm{h}_{\mathrm{pos}}\right) \cdot \frac{100}{\mathrm{d}_{\mathrm{tot}}} \qquad [\mathrm{m}/100\mathrm{km}] \\ d\mathrm{h}_{\mathrm{pos}} & = (h(i) - h(i-1)) > 0 \end{cases}$$

with h(i) being the altitude at instant i and $\mathrm{d}_{\mathrm{tot}}$ the total distance of the trip. The results are reported in Table 4.7. Actually, none of them respect the condition imposed by the RDE. The results might be influenced by the fact that the data here were not smoothed out as they are in the RDE methodology. This is a very interesting point to study in further researches.

Table 4.7: Cumulative altitude gain for five trips of the eighth driver.

N° of the trip	Distance of the trip [km]	Cumulative altitude gain [m/100km]
1	3.71	4277
13	8.69	4541
25	8.34	4869
36	8.14	4551
41	6.45	9131

4.9 Maximal speed

The maximal permitted speed in the RDE test is 145 km/h and 160 km/h for 3% of motorway driving time. FIGURE 4.18 represents the maximal speed for each trip with the blue line being the maximal limit permitted and the green dotted line the maximal limit at 160 km/h. It proves that all the trips were mainly in the city since the maximal speed of all the trips is below 100 km/h. The mean maximal speed over all the trips is 55.28 km/h. This criterion is not well represented in this study since no part on the motorway is required. Moreover, the maximum speed limit is 120 km/h in Belgium so in theory, this condition should not be exceeded. The histogram in FIGURE 4.19 shows the distribution of all speeds. Out of all the points of data, one quarter lies in the range

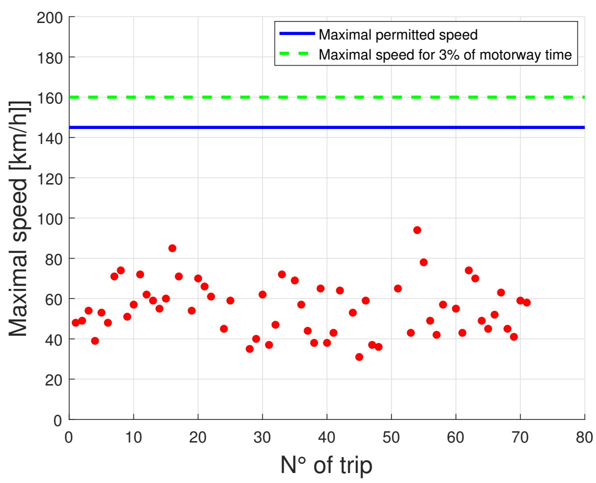

Figure 4.18: Maximal speed for each trip of the subject 8.

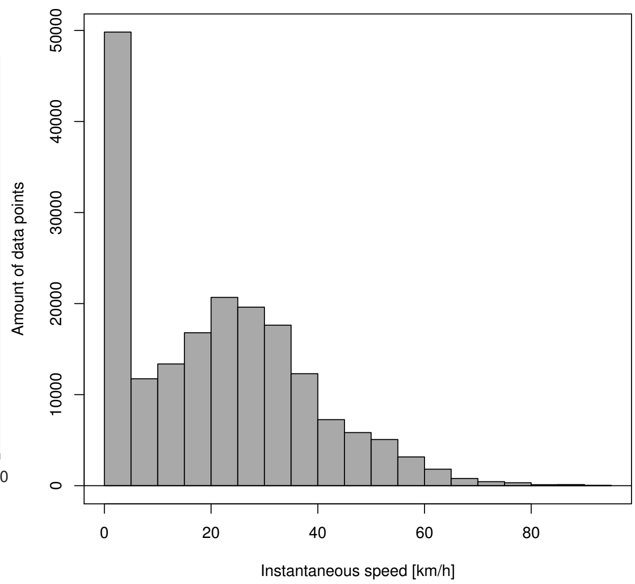

Figure 4.19: Histogram of instantaneous speeds for all data points of driver 8.

[0 km/h - 5 km/h]. This is explained by the fact that the device is activated by the time the engine is turned on. So if the person stays in the car for a while without turning the engine off, a lot of useless data is recorded. The biggest example of this case is waiting at a red traffic light. Moreover, the Arduino takes time to be operational. Hence, if the driver starts the engine and goes directly on the roads, all recorded data are zero until the Arduino is fully functioning. Few data have a speed higher than 65 km/h.

4.10 Temperature

In the RDE methodology, the vehicles start with their engine cold. The cold start encloses the first five minutes of driving or stops when the engine coolant temperature reaches 70°C. This section aims at searching if this Brussels driver always starts with a cold engine. The maximal and minimal temperatures for each trip of the eighth driver are displayed in FIGURE 4.20. Twenty-seven trips out of sixty-one have an engine coolant temperature

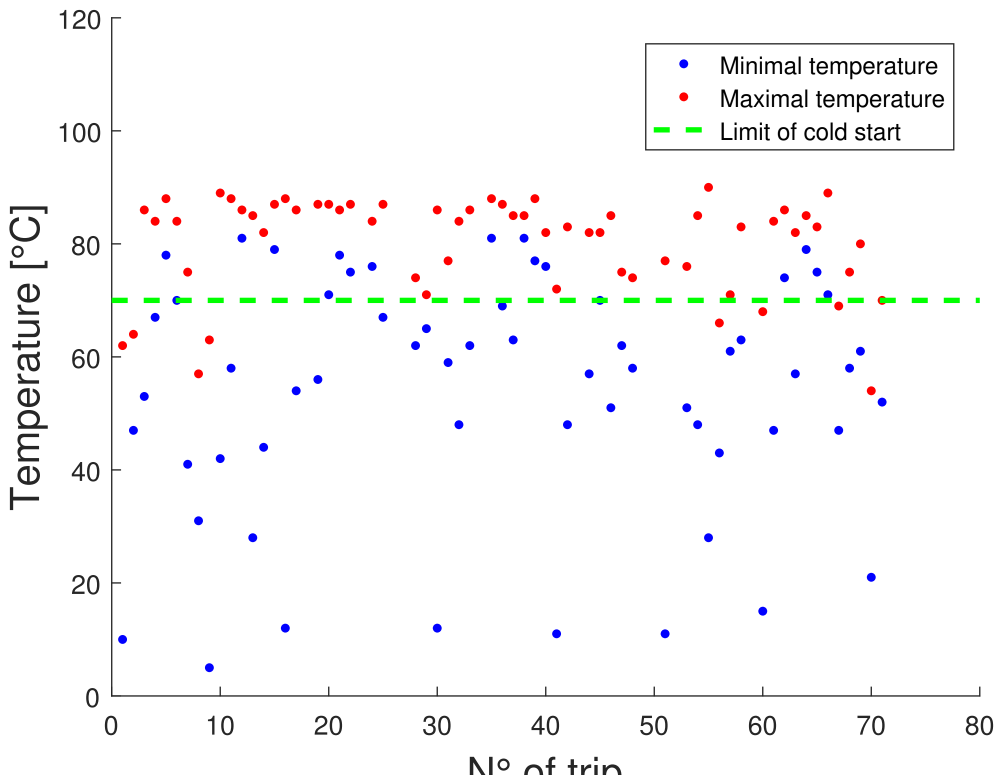

Figure 4.20: Maximum and minimum engine coolant temperature for each trip of the driver n°8.

larger than 70 °C when the car starts. This means that 44% of the trips started with a hot engine. The other way around, eleven trips did not have the time to warm up and their maximal temperature stayed below 70°C. In total, 99.3 km were ridden with the coolant temperature below 70°C which accounts for 30.66% of the total distance travelled. On average, 37.11% of the distance of each trip is ridden with a coolant engine temperature lower than 70°C. The mean time on all trips to reach 70°C is 4 minutes and 30 seconds.

4.11 Overall analysis

TABLE 4.8 summarises the conditions approached in this chapter (average speed, stop proportion, upper and lower dynamic conditions, maximal speed and distance). The *"I"* stands for invalid and the *"V"* for valid regarding the RDE condition analysed.

Out of 71 trips, ten were removed because they were not in Brussels and twenty were invalid regarding some of the five conditions, without taking the distance into account. All 71 trips are invalid due to the condition on the distance. This condition has not been taken into account in the elaboration of this list since all of the data would have been invalid and all the trips would have been reported in this table. The condition on the altitude has also been removed due to its inconsistency. The biggest unfulfilled condition after the distance is the stop proportion, followed by the average speed. Indeed, by considering the trips in Brussels only, almost 23% of them are invalid regarding the stop proportion and 13% for the condition on the average speed. Trips are invalid regarding two conditions at most besides the condition on the distance; four of them were invalid with respect to the conditions on the stop proportion and the average speed together and one trip was invalid because of the stop proportion and RPA at the same time. The fifteen others have only one condition unfulfilled (other than the distance). The criterion on the temperature has not been included in the overall validity of the trips since it is not a condition.

Table 4.8: Invalid trips regarding some of the RDE conditions.

N° of trip	**Average speed**	**Stop prop**	**$v \cdot a_{pos}$**	**RPA**	**v_max**
2	V	I	V	V	V
5	I	I	V	V	V
7	V	I	V	V	V
16	I	I	V	V	V
22	V	V	V	I	V
29	V	I	V	V	V
31	I	V	V	V	V
32	I	I	V	V	V
38	I	V	V	V	V
41	I	V	V	V	V
42	V	I	V	I	V
44	I	V	V	V	V
45	I	I	V	V	V
48	V	I	V	V	V
54	V	I	V	V	V
57	V	I	V	V	V
62	V	V	I	V	V
68	V	I	V	V	V
69	V	I	V	V	V
71	V	I	V	V	V

CHAPTER 5

Statistics for all subjects

This chapter covers the global results of the study. The influence of seven quantitative parameters and twelve qualitative ones on the driving behaviour are looked at.

First, the parameters are explained, then descriptive statistics present them. Statistical hypothesis tests expose if those quantitative parameters (chosen based on the conditions imposed by the RDE) have an impact on the validity (regarding the RDE) of the test. The goal of this part is to see if the variables alone influence the validity of the trip, without taking the other parameters into account. These results define if the RDE is really representative of the driving patterns of Brussels drivers or not. Then, twelve qualitative parameters are analysed and their impact on the validity of the trips and on the mean of each quantitative variable is studied. This helps to visualise which factors impact the most the conditions of the RDE.

In the previous chapters, it has been noticed that the rural and motorway categories were underrepresented. Moreover, all data points were mainly recorded in Brussels and not on real rural or motorway roads so these categories were artificial. These data negatively influenced the statistics. Therefore, it has been decided to only consider the data from the urban category in the computations. This choice is questionable, since all the data points with an instantaneous speed greater than 60 km/h are exploited in this chapter. However, it does not make sense regarding the RDE to work with all data together and working with the three categories is not the most appropriate solution with few data.

5.1 Parameters studied

In total, 899 trips have been recorded. However, only 584 were kept and the 315 others were deleted because they were out of Brussels or less than 1 km long.

The main parameters computed are the following: the average speed, the stop proportion, the values of the lower and upper dynamic conditions, the distance of the trip and its duration and the maximal speed. Binary values determining if the driver rode during the peak hours and the weekend are also returned, along with whether the car started with a cold engine. These are not explicit conditions in the RDE methodology but since they are important factors, it seemed legit to add them in the analysis.

Finally, a last variable about the overall validity of the trip is given. It takes into account the average speed, the RPA, $(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$, the stop proportion and the maximal speed. Since the condition on the distance is too severe and the condition on the altitude is inconsistent, those two were not considered in the computation of the overall validity. Actually, after analysing the results of the altitude, it was decided to completely remove this condition from the study since the failure of the recording of data was too significant (as partly seen in SECTION 4.8). If the trip is invalid regarding one of these parameters, the overall validity coefficient is zero, meaning that the trip is invalid. It is important to understand that these invalid trips were well recorded and well driven but they are invalid regarding the RDE conditions. The definition of invalidity in this chapter is linked to the conditions of the test.
Among the 584 trips, 190 were invalid to the test conditions (without taking the distance into account) and 394 were valid regarding the overall factor. This means that one third of the trips does not respect the conditions imposed by the RDE test procedure even when the condition on the distance is not taken into account. Brussels people's driving behaviour does not match with the requirements of the RDE. If the distance is considered in the overall validity factor, only 13 trips out of 584 are valid regarding the RDE conditions studied in this work which accounts for 2.2% of the trips. The summary of the valid and invalid trips regarding some of the RDE conditions are listed in TABLE 5.1. The overall results come from the overall validity factor with and without the distance.

Table 5.1: Summary of the valid and invalid trips regarding the different conditions of the RDE.

Variable	**Valid**	**Invalid**
Average speed	516	68
Stop proportion	412	172
RPA	582	2
$(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$	584	0
Maximal speed	583	1
Distance	14	570
Overall result (without the distance)	394	190
Overall result (with the distance)	13	571

In this chapter, the influence of all factors are studied on the overall validity of the trip. It is computed based on the overall validity factor as defined here above, without the distance.

On top of the parameters listed before, information about the car has been reported and are included in the global analysis. TABLE 5.2 outlines the parameters under investigation.

They are classified in two categories: quantitative or qualitative. Quantitative values represent the quantities, for example the exact value for the RPA. Qualitative variables depict the nature of something, for instance the type of fuel.

Table 5.2: Quantitative and qualitative parameters adopted in the statistical analysis.

Quantitative	**Qualitative**
Average speed	Cold start or not
RPA	Overall validity factor
$(\mathrm{v} \cdot \mathrm{a}_{\mathrm{pos}})_k[95]$	Weekend - week
Stop proportion	Peak - Off-peak hours
Maximal speed	Type of fuel (Diesel - Gasoline)
Distance of the trip	Car usage (Company - private car)
Duration of the trip	Brand of the car
	Displacement (2 classes)
	Year (2 classes)
	Power (2 classes)
	Type of gearbox (manual - automatic)
	in Brussels - outside of Brussels

The displacement, power and year have been changed from numeric variable to categories. Indeed, they were each divided in two classes. This is useful for the analysis of their impact on the variables studied by the RDE, otherwise there would be as many classes of variables as there are different values for these parameters.

Data recorded outside of Brussels are compared to data from Brussels to spot any similarities or differences. However, data outside of Brussels are not integrated in the computations of the other parameters studied.

The brands of the cars have been analysed in APPENDIX A.2.1.

5.2 Introduction to statistical hypothesis testing

Statistical hypothesis testing is used in the next sections, it is therefore appropriate to introduce this subject here.
Statistical hypothesis testing is a decisional process between two hypotheses: the null hypothesis H_0 and the alternative hypothesis H_1. The principle is to fix a value or a statement and then this value (H_0) is confirmed or rejected based on the sample data. The null hypothesis generally affirms that the variable is equal to a certain value and the alternative hypothesis usually denies the null hypothesis. Non rejection of H_0 does not automatically mean that the null hypothesis is true. Multiple relevant hypothesis tests exist, such as the Student's t-test, or the Mann–Whitney–Wilcoxon test. These are approached in the next sections.

The null hypothesis is rejected in favour of the alternative one if and only if the p-value is smaller than a certain threshold. This threshold is conventionally set at 5% so this value is used as well in this work. The p-value is the smallest level of significance for which H_0 is rejected. In other words, it is the probability of reaching, under the null hypothesis, the value of the statistical test somewhat as extreme as the value computed from the sample. If the p-value is greater than the significance value set, H_0 cannot be rejected (this does not mean that it is accepted) so no conclusion can be drawn due to lack of evidence.

5.3 Analysis of numeric variables

Two kinds of statistics are conducted on the numeric variables from the RDE conditions. First, they are graphically and numerically described and then, a statistical hypothesis test is taken to check if their mean values are the same for valid and invalid trips.

Descriptive statistics Descriptive statistics, either quantitative or easily understood graphs, are useful to summarise, represent and describe statistically a large quantity of data. The goal is to observe and give as much information as possible about the dataset, making the most efficient use of it. The boxplots and histograms permit analysing datasets without asking any statistical question; the goal is to describe the variables and not to learn about the population that the data represent. Descriptive statistics are nonparametrics which means that they are distribution-free (or they have a distribution with undefined parameters).

