UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

MASTER IN INNOVATION AND RESEARCH IN INFORMATICS
DATA MINING AND BUSINESS INTELLIGENCE

MASTER THESIS

Facility Location Models for Electric Vehicle Charging Infrastructure

Author: Damià Valero Bover
Advisor: Josep Lluis Larriba Pey
Department: Arquitectura de Computadors (DAC)

Barcelona, May 2017

Acknowledgements

I would like to express my acknowledgement to all those who encouraged and gave me support during my Master.

To my parents, Francesc and Lourdes, for their unconditional support during all my academic career. To Nerea, for encouraging and helping me all this years.

This thesis would not have been possible without the advice and ideas from my advisor, Josep Lluis Larriba. Also to Roberto Villafàfila for his vision and expertise in the electric vehicles field.

Abstract

This thesis deals with the study of current charging infrastructure availability in highways, as well as proposing optimal allocations for new stations. First, a Machine Learning model is trained in order to estimate the actual range of an electric vehicle. This model will be constructed using heterogeneous data sources and variables that influence the total autonomy, such as speed, temperature, degradation or elevation, among others. Second, this model is used in combination with geospatial data regarding French highway and charging infrastructure locations, in order to propose a methodology for analyzing the availability level of charging stations in highways for electric vehicles. Finally, an optimization framework is implemented to decide the opening of several charging stations inside a highway, providing as possible locations rest and service areas already built, and considering current highway operational charging points.

Keywords: Electric vehicle, charging infrastructure, range estimation, facility location problem.

Contents

Chapter 1

Introduction

1.1 Context

Despite being at a relatively embryonic stage in terms of market penetration, Electric Vehicles (EV) present important benefits respect the Internal Combustion Engine (ICE) vehicles. The main advantages are in terms of reduction of greenhouse gas emissions in transportation, as well as improvements in air quality and reductions in noise pollution. Other potential benefits involve providing flexibility to the power system by vehicle-to-grid (V2G) , and therefore contributing to the integration of renewable energy. However, several challenges remain unsolved, and as a consequence EV is far from being chosen as the first choice by users.

The low number of charging infrastructures is one of the main problems, even though there exists a growing exponential trend since 2010, as shown in Figure 1. In addition to this problem, there exists a problem of standardization, including charging protocols, plug designs and billing systems. In fact, even within many European countries, various networks with proprietary identification and billing systems have emerged that do not yet allow EV drivers to roam between these networks. The diversity and incompatibility among these networks makes that EV owners can not use the full potential of EVs and that cross-border trips are virtually impossible [1].

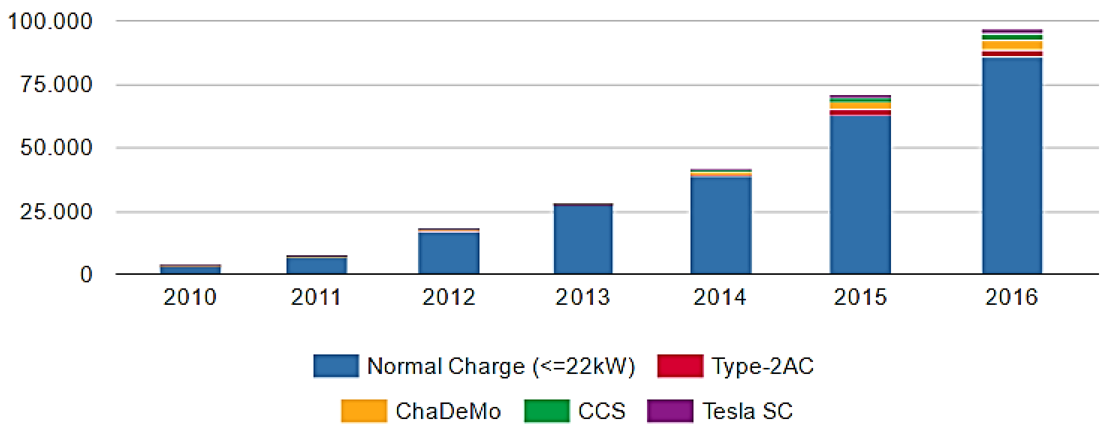

Figure 1: Evolution of the total number of EV charging positions in Europe. Source: [2]

What is more, the reduced autonomy of the EV respect to the ICE in terms of total range, makes more essential stations to stop and recharge the batteries. Both the number of available charging points and the EV range influence what it is defined as *range anxiety*, which can be described as the anxiety felt by many drivers about the remaining driving range their vehicle can run before the next charge. For this reasons, the electric vehicle has been positioned as a city car, primarily used in urban areas, although some manufacturers like Tesla Motors have cars that can reach up to 540 km.

On the other hand, data-driven strategies have come into view recently as they have the advantage of being economic and realistic comparing to traditional improvement on charging infrastructure and battery technologies, thanks to the dropping price of deploying Internet of Things (IoT). The increasing amount of sensory data can be gathered from in-vehicle networks and transmitted to the Cloud, where Machine Learning (ML) algorithms can be applied to useful applications such as range prediction, smart route planning, finding the nearest empty charging station or recommendations on energy harvesting.

In terms of sold units, the year 2015 saw the global threshold of 1 million electric cars on the road exceeded, closing at 1.26 million. This symbolic achievement highlights the efforts deployed jointly by governments and industry over the past ten years. In 2014, only about half of today's electric car stock existed. In 2005, electric cars were still measured in hundreds [3].

1.2 Objectives

The analysis performed on the optimal location of charging infrastructures for electric vehicles in highways respond to the following objectives, each of them presented in different chapters:

- **Model a electric vehicle range estimator**: Construct a Machine Learning model using different data sources to predict the actual range of an electric vehicle. This model will be used in posterior analysis to study the availability of charging infrastructure in French highways.

- **Propose a methodology for analysing the availability of charging infrastructure in highways**: Considering geospatial data from a given highway, and its nearest charging points, deploy a method in order to study the degree of electric vehicle penetration in highways. The obtained results can be used to decide whether or not it is necessary to invest in new charging points for a given motorway.

- **Construct a Location model of charging infrastructure in highways**: Build a model to find the optimal allocation of EV fast-charge charging points in highways, taking into account the availability study, the existing infrastructure and possible locations such as rest and service areas within the highway. The result should offer a grid that prevents a driver running out of battery, and at the same time reducing the number of points needed.

As a secondary objectives, state-of-the-art of EV applications used for facilitating the usage of electric vehicles will be revised, as well as the collection and integration of various data sources in order to conduct the described objectives.

Chapter 2

Background

In this chapter, electric vehicles (EV) and EV charging infrastructure are described as being the main features to consider for modelling the location of new infrastructures, as well as constructing a model capable of predicting the actual range of a vehicle given certain initial trip conditions. This chapter is structured as follows: first, the electric mobility, including EVs and charging infrastructure, is analyzed. The second section consists of a study of current state-of-the-art in range estimation is presented. The chapter ends with an analysis of current studies for locating new charging points in highways.

2.1 Electric Mobility

This section encompasses the main aspects of electric mobility, including electric vehicles, batteries for electric vehicles, data transmission approaches in this area, charging modes and finally charging infrastructure in France.

2.1.1 Electric Vehicles

Although the electric vehicle is considered a current concept, the first models emerged in the mid-nineteenth century. What is more, the EV appeared first rather than the conventional thermal vehicles, powered by internal combustion engines (ICE). According to [5], the first prototype was designed in 1835 at the hands of the Dutchman Sibrandus Stratingh. However, in the beginnings of XX century the ICE vehicle prevailed against the EV with improvements such as Ford cost reduction and limited access of electricity.

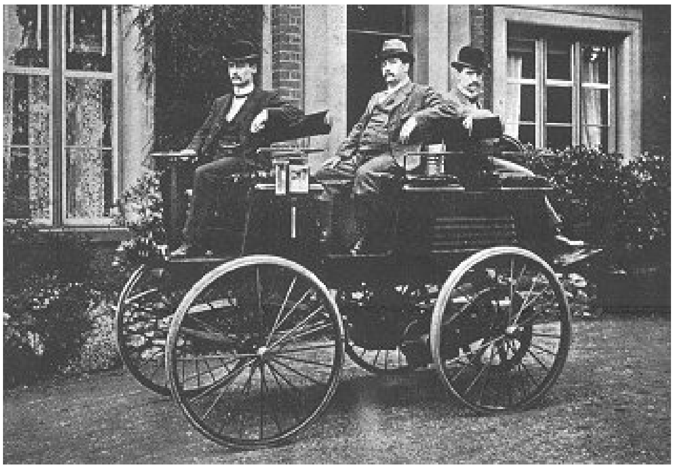

Figure 2: Thomas Parker Electric Car, 1880s. Source: [6]

This type of cars are powered by an electric motor, which converts battery electricity into kinetic motion. Essentially, this engine is composed by a fixed part, where the magnetic field is produced, and the rotor, a mobile component which moves within the induced magnetic field. Through this exchange of the two magnetic elements, the rotor begins to move. This movement is finally connected to the transmission system in order to transfer the movement to the wheels.

The advantages of this engine compared to the conventional internal combustion engine (ICE) can be summarized in three aspects: Recovery, efficiency and greenhouse gas reduction. First, recovery due to it can be used as a motor and also as a generator, allowing systems such as braking energy recovery. Regarding the efficiency, electric motors have a performance of between 93-99%, exceeding three or four times the ICE [4]. However, this rate of efficiency does not take into account that ICE takes energy from a primary energy source, whereas the electricity from the battery is obtained by national electric grids, i.e. it is generated in Power plants. Therefore, the conversion efficiency varies according to the type of central. Finally, EV emissions are given by how this electricity was generated: renewable energy provides a zero-emission type, whereas thermal centrals have its own rate of emissions given the type of fuel.

The electric vehicle can be classified in three main groups: hybrids, plug-in hybrids and battery electric vehicles. It is important to note that this project is focused on the thrid type, Battery Electric Vehicles, henceforth EVs.

- **Hybrid Electric Vehicles (HEVs)**: Powered by both petrol and electricity, even though the ICE is the principal engine. The electric part is only powered by systems such as regenerative brakings.
- **Plug-in Hybrid Electric Vehicles (PHEVs)**: It is a subtype of HEVs, but with two main differences: PHEVs have a higher autonomy than conventional HEVs, and more interesting, it is possible to plug them and charge the battery. The principal engine in this case is the electric
- **Battery Electric Vehicles (BEVs)**: Full electric cars, only powered by electricity. The disadvantage of this type of electric vehicles is its total range, significantly slower than the hybrid approach.

According to the International Energy Agency (IEA), the future prospects show a greater adaptation of the hybrid vehicles rather than BEVs in a scenario of short and medium term. Starting in the year 2030 sales of electric vehicles, with fuel cells and plug-in hybrids will be fired and will replace the ICE cars in 2045, as it is presented in Figure 3.

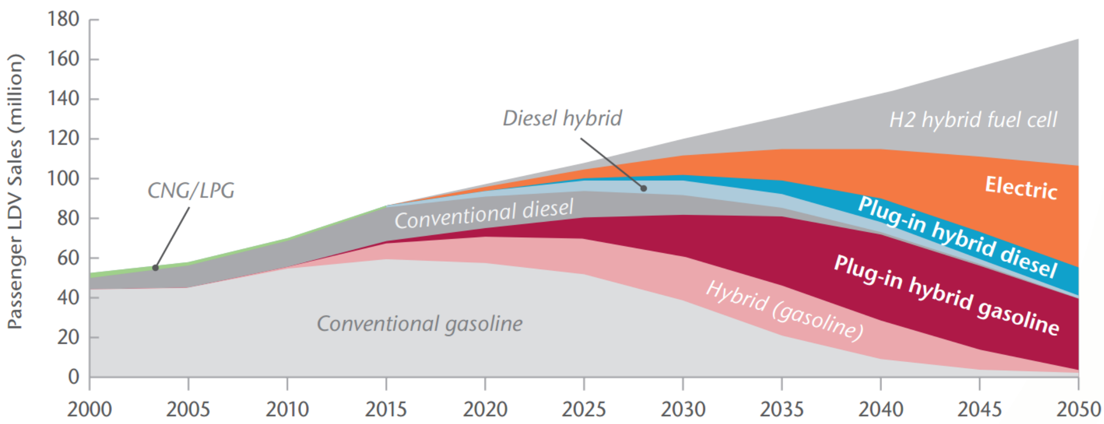

Figure 3: Future prospective of sales for different vehicle types. Source: [7]

The keys aspects to achieve these objectives include the following points:

- The cost and energy density of the batteries should decrease significantly for making the EV technology competitive against conventional ICE cars.
- Fundamental regulatory support in two areas: provision of adequate and standardized charge infrastructure; and assure a competitive cost of EVs versus internal combustion via financial aids.
- Collaboration of the involved stakeholders in order to accelerate VE implementation: industry and governments must work together on research, regulatory and adaptation programs to the grid infrastructure.

Finally, Table 1 presents a comparison between popular full electric vehicles currently available in the market. The range is indicative, as it can be influenced by a multitude of parameters (see section 2.2).