In this section, numeric variables are studied along with an univariate analysis. Single variables are described with histograms (briefly introduced in SECTION 4.3) and boxplots. For each variable, the graph reporting the data with respect to the conditions from the RDE methodology is presented. Then, as stated before, a histogram helps visualise the distributions of the quantitative data for invalid trips and all trips. This permits spotting the differences and the influence of the parameters regarding the validity of the trip. After that, a boxplot investigates the values for both invalid trips and all trips.

A boxplot, displayed in FIGURE 5.1, is used in descriptive statistics to represent the profile of a quantitative series of data. This profile is depicted by means of quartiles. There are three quartiles dividing the data into four parts. The first one (noted the 25th percentile, $q_{0.25}$ and sometime called Q1), separates the lowest 25% of data from the rest.

The second one is the median and as briefly evoked in SECTION 4.3, it cuts the data set in half. Finally, the third one, the 75th percentile or Q3, has 75 % of the sample taking values smaller than it and 25% taking values larger than it.
The interquartile range (IQR) is defined as the difference between the values of the 75th percentile and the 25th percentile and it measures the statistical dispersion. Mathematically, $\boxed{IQR = Q_3 - Q_1}$. The minimum and maximum values of the boxplot are the values within 1.5 time the interquartile range. All values outside these ranges are called outliers.

In other words, a point of data is considered as an outlier if it is below $q_{0.25\text{ - }1.5 \cdot \mathrm{IQR}}$ or above $q_{0.75\,+\,1.5 \cdot \mathrm{IQR}}$, with q being the quantile. They are points of data significantly different and distant from the other values.

In boxplots, the spaces between the different parts represent the skewness (which measures the asymmetry of a probability distribution) and the dispersion of the values. If the data are very spread, the boxplots become larger.

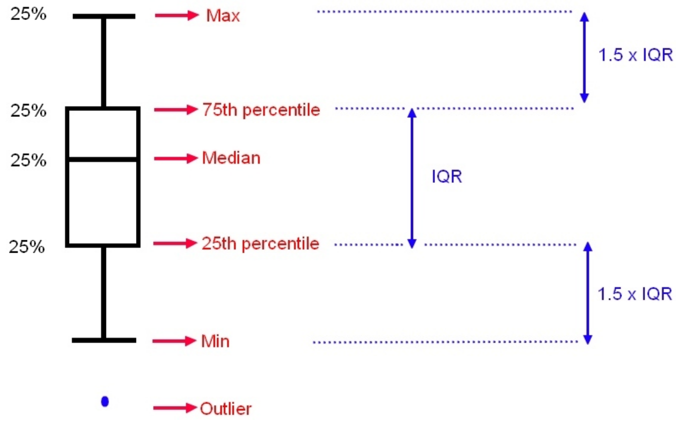

Figure 5.1: Parts of a boxplot.

The boxplots are usually presented abreast to compare the distribution of different groups of datasets. In this case, the two groups are invalid trips and all trips. This has been chosen to have an idea of the trends of the abnormalities compared to all data together. The descriptive statistics are conducted on the trips grouped regarding their overall validity.

Statistical hypothesis testing Statistical hypothesis testing have been introduced in SECTION 5.2. These tests are used on parametric variables to see if their means are the same for the valid and invalid cases. In order to prove that or the contrary, the best test must be chosen. If the data are normally distributed, the Student's *t*-test is necessary. Otherwise, the Mann–Whitney–Wilcoxon test are needed. Quantile-quantile plots (Q-Q plots) are analysed to determine whether the variable is normally distributed.

The Q-Q plot compares the observed data to the one following a normal distribution. This permits knowing which test to conduct. The normal distribution is the most common law for continuous quantitative variables. It is written as $\mathcal{N}(\mu, \sigma^2)$ with μ the expectation and σ^2 the variance. The variance is the expectation of the squared deviation of a random variable from its mean value. The normal distribution is represented as a bell curve. In a Q-Q plot, the data obtained are compared to the ideal data with respect to a normal distribution. The y-axis represents the observations and the x-axis the normal theoretical quantiles. Each recorded data point (the parameter under study for each trip) is a quantile and the normal curve is divided in the same amount of quantiles than the amount of data points. Then, each couple made of the observed point and the ideal point is represented on the graph. If the points are on the straight red line in FIGURE 5.2, it means that they follow a normal distribution. FIGURE 5.2 explains how the Q-Q plot is made. The observations in black are plotted as a function of what their value would have been if the dataset followed a normal distribution. In the case of the picture, since the points are close to the red line, the data are normally distributed.

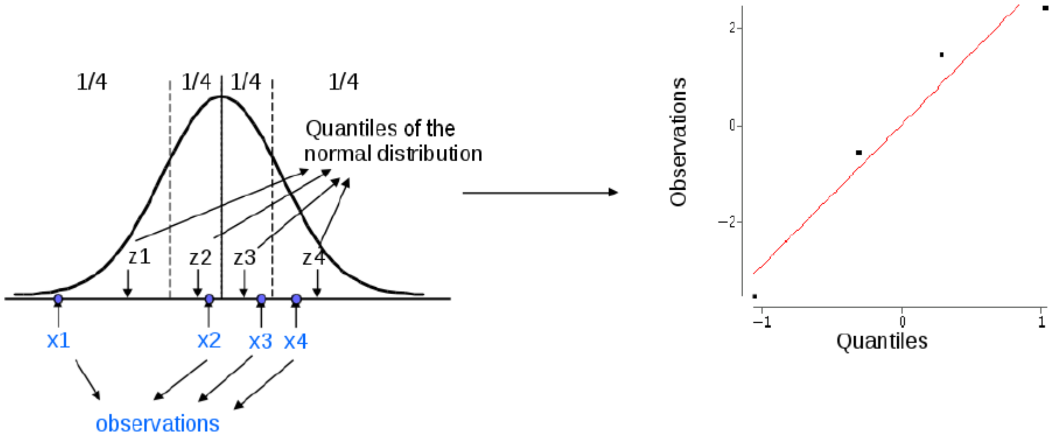

Figure 5.2: Construction of a Q-Q plot, from R. Von Sachs, 2016 [51].

Once the Q-Q plot has been analysed, two possibilities arise: the Student's t-test for normally distributed dataset and the Mann–Whitney–Wilcoxon test for the others. For both tests, H_0 states that the mean of the variable under investigation for a valid trip is the same as the mean for an invalid trip. The alternative hypothesis is thus that these means are different. Both tests raise the same question but use different computations to answer it because their assumptions are different.

Statistical hypothesis testing of each parameter makes sense to determine whether it affects the validity of the test. All 7 parametric variables are described and analysed one by one.

5.3.1 Average speed

Out of all trips, 68 were invalid regarding the average speed imposed by the RDE test methodology and 516 were valid. Figure 5.3 shows the stop proportion as a function of the average speed, with the vertical red lines being the conditions imposed on the average speed at 15 km/h and 40 km/h. The average speed must stay within this interval and it can be seen that the upper limit of the average speed has never been exceeded.

The histograms in Figure 5.4 are represented for the average speed of invalid trips regarding the definition of the overall validity and for all the trips together. Almost 30% of invalid trips have an average speed within 10 km/h and 15 km/h. This makes sense since most of invalid trips regarding the average speed are around that range. Concerning all trips together, most of them have an average speed around 20 and 25 km/h.

The boxplots for invalid and all trips are displayed in Figure 5.5 and the accurate values are reported in Table A.4 from Appendix A.3. The values for the valid trips only are also presented in that table. It seems logical that the mean speed of invalid trips is smaller than the one of all trips, since there are more valid trips and their average speed must be greater than 15 km/h. The comparison of IQR shows that the data of invalid trips are more spread than for all trips.

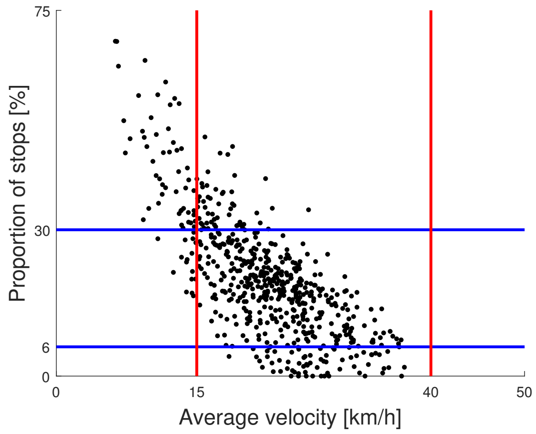

Figure 5.3: Proportion of stops in function of the average velocity for the urban category.

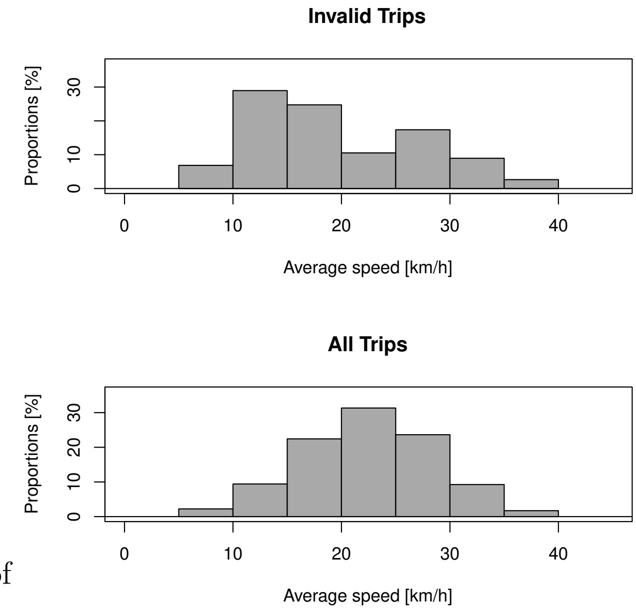

Figure 5.4: Histogram of the average speeds for all drivers.

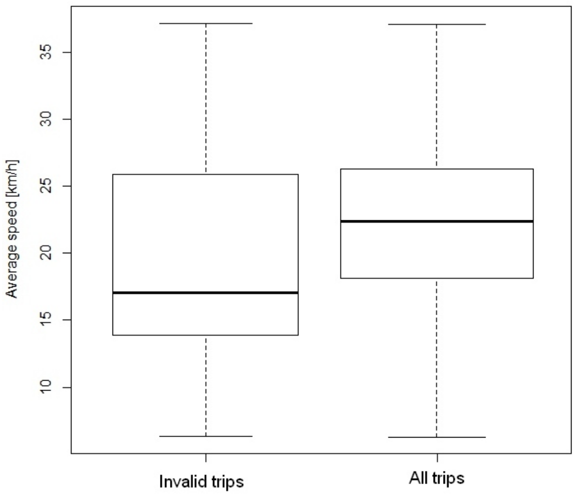

Figure 5.5: Boxplot of average speeds for invalid and valid trips.

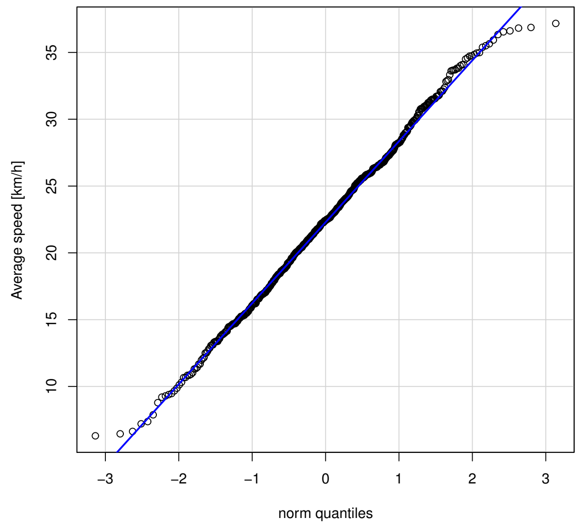

Figure 5.6: Q-Q plot for the average speed.

In order to see if the means are significantly different, a statistical hypothesis test must be conducted. In this case, all the data together (with no distinction about valid or invalid trip) are normally distributed because they follow the line on the Q-Q plot as displayed in FIGURE 5.6 so a t-test can be applied.

The null hypothesis claims that the mean value of average speed is the same for invalid and for valid trips. The t-test rejects the null hypothesis because the p-value is

$5.9 \cdot 10^{-11}$ which is smaller than 0.05. Therefore, the mean average speed of an invalid trip is significantly smaller than the mean average speed of a valid trip. The average speed is thus a cause for invalidity.

5.3.2 Stop proportion

172 trips were invalid regarding the condition on stop proportion and this accounts for almost 30% of the trips. Out of the 412 remaining trips, only 17 were invalid from another cause than the stop proportion which makes it the most demanding condition after the distance. The condition is shown in FIGURE 5.3, with the horizontal black lines representing the lower limit at 6% of the urban time and the upper limit at 30% of the urban time. All points outside this range are invalid. Unlike the average speed, both lower and upper limits have been outreached.

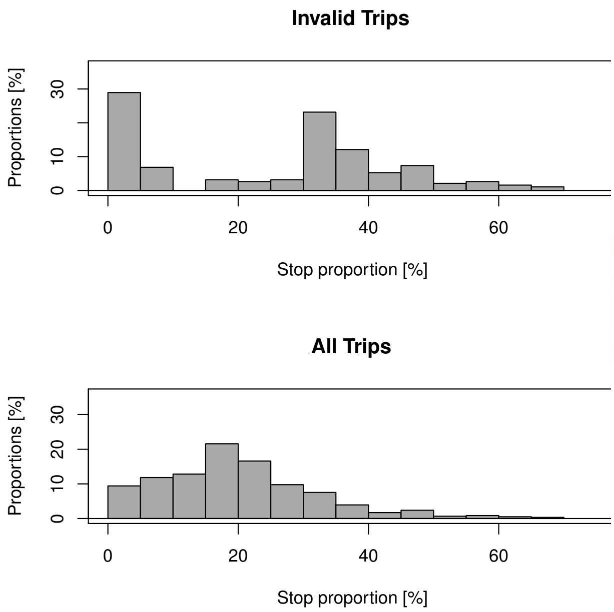

Figure 5.7: Histogram of the stop proportions for all drivers.

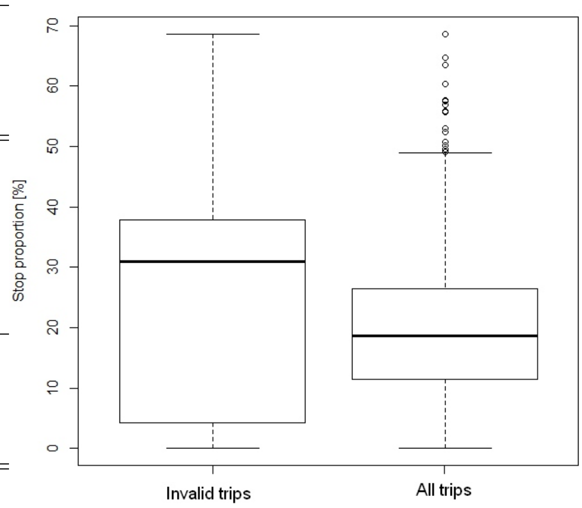

Figure 5.8: Boxplot of stop proportion for invalid and valid trips.

The histogram in FIGURE 5.7 shows that the ranges of stop proportion for which trips are mostly invalid are [0%, 5%] and around [30%, 35%]. One fifth of all the trips sits in the range [15%, 20%]. The boxplot in FIGURE 5.8 is very spread for the invalid trips, with an interquartile range of 33.55% compared to 9.46% for the valid trips only and to 14.94% for all trips together. TABLE A.5 from APPENDIX A.3 contains the detailed values of the boxplots. The outliers above 50% of stop proportion on the boxplot for all trips together prove that having a stop proportion above 50% is an exception for a normal trip.

The Q-Q plot in FIGURE 5.9 shows that the data set is not normally distributed if all the data are taken into account. For that reason, the Mann–Whitney–Wilcoxon test determines whether the means of stop proportion are the same for valid and invalid

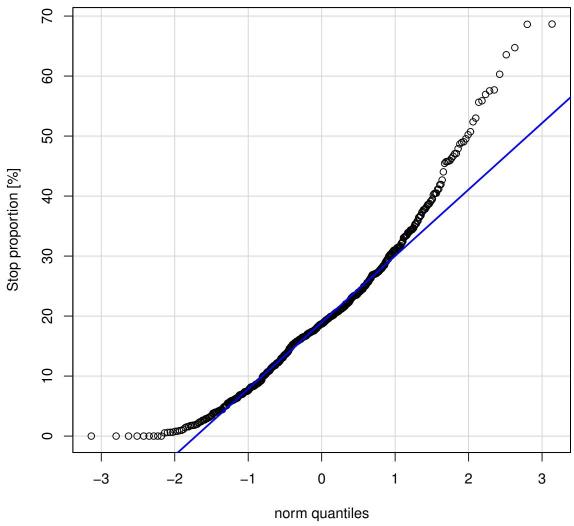

Figure 5.9: Q-Q plot for the stop proportions.