Table 1: Comparison between battery technologies. Sources: Vehicle manufacturers

	Battery type	Capacity [kWh]	Power [kW]	Top speed [km/h]	Range [km]
BMW i3	Lithium-ion	19	125	150	190
Citroen C-Zero	Lithium-ion	16	49	130	150
Mitsubishi i-MiEV	Lithium-ion	16	49	130	170
Nissan Leaf	Lithium-ion	30	80	150	200
Renault Fluence Z.E.	Lithium-ion	22	70	135	175
Renault Twizy	Lithium-ion	6.1	4	45	100
Tesla Model S 60	Lithium-ion	60	225	193	335
Tesla Model S 85	Lithium-ion	85	225	225	426
Th!nk City	ZEBRA	24	37	110	160

There is a clear difference between the models, as it is also in the prices. Some electric vehicles are designed to a city-level usage, as Renault Twizy, whereas on the other hand we can find the Tesla Model S, which can compete with ICE cars in terms of speed and range. However, this has a price: at the time of writing this thesis, the price of a Tesla Model S with 85 kWh battery pack is around 80000€and 120000€[13].

2.1.2 Electric Vehicle Batteries

The battery of the electric vehicle is possibly the most important component, in terms of functionality and cost. Its aim is to provide electricity to the mechanical engine. Lots of efforts are being made in order to reduce the energy density (Wh/kg) and energy density by volume (Wh/l), in order to fit the maximum energy in the minimum space and weight. The batteries also play an important role in terms of EV integration in the power systems, as it can be used to both store and provide energy to the electric grid with the objective of flatting the energy curve. This procedure is called vehicle-to-grid, also known as V2G [8].

Batteries are classified according to their chemical composition. The following list collects the features that differentiate the most outstanding battery types

- **Lead-acid**: Typical batteries from ICE cars. The advantages over other batteries are its large-scale industrial production, becoming a mature technology used more than 50 years ago, and its low cost. In contrast, it has problems such as not accepting discharges of more than 20% of its capacity without being affected its life cycle. It also has a low energy and power density due to the weight of the lead, besides being corrosive for the environment [9].
- **Lithium-ion**: Most used type in EVs. This type presents higher performance and better future prospects for EVs. They have a high energy density, good performance at high temperatures, have low self-discharge and are recyclable. In addition, they have a very low memory effect and a considerable power density with good life cycle. The main drawback of this type of battery is its price per kWh [9].
- **Nickel metal hydride (Ni-MH)**: It has twice the energy density than lead-acid and its components do not harm the environment. It also stands out for its high life cycle, the possibility of recycling, a high operating temperature range and resistance to loading/unloading. On the other hand, repeated discharge of high current reduces its life cycle and if it works discharging more than 20-50% loses capacity. Another negative aspect is its memory effect, which over time causes the battery not to reach 100% state of charge. [10]

- **Molten salt (ZEBRA)**: Relatively mature technology with a higher energy density than lead acid and Ni-MH. They have a high life cycle, have no memory effect and are not polluting. The main disadvantage is its average operating temperature, which is around 250ºC, having to isolate the system and always be charging or running to maintain the temperature [11].

The main characteristics of the described battery technologies are presented in Table 2.

Table 2: Comparison between battery technologies. Sources: [4], [12]

	Lead-acid	Lithium-ion	Ni-MH	Molten salt
Energy Density [Wh/kg]	20-50	80	110	100-120
Power density [W/kg]	80-100	200-1300	500	110-115
Operational Temperature [ºC]	-10 to 40	-20 to 60	-20 to 60	>240
Life cycle	700-800	2000	>2000	>600
Cost ($/kWh)	50-150	200	250 [12]	200

2.1.3 Electric Vehicle Data transmission

Data transmission is one of the key features for EV industry. Just to give an example, Intel's proposed $15.3 billion acquisition of Mobileye, an Israeli company that supplies car manufacturers with a computer-vision technology and advanced driver assistance systems. It uses a single camera, together with a proprietary computer chip and some clever software, to provide various advanced driver assistance features. Its systems can, for example, identify the speed limit from road signs, or identify vehicles and pedestrians for an automatic braking system [14].

At this moment, it is difficult to obtain vehicle variables such as the State Of Charge, SOC [%], and other real time data directly from the vehicle. A typical workaround is to physically connect with the CAN bus of the vehicle, but it is impracticable for inexperienced users. Several studies are based on monitoring a set of electric vehicles, as the one presented by Pavel Brandstetter et al. [19], or the performance analysis conducted by this author and CITCEA university group [18].

On the other hand, some car manufacturers such as Volkswagen offer their own private applications to obtain these type of data. The following list illustrates the different methods that the manufacturers are applying at this time:

- **Volkswagen**: Data transmission technology is called *Car-Net*, and it allows to inform about current SOC, real time traffic information and to search for interest places (restaurants, charging stations, etc.) [15].
- **Tesla Motors**: Their users have the Tesla unofficial Mobile API, with its own Github repository provided by Tim Dorr. It features functionality to monitor and control the car remotely. It consists of three main categories: Login/ Authentication module, Vehicle List and Information, and Vehicle Command. You can also relatively easily retrieve various settings and the status of the user's vehicle, including the charge state of the car, climate settings, the vehicle state, the driving and position features and finally it provides mobile access. However, the company does not claim this API as official [16], [17].
- **Ford**: MyFord mobile app, launched for hybrid vehicles, it supports the C-MAX Enrgi together with Fusion Energi hybrid cars. More features have been added to find charging stations, and can be used in combination with PlugShare. Data from Plug Share shows the station's status, like if it is free (currently not being used), the chargers at the charging station and the type of charger used there. The driver may also search various places of interests like restaurants, shopping malls or even gymnasiums using the MyFord app.
- **BMW**: The BMW i Remote App for iOS and Android, it shows you detailed information on the current status of your BMW i – e.g. car location, range, and battery charge. It is compatible with the BMW i3 and i8.
- **Nissan Leaf**: The Nissan LEAF app is designed to help in the management of Leaf vehicle and control many most features directly from the smartphone.
- **Audi**: Audi A3 e-tron Connect App connect services specially developed for the e-tron that enable the user to call up specific information and manage individual functions via smartphone and web portal. Using a smartphone app and the A3 e-tron platform, data on the vehicle status can be called up on aspects such as the current level of charge.
- **Toyota**: Toyota Entune App is a collection of popular mobile apps and in-car

data services accessible from a Toyota EV. These services are delivered via most smartphones using hands-free Bluetooth 9 and a cellular data connection.

2.1.4 Charge Modes

The process of charging EVs is one of the properties that most condition their development. Firstly, the time to charge the battery depends directly on the charging power, and it takes usually several hours if it is charged with standard household outlets. This fact is far from the few minutes that the ICE cars takes to fill the fuel tank. On the other hand, there are few fast charging points at this time. It is therefore necessary to develop a rapid charging mode that allows the waiting time and an investment in infrastructure to be used to boost the purchase of EVs. The most common way to charge the batteries is to connect the vehicle to the mains. However, loading methods are not fully standardized and therefore there are several options for loading vehicles.

There are four modes of charging technology, described in the IEC 61851. Each of them involve different combinations of power supplied by the electric grid, the plug types and the electric mode used: Direct current (DC) or Alternating current (AC). Note that batteries work with DC, therefore the owners may use a on-board AC/DC converter or a converter integrated into the charging point itself. The power level (kW) of the charging infrastructure depends on both the voltage and the maximum current of the power supply. The power level of charging points ranges widely, from 3.3 kW to 120 kW. Lower power levels are typical of residential charging points [20], [21], [1].

- **Mode 1 (Slow charging)**: Charge using conventional household sockets and cables, with AC current. Commonly found in domestic and office buildings.
- **Mode 2 (Slow or semi-fast charging)**: Again, it uses non-dedicated sockets with AC current. The difference lies in a special charging cable provided by the car manufacturer, with a protection device for electrical installations.
- **Mode 3 (Slow, semi-fast or fast charging)**: Also in AC current, but this mode uses a special socket and a dedicated circuit which allows charging at higher power levels, and ensures safe operation. It is typically found in stand-alone poles in public locations. The slow mode 3 includes supply power from 2.3 to 3.6 kW, the semi-fast at 22 kW and 32 A, and the fast with 43 kW and 63 A.

- **Mode 4 (Fast charging)**: This is the unique mode with direct current, DC. An AC/DC converter is located in the charging equipment, instead of inside the vehicle as for the other levels. The only drawback of this method is that it has lower efficiency, due to the double conversion DC/AC/DC. The supply power is typically equal or higher than 50 kW.

Just to give some numbers, charging 100 km with mode 4 and power supply of 120 kW will took around 10 minutes, whereas a power supply of 50 kW will account for 20-30 minutes. Using the other modes in AC current, the time can be increased to 6-8 hours for household sockets (3.3 kW), and 1-2 hours with public stations (22 kW) [20].

2.1.5 Charging Infrastructure in France

One of the keys to success for EV is to ensure user confidence in their driving range and safety. Reliable charging infrastructure that is backed by a national installation strategy is required to ensure sufficient driving range. France is planning to deploy this infrastructure in all sectors of daily life, in particular for the following groups [22]:

- **Enterprises**: Charging infrastructure will be installed for captive fleets of plug-in vehicles, such as corporate fleets. The possibility of "plug-in benefits" will be considered, such as allowing employees to recharge their personal or company cars at their place of work with low or no cost. Added power demand for charging would be managed.
- **Public domain**: Plug-in vehicles and charging infrastructure will also be deployed in public areas, such as roadways and public parking garages. Suitable options for use are being developed, such as shared vehicles and vehicles on demand.
- **Residential sector**: Plug-in vehicles and charging infrastructure will be made available to individual users, with or without vehicle ownership.

As of January 2016, France vehicle fleet was composed by 57.2% of passenger vehicles had diesel engines and 38.6% had gasoline engines. By the end of 2015, France was second behind Norway in Europe in registration of hybrid and electric passenger vehicles. The market share for plug-in hybrid electric passenger vehicles reached

3.3% (of which 0.3% are rechargeable), and 0.9% for fully electric vehicles. This is an increase of 27% compared to 2013. Regarding the vehicle manufacturers, in 2015 the Renault ZOE has a 60.4% of market share, the second position is for Nissan Leaf, with 12.9% of market share. Tesla has around 4.3% of market share [23].

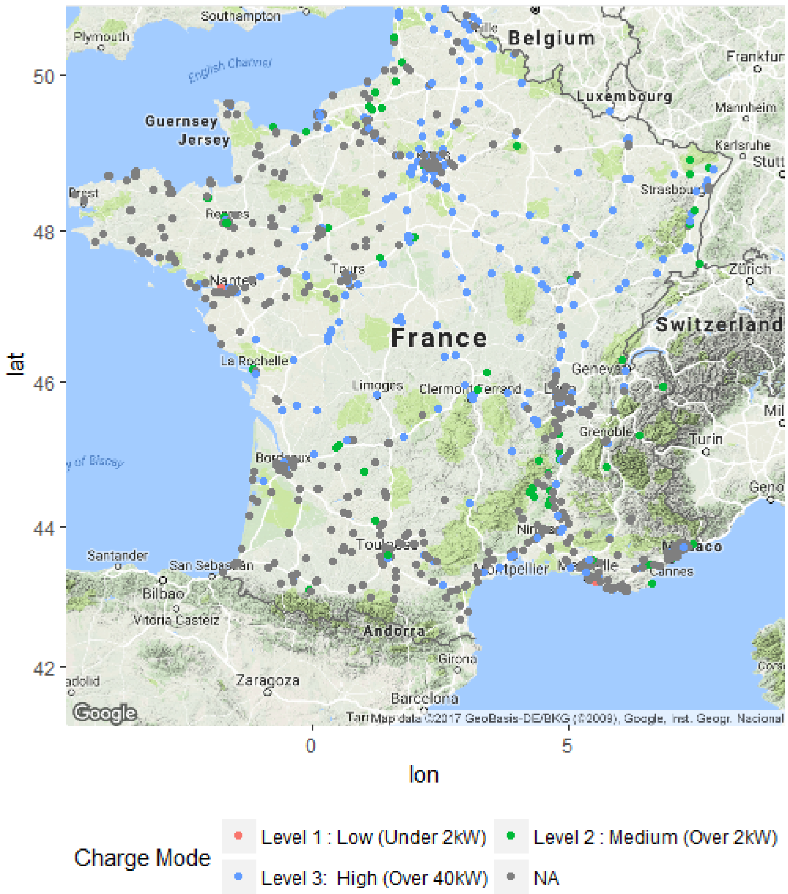

Figure 4: French charging infrastructure by charge type. Source: [25]

About charging infrastructure, the French government's objective is to reach 7 million charging points for hybrid and electric vehicles by 2030. Individuals who install a charging station at home will benefit from a 30% tax rebate. Table 3 shows the percentage of charging stations by location, according to ChargeMap.

According to European Alternative Fuel Obsevatory, there are actually 14.800 charging stations in France, where 1.543 of them are considered as high power (>22kW) [2]. However, it is not easy to have access about the current coordinates of these points.