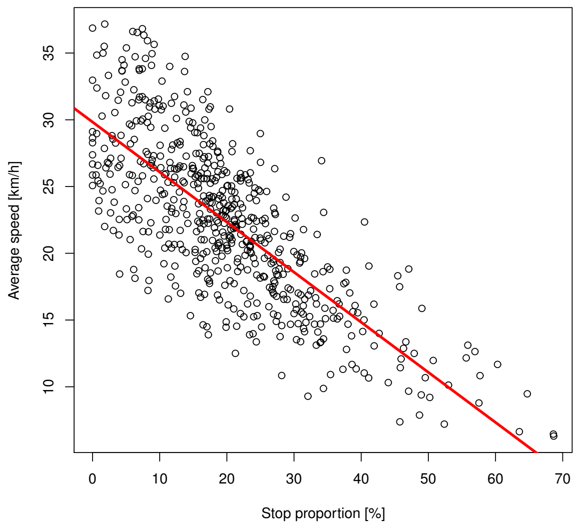

Figure 5.10: Average speed in function of the stop proportion.

trips. The p-value obtained with this test is $2.6 \cdot 10^{-6}$. Therefore the null hypothesis is rejected which implies that the means are different and, this factor being included in the computation of the overall validity, it has an impact on the overall validity of the trip. The mean of the stop proportion of the invalid trips are significantly higher than the mean stop proportion of the valid trips.

The stop proportion is obviously strongly linked to the average speed as seen in FIGURE 5.10. In fact, the red line of equation:

$$y = -0.375 \cdot x + 29.833$$

approximates the data, with x being the stop proportion and y the average speed. This explains why the results on the stop proportion and the average speed suggests some same trends.

5.3.3 Relative positive acceleration

Two trips were invalid regarding the RPA for the urban category. However, 25 invalid cases were reported in the two other categories as shown in FIGURE 5.11. The size of the dots is proportional to the distance of the trip. This permits seeing that the invalid trips in the rural part are quite small and therefore not representative. However, one trip in the urban category was indeed consequent and did not respect the condition.

The histograms in FIGURE 5.12 and the boxplots in FIGURE 5.13 with the data from the urban part have the same shape for the invalid and all trips. The accurate values of the quantiles are reported in TABLE A.6 from APPENDIX A.3 and the difference of median values between all trips and invalid trips is of 0.004 m/s^2. For both cases, more than one third of the trips stands in the range [0.2 m/s^2, 0.25 m/s^2]. A few outliers in the boxplots show that the values under 0.13 and above 0.4 are exceptional. The Q-Q plot in

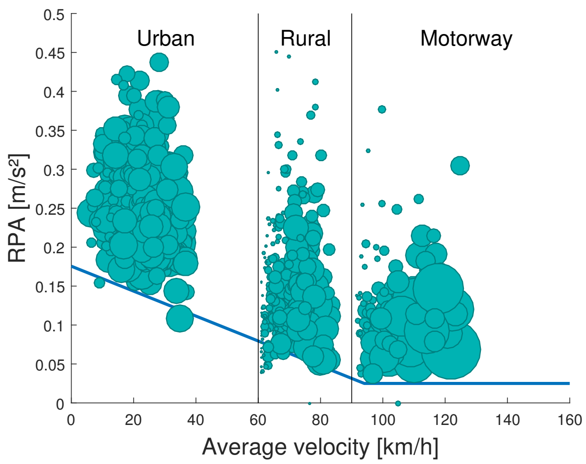

Figure 5.11: RPA in function of the average velocity for the three categories.

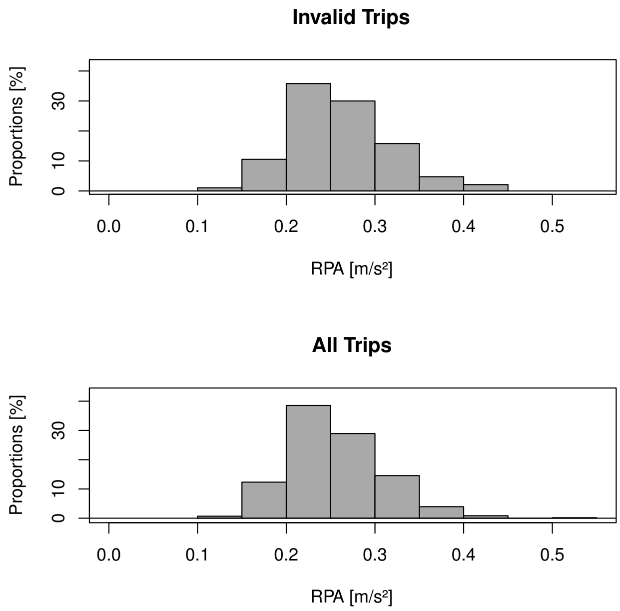

Figure 5.12: Histogram of the RPA for all drivers.

FIGURE 5.14 proves that the dataset is normally distributed which enables to conduct a *t*-test. The p-value is 0.056 which is greater than 0.05. Therefore, it cannot be stated that the null hypothesis is rejected. It means that the mean values could be the same for the valid and invalid trips regarding the RDE conditions and that the RPA is not a determining parameter regarding the overall validity. However, not rejecting the null hypothesis does not mean that it is accepted for sure.

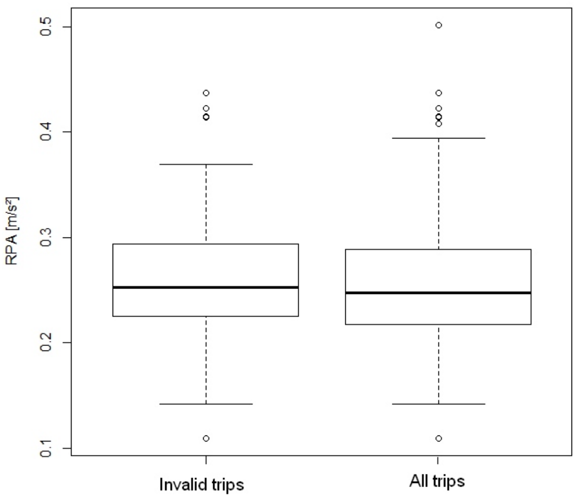

Figure 5.13: Boxplot of the RPA for invalid and valid trips.

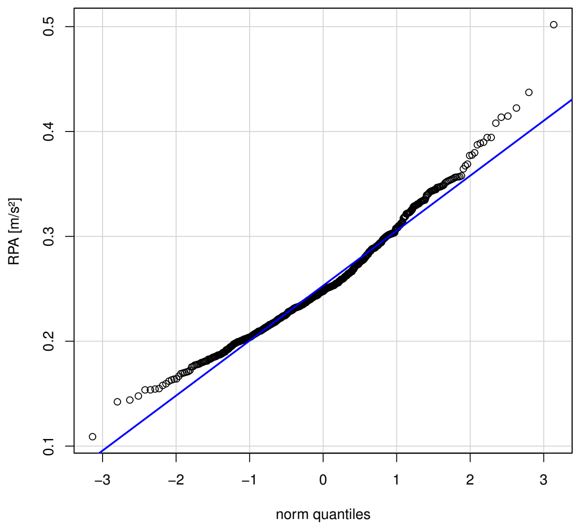

Figure 5.14: Q-Q plot for the relative positive acceleration.

5.3.4 Upper condition: $(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$

This condition has been well respected in the urban part. Indeed, all trips were valid regarding the upper dynamic condition. However, 13 trips were invalid with respect to that condition in the two other categories as shown in Figure 5.15. Yet, the size of the circles is proportional to the distance of the trips and all the invalid trips are quite small. Hence, they can be considered as not representative.

Considering the urban data only, the histograms in Figure 5.16 along with the boxplots in Figure 5.17 show that invalid trips tend to a have lower value for $(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$. The invalid trips are more spread than all trips together. One third of all the data has a value of $(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$ between 5 $\mathrm{m^2/s^3}$ and 6 $\mathrm{m^2/s^3}$. The majority of the data lies between 4 $\mathrm{m^2/s^3}$ and 6.5 $\mathrm{m^2/s^3}$ for invalid and all trips. This variable is rarely greater than 9 $\mathrm{m^2/s^3}$. The interquartile ranges are quite close, with 1.90 $\mathrm{m^2/s^3}$ for the invalid trips against 1.59 $\mathrm{m^2/s^3}$ for valid trips and 1.65 $\mathrm{m^2/s^3}$ for the whole set trips together as reported in Table A.7 from Appendix A.3.

Figure 5.18 proves that the data are normally distributed and that a *t*-test can be applied. The p-value is $1 \cdot 10^{-3}$ which is smaller than 5%. Therefore, the null hypothesis is rejected. The $(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$ values are thus different regarding the overall validity of the trip and the mean value for the valid trip is significantly higher than the value for the invalid trips. This could seem counter intuitive knowing that all the trips were valid regarding the $(\mathrm{v} \cdot \mathrm{a_{pos}})_k[95]$ factor. However, the test takes all the real values of the variables into account whereas the overall validity has been computed based on only two possible numbers (0 or 1) which are the binary values stating if the condition was fulfilled or not.

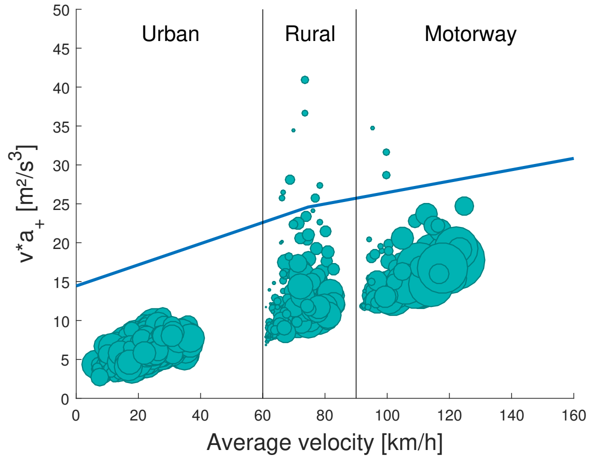

Figure 5.15: $(\mathrm{v} \cdot \mathrm{a}_{\mathrm{pos}})_k[95]$ in function of the average velocity for every category.

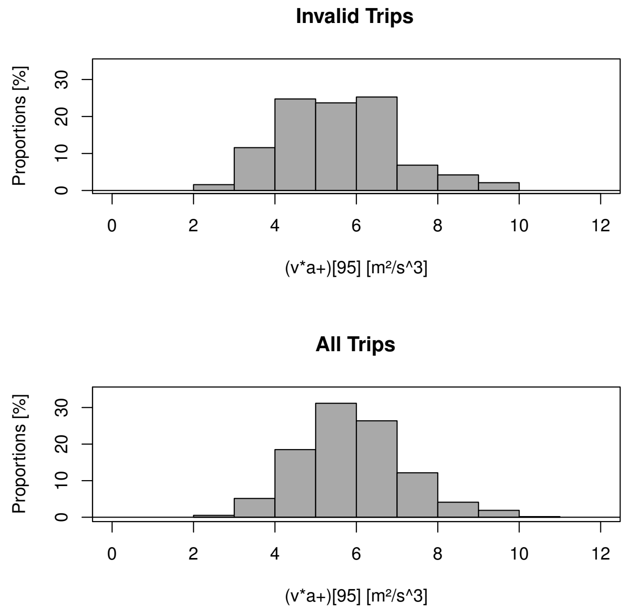

Figure 5.16: Histogram of $(\mathrm{v} \cdot \mathrm{a}_{\mathrm{pos}})_k[95]$ for all drivers.

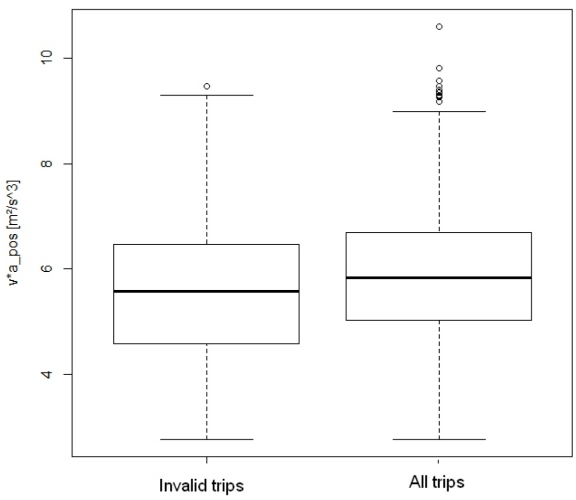

Figure 5.17: Boxplot of $(\mathrm{v} \cdot \mathrm{a}_{\mathrm{pos}})_k[95]$ for invalid and valid trips.

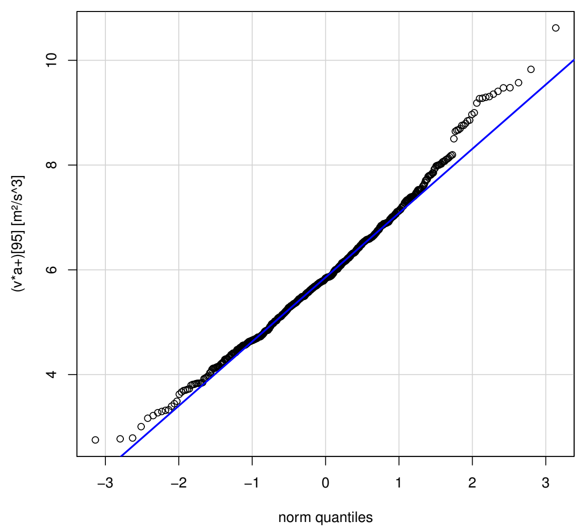

Figure 5.18: Q-Q plot for $(\mathrm{v} \cdot \mathrm{a}_{\mathrm{pos}})_k[95]$.

5.3.5 Maximal speed

Since the urban category only covers speeds up to 60 km/h, the maximal speed of each trip, all classes included, has been taken in this case. These speeds are represented in Figure 5.19. All points of speed greater than 160 km/h and all trips with more than 3% of motorway speed greater than 145 km/h led to the total invalidity of the trip regarding the RDE methodology. For this study, only one driver outreached once the 160 km/h (represented by the green dotted line). Two exceeded 145 km/h (represented by the blue line) but only for a short time so, since they had less than 3% of the motorway time above 145 km/h, those two trips are still considered as valid.

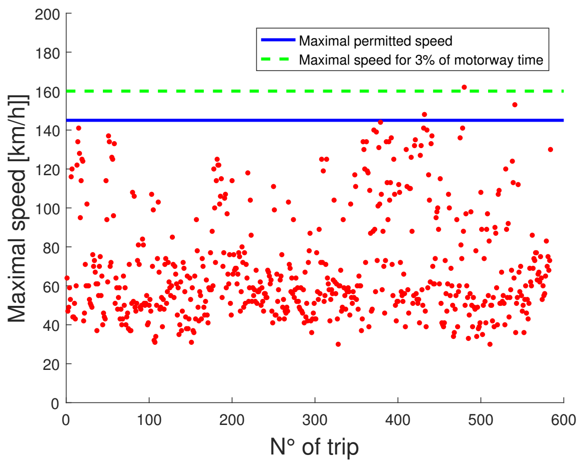

Figure 5.19: Maximal speed for each trip.

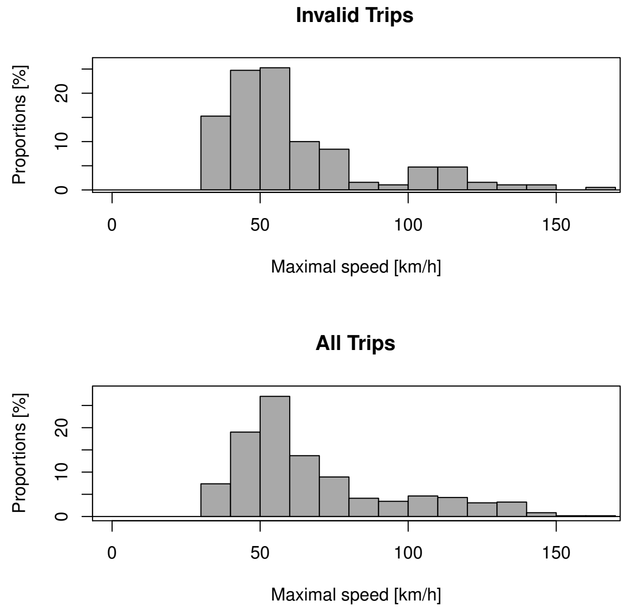

Figure 5.20: Histogram of the maximal speeds for all drivers.