Table 3: Breakdown of Charging Stations by Location. Source: [24]

Location	Relative frequency
Parking	23.4%
Road	23.0 %
Car Dealership	12.0 %
Home	11.1 %
Shop	9.4 %
Hotel	5.6 %
Company	4.7%
Fuel Station	2.7 %
Train Station	1.5%
City Hall	1.8%
Restaurant	1.2%
Others	4.1%

As it will be explained in section 3.1.3, OpenChargeMap [25] is used to gather this information. Figure 4 shows the location of the obtained infrastructure, also with the different modes of charge. Notice the high level of Not Available (NA) data in this feature.

2.2 Electric Vehicle Range Estimation

Even though there are lots of related work focused on the development and study of EVs technology and their operation with smart grids (for example [26], [27], [28]), the investigation of EV related data mining applications is practically non-existent. Another important field involving EV can be seen from the view of intelligent transportation systems and urban computing a, where sophisticated mechanisms to process heterogeneous data is required when the number of EVs that join into the transportation networks and connect to the cloud increases. Additionally, the security and privacy on EV-based cloud environment must be considered during mining EV related big data [29].

This section presents the current state-of-the-art for EV range prediction, as well as current commercial approaches, basically apps, that are used to improve the driver's experience.

2.2.1 Commercial Applications

There are several available applications for mobile users, not only for range estimation but also for charging infrastructure location, among other purposes. For that reason, this section tries to encompass all the apps that are being deployed for electric vehicle users.

Therefore, the applications for EVs can be divided in three types. The first one helps owners to find nearby charging points and are called charging apps. The second group are the monitoring apps, which helps EV owners to update the current state of their vehicles, and also control them for some applications. The last type are the informative apps, and give general information about electric cars [32].

There are several applications provided by the manufacturers, presented in section 2.1.3. The following list excludes these apps, as they are defined previously.

- **Better Place Oscar**: In case you want to contact customer service, the Better Place Oscar is the app for you as it enables you get assistance anytime. Additionally, the apps offers smart navigation, range forecasting and customised energy management options that will make you a proud owner of that electric car or plug-in hybrid vehicle [32].
- **GreenCharge**: enables you to check the range of the current battery before a charge is required. In addition, the app offers insightful metrics and battery data. Since this information can be accessed daily, weekly and monthly, you can effectively monitor your driving habits to further bolster your cost-saving efforts [30].
- **NFPA EV EFG**: Mobile emergency field guide used for electric and hybrid vehicle incidents involving damaged high voltage batteries, battery fires, extrication challenges, submersion, and charging stations. This app covers the vital aspects of EV/hybrid hazard awareness and procedures.
- **PlugShare**: Operates through a social network of electric-vehicle driving individuals. Through this network, drivers exchange information regarding the

18000 electric charging stations that are stationed around the United States and listed in the PlugShare network [31].

- **EV Ping**: EV charging stations are often under-utilized because a given station is being occupied by another EV driver. EV Ping enables drivers to easily communicate with each other to maximize EV charging. It provides electric vehicle (EV) drivers who need to charge their cars an easy way to communicate with each other to optimize charge station utilization using SMS and a simple mobile interface [31].
- **Open Charge Map**: This app also helps you find charging stations. This gives detailed information about what kind of charging station it is with pictures and other details so you can locate it easily. This app also gives you directions to the charging stations better than the other apps since it has an actual navigation function built-in.
- **Recargo**: Locate electric vehicle charging stations, check availability, share photos and tips.

Notice that the main part of applications can be classified as charging apps, whereas the monitoring apps are reserved to manufacturers private applications, as they ensure a reliable and safe connection between the vehicle and the device.

2.2.2 State-of-the-Art of Range estimation

Range anxiety, also known as the fear of losing power and seeing EV car shut down in the middle of a long-distance drive, is one of the biggest factors preventing more widespread adoption of electric vehicles [29]. A quick way to overcome range anxiety is through the wide deployment of battery charging points in a country or increasing the EV battery capacity [33].

There are several factors that may affect to the overall performance of an EV in terms of reducing the actual autonomy or range, and can be divided in four categories:

- **Vehicle status**: Initial State Of Charge (SOC) in %, speed, degradation of the battery, car model, vehicle weight, etc.
- **Driver behaviour**: Type of acceleration, braking, charging habits, etc.

- **External conditions**: Temperature, weather, wind, etc.
- **Trip characteristics**: Elevation, traffic, type of road, etc.

One example is the work presented by Habiballah Rahimi-Eichi and Mo-Yuen Chow [37], which integrates various data sources to construct the range estimator model, as it can be seen in Figure 5.

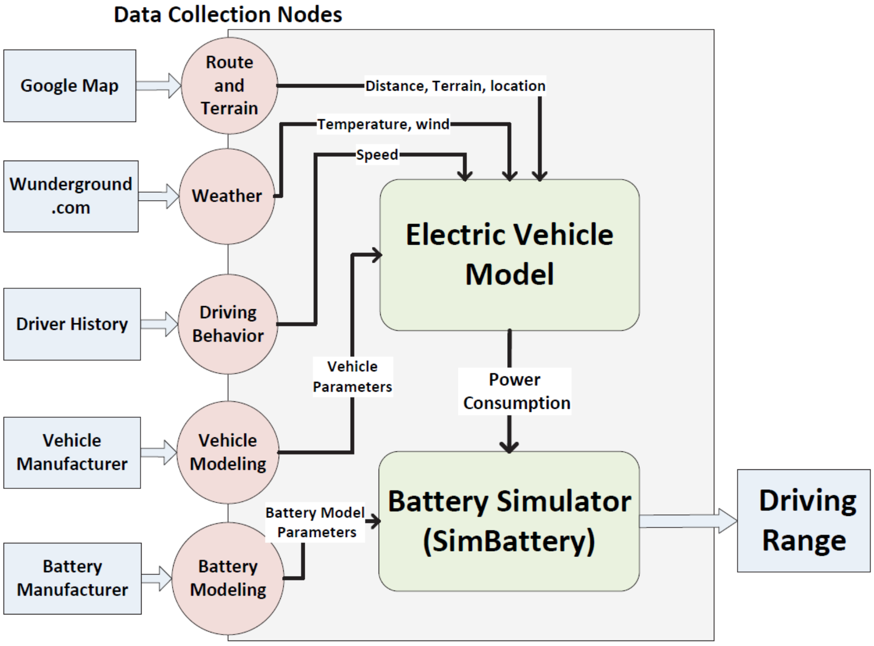

Figure 5: An example of range estimation framework. Source: [37]

Range estimators can be divided in two groups, depending on the time of prediction. The first group belongs to route planners, with the aim of estimating the autonomy of the vehicle given certain a-priory conditions of the trips. On the other hand, real-time range estimator models are used in on-board vehicle dashboards in order to vary the output of the model as the trip advances. The following list presents a literature review of the diverse factors that are considered when predicting this range:

- Bohan Zheng et al. [34] propose a hybrid Model built with modified Self-Organizing Maps (SOM) and Regression Trees to predict the power consumption of EV trips. They use a time-series dataset of EV drivers in North America and Europe with real-world commutes in terms of speed, distance, traffic conditions, hills and driving behaviour, containing 421 EV trips.

- Joonki Hong et. al [35] consider a two consecutive steps model:a driving profile estimation and a power consumption estimation using the power model. Among the variables, they introduce route information, slope and driving patterns, as well as traffic, current speed and weather.
- Daniela Chrenk et al. [36] propose a methodology to estimate the energy consumption of an electric vehicle based on road planning software. They introduce some trip data such as elevation using Google Directions API as well as vehicle current speed.
- Habiballah Rahimi-Eichi and Mo-Yuen Chow [37] presented a Big Data framework considering five types of public data for EV range estimation with various structures and resources, where the calculation relied on basic physical mechanisms.
- C. H. Lee and C.H. Wu [38] propose a big data analysis method with Machine Learning algorithms to process related data, but basically speed-energy consumption ratio data is fed to the cluster and the other related variables are implied by the speed time series.
- Eugene Kim et al. [41] present a real-Time model, with diverse data such as history of users power consumption, speed, and acceleration, as well as the road information from pre-downloaded map. The objective is to use the predicted power requirement to prevent the damage of battery cells that might result from high discharge rates.
- Y. Zhang et. al [39] considered several factors in their estimation method, including driving style (quick acceleration and fast driving), aggressive braking, temperature, weight, inclines and weather.
- A. Bolovinou et al. [40] proposed an online prediction system based on regression analysis methods focusing on time series data including distance, velocity and elevation.

There are other approaches that, even though the objective does not involve predicting the range, use Machine Learning (ML) models to facilitate this task. In [42], a pattern recognition approach is proposed to model the driving pattern according to the energy consumption of an EV. The growing hierarchical self-organizing maps (GHSOM) is applied to learn driver's behaviours gradually in the offline process, and the clustered neurons are used as the training sets for implementing online classifiers based on support vector machine (SVM).

2.3 Location of Charging Infrastructures in highways

This section presents an overview of the current projects and papers that tackle the root of this problem. It is important to consider that building a charging stations around highways from scratch would be too expensive, as the cost may include new roads, electrical installation and facilities, among others. For that reason, it is desired to adapt existing areas of service (fuel stations, rest areas, etc.) in order to include charging points, reducing the total cost of the installation.

In contrast to range estimation for electric vehicles, the location of charging infrastructures in highways is not a mature field. Hence, few literature was found at the time of writing this project. The main projects that have been found are presented below, analyzing not only the description and the approach, but also the critical issues, software used and input data, among other features.

European Commission: Optimal allocation of EV charging infrastructure in cities and regions, 2016

- **Description**: Propose a GIS methodology to provide locations for EV charging infrastructure, both in cities and in national networks (highway and rural roads). In the case of highways, only model one highway at a time.
- **Approach**: In highways use already build areas and gas stations to minimize additional investment costs. Very simple case, only 2 gas stations available per direction, without analyzing inter-correlation between highways
- **Critical issue**: Data availability
- **Software**: Open Source GIS tools, particularly QGIS along with the QChainage and MMQGIS open source plugins
- **Input data**: Road network (Open Street Maps) and Service areas.
  Real driving range: They use a JRC study which concluded that the maximum distance between charging stations should be 60 km, the two directions of the highway being independent.

- **Modelling**: Directions are independent, placement on already existing gas stations
- **Algorithm**:
  1. Create a geospatial file (shape-file) with the highway network, including direction.
  2. Create the shape-file with the location of gas stations along the highway (discard the ones that need to exit the highway).
  3. Calculate the distances between consecutive gas stations
  4. Check first whether the distances between gas stations are less than 60 km
  5. Additional condition: charging infrastructure has to be build at the first and the last gas station of the highway, since it is unknown if there is any charging station before and after the studied area.
  6. Check distance between consecutive gas stations. An area is marked as "suggested" every time the limit of 60 km from the last "suggested" area is exceeded.
  7. As a decision criteria, they use the minimum distance between stations

Deployment Methods For Electric Vehicle Infrastructure

- **Description**: Methodology for the spatial deployment of public charging stations for electric vehicles
- **Approach**: Three cases: public slow charging in a city area, public fast charging in a city area, and charging station along a major road.
- **Critical issues**: Much work will have to be invested in finding adequate GIS input data. Uncertainties in predicting the future electric vehicle market penetration, charging behaviours and market models
- **Software**: GIS-software ArcGIS (proprietary software)
- **Input data**: Road grid, electric grid, suitable stops, EV range

GIS-driven analysis of e-mobility in urban areas: An evaluation of the impact on the electric energy grid

- **Description**: Using GIS software, evaluate the potential of the BEVs to meet the urban mobility demand. Based on driving patterns collected from conventional fuel vehicles by means of on-board GPS systems. Only focuses on city level.

- **Results**: More than 80% of the urban trips are compatible with the performance of the current BEVs and HEVs, and that an urban fleet share ranging from 8% to 28% could be replaced by BEVs without significant changes n their driving patterns, resulting to be no more than 5% of the total electricity demand in the analysed areas.
- **Other publications**: Customer-driven design of the recharge infrastructure and Vehicle-to-Grid in urban areas: A large-scale application for electric vehicles

Chapter 3

Data Framework

This chapter presents the several data sources considered to develop this project. Due to the wide range of fields treated in this this work, it was a must to search and process different kind of data, such as electric vehicle trip information, battery behaviour over temperature or elevation, highway geospatial data and available charging stations, among others. Considering that this project had no previous work done, data gathering could be considered the critical phase: first for the difficulty of finding open and reliable data, and second for directly affecting to the accuracy of the present models and simulation.

The case of study of this project is focused on France highways, concretely in Autoroute A10. For this reason, highway data and charging infrastructure are referred to France national level.

All the steps involving data gathering, data cleaning, and data integration are done using the statistical open software R.

3.1 Data Collection

This section is divided in the three main data sources required for developing this project. Location data about EV infrastructure and highway areas are used in order to perform the simulation to study the situation of EVs in highways, and also to perform the optimization model for allocating new charging stations. On the other hand, electric vehicle usage data are needed in order to implement the range estimator model.

3.1.1 Highway geospatial data

Highway data were acquired via web scraping from Saratlas website. Saratlas [43] is a free database that provides information about the French expressway and motorway, called Autoroutes, such as their length, their exits and infrastructures. Again, data is entered by volunteers, so there is no guarantee of its correctness. However, it is an active forum and there exists a community behind, so the information is constantly updated and verified.