The histograms in FIGURE 5.20 show that almost 60% of the invalid trips and almost half of all the trips have a maximal speed lower than 60 km/h. The boxplots in FIGURE 5.21 show that the invalid trips slightly cover a smaller range of maximal speed than all trips together, with an IQR of 25 km/h compared to 27 km/h for all trips together. TABLE A.8 from APPENDIX A.3 reports all quantiles for invalid, valid and all trips. The mean value for the valid trips is bigger of 10 km/h than the mean value of the maximal speed for the invalid trips. This is explained by the fact that trips with smaller maximal speed tend to have lower overall speed so the condition on the average speed or stop proportion have more chance to be unfulfilled and this would lead to the overall invalidity of the trip.

The Q-Q plot in FIGURE 5.22 shows that the data have a logarithmic shape. Therefore two options arise, either the Mann–Whitney–Wilcoxon test is conducted, either the logarithm of the data is taken and if it is normally distributed then a *t*-test can be applied on the new dataset. In this case, the results are not better by transforming the data as displayed in FIGURE A.1 in APPENDIX A.4. Thus, Mann–Whitney–Wilcoxon test is used and the p-value computed is $6 \cdot 10^{-8}$ which rejects H_0. Therefore, the mean values of maximal speed are different regarding the overall validity. The maximal speed is significantly higher for valid trips.

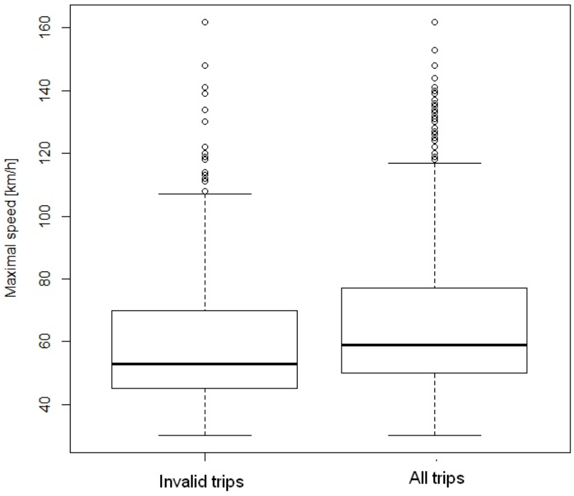

Figure 5.21: Boxplot of maximal speeds for invalid and valid trips.

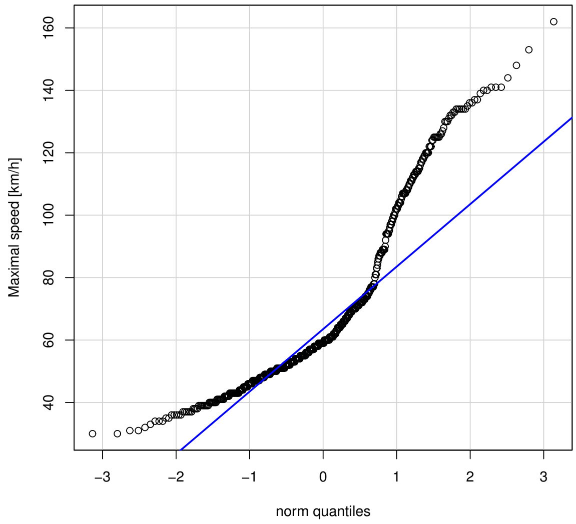

Figure 5.22: Q-Q plot for the maximal speed.

5.3.6 Distance

The distance of each trip has been computed for each category. The distance shown in FIGURE 5.23 is solely the distance driven in the urban category. Only 14 trips were valid regarding the 16 km required. This condition has not been included in the overall validity since it is obviously too constraining.

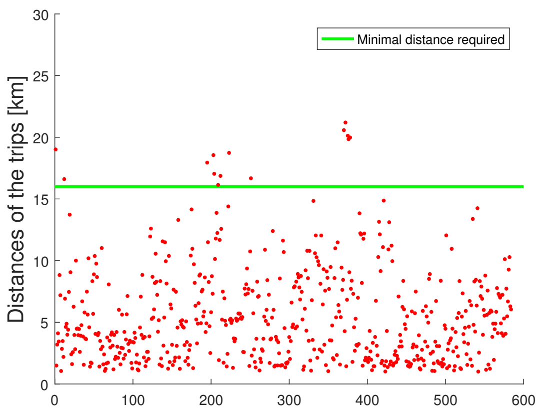

Figure 5.23: Distance of each trip in the urban category.

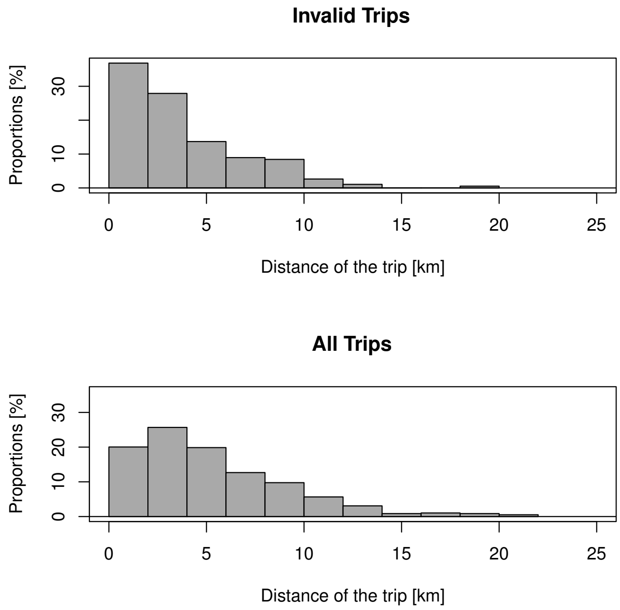

Figure 5.24: Histogram of the distances for all drivers.

Figure 5.24 proves that most of invalid trips are small trips in term of distance. Indeed, around seventy percent of the trips among the invalid trips are smaller than five kilometres. This could be due to the fact that small trips have more often lower average speed which leads to the overall invalidity of the trip. The boxplots in Figure 5.25 show that the range of distances covers less data points for the invalid trips than for all of them. A trip of more than 15 km is quite exceptional in the urban category and is considered as an outlier. The mean distance of a trip ridden in Brussels is 5.39 km and the median is 4.48 km as reported in Table A.9 from Appendix A.3. This is far from the RDE constraint. This is obviously not representative at all.

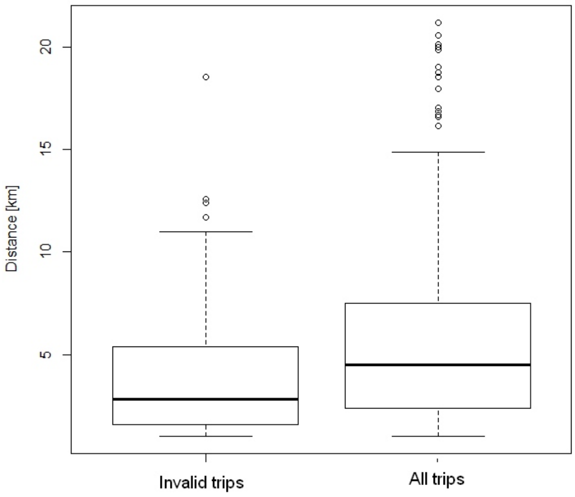

Figure 5.25: Boxplot of distances for invalid and valid trips.

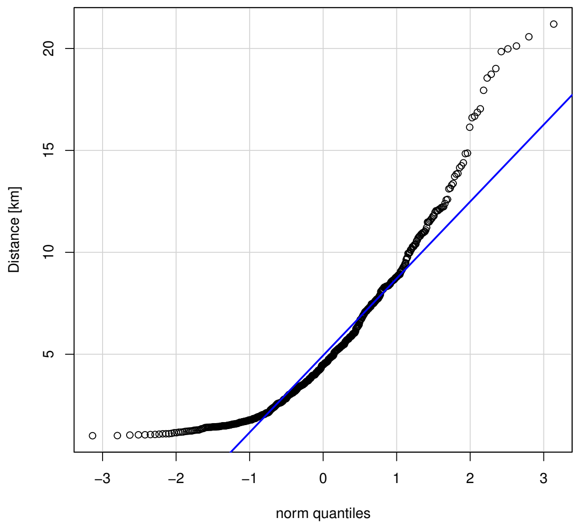

Figure 5.26: Q-Q plot for the distance.

The Q-Q plot in Figure 5.26 shows that the data are not normally distributed. However, by taking the logarithm of the distance, the shape of the curve looks more like a normal distribution as shown in Figure A.2 in Appendix A.4. Thus, a t-test on the logarithm of the data points is possible. The t-test is more powerful but transforming the data can raise the question of representativeness. Indeed, a graph with a logarithmic scale gives a very different sight of the data, because if the points are slightly apart from each other on a graph with a logarithmic scale they could actually be far away from each other. Thus, for the further analyses in Section 5.4, the Mann–Whitney–Wilcoxon test is used.

The p-value resulting from the t-test with the logarithm of the distances is $2.9 \cdot 10^{-15}$. The Mann–Whitney–Wilcoxon test gives a p-value of $1.1 \cdot 10^{-14}$. Thus the mean of the distance is significantly higher for valid trips than invalid trips. However, since the criterion on the distance has not been included in the computation of the overall validity, it cannot be said regarding the statistics that this factor alone has a significant influence on the overall validity. However, it has an impact due to its link with the average speed.

5.3.7 Duration

No condition is imposed on the duration of the trip in the RDE methodology. However, it is interesting to see how this parameter is situated with respect to the overall validity. Figure 5.27 shows the duration for each trip. The 584 trips are represented in the urban category. Then, some of these 584 trips had values of speed entering the other categories so they are displayed on the graph as well. These points are much closer to the x-axis than the ones in the urban class. It proves that the duration of the data points for those categories is smaller than the duration in the urban category. This is normal since the data come from Brussels.

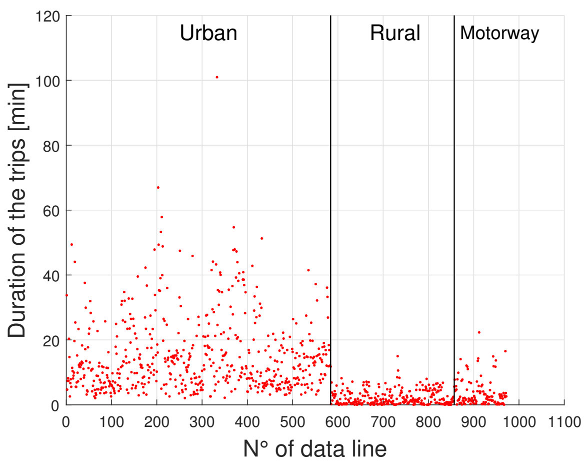

Figure 5.27: Duration of each trip for each category.

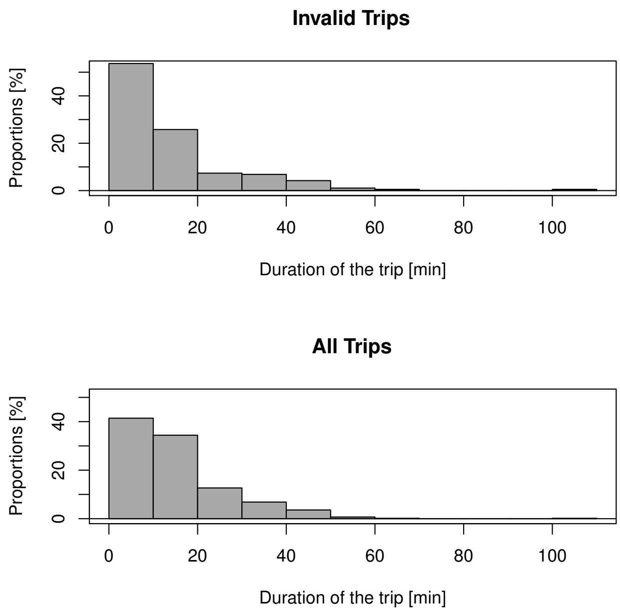

Figure 5.28: Histogram of the duration for all drivers.

The histograms in Figure 5.28 show that 80% of the trips last less than 30 minutes. 50% of the invalid trips last less than 10 minutes. This encourages the conclusion drawn before. The median value is 9 minutes driven for invalid trips whereas it reaches 13 minutes for the valid trips. The valid trips tend to last longer than the invalid ones as shown in Figure 5.29. Considering all trips together, the mean duration is 15 minutes and the median one almost reaches 12 minutes as reported in Table A.10 from Appendix A.3. All trips lasting more than 40 minutes are considered as exceptions.

The dataset does not follow a normal distribution as shown in Figure 5.30. Yet, the logarithm of the dataset is normally distributed as shown in Figure A.3 in Appendix A.4. The *t*-test on the logarithm of the data gives a p-value equal to $2 \cdot 10^{-4}$ and the Mann–Whitney–Wilcoxon test gives a p-value of $4 \cdot 10^{-5}$. Again, the null hypothesis is rejected and the mean duration is significantly higher for valid trips than for invalid trips.

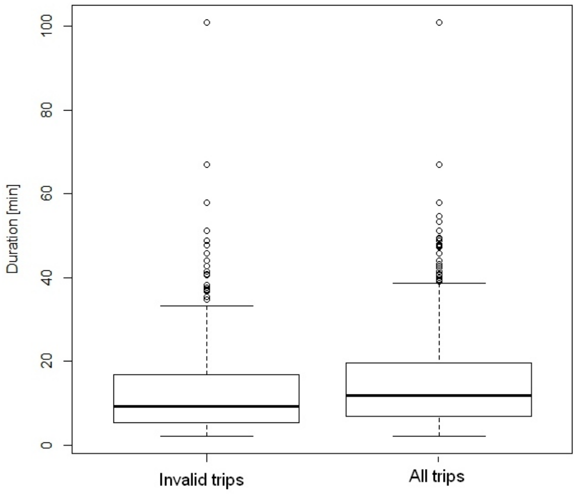

Figure 5.29: Boxplot of the duration for invalid and valid trips.

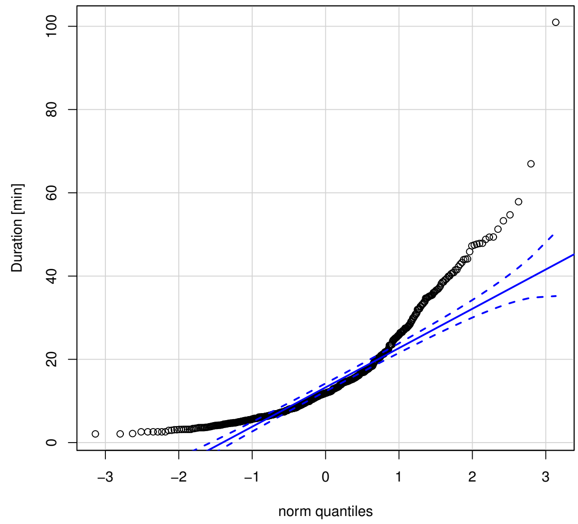

Figure 5.30: Q-Q plot for the duration.

5.4 Statistics on categorical variables

Descriptive statistics In this section, descriptive statistics defined in Section 5.3 are applied on categorical variables. In this case, a bivariate analysis is conducted with contingency tables. The goal is to describe the relationship between the variable of interest and the overall validity of the trip regarding the RDE conditions. A contingency table is a representative method of frequency distributions of variables. In this case, the variable of interest (e.g., the fuel type) is analysed along with the overall validity of the trip so the table displays the amount of valid and invalid trips for this criterion. This number divided by the total amount of trips gives proportions that can be seen as probabilities since the sum of all proportions is equal to 100%. If the proportions of different rows of the contingency table do not significantly vary between columns, the parameter studied is independent of the overall validity of the trip and this can be proven with a statistical hypothesis test.