The R package *xml2* was used to retrieve the data from Saratlas website. The first step was to obtain all available highways, identifying the desired node returned from XML response. However the smallest branches were discarded due to its short driving lengths. Next, a function was build to copy and clean the table that Saratlas presents for each Autoroutes. Text preprocessing and data cleaning have been applied, as not only the table is gathered from Saratlas, but also the other information contained in the page. Then, the function is applied to all the retrieved Autoroutes, and the data is grouped into a unique matrix.

Finally, all the tables were combined in order to retrieve every rest area and service area from France highways. The final table include the following features:

- **Identifier**: Name of the Point Of Interest (POI)
- **Type of Area**: Categorical data. Two main classes, service area and rest area. It also include the direction of the highway, as some areas are only accessible in one way, left or right, and others are accessible by both ways
- **Location**: Kilometre of the highway where the area is located
- **Latitude**: North-south geographical coordinate
- **Longitude**: Est-west geographical coordinate
- **Highway**: French Autoroute name (e.g. A1)

Apart from areas, it is also interesting to gather information about the other Points Of Interest (POIs) from the highways, in order to use it in the charging infrastructure availability study. For that reason, an extended version of the previous table is created, incorporating junctions, which represents the entries and exits for a given highway.

3.1.2 Electric Vehicle Range

This data source is crucial in order to construct a model to predict the EV range, and at the same time it is the less accessible data. The search across Internet and well-known data repositories, such as USA government [45], was not successful.

The first approach to solve this issue was to use the iThink EV ZEBRA dataset, studied in detail during the author's Bachelor thesis. However, this dataset belongs to a EV fleet of car-sharing, and the majority of the trips are in a city level, with low percentage of discharge (initial SOC minus final SOC), and small distances. Bearing in mind that the objective is to predict the full range of a vehicle for doing long trips in highways, the final decision was to discard this dataset.

Instead, data from Nissan Leaf owners was used to create the range model. Two different data sources were used to build the model:

1. **MyNissanLeaf** [47]. Forum with approximately 20.000 users. The data was initially collected by Tony Williams and QuickChargePower LLC. As this data is only collected by one user, it can not be assured that it applies to all Nissan Leaf vehicles, and other EV models. The data was compiled with empirical methods over a multi year period. It consists of a four tables, representing the degradation of the battery (0%, 10%, 20% and 30%). For each table, the rows represent the level of SOC [%], and the columns the speed level, with minimum of 72 km/h and a maximum of 136 km/h. All the values contained in the table are the actual range.
2. **Fleet Carma** [48]: This data contains the actual driving range of the electric vehicles given several levels of temperature, from -25ºC to 35ºC. All of the results were created from FleetCarma real-world vehicle loggers installed on vehicles all across North America. The data corresponds to the study of 7.375 trips of

Nissan Leaf cars.

3.1.3 Charging Infrastructure

There are several applications, web-sites and communities that offer world-wide geospatial data about location of charging stations. However, at the time of doing this thesis there is only one repository that actually allows to download the data under an Open Data license, called Open Charge Map [25].

Given that the information from Open Charge Map is provided and validated by its users, it is easy to see that their system does not include all charging stations. This problem could be partially solved crossing data from other sources, but no other application offers the possibility of downloading data without paying a variable fee.

There are actually 14.800 charging stations in France, where 1.543 of them are considered as high power (>22kW) [2]. However, Open Charge Map only includes 1.184 stations, whereas one of its competitors, Charge Map, collects 7.679 points, even though this data is not accessible. This huge difference may affect the model accuracy, as it will not be possible to include all the current charging infrastructure. On the other hand, improving the model with more data would be trivial considering that the model primarily focuses on the geolocation of the stations.

Data were obtained through an Open Charge Map API GET request, using *R* as a programming language and its package *httr*. Even though Open Charge Map provide several features of each charging station, only the relevant ones were kept, listed below:

- **ID**: Identifier of the charging station
- **Address**: Street, postal code and city
- **Latitude**: North-south geographical coordinate
- **Longitude**: Est-west geographical coordinate
- **Number of points**: Discrete data. Quantity of charging points per station
- **Usage**: Categorical data. Levels include Private (costumers), Private (restricted access), Public, Public (membership required) and Privately Owned

- **Charging Type**: Categorical data. Levels include Low (under 2 kW), Medium (over 2 kW) and High (over 40 kW)
- **Status**: Categorical data. Levels include operational, not operational, planned for future, partly operational and unknown
- **Last status update**: Date of the last update

3.1.4 Trip information

To build the simulation data, as well as combining data from Autoroutes and charging stations, it is necessary to obtain the distance, duration and elevation between a pair of coordinates. For this purpose, Google APIs where used, with a limit of 2.500 calls/day (75.000 calls/month). Concretely, the following features where extracted:

1. **Elevation between coordinates**: Obtained using Google Elevation API [49]. The API response has JSON format, and it is processed with R to obtain the elevation in meters.
2. **Distance between coordinates**: Obtained using Google Distance Matrix API [49]. There are several parameters to be specified in the API call, in this case the *type* was set to *driving*. It also cames in JSON format.
3. **Duration of a trip**: Obtained using Google Distance Matrix API [49]. This variable is obtained together with Distance between coordinates, taking advantage of the same call to gather both components.

3.2 Data Cleaning

The next step is to transform the data adding or deleting variables in order to prepare it for the modelling and simulation phases.

Charging infrastructure data contains all the charging points of France. However, giving the case of study of this project, home and slow charging points are not suitable for charging an EV in a highway. For that reason, the next considerations were applied to filter this dataset:

- The NAs in the variable number of points were set to 1, as it is the most frequent value in this feature.
- Remove points with status "Planned for future date" and "Not operational".
- Remove points with usage "Private - Restricted Access".
- Remove points with charging type "Level 1 : Low (Under 2kW)" and "Level 2 : Medium (Over 2kW)". Then we are only considering fast charging, with Level 3: High (Over 40kW).
- Remove points inside Paris, as they are far from highways. This is done to reduce the computational cost and avoid Google API limits in the data integration phase among charging points and highway information.

3.3 Data Transformation

The data sources described in section 3.1 are combined and/or converted in order to develop the proposed range estimation model, the study of charging points availability for EVs in highways and the location of new charging infrastructure. Due to the nature of the data sources, data transformation is divided in two parts: one regarding the geospatial data from highways, and the other to electric vehicle data.

3.3.1 Available Charging Infrastructure in Highways

This section describes the procedure followed to combine the charging point and highway data sources. The idea is to obtain for a given highway not only the within charging infrastructure, but also near infrastructure that could be used in long trips. Figure 6 shows the location of the obtained infrastructure inside France, and the highway points of interest (POI). Plotted using R packages ggmap and ggplot2.

The distance between pairs of highway Point of Interests (POI) and charging points is used to decide whether or not a charging point is considered accessible for a given highway. Given the size of the data, it is not feasible to calculate the distance for all the possible combinations, as it will result in more than 150.000 API calls, with a limit of 2.500 calls per day. To solve this issue, first only junctions of the highway are

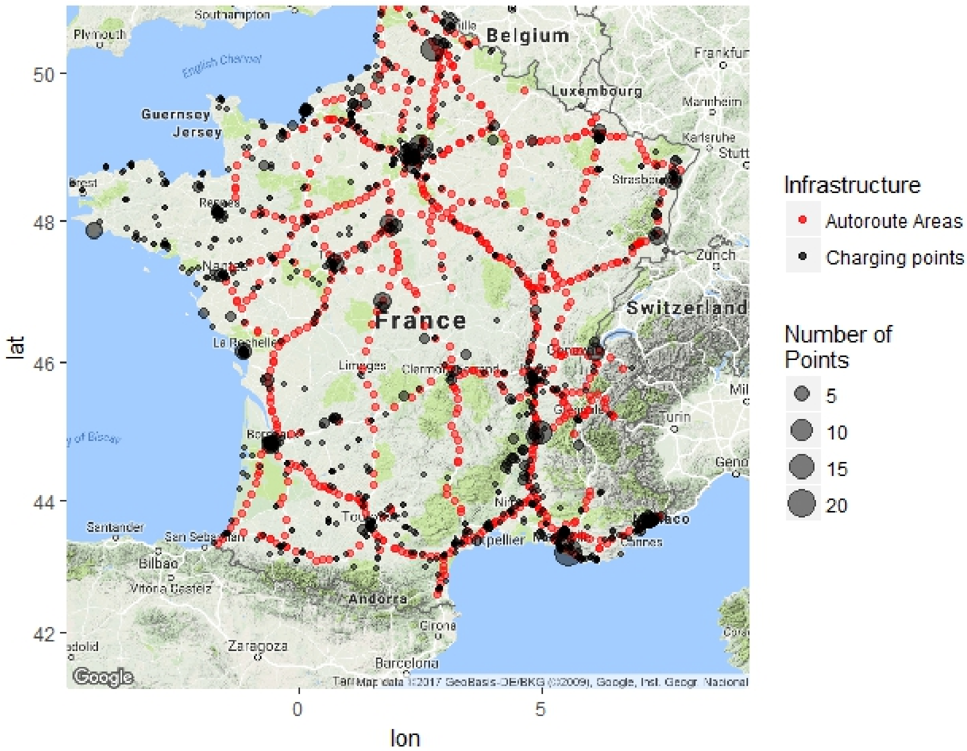

Figure 6: France charging points and Autoroute POIs. Source: Own

considered, and second the Haversine distance is computed using *distm* function in geosphere R library. Haversine distance considers the great-circle distance between two points on a sphere given its coordinates. For this reason, the distance computed with this method will be always smaller or equal to the driving distance given by Google Distance Matrix API.

Next, the charging points are filtered to consider the points that are near to highway POIs. The threshold is defined in 5 km, taking into account that the driver would not want to deviate that much from the initial route. This process will discard the major part of charging points, obtaining a reduced subset. Then, the same procedure of comparing pairs of Autoroute POIs and charging points is applied, but in this case with the real driving distance provided by Google Distance Matrix API. Finally, the results are filtered again by the 5 km threshold. Figure 7 shows the described procedure, where the A_i represents a concrete Autoroute.

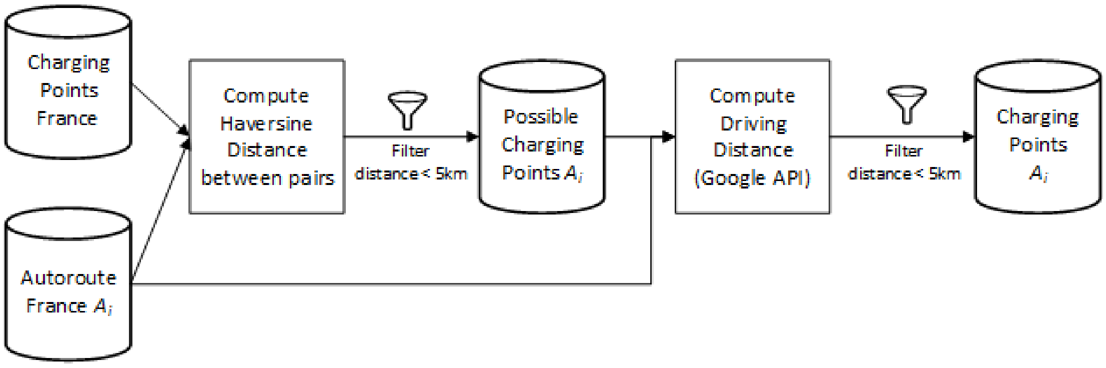

Figure 7: Integration of charging points and Autoroute POIs. Source: Own

3.3.2 Electric vehicle data

Data collected from Nissan Leaf forum does not present a suitable format to apply a statistical method for predicting the range given more features rather than initial State Of Charge and mean speed. For this reason, the data is transformed into a matrix with multiple variables, where each row stands for a journey. The columns will represent the different features of each trip, including degradation, available State Of Charge [%], speed, temperature, elevation and range.

The tables of equivalences between SOC and speed does not consider the effect of temperature nor elevation. The data will be adjusted in order to consider these variables. Furthermore, the four tables respect the levels of degradation will be combined in an unique matrix.

The first step is to build models in order to adjust the range with a given temperature and elevation:

- **Elevation**: Unfortunately, after reviewing several papers and other sources, presented in subsection 2.2.2, it was not possible to find data or determine a function to model the range given an elevation. For that reason, the adjustment empirically calculated from Tony Williams provided with the documentation of the data, will be used. It recommends to use a model that reduces 1.5 kWh from the battery every 300 meters of elevation. For this reason, a linear model is created to consider the elevation effect, considering that the results may not have a high degree of accuracy.

- **Temperature**: In this case, it was possible to collect data from Fleet Carma [48]. Normalized Root Mean Squared Error was used to decide the final model.

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{(N-1)var(y)}}$$

  Polynomial linear regression is selected to model the effect of temperature in range, as we only need to make an adjustment to the range before building the final model, and more sophisticated methods can lead to over-fitting or add to much complexity. Table 4 present the obtained results.

  Table 4: Results of modelling Range against temperature

Degree	1	2	3	4	5	6	7	8
NRMSE	0.621	0.508	0.359	0.317	0.308	0.308	0.307	0.3039

  It can be seen that after polynomial degree 4 the model does not get significantly better, and including such irrelevant variables leads to unnecessary complexity in the resulting model [46]. For that reason, degree 4 is chosen. Figure 8 illustrates the fitted model.