Statistical hypothesis test The Pearson's Chi-squared test is associated to the contingency table to assess if conclusions can be drawn about the role played by the categorical variables in the overall validity of the trip. This statistical test, applicable for independent data from large samples, evaluates the probability that the differences between the sets are random. In the case of the Pearson's Chi-squared test, the null hypothesis states that the frequency distribution of the data did not arise by chance and follows a certain theoretical distribution. In other words, H_0 affirms that the overall validity of the test is independent of the categorical variable studied whereas H_1 specifies they are dependent. Thus, if the p-value is smaller than 0.05, the two factors studied are dependent and if not they might be independent. This could thus bring to light the variables having an impact on the overall validity of the test.

Statistical hypothesis test on the parametric variables The influence of each categorical parameter on the mean of the parametric variables has also been studied. This pushes the analysis further and permits seeing if those categorical factors have an influence on the variables studied in the RDE. This is interesting to understand if the variables studied by the RDE are relevant and impacted by many factors or not. A *t*-test or a Mann–Whitney–Wilcoxon test has been conducted for each quantitative variables. The choice of test for the parametric variables has been made based on their Q-Q plot in Section 5.3. Indeed, if data of the variable were normally distributed in Section 5.3, a *t*-test has be conducted and if not, a Mann–Whitney–Wilcoxon test has be used. The p-value has thus been computed and if it is smaller than 0.05, the factor has an influence on the mean of the variable studied. This is also presented in this chapter.

5.4.1 Links between the categorical variables

Before studying the impact of the external factors, it is important to see the potential relationships between them. Indeed, the amount of cars studied is very small so this study is not representative of the Brussels car fleet nor of the typical drivers. It should be conducted on thousands of people and the proportion of brands, type of car, type of fuel and of all other factors should be respected.

A Pearson's Chi-squared test has been conducted on the tables of contingency relating all categorical factors with respect to each other. The p-values obtained are reported in Table A.11 from Appendix A.5. Table 5.3 shows that the intrinsic parameters related to the cars are all linked together. The "V" in the blue cells means that the factors are significantly related and the "X" in the grey cells represents the opposite. These influences should be mitigated with a larger car population.

Given the p-values, the fuel type, the class of year, the class of power, the class of displacement and whether the car is a company car are the most related factors. Some links are instinctive such as the relationship between the weekend or week and the peak or off-peak hours. But some are unusual and mostly due to the small car population such as the relationship between the power of the car and whether the trip was at the weekend.

Table 5.3: Summary of the influences between categorical factors.

	Fuel	Company/private car	Displacement	Year	Power	Gearbox	Cold start	Week/weekend	Peak/Off-peak
Fuel		V	V	V	V	X	V	X	X
Company/private car	V		V	V	V	V	X	V	X
Displacement	V	V		V	V	V	X	X	X
Year	V	V	V		V	X	V	V	X
Power	V	V	V	V		V	X	V	X
Gearbox	X	V	V	X	V		X	X	X
Cold start	V	X	X	V	X	X		V	X
Week/weekend	X	V	X	V	V	X	V		V
Peak/Off-peak	X	X	X	X	X	X	X	V	

5.4.2 Type of fuel

Out of the thirty-two cars useful to the study, fifteen have a Diesel engine and seventeen ran with a gasoline engine. Each trip has been assigned to the data of the car used and the amount of trips regarding the type of fuel is given in Table 5.4. 47.6% of the trips were driven by a car powered by a Diesel engine.

Table 5.4: Contingency table for the type of fuel.

Type of fuel	Valid trips	Invalid trips	Total
Diesel	97	181	47.6%
gasoline	93	213	52.4%
Total	32.5%	67.5%	100%

The p-value resulting from the Pearson's Chi-squared test is 0.25. Since it is greater than 5%, it is not possible to reject the null hypothesis. This means that there is no apparent link between the overall validity of the trip and the fuel of the vehicle.

However, the impact of fuel has been analysed on the seven quantitative parameters in TABLE 5.5.

Table 5.5: Impact of the type of fuel on the quantitative parameters.

Variable	p-value	Comment
Average speed	$1.2 \cdot 10^{-7}$	The average speed is significantly smaller for Diesel engines (20.97 km/h against 23.60 km/h).
Stop proportion	$1.5 \cdot 10^{-4}$	The stop proportion is significantly smaller for gasoline cars (18.16 % for gasoline cars and 21.94% for Diesel ones).
RPA	0.26	The RPA is not significantly affected by the type of fuel.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k$**[95]**	0.18	$(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is not affected by the type of fuel.
Maximal speed	0.18	The type of fuel does not significantly change the maximal speed.
Distance	$1.8 \cdot 10^{-3}$	The distance is considerably impacted by the type of fuel of the car (4.78 km on average for gasoline cars and 6.07 km for Diesel ones).
Duration	$9.9 \cdot 10^{-7}$	The trips with a gasoline car last 12 minutes on average compared to trips with a Diesel engine which last 18 minutes on average.

The p-value has been calculated thanks to the statistical tests used. The result is that the fuel has an influence on the average speed, the stop proportion, the distance and the duration of the trips. The impact on the distance and the duration of the trips can be explained by the fact that the Diesel engine is said to be unsuited for small trips or urban trips since the engine takes more time to heat up. Therefore, most of drivers tend to buy a Diesel car if their goal is to drive bigger distances. However, the differences in average speed and stop proportion are hardly explainable. The results would suggest that the vehicles with a Diesel engine spend more time in the traffic jam or at red lights. Following the previous point, one reason could be that the drivers of a gasoline car bought it because they live in the city but they perhaps go to work with public transport. Therefore, they do not use their car as much in the traffic jams. In fact, 5070 minutes were recorded in the Diesel cars and 3826 minutes in gasoline cars. This proves thus that most time is spent in Diesel cars in the scope of this study.

5.4.3 Company - private car

There are two types of company cars: the ones used as salary and the ones used for the professional activity. The distinction has not been made in this research but the results might vary from one to the other.

In this case, seventeen cars came from the driver's company whereas the fifteen others were private. The contingency table summarising the amounts of trips for each category is shown in TABLE 5.6. The p-value from the Chi-squared test is 0.07 so the null hypothesis cannot be rejected. There is thus no link between the overall validity of the trip and whether the vehicle is a company car.

Table 5.6: Contingency table for the type of car usage.

Type of car	**Valid trips**	**Invalid trips**	**Total**
Company car	118	213	56.7%
Private car	72	181	43.3%
Total	32.5%	67.5%	100%

The impact of the function of the car has been looked at for each parameter in TABLE 5.7.

Table 5.7: Impact of the car function on the quantitative parameters.

Variable	**p-value**	**Comment**
Average speed	$1.1 \cdot 10^{-4}$	The average speed is significantly smaller for company cars (21.50 km/h on average for company cars and 23.46 km/h for private cars).
Stop proportion	$3 \cdot 10^{-3}$	The stop proportion is significantly smaller for private cars (18% against 21%).
RPA	$1.2 \cdot 10^{-4}$	The RPA for the private cars is significantly smaller than the RPA for the company cars (0.245 m/s^2 versus 0.262 m/s^2).
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[\mathbf{95}]$	0.587	$(v \cdot a_{pos})[95]$ is not significantly affected by the function of the car.
Maximal speed	0.55	The maximal speed is not significantly impacted by whether the vehicle is a company car.
Distance	0.08	The distance is not impacted by the function of the vehicle.
Duration	$5.3 \cdot 10^{-3}$	The drivers with a company car tend to stay longer in their vehicle than the other (trips of 13 minutes on average for private cars and 17 minutes on average for company cars).

This table shows that the average speed, the stop proportion, the RPA and the duration of the trips are influenced by the function of the car. The drivers with a company car tend to drive slower and have bigger stop proportion than the people with their own car. The duration is impacted and not the distance which comforts the idea that company cars are more often stuck in traffic or idling than the private cars. Along with the explanation for the impact of the type of fuel on the average speed and stop proportion, the reason of this difference could be that people with a company car tend to use it to go to work (thus during peak hours) whereas people with their own car may use public transports or other means of transport. Since the average speed is lower and the stop proportion is bigger for company cars, it makes sense that the duration of these trips tend to be larger than for private cars.

The RPA tend to be lower for company cars which means that they have the trend to be more passive than the private cars but this can be linked with the fact that they ride slower more often.

5.4.4 Displacement

As explained before, the displacement has been divided in two categories. The first category comprises cars with a displacement from 898 cm^3 to 1461 cm^3 and the second category from 1499 cm^3 to 1995 cm^3. There are 16 cars in each category of displacement. The contingency table is presented in TABLE 5.8.

Table 5.8: Contingency table for the displacement of the car.

Displacement range [cm^3]	**Valid trips**	**Invalid trips**	**Total**
1st class of displacement [898 cm^3 - 1461 cm^3]	91	199	49.7%
2nd class of displacement [1499 cm^3 - 1995 cm^3]	99	195	50.3%
Total	32.5%	67.5%	100 %

The p-value is 0.55, thus the null hypothesis cannot be rejected and the overall validity of the trips is independent of the displacement of the car.

The statistical hypothesis tests show in TABLE 5.9 that the displacement influences the average speed, the stop proportion, the RPA, the maximal speed and the duration.

The influence of the displacement on the average speed, the stop proportion and the duration is hard to understand. TABLE 5.3 shows that the displacement of the car and the function of the vehicle (company or private car) were linked. Indeed, most of company cars were in the second category of displacement. The car fleet of this study being small, the influence of the displacement on the three variables could be explained by the close relationship between the displacement and whether the vehicle is a company car.

Concerning the influence of the displacement on the RPA, it can be explained by the fact that the bigger the displacement, the bigger the torque produced and thus the better the acceleration. Thus, the RPA is bigger for higher displacement. However, the upper dynamic condition is not affected by the change of displacement. This might be explained by the fact that for this condition, the 95th percentile is taken so the accelerations that are higher and that could make the difference are not taken into account. Most of the time, cars with a higher displacement produce a higher power and thus higher speeds can be reached. This would explain the influence of displacement on the maximal speed.

Table 5.9: Impact of the displacement on the parametric variables.

Variable	p-value	Comment
Average speed	0.03	The average speed is significantly smaller for cars with a higher displacement (21.81 km/h for the category [1499 cm^3 ; 1995 cm^3] and 22.89 km/h for the other).
Stop proportion	$6 \cdot 10^{-3}$	The stop proportion is significantly smaller for cars with a smaller displacement (18.7% against 21%).
RPA	$2 \cdot 10^{-3}$	The RPA is significantly smaller for cars with a smaller displacement (0.248 m/s^2 versus 0.261 m/s^2).
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	0.31	$(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is not significantly affected by the class of displacement.
Maximal speed	0.02	The mean value of maximal speed is significantly smaller for cars with a smaller value of displacement (66.44 km/h compared to 70.92 km/h for cars in the second category of displacement).
Distance	0.10	The displacement of the car does not considerably impact the distance ridden.
Duration	0.02	The duration of the trips is significantly smaller for cars with a lower displacement (14 minutes for the first class of displacement compared to 16 minutes for engines with a higher displacement).

5.4.5 Year of construction

Due to the small variety in terms of construction years and for the purpose of the hypothesis tests, only two categories were made for this factor. The first category of years comprises vehicles from 2008 to 2016 and the second one contains cars from 2017 to 2019. The p-value resulting from the Pearson's Chi-squared test made on the contingency table in Table 5.10 is 0.74. No link is apparent between the overall validity of the trip and the year of construction.

Table 5.10: Contingency table for the year of construction of the car.

Year range	Valid trips	Invalid trips	Total
1st class of year [2008-2016]	97	207	52.1 %
2nd class of year [2017-2019]	93	187	47.9 %
Total	32.5 %	67.5 %	100 %

Table 5.11 shows that the year of construction is a significant factor for many aspects. Indeed, it influences the average speed, the stop proportion, the RPA, the 95th percentile of $(\mathrm{v} \cdot \mathrm{a_{pos}})$ and the duration of the trips. Concerning the influence of the year on the

Table 5.11: Impact of the year of construction of the car on the parametric variables.

Variable	p-value	Comment
Average speed	$8.9 \cdot 10^{-4}$	The average speed is significantly smaller for more recent cars (21.48 km/h against 23.15 km/h).
Stop proportion	0.01	The stop proportion is significantly smaller for older cars (18% vs 21%).
RPA	$1 \cdot 10^{-3}$	The RPA is significantly smaller for more recent cars (0.247 m/s^2 compared to 0.261 m/s^2 for cars constructed before 2017).
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	$1.1 \cdot 10^{-3}$	The mean value of $(v \cdot a_{pos})[95]$ is significantly smaller for more recent cars with a value of 5.75 m^2/s^3 compared to the value of 6.07 m^2/s^3 computed for cars older than 2017.
Maximal speed	0.06	The year of construction of the vehicle does not influence the maximal speed.
Distance	0.30	The age of the car is not a significant factor for the distance driven on road.
Duration	0.03	The duration of cars older than 2017 tends to be smaller than for more recent cars (14 minutes on average for old cars and 16 minutes for more recent ones).

average speed, the stop proportion and the duration, as for the case of the displacement, it can be explained by hidden links between variables. Indeed, TABLE 5.3 proves that there is a significant link between company cars and the year of construction. In fact, company cars are often replaced and most of them were indexed in the category of recent vehicles for this study. Concerning the dynamic conditions, since they both are lower for more recent cars, it would mean that drivers of new cars are less aggressive than drivers of the old ones, or that they drive in conditions that do not allow to be aggressive.

5.4.6 Power

The cars are divided in two categories regarding the power of their engine: the first category is from 51 kW to 80 kW and the second one from 81 kW to 140 kW. Sixteen cars are reported in each category. The division in categories and the fact that the analyses are made based on those categories explain the differences between the results concerning the power and the displacement of the car. The contingency table is shown in TABLE 5.12. The p-value is 0.12. Therefore, the null hypothesis is not rejected and the power is not linked to the overall validity of the trip. It is important to bear in mind that this is a conclusion based on the two categories made. The choice of two categories has been made to be able to conduct statistical hypothesis tests but by considering other classes of

Table 5.12: Contingency table for the power of the car.

Power range	Valid trips	Invalid trips	Total
1st class of power [51 kW - 80 kW]	99	178	47.4%
2nd class of power [81 kW - 140 kW]	91	216	52.6%
Total	32.5%	67.5%	100%

power, this factor could have an effect.

TABLE 5.13 expresses that the power of the vehicle has a significant influence on the stop proportion, the maximal speed, the distance and the duration of the trips. The

Table 5.13: Impact of the power of the engine on the parametric variables.

Variable	p-value	Comment
Average speed	0.13	The average speed is not significantly influenced by the power of the car.
Stop proportion	0.03	The mean of the stop proportion is significantly smaller for cars with less powerful engine (19.12% vs 20.72%).
RPA	0.31	The influence of the class of power of the engine has no significant effect on the RPA.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	0.99	The $(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is not significantly influenced by the class of power of the engine.
Maximal speed	$6 \cdot 10^{-3}$	The vehicles with a smaller power engine tend to reach a smaller maximal speed (mean value: 65.51 km/h) than the vehicles with a bigger power (mean value of maximal speed: 71.57 km/h).
Distance	$5.9 \cdot 10^{-3}$	The distance is significantly smaller for less powerful cars (4.81 km on average) compared to cars in the second class of power (5.92 km).
Duration	$2.1 \cdot 10^{-3}$	The duration is significantly smaller for less powerful cars (14 minutes on average) compared to 17 minutes ridden by cars in the second category of power.

power influences the stop proportion and the duration but not the average speed. However, those three parameters are strongly linked and especially the average speed and the stop proportion. The power of the vehicle should not significantly affect the stop proportion since this parameter is more related to the traffic rules and conditions and not the car itself. However, as explained before, there are be other latent bonds due to the small amount of vehicles covering the study. Some bonds are stronger than others which would explain why the average speed is not affected. Concerning the distance travelled, the results state that more powerful cars tend to drive longer distances. This might be associated to the fuel type, since it has a noticeable influence on the distance of the trip. Indeed, the Pearson's Chi-squared test on the contingency table comparing the type of fuel and the power of the vehicle proves these two factors are linked in this study (see TABLE 5.3). Therefore, this influence is more dictated by the type of fuel and the displacement than by the power of the car itself. Finally, the bigger the power, the higher attainable speeds which is well proven by the results.

5.4.7 Type of gearbox

From the car fleet, 10 vehicles were equipped with an automatic gearbox and 22 with a manual one. The contingency table is shown in TABLE 5.14.