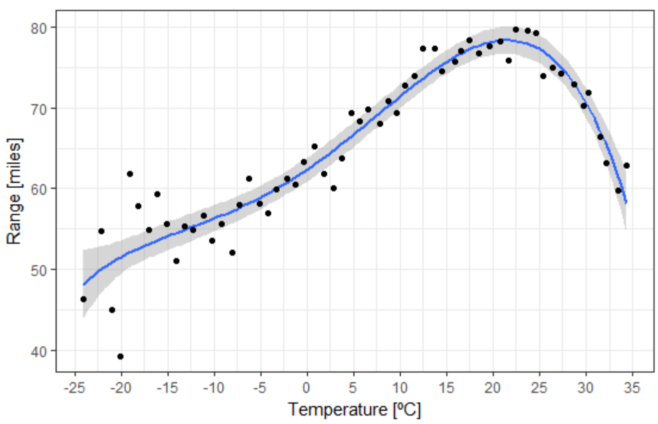

Figure 8: Range vs Temperature and the fitted Polynomial degree 4. Source: Own

  Finally, the model is not fitted using the original range, but using a normalized version. The reason of this adjustment is to avoid the both the units (in miles) and differences between the data collected by Fleet Carma and Tony Williams. Range is normalized dividing by the maximum range, considering that the minimum value is 0 miles, getting a percentage between 0.49 and 1, meaning the reduction of range for a given temperature level, with the maximum located at 22ºC.

After creating the models to adjust the range provided by Tony Williams using different values of temperature and elevation, it is possible to transform the data into the unique matrix.

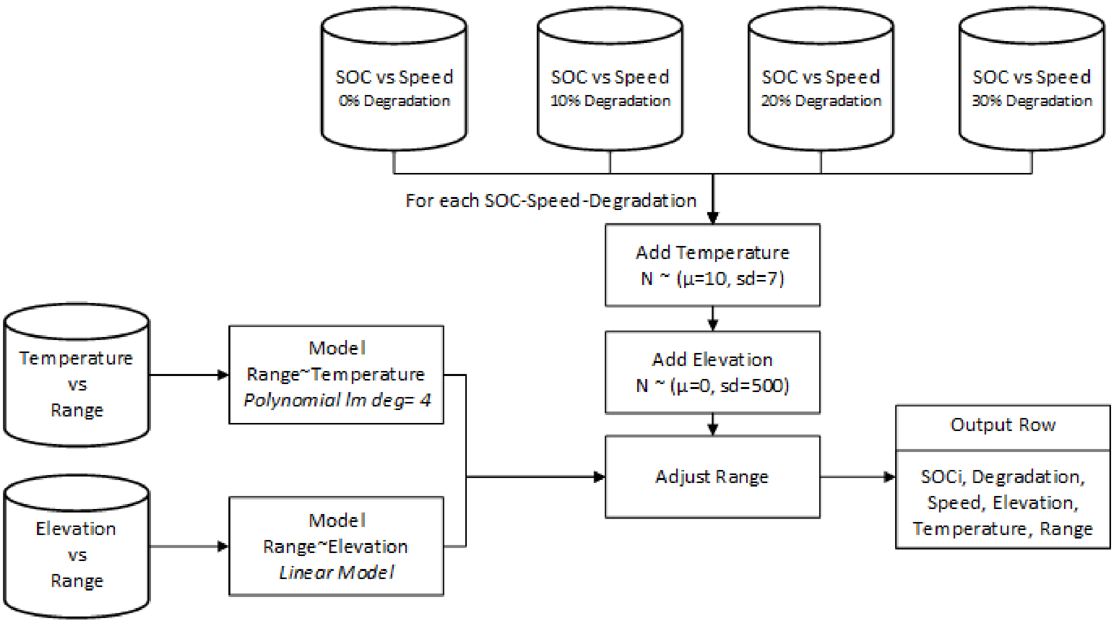

Figure 9: Transformation of EV data. Source: Own

The procedure is described in Figure 9, and has the following steps:

1. Iterate through the four tables, taking one row at a time
2. Iterate through the columns of each table, gathering the speed level
3. Write the variables into the new matrix: Degradation, initial SOC, speed and battery capacity (kWh).
4. Generate values for temperature with a random variable with Normal distribution, with mean = 10 and standard deviation of 7 (values from -11ºC to 33ºC).
5. Generate values for elevation with a random variable with Normal distribution, with mean = 0 and standard deviation of 500 (values from -1300 to 1500 meters). This values are feasible given the elevation extremes of France [52], considering

that the elevation represents the difference between origin and destination height from sea.

6. Modify the actual range applying the presented methods for temperature and elevation. In the elevation case, an interpolation to get the range is applied, as the model predicts the capacity variation (kWh) against elevation.
7. Write Temperature, Elevation and Range row into the new matrix.

Chapter 4

Electric Vehicle Range Estimation

This chapter presents the methodology for predicting the range of a vehicle given a set of initial conditions for a trip. The initial idea of this project was to consider various vehicles and users to develop the model, in order to get a prediction that considers not only trip variables, but also the type of the vehicle and user behaviour. However, the only available data that was found come from just one user and one vehicle, the Nissan Leaf.

4.1 Exploratory Analysis

The data obtained from the transformation phase is in the desired format to apply an statistical method in order to predict the range. The following features are used:

- **Degradation**: Percentage of degradation of the battery. Categorical variable with four levels: Zero (0%), Low (10%), Medium (20%) and High (30%).
- **SOCi**: Initial State Of Charge. The available range represents the full distance until the vehicle runs out of battery. Continuous variable, from 100% to 7%.
- **Speed**: Mean speed of the trip. As the study is focused in highways, the range of this variable is considered from 80 km/h and 130 km/h (minimum and maximum speed limits in french Autoroutes [51]).

- **Temperature**: The initial data does not contain information about temperature. The range is from -11ºC to 33ºC. The seed was set to 27 for reproducible purposes.
- **Elevation**: The initial data does not contain information about elevation. Range from -1300 to 1500 meters.
- **Range**: Label to predict. Continuous variable with range from 2 to 215 km. The maximum value is lower than the maximum range of 250 km provided by Nissan [53] but higher than the range provided by the United States Environment Protection Agency (EPA) [54]. The values are adjusted to consider the effect of the temperature and elevation in each case.

Figure 10 presents a grid of plots relating the different levels of degradation and mean speed with the range and initial SOC. The elevation is also differentiated using three levels: *Descent* from -1500 to 200; *Flat* from 200 to 200; and *Climb* from 200 to 2000 meters. It is easy to see graphically the effect of the variables to the final range, which decreases with more degradation, elevation and speed.

The maximum likelihood framework can yield unstable parameter estimates if the explanatory variables are highly correlated. A correlation analysis is also conducted between pairs of continuous variables, obtaining a correlation of 0.996 between initial SOC and battery capacity (kWh). This result was expected, as we are only modelling one vehicle model, Nissan Leaf. For that reason, **capacity was discarded** to fit the model. The other continuous variables have correlations lower than 0.13.

4.2 Metric Selection

Normalized Root Mean Squared Error (NRMSE) is used as a metric to decide the final model. NRMSE is a risk metric that corresponds to the normalization of the expected value of the squared (quadratic) error loss. Exists the assumption that errors are unbiased, and follow a normal distribution. It is considered as an excellent general purpose error metric for numerical predictions [55]. The RMSE without normalization is used in competitions such as Kaggle Home Depot Product Search Relevance [56].

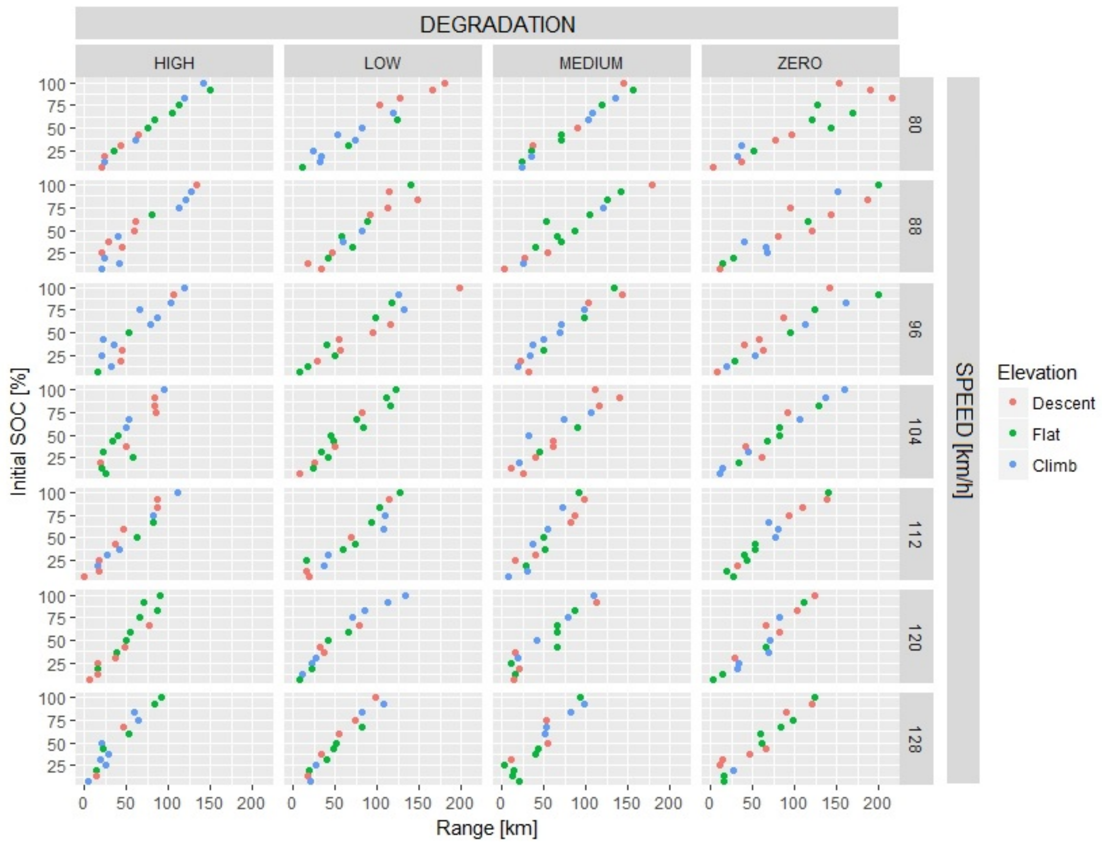

Figure 10: Effect of degradation, speed and initial SOC to range. Source: Own

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{(N-1)var(y)}}$$

Normalizing the RMSE we gain interpretability, as the error lies between 0 and 1, where 1 stands for a model that always outputs the average of the target, and 0 a model with no error.

4.3 Model Selection

The models that will be used to fit the data are Support Vector Machines (SVM), and Multiple Linear Regression (MLR):

4.3.1 Multiple Linear Regression

Is one of the most commonly used statistical techniques. Some of the advantages are that is a flexible model, and it is relatively easy to interpret the model coefficients. There are four assumptions. First, it assumes that the errors are normally distributed; second, that the errors have constant variance; third, that the mean of the errors is zero; and finally that the errors are independent [57] [55]. If Y is the response variable, and X [1..p] the predictors, it can be formulated as:

$$Y = \beta_0 + \beta_1 * X_1 + ... + \beta_P * X_P + \epsilon$$

4.3.2 Support Vector Machines

Introduced by Vladimir Vapnik and his co-workers at the Computational Learning Theory (COLT) 1992. This method has several advantages: it contains a regularisation term, which may help to avoid over-fitting; it uses the kernel trick, so it is possible to model non-linear relations; SVM is defined by a convex optimisation problem, i.e. it has no local minima; SVM also presents sparseness of the solution and capacity control obtained by acting on the margin, or on number of support vectors [58]. Even though SVM is sometimes only related to classification problems, it also works in regression, where a non-linear function is supported by linear mapping into a high dimensional kernel induced feature space. SVM for regression are based on defining the loss function that ignores errors, placed within the certain distance of the true value. This type of function is often called epsilon intensive loss function [59].

Formulation: The input X is mapped into a m-dimensional feature space using a kernel mapping, which can be non-linear, bext a linear model is constructed in this feature space. The linear model $f(w, \omega)$ is given by:

$$f(x, \omega) = \sum_{i=1}^{m} w_i \phi(x) + b$$

Where $\phi(x)$ denotes the set of non-linear transformations, and b the bias term. Assuming a zero mean in the data (achived by scale=T in R), the bias term is dropped.

The quality of the estimation is measured by the following loss function, called ϵ-intensive function, proposed by Vapnik:

$$L(y, f(x,\omega)) = \begin{cases} 0 & |y - f(x,\omega)| \leq \epsilon \\ |y - f(x,\omega)| - \epsilon & otherwise \end{cases}$$

And the empirical risk:

$$R_{emp}(\omega) = \frac{1}{n}\sum_{i=1}^{N} L(y_i, f(x,\omega))$$

Finally, SVM regression also tries to reduce model complexity by the minimization of $||\omega||^2$, measuring the deviation of training samples outside ϵ-intensive zone with the non-negative slack variables ζ_i amd ζ_i^*. The minimization problem can be formulated as:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{N}(\zeta_i + \zeta_i^*) \\ \text{subject to} \quad & y_i - f(x_i,\omega) \leq \epsilon + \zeta_i^* \\ & f(x_i,\omega) - y_i \leq \epsilon + \zeta_i \\ & \zeta_i, \zeta_i^* \geq 0, i = 1, \ldots N \end{aligned}$$

As it can be seen, one of the drawbacks of SVM Regression against MLR are the number of hyper-parameters to be tuned, in this case 3: The penalty factor or cost, C; the epsilon, which controls the width of the ϵ-insensitive zone; and gamma (γ), which controls the trade-off between error due to bias and variance in your model. However, it is also necessary to specify the basis function to perform the kernel transformation in the feature space.

To compare the performance of different basis functions, SVM will be tested modelling with four kernels:

- Radial Basis Function kernel (RBF): Gaussian kernel, polynomial of infinite degree with the form $k(x, y) = exp(-\gamma||x - x'||^2)$

- Linear: Simplest kernel function. It is given by the inner product $< x, y >$
- Polynomial: Cubic transformation will be tested. Kernel with the form $k(x, y) = (\alpha x^T y + c)^d$. Adjustable parameters are the slope alpha, the constant term c and the polynomial degree d.

4.3.3 Validation Model

In order to measure the performance of the models, the data is divided in training set with 2/3 of the total data, and test set with the remaining information. MLR and SVM will be trained using the first set, but the criteria to chose the best method will take into account the achieved NRMSE in the test set. The idea is to try to avoid over-fitting, i.e. models that adapt very well to the training data but do not achieve the same accuracy predicting new data points.

4.3.4 Cross-Validation

Support Vector Machines for Regression are a more flexible method, which allows to apply nonlinear transformations. However, there are several parameters to be tuned, such as the cost C, γ and the ϵ. For that reason, a cross-validation to the training set will be applied to select the best combination of parameters. In order to define the best hyper-parameters of the SVM, a k-fold cross-validation is performed. The number of folds is set to 5 to reduce the computational time, even though a smaller value of k could lead to biased results [60].

Table 5: Search range of SVM parameters

Cost	γ	ϵ
0.25	0.008	0.008
1.00	0.031	0.031
4.00	0.125	0.125
16.00	0.500	0.500
64.00	1.000	1.000

Table 5 shows the selected values to tune the SVM. The range of values was chosen considering the projects done during the MIRI Master, and literature review such as [61].

4.4 Results

4.4.1 Multiple Linear Regression

This method is applied in R using the base R function lm. This functions directly deals with categorical variables, so there is no need to dummy code the Degradation factor. We can not discard the hypothesis that all the predictors affect to the response variable, as p-values are in the worst case lower than 7.15e-15. The obtained Multiple R-squared is 0.948, and the adjusted R.squared 0.9467. The model can be validated looking at its residuals, which may have zero mean, independent, normally distributed and with same variance. For that reason, Figure 11 is presented.

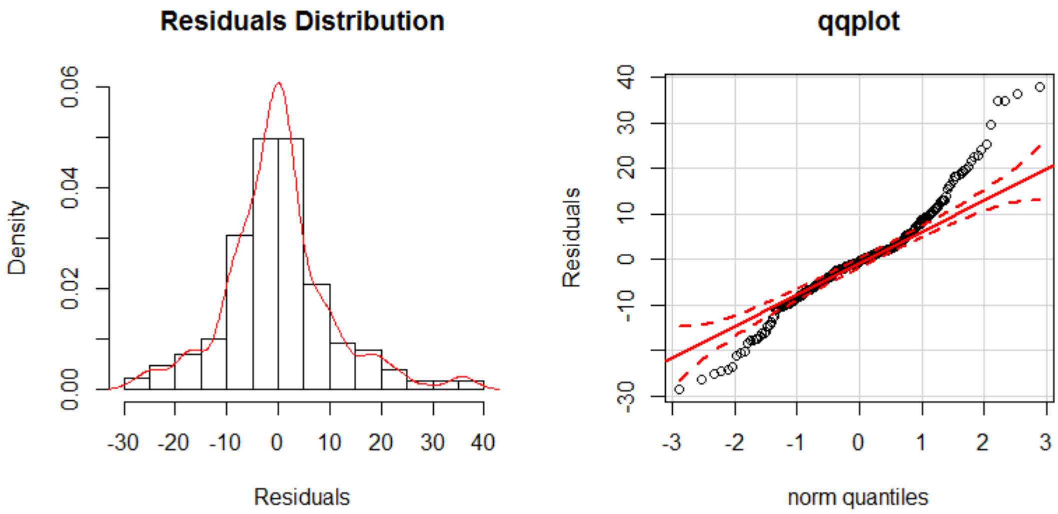

Figure 11: Validity check of Multiple Linear Regression. Source: Own

Even though the residuals may look Gaussian, the tails of the quantiles plot (qqplot) are far from the assumption.

4.4.2 SVM

Support Vector Machines are trained using the *e1071* R package. Several assumptions are taken into account:

- The categorical variable degradation is dummy coded into binary (0,1) features, as this function only allows continuous predictors
- The predictors are scaled to have zero mean
- *tune* function from the same package is applied to do the grid search of the values presented in section 4.3.4
- *tune.control* function is used to perform the 5-fold cross-validation, inside the grid.search part.

4.4.3 Model comparison

After selecting the best hyper-parameters for each SVM kernel model, it is possible to compare the results in terms of the Normalized Root Mean Squared Error. Results are presented in Table 6 .

Table 6: Search range of SVM parameters

Model	Train NRMSE	Test NRMSE
Multiple Linear Regression	0.22783	0.21829
SVM Radial Kernel	0.02466	0.04227
SVM Linear Kernel	0.23015	0.21988
SVM Polynomial degree 3	0.15805	0.24385

As it is explained in section 4.2, test NRMSE will be used to chose the best model. On the other hand, comparing the traning and the test errors we can have an idea about the generalization of the model, i.e. if it tends to over-fitting. This phenomena can be detected in the cubic SVM (polynomial degree 3), where the test error is significantly higher than the training error. The other fitted models have a theoretically good generalization, even though the RBF test error is the double than the training one, but it has to be taken into account that the achieved precision with this method is really high. Another aspect to comment is that, as it was expected, the results from the multiple linear regression and the linear SVM regression are very similar.

In conclusion, RBF is the model with the best metric with difference. In fact, the reached NRMSE is really low compared with the other models.

4.4.4 Final Model

The selected model is the RBF SVM, which got the best metric compared to the other evaluated models. It has the following parameters:

- Type: ϵ-regression
- Kernel: Radial
- Cost: 64
- Gamma (γ): 0.0312
- Epsilon (ϵ): 0.0078
- Number of Support Vectors: 203

One particularity is that the number of support vectors is significantly high. This could be explained as we are training the model departing from a table of relations of SOC versus speed, i.e. not real-world data with similar cases. However, analysing the prediction in the test set, we can see that the model generalize really well, as it can be seen in Figure 12.

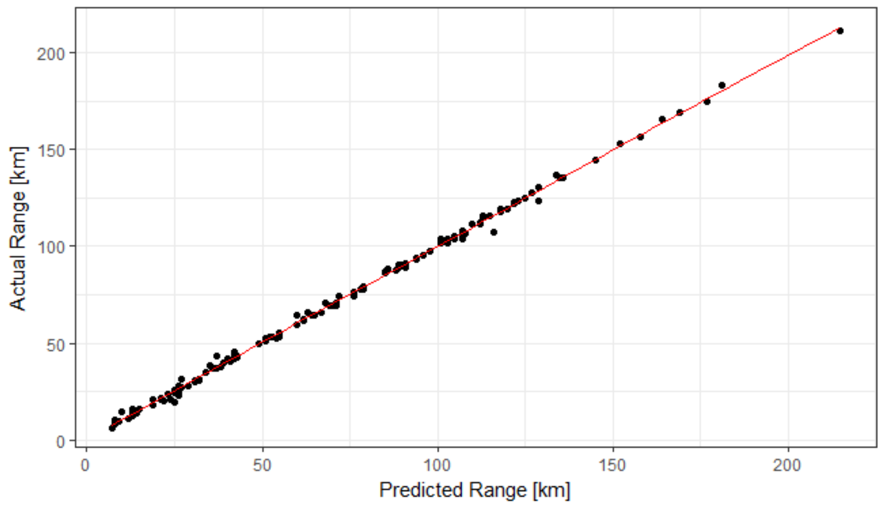

Figure 12: Scatter plots of Actual vs Predicted Range values from RBF SVM. Source: own

Scatter plots of Actual vs Predicted are one of the richest form of data visualization. Ideally, all the points should be close to a regressed diagonal line, marked in red, somehow visualizing the Goodness of fit of the model: the more foggy or dispersed the points are (away from this diagonal line), worse is the prediction. It can also be seen that the residuals are homoscedastic, i.e. they have the same variance across various levels of the dependent variable.

Chapter 5

Charging Infrastructure Availability Analysis in Highways

This chapter presents a methodology for evaluating the availability of charging stations in highways for electric vehicles. One of the the major impediments for EV owners is the limited number of fast chargers, also known as Level 3 and Level 4 chargers, capable of providing a full charge in question of minutes. This issue combined with the reduced range of EVs respect to internal combustion engine vehicles, aggravate the problem. For this reason, it is important to know the state of a given highway with the objective of deciding if it is necessary to invest in new charging stations, and, of course, where to place them.

This study is focused in launching a simulation of multiple possible real trips between two points in a highway, in order to detect whether or not the electric vehicle will reach its destination, considering the current charging infrastructure in that highway. For that reason, not only the charging points within the highway are considered, but also near points with a threshold of 5 km, as it is described in section 3.3.1.

5.1 Simulation Setup

This section presents the different considerations that are taken into account to configure the simulation, as well as the different data sources. The following assumptions are made:

- The availability of a given charging point is total, i.e. the number of charging points will be enough to charge the vehicle without queues .
- Initial State Of Charge (SOCi): Even though the range estimation model was trained with values of SOCi in the interval [7, ..., 100], it is not likely for a user to plan a long trip and begin the journey with a practically discharged battery. For this reason, the range of values of SOCi is set from a minimum of 40% and a maximum of 100%.
- Trip distance: The generated trips will belong the origin - destination pairs of POIs between a given highway. However, only the pairs that represent distances greater than 20 km will be taken into account, as we are interested on studying the long trip scenario.

5.1.1 Trip data generation

It is not possible to study the availability of charging infrastructure for EV owners without analyzing different kind of trips between a given highway. Due to the absense of real data, the decision is to create possible trips given two locations of a concrete highway.

The departure point is the full Autoroutes data source presented in section 3.1.1, which contains information about areas and junctions. Here we are interested in generating all possible combinations of each pair of coordinates $\in$ P [1,...,p], that will be taken as origins and destinations for the simulated trips. Figure 13 shows the followed procedure.

The combination of all origin - destination points can be easily achieved in R with the function *expand.grid*, which returns a matrix object with dimensions [p,p]. Next, the Haversine distance is computed between all pairs, in order to avoid Google API

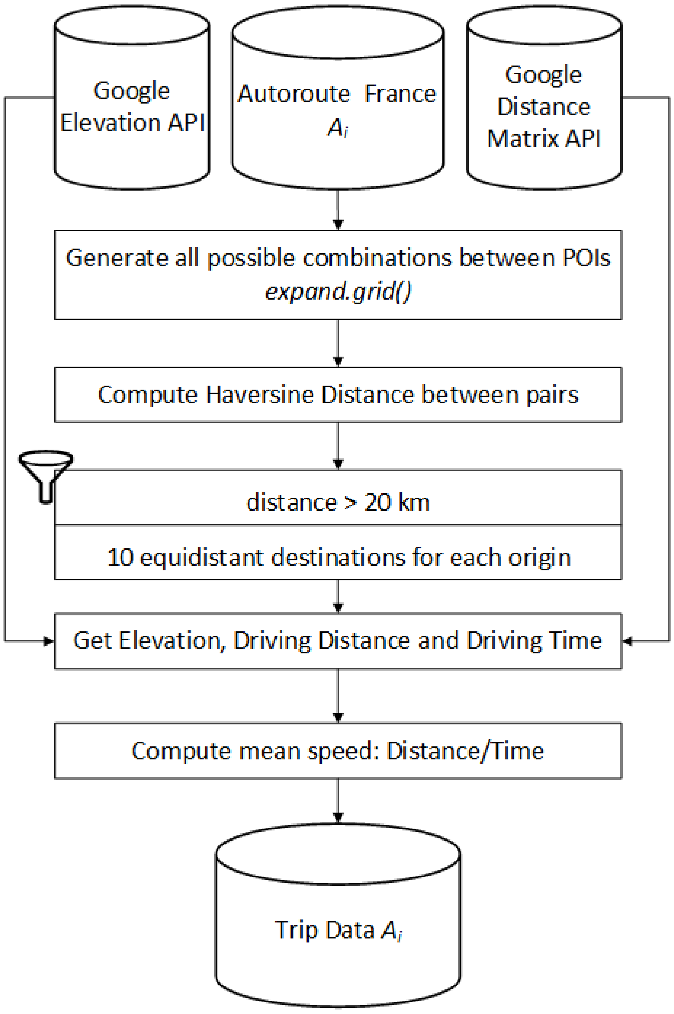

Figure 13: Flow chart of trip data generation. Source: Own

request limit. After that point, a filtering is applied to reduce the number of trips to simulate, due to computational cost: we consider trips with distance greater than 20 km, and for a given coordinate p, we select 10 equidistant points (i.e. in the range between minimum and maximum distance between trips). The efficiency of this last filter can be checked with Figure 16, in the final section of this chapter, which shows the two histograms of the distance between POIs in the case study of one Autoroute (A10), for the total population (left) and the reduced sample (right).