Table 5.14: Contingency table for the type of gearbox.

Type of gearbox	**Valid trips**	**Invalid trips**	**Total**
Automatic	75	142	37.2%
Manual	115	252	62.8%
Total	32.5%	67.5%	100%

The p-value is 0.42. The null hypothesis cannot be rejected. There is thus no noticeable link between the overall validity of the trip and the type of gearbox.

The influence of the type of gearbox on the parametric variables is discussed in TABLE 5.15.

The results are that the type of gearbox does not considerably influence the conditions except for the maximal speed. In this study, the type of gearbox is strongly linked with the power of the car as seen in TABLE 5.3 so this conclusion can come from this close relationship.

Table 5.15: Impact of the type of gearbox on the parametric variables.

Variable	p-value	Comment
Average speed	0.11	The average speed is not significantly influenced by the gearbox of the car.
Stop proportion	0.31	The stop proportion is not significantly affected by the type of gearbox.
RPA	0.70	The effect of the type of gearbox on the RPA is not consequent enough.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	0.18	The $(v \cdot a_{pos})[95]$ does not considerably change with the type of gearbox.
Maximal speed	$1.6 \cdot 10^{-3}$	The maximal speed is significantly smaller for manual cars (64.73 km/h compared to 75.41 km/h for automatic gearbox).
Distance	0.61	The type of gearbox is not a key factor influencing in the distance ridden.
Duration	0.88	The duration is not influenced by the type of the gearbox.

5.4.8 Cold start

All cars begin the RDE test procedure with a cold engine. Therefore, it is interesting to see the proportion of trips with a cold start. Cold start period is considered for values of coolant temperatures smaller than 70°C. In total, 418 trips started with a cold engine which is equivalent to 75.9% of the total amount of trips. The p-value from the contingency table in TABLE 5.16 is 0.24 which does not allow to draw any conclusion and no link has been found between the temperature of the engine and the overall validity of the trip.

Table 5.16: Contingency table for the cold start.

Cold start	Valid trips	Invalid trips	Total
Warm start	60	106	28.4%
Cold start	130	288	71.6%
Total	32.5%	62.5%	100%

TABLE 5.17 analyses the effect on the variables computed for the RDE test. The results show that the coolant temperature affects the average speed, the upper dynamic condition, the maximal speed, the distance ridden and the duration of the trips. All these variables are smaller for hot engine. It would be interesting to know in which conditions the drivers were at the moment they started their engine again. For instance, if the driver is stuck in the traffic jam and decides to stop the engine, a new trip is recorded when he starts it again. Therefore, if all those trips were recorded in these conditions and if they

Table 5.17: Impact of the cold start on the parametric variables.

Variable	p-value	Comment
Average speed	0.02	The average speed is significantly smaller for trips beginning with a hot engine (21.38 km/h for engine with a temperature greater than 70°C against 22.73 km/h for cold engines).
Stop proportion	0.32	The stop proportion is not significantly affected by the fact that the car started with a cold or hot engine.
RPA	0.57	The RPA is not considerably influenced by whether the car started with a cold engine.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	$6.9 \cdot 10^{-3}$	The value of $(\mathrm{v}{\cdot}\mathrm{a_{pos}})[95]$ is significantly smaller for cars starting with the temperature of the engine greater than 70°C (6.01 m^2/s^3 compared to 5.68 m^2/s^3 for the cold start).
Maximal speed	$1.2 \cdot 10^{-6}$	The maximal speed reached is considerably smaller for cars with a hot engine (59.76 km/h compared to 72.25 km/h for cold start).
Distance	$3.9 \cdot 10^{-7}$	The trips with the engine already hot are significantly smaller than the ones with a cold engine (4.24 km and 5.85 km respectively).
Duration	$3.8 \cdot 10^{-5}$	The duration of the trips are significantly smaller for trips starting with hot engine compared to cold start engine (12 minutes and 16 minutes respectively).

are longer than 1 km, they are kind of "sub" trips and it is thus normal that all these variables are smaller.

5.4.9 Time of the week

The RDE methodology only considers trips during the week as seen in Section 2.4.1. Therefore, it is interesting to see the proportion of trips ridden during the week and at the weekend. In total, 443 trips happened during the week and 141 at the weekend. Table 5.18 shows the contingency table associated. The p-value is 0.42. The variable might thus

Table 5.18: Contingency table for the time of the week.

Time of the week	Valid trips	Invalid trips	Total
Week	148	295	75.9%
Weekend	42	99	24.1%
Total	32.5%	67.5%	100%

be independent of the overall validity of the trip.

Table 5.19: Impact of the time of the week on the parametric variables.

Variable	p-value	Comment
Average speed	$1.5 \cdot 10^{-8}$	The average speed is significantly smaller during the week (21.5 km/h during the week compared to 24.8 km/h at the weekend).
Stop proportion	$1.8 \cdot 10^{-4}$	The stop proportion is significantly smaller for trips at the weekend (16.44%) compared to trips during the week (21.08%).
RPA	0.12	The RPA is not significantly affected by the time of the week.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	$1.5 \cdot 10^{-3}$	The value of $(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is significantly smaller during the week (5.80 m^2/s^3) compared to the weekend (6.28 m^2/s^3).
Maximal speed	0.23	The maximal speed is not influenced by the day of the week.
Distance	0.22	The distance driven is not influenced by the time of the week.
Duration	$1.5 \cdot 10^{-3}$	The trips at the weekend last 12 minutes on average compared to 16 minutes during the week.

Table 5.19 shows the influence of the time of the week on the parameters computed for the RDE test. The average speed, the stop proportion, the upper dynamic boundary and the duration of the trip are different during the week compared to the weekends. The average speed is smaller during the week whereas the stop proportion and the duration are higher and this is logical since more cars are on the roads during the week. The influence on the upper dynamic condition suggests that people have more peak accelerations at the weekend. This could make sense, since less people circulate on the roads, the driver can afford to accelerate at higher pace.

5.4.10 Time of the day

Off-peak and peak hours are not analysed in the RDE test but it might be interesting to look at that factor to see if some relations can be extracted. In total, 315 trips were ridden during peak hours whereas 269 were driven off-peak as seen in Table 5.20. The p-value is 0.93. There is no link between the peak hours and the overall validity of the trip. This is quite surprising because it could have been thought that the trips during peak hours would have greater stop proportion and lower average speed which would bring to more invalid trips. However, by conducting a Pearson Chi-squared test on the table of contingency with the validity of stop proportion and the peak hours, the p-value

Table 5.20: Contingency table for the time of the day.

Time of the day	**Valid trips**	**Invalid trips**	**Total**
Off-peak	87	182	46.1%
Peak	103	212	53.9%
Total	32.5%	67.5%	100%

is 0.8233 and for the table with the validity of average speed and the peak hours the p-value is 0.1684. There is thus no statistical link between the validity of those factors regarding the RDE procedure and the peak hours.

Table 5.21 shows that the time of the day has a significant impact on the means of the average speed, the stop proportion, the upper dynamic boundary, the distance and the duration of the trips.

Table 5.21: Impact of the time of the day on the parametric variables.

Variable	**p-value**	**Comment**
Average speed	$2 \cdot 10^{-3}$	The average speed is significantly smaller for trips during peak hours (21.62 km/h compared to 23.18 km/h off-peak).
Stop proportion	0.03	The stop proportion is significantly smaller for trips off-peak hours (18.78% compared to 20.97% for peak hours).
RPA	0.96	The RPA is not considerably affected by the time of the day.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	$5.7 \cdot 10^{-3}$	The value of $(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is significantly smaller for trips during the peak hours (5.77 m^2/s^3) compared to off-peak hours (6.09 m^2/s^3).
Maximal speed	0.08	The maximal speed is not impacted by the time of the day.
Distance	$1.2 \cdot 10^{-3}$	The distance is significantly smaller for trips during peak hours (4.79 km) compared to off-peak hours (5.91 km).
Duration	$8.3 \cdot 10^{-6}$	The duration of the trips during peak hours is significantly smaller than the one during off-peak hours (13 minutes on average compared to 17 minutes).

Thus, the time of the day does not influence the overall validity regarding the average speed and the stop proportion but it influences the means of these factors. This makes sense since at peak hours more vehicles circulate on the roads so the speed and the duration are smaller and the stop proportion higher. These results suggest thus that the peak hours have an impact on these two factors but not enough to interfere with the overall validity defined by the RDE test procedure. The explanation for the influence on the upper dynamic condition could join the one for the influence of the time of the week on the same factor. Indeed, during off-peak hours, the ability to reach higher accelerations increases and the driver can be more aggressive than during peak hours. The influence on the distance can have multiple explanations. One of them could be that people turn their engine off in traffic jam which stops the recording of the trip or the driver may try to avoid using its car too much during peak hours.

5.4.11 Comparison with data outside of Brussels

The study took part in Brussels but the impact of the location is studied in this section to confirm or overturn the ideas about the Brussels mobility and to analyse the influence on the RDE conditions. 60 trips have been driven outside of Brussels and were compared to the 584 trips from Brussels. The driver rode mainly in Wallonia, covering the provinces of the Hainaut, the Walloon Brabant, Namur and Liège as represented in FIGURE 5.31. The driver went also through Brussels and the three trips covering more than 30% of distance in Brussels were categorised as trips from Brussels.

Figure 5.31: Data points outside of Brussels.

To be consistent, the data of the rural and motorway categories have not been considered in the analysis. Therefore, all points of data have an instantaneous speed smaller than 60 km/h. However, for the record, the rural and motorway categories are better represented in this case: 55.25% of the duration of the trip lies in the urban class, 16.52% in the rural class and 28.23% in the motorway class.

It is important to note that points of data for the driver under study are only used for the purpose of this section and have not been integrated in the computations of the previous sections. As for the other categorical factors, Table 5.22 reports the influence on the quantitative parameters. The data outside of Brussels are represented by 57 trips and the data from Brussels by 587 trips. This proves that the following results should be taken with a grain of salt.

Table 5.22: Impact of the location on the quantitative parameters.

Variable	**p-value**	**Comment**
Average speed	$7.35 \cdot 10^{-3}$	The average speed for trips in Brussels is significantly smaller than trips out of Brussels (22.34 km/h against 25.98 km/h).
Stop proportion	0.94	The stop proportion does not seem to be impacted by the localisation.
RPA	$1.57 \cdot 10^{-6}$	The location of the trip has an impact on the relative positive acceleration. Its mean is equal to 0.255 m/s^2 in Brussels compared to 0.313 m/s^2 outside of Brussels.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k$**[95]**	$3.75 \cdot 10^{-13}$	The value of $(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$ is significantly smaller in Brussels (5.93 m^2/s^3) than outside of Brussels (8.21 m^2/s^3).
Maximal speed	$6.63 \cdot 10^{-7}$	The mean value of maximal speed is significantly smaller for trips in Brussels (68.75 km/h on average) than for trips out of Brussels (90.67 km/h on average).
Distance	0.541	The distance is not significantly influenced by the location of the trip.
Duration	0.020	The duration of the trips out of Brussels is 11 minutes on average which is significantly smaller than the 15 minutes that the Brussels drivers.

Nevertheless, these results are interesting: the localisation has an impact on the average speed, the RPA, $(\mathrm{v} \cdot \mathrm{a_{pos}})[95]$, the maximal speed and the duration of the trips. Only the stop proportion and the distance are not differing with respect to the localisation. The effect on the average speed is straightforward since Brussels is known to be very congested and the speed limits are smaller in the city. This also explains the higher maximal speed outside of Brussels. The effect on the dynamic conditions suggests that drivers in Brussels have a calmer way of driving. In Brussels, there are many traffic lights, more people on the roads and more traffic jam. The Brussels drivers cannot always afford to reach higher acceleration changes compared to others. This is not completely representative since only one driver is analysed for the data outside of Brussels and the driving behaviour is personal. The difference in the duration is mostly explained by the amount of drivers on the road. The driver rode the same distance on average than a Brussels driver.

In this case, the stop proportion is 22% for data outside of Brussels and 20% for data inside of Brussels. This difference is not significant regarding the localisation. However, this is unusual since the average speed is linked to the stop proportion as explained in SECTION 5.3.2 and a higher average speed often results in a lower value of stop proportion. The stop proportion should thus be smaller outside of Brussels. As said before, the data are continuously recorded from the moment the engine has started. Hence, an unusual long period with the engine on while the car is stopped could explain this abnormality. Since the amount of trips outside of Brussels is limited, every small flaw influences a lot the results compared to the case of data in Brussels where all the singularities are drowned in the quantity of data points.

The relationship between the location and the parameters of the car (such as the displacement or the year of construction) has not been studied since the data out of Brussels are only covered by one car. However, the potential links between the location and whether the trip was ridden at the weekend, during the peak hours or with a cold start. These relationships have been analysed with a Pearson's Chi-squared test on the relative tables of contingency. The results are reported in TABLE 5.23. The p-values are all larger than 0.05 so those three factors are not related to the localisation. In fact, the opposite would have been curious. This section emphasises the interest of studying

Table 5.23: p-value from a Person's Chi-squared test on the links between the location and three external factors.

Factor 1	**Factor 2**	**p-value**
Location	Weekend - week	0.557
Location	Peak - off-peak hours	0.572
Location	Cold start	0.655

driving patterns outside of Brussels as well in further researches.

5.5 Summary

To sum up, the goal of this section was to understand the influence and the links between many factors. The parameters computed and used as RDE conditions have indeed a significant impact on the overall validity of the trip except for the RPA as summarised in TABLE 5.24. This proves that the conditions are well chosen and are indeed relevant to look at. The RPA might not have any impact on the overall validity of the trip but redoing the same analyses with a bigger data set could be interesting to confirm or overturn that conclusion. The stop proportion is the second most demanding boundary set by the RDE methodology after the condition on the distance. This study has been conducted in Brussels which is known to be in the top ten of the most congested cities in Europe [52]. Therefore, the condition on the stop proportion could be accurate for other places in

Europe but not in Brussels. The condition on the average speed is often unfulfilled as well. However, the dynamic boundaries and the maximal speed were well respected.

Table 5.24: Summary of overall validity for the condition, for the urban category.

	Valid	**Invalid**	**Null hypothesis**
Average speed	516	68	Rejected: significant impact.
Stop proportion	412	172	Rejected: significant impact.
RPA	582	2	Not able to reject it.
$(\mathbf{v} \cdot \mathbf{a_{pos}})_k[95]$	584	0	Rejected: significant impact.
Maximal speed	583	1	Rejected: significant impact.
Distance	14	570	Rejected: significant impact.

Table 5.25 represents the links between categorical and parametric variables. The "X" in the grey cells means that the quantitative variable is not influenced by the categorical factor and the "V" in the blue cells represents the opposite. This table helps to understand that the overall validity of the trip (without the distance) is not influenced by these external factors.
However, the variables used to compute the RDE conditions themselves are impacted by these categorical factors. The duration is influenced by all factors except for the type of gearbox. It is important to bear in mind that the car fleet was very restricted. Therefore, some intrinsic bonds play a role although they should not. Nonetheless, this table teaches that the gearbox is not an essential factor since it has no influence on the variables computed for the RDE conditions except on the maximal speed. Thus, further studies could afford to leave out this factor in the choice of their car fleet.

The external factors (except the type of gearbox) influence on average four to five variables. The factors on the weekend and peak hours influence four and five variables respectively. This proves the relevance of conducting a test on public roads and not only on a dynamometer. Even if not entirely representative, the comparison between trips in and out of Brussels showed that it was a significant factor on most of parametric factors (except the stop proportion and the distance).

Table 5.25: Summary of influences between parametric and categorical factors.

	Overall validity of the trip	Average speed	Stop proportion	RPA	v*a_pos	Maximal Speed	Distance	Duration
Fuel	X	V	V	X	X	X	V	V
Company/private car	X	V	V	V	X	X	X	V
Displacement	X	V	V	V	X	V	X	V
Year	X	V	V	V	V	X	X	V
Power	X	X	V	X	X	V	V	V
Gearbox	X	X	X	X	X	V	X	X
Cold start	X	V	X	X	V	V	V	V
Week/weekend	X	V	V	X	V	X	X	V
Peak/Off-peak	X	V	V	X	V	X	V	V

CHAPTER 6

Conclusion

The goal of this thesis was to analyse if the conditions imposed by the RDE test procedure were relevant and representative of Brussels drivers behaviour. In that sense, 51 drivers in Brussels have been chosen and the data of 32 of them that were well recorded were used to compare some parameters with the existing conditions in the RDE procedure. A recording device developed by the research team has been installed in their car for one to two weeks and the subjects of the study were asked to drive as usual.