The next step involves calculating the real elevation between POIs, as well as its driving distance and driving time, since the reduced sample has a suitable dimension to don't exceed the Google API limits. Finally, mean speed is calculated dividing the driving distance and the driving time.

5.1.2 Electric vehicle data generation

At this point, the first part of simulation data is ready. However, the range estimator model is trained bearing in mind other variables that significantly affects to the total autonomy of the vehicle, such as initial State Of Charge, temperature and degradation of the battery. Consequently, the approach is to simulate different combinations of these variables for each generated trip. This strategy will allow to study the availability of charging infrastructure in highways under certain conditions, and extract more accurate conclusions.

Considering both a representative range of values for each variable and the feasibility in terms of computational time, the following configuration is presented.

- **SOCi**: 4 levels, from 40% to 100% [40, 60, 80 and 100].
- **Temperature**: 4 levels, from -5ºC to 30ºC [-5, 10, 20, 30].
- **Degradation**: 4 levels [High (30%), Medium (20%), Low (10%) and Zero (0%).
- **Speed**: 2 levels, the real driving mean speed calculated from Google Distance Matrix API values, and the minimum speed in highway, 80 km/h, which would allow the maximum range for the Nissan Leaf.

Therefore, each trip will be tested under 128 different combinations of variables.

5.1.3 Distance between charging infrastructure

Last but not least, it is important to pre-compute the distance between all the charging points of a given highway in order to avoid costly computations inside the simulation loop. For this reason, first the charging points are sorted by its coordinates and then the Google Distance Matrix API is called to retrieve the driving distance between them.

In addition, to relate the charging points location and the highway POIs, an index is created. This index takes into account the nearest POI of a given charging infrastructure.

This index will be used again to avoid calculating inside the simulation loop which charging points can be found inside each trip, notably reducing the computational time.

5.2 Methodology

This section presents the methodology for simulating the availability of charging infrastructure for electric vehicle owners in large trips within highways. The different cases are presented in the previous section, with a list of possible journeys using pairs of POIs from a given highway. Each of the trips will be tested under certain conditions, concretely 128. Therefore, the size of simulations increases to the number of pairs origin-destination (OD) multiplied by the number of possible configurations (128).

5.2.1 Initialization

The main objective of the algorithm is to decide whether or not an electric vehicle user could reach its destination, and if it makes uses of any of the infrastructures that could be found within the trip route, given the generated trip data and the vehicle configuration. During this study, several variables will be stored for each case, described in table 7.

Table 7: Variables from the charging infrastructure availability simulation

Variable	Description
Success	Boolean that indicates if the vehicle would reach the destination
anyInfra	Boolean to know if there was available infrastructure inside the trip route. It only needs to be calculated once (for each trip).
useInfra	Boolean that shows if any charging station was used
DistToDest	Distance from the destination or the nearest charging point

Taking into consideration that the data that will be used in the simulation is already created, the range estimator model (RBF SVM) is applied previous to the algorithm in order to save computational time.

All the variables will be written in the matrix of simulation setup, in order to facilitate the posterior analysis of results.

5.2.2 Procedure

At this point all the data is prepared to run the simulation, including the generated trips with the vehicle configurations and the RBF SVM range estimator model. The algorithm is described in the flow chart of Figure 14.

The process starts with a loop for each trip *t*, representing a pair of origin - destination coordinates. For each case, a temporary data matrix of the 128 configurations is saved in memory, as well as some calculations that do not affect the inner loops: the direction; the presence of any infrastructure in the route, which is easily to check using the pre-defined indexes; and the *Vdist* vector. This last structure is created if only *anyInfra* variable is true, and collects all the distances between the origin, the charging stations that could be found inside the route, and the destination. The distances between the different charging points are pre-calculated (see subsection 5.1.3), but the distances of the nearest charging stations from origin and destination must be calculated through Google Distance Matrix API. The direction of the trip is taken into account, as the distances may change.

Once all the static data from the outer loop is loaded, the algorithm enters to the inner loop that iterates through all the different configurations *c* of the trip *t*. The first if statement compares the real driving distance versus the predicted range. As it is commented in the previous section, all the range estimations are pre-calculated outside the algorithm. If the predicted range is enough, the inner procedure is ended by writing the results: Success true, useInfra false and DistToDest equal to zero. If distance is greater than the range, the algorithm passes to the next statement, which checks if exists any infrastructure within the route. In the negative case, the simulation *c* is ended, writing the corresponding result (Success false, useInfra false, DistToDest equal to the difference between real distance and range). In case the statement is true, the range estimator will be applied again changing the value of SOCi to 100, as we will need this value considering that if the vehicle reaches one charging point, it will get out of it with full battery.

Finally, the algorithm enters to the second inner loop, for each value of trip t and configuration c, which iterates through the values of the vector *Vdist*. The objective is to check:

- **Case 1**: If the vehicle could reach the nearest charging point from origin. If false, end the two inner loops and write Success false, useInfra false, DistToDest equal to the distance origin - charging point minus the predicted range.
- **Case 2**: If the vehicle can reach the k^{th} charging stations from the $k^{th}-1$ charging point (or the origin if there is only one station within the route). If false, end the two inner loops and write Success false, useInfra true, DistToDest equal to the distance k^{th} - point $k^{th}-1$ minus the predicted range with 100% of SOCi.
- **Case 3**: If the vehicle can reach the destination from the last charging point. If true, write Success true, useInfra true, DistToDest zero. Otherwise, write Success false, useInfra true, DistToDest equal to the difference of the distance between the last charging - destination, and the predicted range with 100% of SOCi.

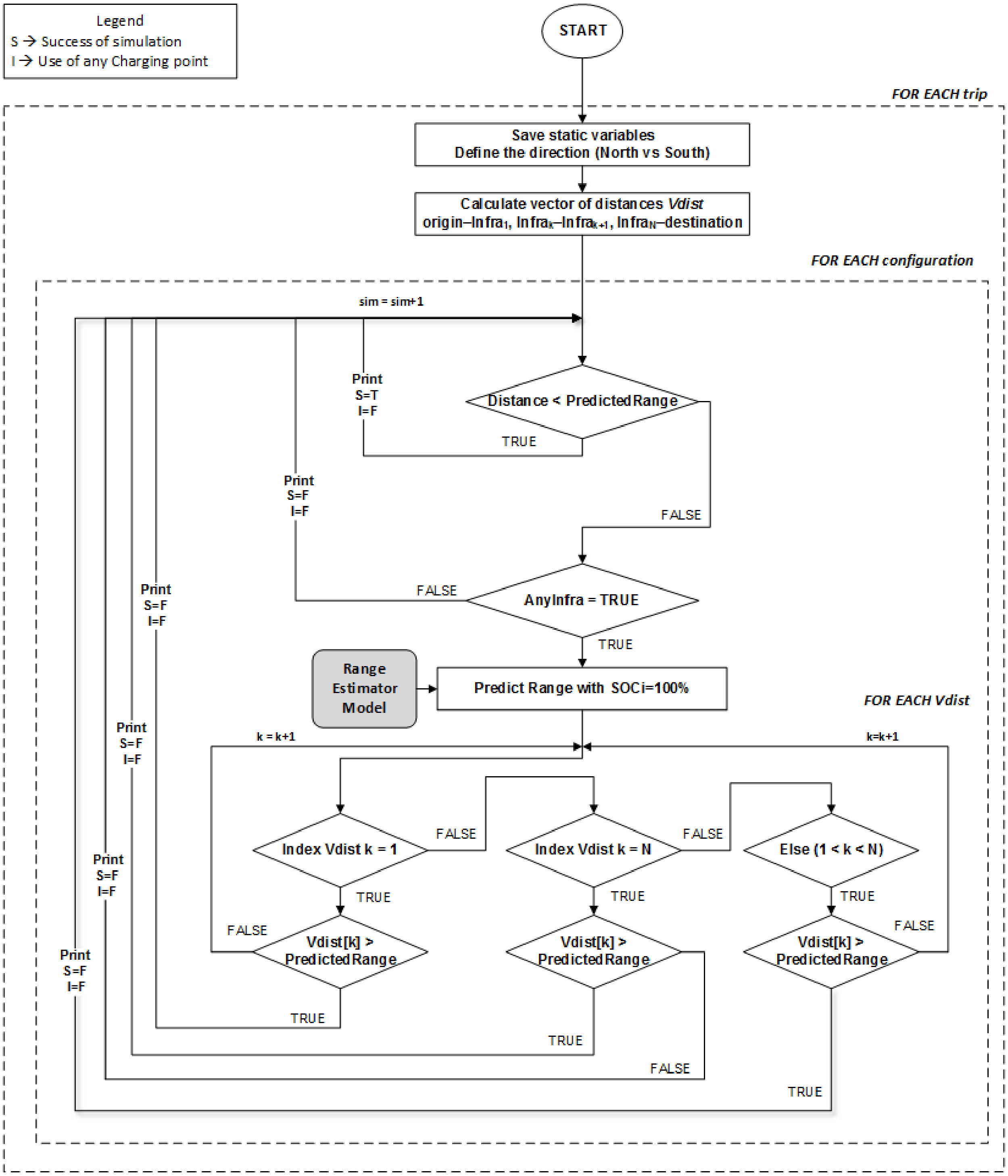

Figure 14: Simulation flow chart. Source: Own

5.3 Case of Study: Autoroute A10

The availability model for a given highway described in the previous sections is tested with the Autoroute A10 from France. The A10, also called L'Aquitaine, is an Autoroute in placed in France, with a total length of 549 km from the A6 south of Paris to the A630 at Bordeaux. It is the longest motorway in France [62]. To have an idea about the placement of this autoroute, Figure 15 shows the plot of A10 charging points, in black, and A10 service and rest areas, in red.

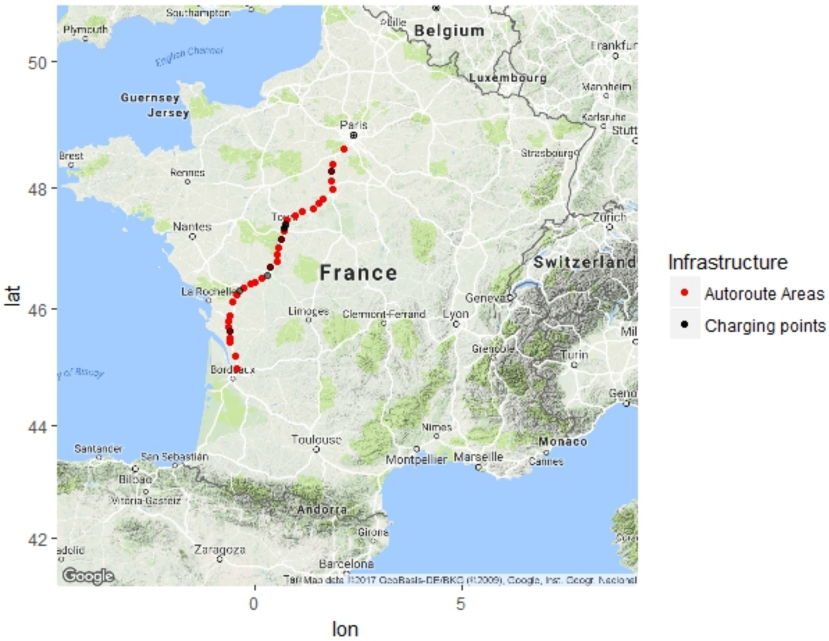

Figure 15: Integration of charging points and Autoroute POIs in the A10. Source: Own

5.3.1 Simulation setup

Trip and electric vehicle data is generated with the method described in section 5.1. The following list shows the configuration values for this simulation:

- Number of different trips $T = 963$.
- Number of configurations $C = 128$.
- Total number of cases $T{\cdot}C = 123264$.

- Interval of driving distance from POIs from 22.27 km to 556.26 km.
- Interval of mean speed from 78 km/h to 125 km/h.
- Interval of elevations between origin-destination from -162 meters to 163 meters.
- Predicted range interval betweem 28.03 km to 227.39 km.

As it is commented in subsection 5.1.1, it is not feasible to perform the simulation for each pair of POIs points. For that reason, a subset of the total combination of points is used, trying to preserve the same distribution of data. Figure 16 shows the two histograms of the distance between POIs in the case study of one Autoroute (A10), for the total population (left) and the reduced sample (right). It can be concluded that, even though the number of trips to study is reduced, it preserves a close distribution with the same minimum and maximum values.