Twelve parameters were recorded by the data logger with either a GPS chip, an accelerometer or a connection to the ECU of the car. However, only six of them were used: the date and time, the GPS coordinates, the vehicle speed from the OBD, the altitude and the engine coolant temperature. The data were sorted and all trips smaller than 1 km or with more than 30% of distance out of the zone representing Brussels in this work were removed from the analysis. With this first filtering, 315 trips out of 899 were deleted in total. The GPS chip has been proved to be the most unstable component of the data logger. Therefore, the speed from the OBD and the resulting distance were chosen at the expense of the same variables coming from the GPS.

These six measurements permitted computing parameters used as conditions framing the RDE test procedure. Indeed, conditions have been set for the RDE test and if one of them is not respected, the emissions measured for the trips in question cannot be used to assess the performance of the car. In that case, the trip is said to be invalid regarding the RDE procedure. These conditions are set on many parameters and the ones studied in this work were the average speed, the stop proportion, the altitude, the RPA, the 95th percentile of $(v \cdot a_{pos})$, the maximal speed and the distance. A RDE trip must also start with a cold engine, so the coolant engine temperature has been added to the analysis. Thus, to consider and validate the emissions measurements of a test all parameters must be kept within the conditions imposed.

In the RDE methodology, the data are separated in three categories based on the instantaneous speed: the urban one (with instantaneous speeds lower or equal to 60 km/h), the rural one (with $60 < v_i \leqslant 90$ km/h) and the motorway (with instantaneous speeds strictly higher than 90 km/h). Since the data were recorded in Brussels, the rural and motorway classes were artificially built and barely represented so these data were not used in the overall analysis of the parameters.

The RDE methodology should adapt their condition on the distance ridden because it is not representative of the minimal distance ridden in Brussels. However, this condition might be adapted to other bigger cities. The condition on the altitude was not relevant in this case since Brussels is flat and the GPS was not always accurate. The stop proportion is clearly the second most unfulfilled condition which makes sense since Brussels is one of the most congested cities in Europe. Brussels people's driving behaviour does not match with the requirements of the RDE in terms of stop proportion, since almost 30% of the trips were out of boundaries regarding that factor. The condition on the average speed has been unfulfilled for 12% of the trips. However, the dynamic conditions and the maximal speed were well represented.

An overall validity factor has been built based on the validity regarding the RDE conditions on the average speed, the stop proportion, the dynamic conditions and the maximal speed. 190 trips were invalid regarding this overall validity factor which accounts for 32%. This shows that the RDE procedure is not really representative of the Brussels driving behaviour since one trip out of three should theoretically be removed. If the distance is taken into account in the overall validity factor, only 13 trips can be kept and the whole rest is invalid regarding the RDE procedure. A statistical test showed that the validity of the trip was influenced by the average speed, the stop proportion, the maximal speed, the distance and $(v \cdot a_{pos})_k[95]$. However, the RPA could have no influence on the overall validity of the trip regarding the RDE, so this variable could be irrelevant to the test.

To push the analysis further, the influence of other factors which are the fuel type, whether the vehicle is a company car, the class of displacement, of power and of year of construction, the type of gearbox, whether the engine was cold when it started, the time of the week and the time of the day has been analysed. This examination has only been made with the data from the urban class due to the lack of data points for the two others. It is really important to keep in mind that the car fleet was too small and thus not representative to draw any definitive conclusion. Many factors were invisibly bounded together due to poverty of variability of types of cars. For instance, if there had been only three company cars all running with a gasoline engine, an association would have been made between those two characteristics, even though it does not represent the reality. None of those external factors had an influence on the overall validity of the trip defined earlier. Then, their impact on the means of the parameters used to set the conditions of the RDE were deepened. The type of gearbox is the less influential factor since it only has effect on the maximal speed. All the other external factors had impact on four to five out of eight variables. The RPA is only influenced by the displacement, the year and the function of the car which makes it the second less influenced variable after the overall validity. The duration is affected by all external factors, except for the gearbox.

Finally, a sample of data outside of Brussels has been analysed and compared to data from Brussels. The location impacts the average speed, the factors used as lower and upper dynamic conditions, the duration and the maximal speed. However, data outside of Brussels have only been recorded by one drive so those results could be deepened and confirmed in further researches.

Indeed, this work is the front door to many further researches. In that sense, some points should be taken into account to improve future studies. It is not necessary to record the data at more than 1 Hz since it adds more noise to the signal. Concerning the choice of the subjects, more cars should be represented to draw better conclusions. Moreover, the present study has been biased by the choice of subjects. Indeed, they were all from the same environment and thus not representative of the real Brussels population. The WIFI of the data loggers should be improved since after one month it was not possible to access the data via WIFI for some of the data loggers. Moreover, huge troubles happened with the SD card since hundreds of trips have been lost due to the inability to read the SD card. The battery was supposed to last only thanks to the reloading during the periods when the engine was on. However, it happened that the battery turned off and some people were kind of afraid touching the device, so they kept riding with the data logger turned off. It should be interesting to understand why those devices turned off and to prepare drivers to the possibility that they could have to turn it back on. Doing the study on a longer period would be better since the logistics of all data loggers is hard to manage. Then, it should be interesting to use the accelerometers. If the issue with the loss of GPS signal is settled, it would be interesting to analyse the cumulative altitude gain since it has not been well reviewed in this work and it seems to not fulfil the condition. Finally, another advice is to try to have enough data in the three categories (urban, rural and motorway) by imposing the route of the trip, for instance.

To conclude, this first study with the data loggers presented some interesting results and many more are to come since the automotive mobility is a growing key sector and this question of emission tests is prevailing and omnipresent.

CHAPTER A

Appendices

A.1 Information about the cars for which the data were recorded

Table A.1: Information about the cars used for the study.

N°	StartDate	EndDate	Brand	Model	Year	Displ. *	P**	Fuel	Gearbox	Company
1	17/02/2019	24/02/2019	Renault	Modus	2008	1149	55	Gasoline	Manual	Private car
2	24/02/2019	01/03/2019	Renault	Kadjar	2016	1197	96	Gasoline	Manual	Private car
3	16/02/2019	18/02/2019	BMW	X1 s Drive 18d	2017	1995	110	Diesel	Manual	Company car
4	16/02/2019	22/02/2019	Citroën	Astra sport S tourer +	2017	1598	100	Diesel	Automatic	Company car
5	25/02/2019	03/03/2019	Renault	Clio	2014	898	66	Gasoline	Manual	Private car
6	27/02/2019	03/03/2019	Kia	Ceed	2017	1368	73.6	Gasoline	Manual	Private car
7	15/02/2019	22/02/2019	Kia	Ceed	2017	1368	73.6	Gasoline	Manual	Private car
8	22/02/2019	03/03/2019	BMW	X1 s Drive 18d	2017	1995	110	Diesel	Manual	Company car
9	16/02/2019	02/03/2019	Audi	A1	2016	1422	66	Diesel	Manual	Company car
10	12/02/2019	23/02/2019	Mini	Cooper	2017	1499	100	Gasoline	Automatic	Company car
11	12/02/2019	23/02/2019	BMW	X1 s18d	2016	1995	110	Diesel	Manual	Company car
12	15/02/2019	15/02/2019	Mini	Clubman One	2016	1499	75	Gasoline	Manual	Company car
13	15/02/2019	19/02/2019	Mini	Clubman One	2016	1499	75	Gasoline	Manual	Company car
14	24/02/2019	03/03/2019	Peugeot	2008	2017	1199	81	Gasoline	Manual	Private car
15	25/02/2019	08/03/2019	BMW	218i Active Tourer	2016	1499	100	Gasoline	Manual	Company car
16	26/02/2019	10/03/2019	Volkswagen	Golf	2011	1598	77	Diesel	Manual	Private car
17	26/02/2019	10/03/2019	Volkswagen	Golf	2016	1197	81	Gasoline	Automatic	Private car
18	23/02/2019	16/03/2019	Mercedes	B180D	2017	1461	80	Diesel	Automatic	Company car
19	16/03/2019	21/03/2019	Citroën	Astra sport S tourer +	2017	1598	100	Diesel	Automatic	Company car

N°	StartDate	EndDate	Brand	Model	Year	Displ. *	P**	Fuel	Gearbox	Company
20	16/03/2019	23/03/2019	Kia	Ceed	2017	1368	73.6	Gasoline	Manual	Private car
21	17/03/2019	31/03/2019	Volvo	XC90	2016	1969	140	Diesel	Automatic	Private car
22	22/02/2019	04/03/2019	Volkswagen	Golf	2017	1598	85	Diesel	Manual	Company car
23	04/03/2019	20/03/2019	Mercedes	A180D	2018	1461	85	Diesel	Automatic	Company car
24	07/04/2019	16/04/2019	Peugeot	308	2018	1499	96	Diesel	Automatic	Company car
25	23/02/2019	10/03/2019	Audi	A1	2016	999	66	Gasoline	Automatic	Private car
26	17/03/2019	31/03/2019	Suzuki	Suzuki Swift 3	2012	1242	69	Gasoline	Manual	Private car
27	01/04/2019	16/04/2019	Seat	Ibiza	2014	1199	55	Diesel	Manual	Company car
28	10/03/2019	24/03/2019	Volkswagen	VW Polo	2019	999	59	Gasoline	Manual	Company car
29	28/03/2019	11/04/2019	BMW	218i	2016	1499	100	Gasoline	Manual	Private car
30	19/03/2019	02/04/2019	Volkswagen	Polo	2011	1598	66	Diesel	Manual	Private car
31	09/04/2019	20/04/2019	Opel	Adam	2016	1229	51	Gasoline	Manual	Private car
32	25/03/2019	01/04/2019	Ford	ecosport	2019	998	93	Gasoline	Automatic	Private car

* Displ. is the displacement of the engine in cm^3.

** P is the power of the engine in kW.

A.2 Information about the cars for which the data were not recorded

Table A.2: Information about the cars for which the recording failed.

Start Date	End Date	Brand	Model	Year	Displ.*	P**	Fuel	Gearbox	Company car?	Reason of failure
16/02/2019	22/02/2019	Citroën	Berlingo	2011	1560	66	Diesel	Manual	Private car	Compatibility issue
10/03/2019	21/03/2019	Opel	Astra	2017	/	/	/	Manual	Private car	SD card failed
15/03/2019	01/04/2019	Renault	Captur	2016	1197	87	Gasoline	Automatic	Company car	SD card failed
17/03/2019	01/04/2018	Fiat	Punto	2013	1242	51	Gasoline	Manual	Private car	SD card failed
03/04/2019	16/04/2019	Seat	Ibiza	2016	999	70	Gasoline	Manual	Private car	SD card failed
06/03/2019	13/03/2019	BMW	325D	2008	2993	145	Diesel	Automatic	Private car	SD card failed
22/03/2019	04/04/2019	Fiat	500L	2016	1368	88	Gasoline	Manual	Private car	SD card failed
25/03/2019	01/04/2019	Ford	ecosport	2019	998	93	Gasoline	Automatic	Private car	Not in Brussels
03/03/2019	16/03/2019	Volkswagen	golf break	2017	999	81	Gasoline	Manual	Company car	SD card failed
16/03/2019	01/01/2019	Citroën	c3	2013	1199	60	Gasoline	Manual	Private car	SD card failed

Start Date	**End Date**	**Brand**	**Model**	**Year**	**Disp.***	**P****	**Fuel**	**Gearbox**	**Company car?**	**Reason of failure**
16/02/2019	22/02/2019	Citroën	c3	2006	1124	44	Gasoline	Manual	Private car	Compatibility issue
16/02/2019	01/03/2019	Volkswagen	Polo	2019	999	59	Gasoline	Manual	Company car	Compatibility issue
31/03/2019	18/04/2019	Toyota	Yaris HYBRIDE	2017	1497	54	Gasoline HYBRIDE	Automatic	Private car	SD card failed
14/03/2019	29/03/2019	BMW	118i	2016	1499	100	Gasoline	Manual	Private car	SD card failed
01/04/2019	18/04/2019	Volvo	V40	2015	1969	88	Diesel	Manual	Company car	SD card failed
03/03/2019	20/03/2019	Volvo	v60	2013	2000	88	Diesel	Manual	Company car	Not in Brussels
25/02/2019	26/02/2019	Ford	Ranger 3.2TDCi	2017	3198	147	Diesel	Automatic	Company car	SD card failed
10/03/2019	22/03/2019	Audi	Audi A1	2016	1422	66	Diesel	Manual	Company car	Data not recorded after 6 days
28/03/2019	05/04/2019	Ford	Eco Sport	2016	998	92	Gasoline	Manual	Private car	SD card failed

* Displ. is the displacement of the engine in cm^3.
** P is the power of the engine in kW.

A.2.1 Proportion of brands

From all the car fleet (covering 13 different brands), the main vehicles were first from BMW and Volkswagen (5 cars each) and then from Citroën, Renault, Kia and Mini (3 cars each). The rest of the brands are listed in Appendix A.1. Table A.3 lists the amount of trips ridden by each brand.

Table A.3: Contingency table for the brand of the car.

Brand	**Valid trips**	**Invalid trips**	**Total**
Audi	38	57	16.3%
BMW	46	113	27.2%
Citroën	5	23	4.8%
Kia	7	11	3.1%
Mercedes	19	27	7.9%
Mini	9	14	3.9%
Opel	1	5	1%
Peugeot	12	32	7.5%
Renault	17	44	10.4%
Seat	4	6	1.7 %
Suzuki	2	12	2.4 %
Volkswagen	29	41	12 %
Volvo	1	9	1.7%
Total	32.5%	67.5%	100%

The p-value from the Pearsons' Chi-squared test is 0.13 which does not permit affirming that the means are different. Therefore, the brand might not be a significant factor in the validity of the trip. Since there are more than two brands, it is not possible to conduct the statistical hypothesis test as for the other categorical parameters.

A.3 Detailed quantiles of the parametric variables

Table A.4: Summary of average speed quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	19.53	23.69	22.35
sd	7.68	4.58	6.08
IQR	11.97	6.02	8.17
0%	6.31	15.24	6.31
25%	13.90	20.38	18.20
50%	17.02	23.31	22.41
75%	25.87	26.40	26.38
100%	37.17	36.82	37.17

Table A.5: Summary of stop proportion quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	25.05	17.53	19.96
sd	18.62	6.14	12.24
IQR	33.55	9.46	14.94
0%	0	6.05	0
25%	4.24	12.58	11.47
50%	30.98	17.56	18.69
75%	37.79	22.04	26.41
100%	68.66	29.97	68.66

Table A.6: Summary of RPA quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	0.260	0.252	0.254
sd	0.056	0.051	0.053
IQR	0.068	0.068	0.070
0%	0.109	0.144	0.109
25%	0.225	0.214	0.218
50%	0.252	0.245	0.248
75%	0.293	0.283	0.288
100%	0.437	0.502	0.502

Table A.7: Summary of $(\mathrm{v} \cdot \mathrm{a}_{pos})_k[95]$ quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	5.58	6.08	5.92
sd	1.39	1.20	1.28
IQR	1.90	1.59	1.65
0%	2.75	3.30	2.75
25%	4.58	5.26	5.03
50%	5.57	5.90	5.84
75%	6.48	6.85	6.68
100%	9.48	10.62	10.62

Table A.8: Summary of maximal speed quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	62	71.92	68.70
sd	26.38	27.22	27.32
IQR	25	35.75	27
0%	30	32	30
25%	45	52	50
50%	53	62	59
75%	70	87.75	77
100%	162	153	162

Table A.9: Summary of quartiles of distances for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	3.89	6.12	5.39
sd	2.98	3.94	3.80
IQR	3.79	4.99	5.09
0%	1.00	1.05	1.00
25%	1.59	3.27	2.39
50%	2.80	5.22	4.48
75%	5.38	8.26	7.49
100%	18.55	21.20	21.20

Table A.10: Summary of duration quartiles for invalid, valid, and all trips.

	Invalid	**Valid**	**All trips**
mean	14.20	15.73	15.23
sd	13.80	10.41	11.64
IQR	11.38	12.33	12.74
0%	2.09	2.09	2.09
25%	5.50	8.01	6.89
50%	9.35	13.07	11.88
75%	16.88	20.34	19.63
100%	100.96	54.70	100.96

A.4 Q-Q plots of the logarithm of the maximal speed, the distance and the duration

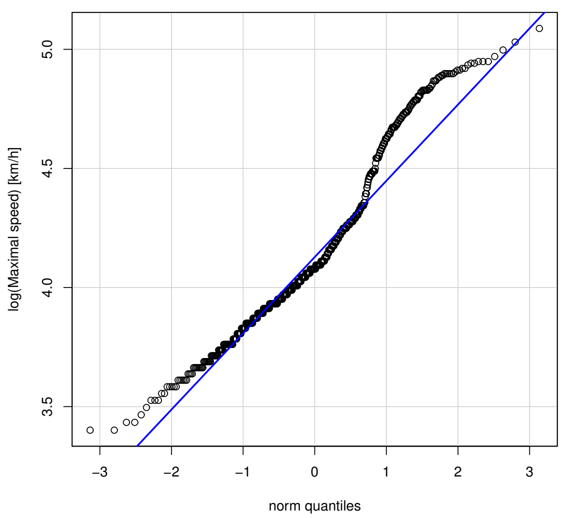

Figure A.1: Q-Q plot of the logarithm of the maximal speed.

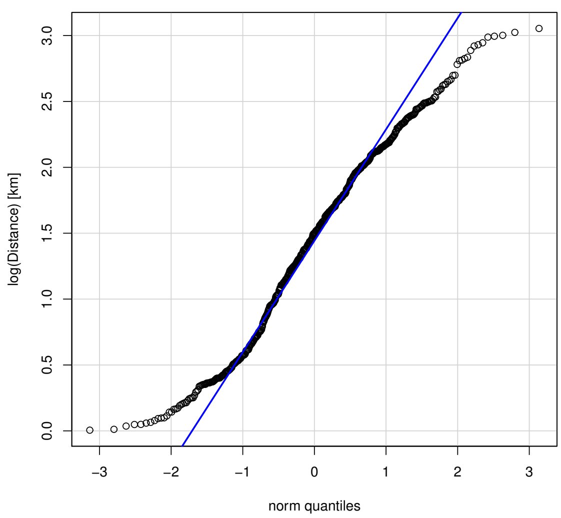

Figure A.2: Q-Q plot of the logarithm of the distance.

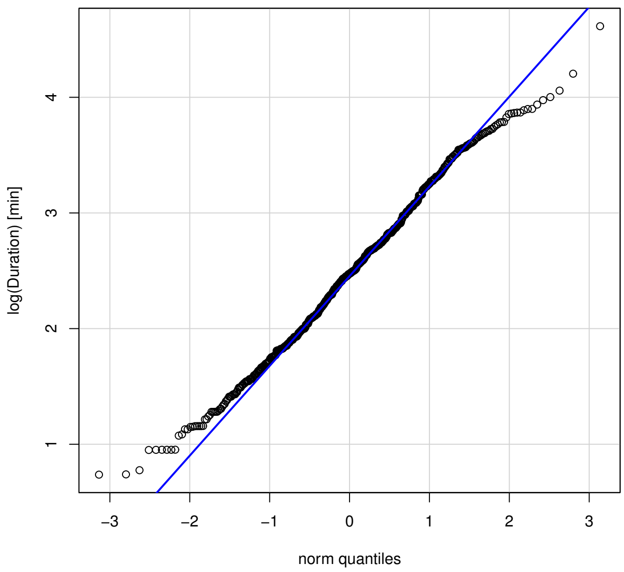

Figure A.3: Q-Q plot of the logarithm of the duration of the trips.

A.5 Links between categorical variables: p-values

Table A.11: p-values from the Pearson's Chi-squared test conducted between the categorical variables.

Factor 1	**Factor 2**	**p-value**
Fuel	Company/Private car	$< 2.2 \cdot 10^{-16}$
Fuel	Displacement	$< 2.2 \cdot 10^{-16}$
Fuel	Year	$< 2.2 \cdot 10^{-16}$
Fuel	Power	$< 2.2 \cdot 10^{-16}$
Fuel	Gearbox	0.959
Fuel	Cold start	0.043
Fuel	Weekend/week	0.056
Fuel	Peak/Off-peak	0.18
Company/Private car	Displacement	$< 2.2 \cdot 10^{-16}$
Company/Private car	Year	$< 2.2 \cdot 10^{-16}$
Company/Private car	Power	$< 2.2 \cdot 10^{-16}$
Company/Private car	Gearbox	0.016
Company/Private car	Cold start	0.723
Company/Private car	Weekend/week	0.002
Company/Private car	Peak/Off-peak	0.561
Displacement	Year	$1.001 \cdot 10^{-12}$
Displacement	Power	$< 2.2 \cdot 10^{-16}$
Displacement	Gearbox	$5.387 \cdot 10^{-10}$
Displacement	Cold start	0.773
Displacement	Weekend/week	0.245
Displacement	Peak/Off-peak	0.463
Year	Power	$< 2.2 \cdot 10^{-16}$
Year	Gearbox	0.395
Year	Cold start	0.021
Year	Weekend/week	$8.85 \cdot 10^{-7}$
Year	Peak/Off-peak	0.996
Power	Gearbox	$6.166 \cdot 10^{-7}$
Power	Cold start	0.059
Power	Weekend/week	0.013
Power	Peak/Off-peak	0.570
Gearbox	Cold start	0.114
Gearbox	Weekend/week	0.379
Gearbox	Peak/Off-peak	0.124
Cold start	Weekend/week	0.031
Cold start	Peak/Off-peak	0.052
Weekend/week	Peak/Off-peak	0.019

List of Figures

List of Tables

Bibliography

[1] International Organization of Motor Vehicle Manufacturers, "World Vehicles in Use (All Vehicles)," tech. rep., OICA, 2017.

[2] European Automobile Manufacturers Association, "ACEA Report Vehicles in use Europe 2018," tech. rep., ACEA, Brussels, 2018.

[3] B. July, "Avec la voiture électrique , le courant passe." Le Soir, 2018.

[4] Fédération Belge et Luxembourgeoise de l'Automobile & du Cycle (FEBIAC), "Datadigest 2018." https://www.febiac.be/public/statistics.aspx?FID=23&lang=FR.

[5] C. Park and M. Allaby, *A Dictionary of Environment and Conservation (3 ed.)*. Oxford: Oxford University Press, 3 ed., 2017.

[6] Greenpeace Belgium, "Même depuis l'espace, on voit que la Belgique étouffe sous le dioxyde d'azote." http://www.greenpeace.org/belgium/fr/presse/Meme-depuis-lespace-on-voit-que-la-Belgique-etouffe-sous-le-dioxyde-dazote/, 2018.

[7] C. K. Gately, L. R. Hutyra, S. Peterson, and I. Sue Wing, "Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone GPS data," *Environmental Pollution*, vol. 229, pp. 496–504, 2017.

[8] European Environment Agency, "Air quality in Europe — 2017 report," tech. rep., European Environment Agency, Copenhague, 2017.

[9] Cellule Interrégionale de l'Environnement (CELINE), "Quelles sont les principales sources d'émissions des PM?." http://www.irceline.be/fr/documentation/faq/quelles-sont-les-principales-sources-demissions-des-pm.

[10] E. O. Okokon, T. Yli-Tuomi, A. W. Turunen, P. Taimisto, A. Pennanen, I. Vouitsis, Z. Samaras, M. Voogt, M. Keuken, and T. Lanki, "Particulates and noise exposure during bicycle, bus and car commuting: A study in three European cities," *Environmental Research*, vol. 154, pp. 181–189, 2017.

[11] A. de Nazelle, O. Bode, and J. P. Orjuela, "Comparison of air pollution exposures in active vs. passive travel modes in European cities: A quantitative review," *Environment International*, vol. 99, pp. 151–160, 2017.

[12] G. Cesaroni, H. Boogaard, S. Jonkers, D. Porta, C. Badaloni, G. Cattani, F. Forastiere, and G. Hoek, "Health benefits of traffic-related air pollution reduction in different socioeconomic groups: The effect of low-emission zoning in Rome," *Occupational and Environmental Medicine*, vol. 69, no. 2, pp. 133–139, 2012.

[13] C. Evans, "Euro 1 to Euro 6: how clean is my car?." https://www.whatcar.com/advice/owning/euro-1-to-euro-6-how-clean-is-my-car/n1190, 2017.

[14] A. Fotouhi, "Tehran driving cycle development using the k-means clustering method," *Scientia Iranica*, vol. 20, no. 2, pp. 286–293, 2013.

[15] N. Romain, "The different driving cycles." http://www.car-engineer.com/the-different-driving-cycles/, 2013.

[16] L. Sileghem, D. Bosteels, J. May, C. Favre, and S. Verhelst, "Analysis of vehicle emission measurements on the new WLTC , the NEDC and the CADC," *Transportation research part D*, vol. 32, pp. 70–85, 2014.

[17] M. André, "The ARTEMIS European driving cycles for measuring car pollutant emissions," *Science of the Total Environment*, vol. 334-335, pp. 73–84, 2004.

[18] T. Donateo and M. Giovinazzi, "Building a cycle for Real Driving Emissions," *Energy Procedia*, vol. 126, pp. 891–898, 2017.

[19] European Commission, "Testing of emissions from cars," *Official Journal of the European Union*, no. September 2018, pp. 2017–2019, 2019.

[20] J. Demuynck, D. Bosteels, M. De Paepe, J. May, and S. Verhelst, "Recommendations for the new WLTP cycle based on an analysis of vehicle emission measurements on NEDC and CADC," *Energy policy*, vol. 49, pp. 234–242, 2012.

[21] S. K. Pathak, Y. Singh, V. Sood, and S. A. Channiwala, "On-Road Vehicle Driving and Energy Requirements and Impact on Unregulated Exhaust Emissions under Urban Driving Conditions," *SAE International Journal of Engines*, vol. 10, no. 4, pp. 1866–1879, 2017.

[22] M. Irwin, H. Bradley, M. Duckhouse, M. Hammond, and M. S. Peckham, "High spatio-Temporal resolution pollutant measurements of on-board vehicle emissions using ultra-fast response gas analyzers," *Atmospheric Measurement Techniques*, vol. 11, no. 6, pp. 3559–3567, 2018.

[23] L. Colart, "Ce qu'il faut savoir sur la triche de Volkswagen." https://plus.lesoir.be/art/996061/article/economie/2015-09-22/ce-qu-il-faut-savoir-sur-triche-volkswagen#__ga=2.98543447.936844624.1546173026-875106071.1544873813, 2015.

[24] UE Commission, "Règlement de la commission du 7 juin 2017," *Journal officiel de l'Union européenne*, vol. 2008, pp. 708–732, 2017.

[25] Car Emissions Testing, “What is the Real Driving Emissions (RDE) test?.” http://www.caremissionstestingfacts.eu/rde-real-driving-emissions-test/, 2018.

[26] Motortalk, “So funktioniert die fahrprüfung für stickoxide.” https://www.motor-talk.de/news/so-funktioniert-die-fahrpruefung-fuer-stickoxide-t6083371.html, 2017.

[27] European Commission, “Commission Regulation (EU) 2016/427,” *Official Journal of the European Union*, no. 692, pp. L 82/1– L 82/98, 2016.

[28] B. Giechaskiel, “Real Driving Emissions (RDE): Particle Number (PN), Portable Measurement Systems (PEMS) calibration,” *Publications Office of the European Union*, 2018.

[29] DieselNet, “EU: Cars and Light Trucks.” https://www.dieselnet.com/standards/eu/ld.php.

[30] UE Commission, “Règlement (UE) 2016/646 de la Commission,” *Journal officiel de l’Union européenne*, vol. 3, pp. L 109/1– L 109/22, 2016.

[31] P. Mock, “Real-driving emissions test procedure for exhaust gas pollutant emissions of cars and light commercial vehicles in Europe,” *The International Council on Clean Transportation*, 2017.

[32] H. Wang, X. Zhang, and M. Ouyang, “Energy consumption of electric vehicles based on real-world driving patterns: A case study of Beijing,” *Applied Energy*, vol. 157, pp. 710–719, 2015.

[33] L. Si and M. Hirz, “Big Data-Based Driving Pattern Clustering and Evaluation in Combination with Driving Circumstances,” *SAE Technical Papers*, pp. 1–11, 2018.

[34] E. Ericsson, “Variability in urban driving patterns,” *Transportation Research Part D: Transport and Environment*, vol. 5, no. 5, pp. 337–354, 2000.

[35] L. Feng, W. Liu, and B. Chen, “Driving Pattern Recognition for Adaptive Hybrid Vehicle Control,” *SAE International Journal of Alternative Powertrains*, vol. 1, pp. 169–179, 2012.

[36] E. Ericsson, “Independent driving pattern factors and their influence on fuel-use and exhaust emission factors,” *Transportation Research Part D*, vol. 6, no. 5, pp. 325–345, 2001.

[37] K. Brundell-Freij and E. Ericsson, “Influence of street characteristics, driver category and car performance on urban driving patterns,” *Transportation Research Part D: Transport and Environment*, vol. 10, no. 3, pp. 213–229, 2005.

[38] J. Höglund, Paul G; Niittymaki, “Estimating Vehicle Emissions and Air Pollution related to Driving Patterns and Traffic Calming,” *Urban transport Systems*, vol. 1, no. 10, pp. 1–11, 1999.

[39] P. Shridhar Bokare and A. Kumar Maurya, "Study of Effect of Speed, Acceleration and Deceleration of Small Petrol Car on Its Tail Pipe Emission," *International Journal for Traffic and Transport Engineering*, vol. 3, no. 4, pp. 465–478, 2013.

[40] A. S. Nagpure, B. R. Gurjar, and P. Kumar, "Impact of altitude on emission rates of ozone precursors from gasoline-driven light-duty commercial vehicles," *Atmospheric Environment*, vol. 45, no. 7, pp. 1413–1417, 2011.

[41] C. Beckx, L. I. Panis, D. Janssens, and G. Wets, "Applying activity-travel data for the assessment of vehicle exhaust emissions: Application of a GPS-enhanced data collection tool," *Transportation Research Part D: Transport and Environment*, vol. 15, no. 2, pp. 117–122, 2010.

[42] Le conseil des communauté européennes, "Règlement (CEE, EURATOM) N° 1182/71 du conseil," *Journal officiel des Communautés européennes*, vol. 124/L4, 1971.

[43] N. Kühn, "Details RDE Legislation Europe," in *Emission controls manufacturers Association*, 2017.

[44] M. Weiss, E. Paffumi, M. Clairotte, Y. Drossinos, and T. Vlachos, "Including cold-start emissions in the Real-Driving Emissions (RDE) test procedure effects," *Joint Research Centre (JRC)*, 2017.

[45] W. T. Hung, H. Y. Tong, C. P. Lee, K. Ha, and L. Y. Pao, "Development of a practical driving cycle construction methodology: A case study in Hong Kong," *Transportation Research Part D: Transport and Environment*, vol. 12, no. 2, pp. 115–128, 2007.

[46] C.-C. Lin, S. Jeon, H. Peng, and J. Moo Lee, "Driving Pattern Recognition for Control of Hybrid Electric Trucks," *Vehicle System Dynamics*, vol. 42, no. 1-2, pp. 41–58, 2004.

[47] M. Nyhan, S. Sobolevsky, C. Kang, P. Robinson, A. Corti, M. Szell, D. Streets, Z. Lu, R. Britter, S. R. Barrett, and C. Ratti, "Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model," *Atmospheric Environment*, vol. 140, pp. 352–363, 2016.

[48] M. Quddus and S. Washington, "Shortest path and vehicle trajectory aided map-matching for low frequency GPS data," *Transportation Research Part C*, vol. 55, pp. 328–339, 2015.

[49] Alex E., "Reading Performance Information Data (PID)." http://www.totalcardiagnostics.com/support/Knowledgebase/Article/View/43/6/reading-performance-information-data-pid, 2013.

[50] Navstar Inc., "Navstar FPS user equipment," *Public release version*, 1996.

[51] R. von Sachs, "FSAB 1105 : Probabilités et Statistique," in *LFSAB 1105*, 2016.

[52] O. Duquesne, “Mobilité / Bruxelles 8e ville la plus embouteillée d’Europe.” https://www.moniteurautomobile.be/actu-auto/mobilite/bruxelles-8e-ville-plus-embouteillee-europe-tomtom-traffic-index.html, 2017.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**