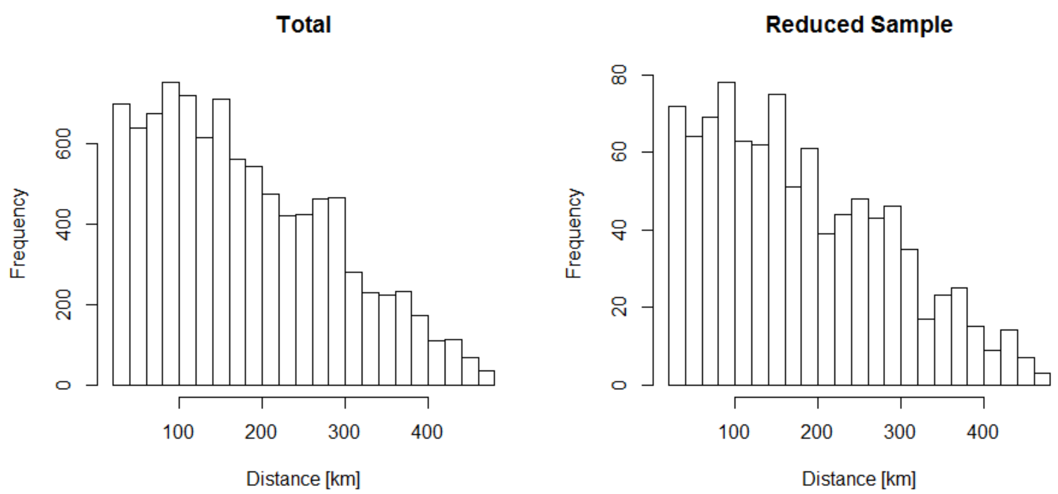

Figure 16: Histogram of trip distances: total and reduced samples for A10. Source: Own

Regarding the predicted range, its maximum value of 227.39 km is far from the highest trip distances, around 550 km. Therefore, to get a success in these long trips will be indispensable to charge the vehicle one or more times within the route.

5.3.2 Overall results

The results regarding the success of each simulation case, understood as the vehicle had enough battery to reach its destination, with or without charging it within the route. Table 8 presents the overall results of the availability study. The average distance to destination is calculated using the cases where *Success* = *TRUE*, considering that for the case, all the values are zero.

Table 8: Overall results from highway availability simulation for A10

Average Distance to Destination [km]	Success [%]	Infrastructure [%]	Used Infrastructure [%]
34.01	62.26	92.10	53.03

For A10 Autoroute, the results of the simulation show that more than 60% of the vehicles would be able to reach its destination. Furthermore, this success cases where mostly possible by using one or more charging points within the vehicle route, concretely 68.51%. The percentage of the presence of charging infrastructure within the routes is significantly high. However, this value may not be taken as an indicator for deciding whether or not building more charging points, as it does not assure that the distance between points is low enough to be reached by all electric vehicles.

Although overall results could be used for stakeholders as Key Performance Indexes (KPIs) to study the availability of charging points in highways, it is important to compare the results by the driving distances of the trips. Figure 17 shows the density percentage of the three binary decision variables: *Success*, *anyInfra* and *useInfra* (see Table 7). Charge during trips only considers the successful journeys. Conclusions:

- Success rate decreases as driving distance increases, and even though for short trips the rate of success in high, for trips longer than 400 km it is less than 25%.
- The presence of charging points within the route are present in 100% of the trips longer than approximately 150 km. However, this does not assure that the distances between charging points are sufficient to finish a trip.
- The simulated short trips, say less than 100 km, practically do not need to charge the vehicle to reach its destination. However, for distances greater than 200 km it is an imperative to stop and charge.

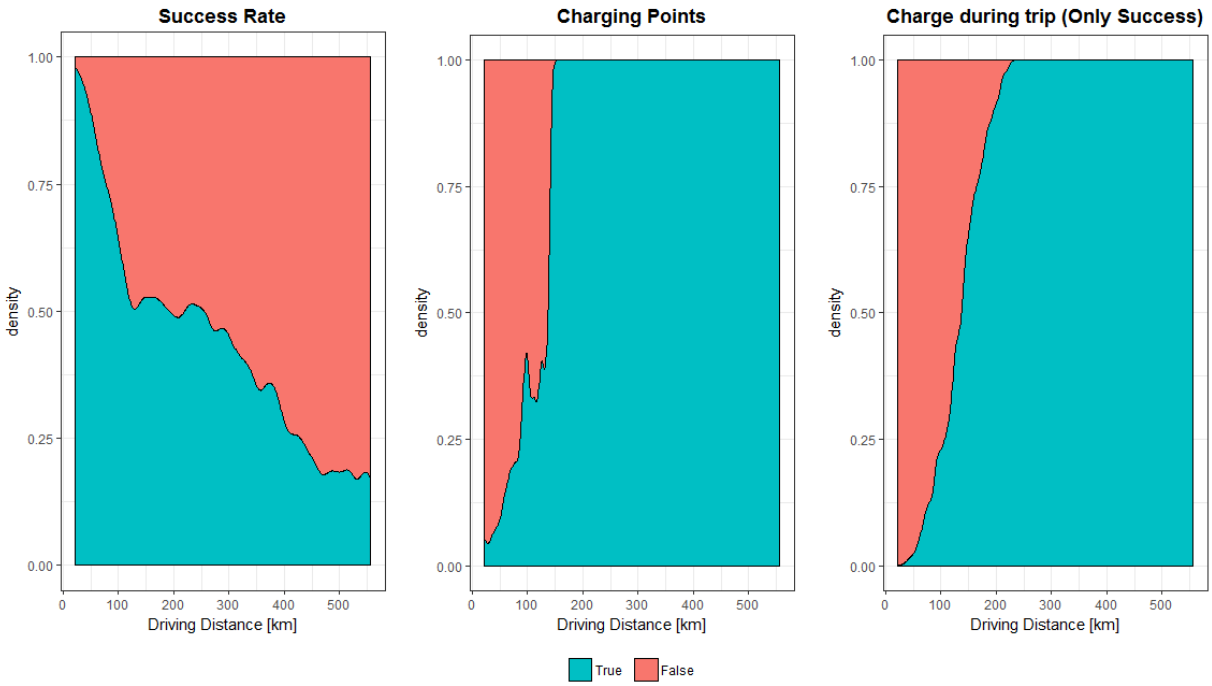

Figure 17: Integration of charging points and Autoroute POIs in the A10.
Source: Own

Figure 18 represents the distribution of trip distances in four different result combinations, concretely the variables *Success* (Success or Failure) and *useInfra* (Charge during trip or No charge). Several conclusions can be drawn for each case:

- *Success* trips are mostly situated in low distance journeys, even though with the usage of one or more charging points the range of distances is increased until maximum trip distances (around 500 km).
- *Failure* trips are distributed across all driving distance range. In comparison with *Success* ones, they have a higher ratio in long distances, as it was expected.
- *Charge during trips* case is essential in *Success* trips for achieving its destination: distances greater than approximately 200 km are not plausible for vehicles if there not exists any infrastructure within the route. However, in some cases is not enough to find a charging point between origin and destination, due to the distances between points.
- *No charge* situation leads to a higher *Failure* rate, especially in long distances.

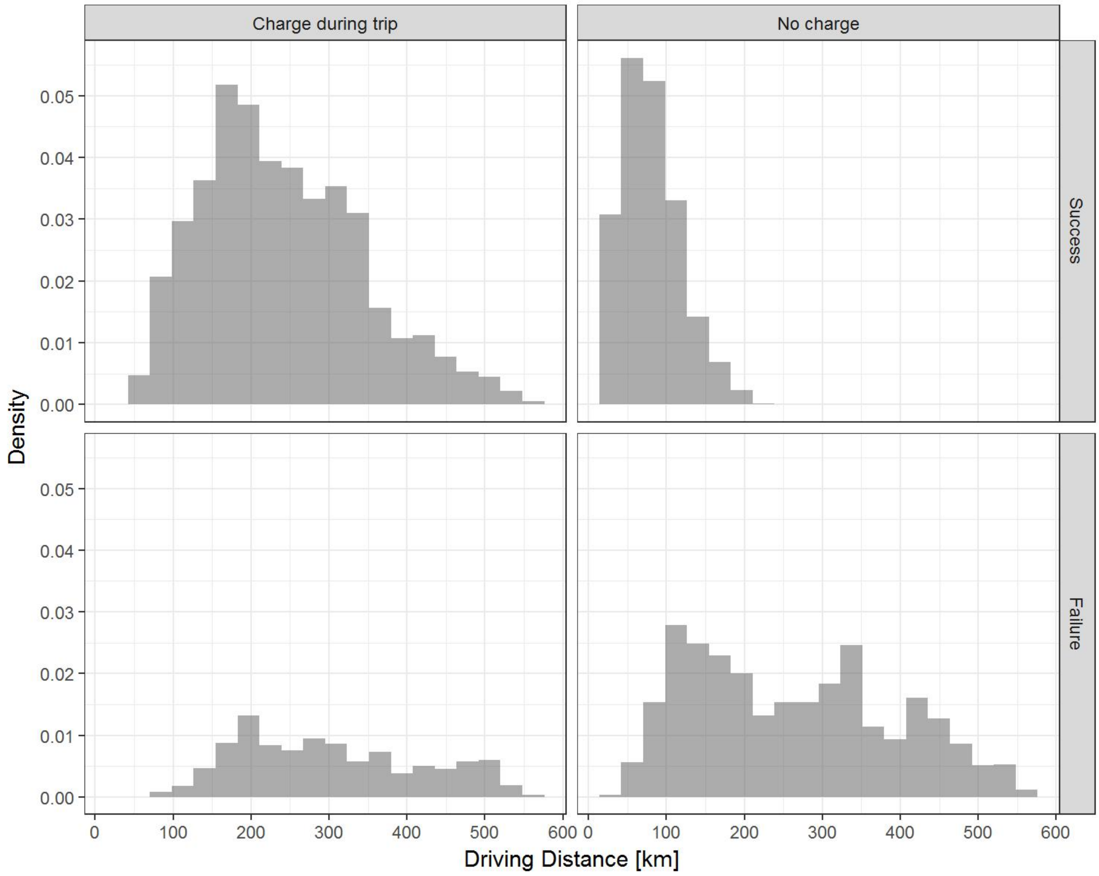

Figure 18: Trip distance distribution for success and used infrastructure cases . Source: Own

5.3.3 Effect of external conditions

This part present how the results are influenced by temperature and elevation between origin and destination. In fact, these effects were modelled in the range estimator chapter (see chapter 4).

For interpretability purposes, the Success rate is compared not only with temperaure or elevation, but also with the association of the initial SOC. This decision is taken because is the variable that mostly influences the available range, as it represents the percentage of battery at the beginning of the trip. Table 9 represents the success rate against temperature and initial SOC. There is a clear influence of temperature, in the worst scenario only between 30% and 57% of the trips reach its destination. However, for higher temperature values, from 10ºC to 30ºC the success rate has lower influence.

Table 9: Contingency table of the Success given the initial SOC and Temperatures

	Temperature			
SOCi	-5ºC	10ºC	20ºC	30ºC
40%	0.3001	0.4524	0.5039	0.4559
60%	0.3982	0.6142	0.6856	0.6292
80%	0.4943	0.7455	0.8132	0.7674
100%	0.5715	0.8149	0.8772	0.8384

Table 10 represents the success rate against elevation and initial SOC. As elevation is a continuous variable, it is discretized in three categories: Descent, representing the elevations from -162 meters to -50 meters; Flat, from -50 meters to 50 meters; and Climb, from 50 meters to 163 meters. Even though temperature has higher influence, elevation also plays an important role in the success rate.

Table 10: Contingency table of the Success given the initial SOC and Elevation

	Elevation		
SOCi	Descent	Flat	Climb
40%	0.4699	0.4290	0.3416
60%	0.6388	0.5651	0.4826
80%	0.7637	0.6821	0.6089
100%	0.8313	0.7592	0.6783

5.3.4 Battery characteristics and driving behaviour

In addition to external features, it is interesting to study how the results are affected by vehicle-specific variables. Concretely, the degradation level of the battery, the initial State Of Charge at the beginning of the trip, and the average speed will be analyzed.

First, the success rate given the different values of initial SOC and degradation battery levels are presented in table 11. There is a clear influence of both features to the success rate. As it is expected, as degradation increases the total range is reduced, affecting to the success of the trip. At the same time, decreasing the level of initial State Of Charge have the same effect. Even though degradation has a significant effect on the success rate, it is the initial SOC the variable that mostly influences the total range of the vehicle.

Table 11: Contingency table of the Success given the initial SOC and Degradation level

	Degradation			
SOCi	0%	10%	20%	30%
40%	0.5308	0.4441	0.3956	0.3415
60%	0.7276	0.6056	0.5354	0.4584
80%	0.8267	0.7280	0.6815	0.5839
100%	0.8939	0.8037	0.7457	0.6586

Apart from the previous variables, the speed of the vehicle is another factor that influences the total range. Figure 19 presents a boxplot of the simulation variable *DistToDest* (see Table 7) against two vehicle-specific variables: the initial SOC and the speed. As speed is a continuous feature, it was discretized in two levels: the maximum distance, calculated by the Google Distance Matrix API values, and the minimum distance, set to 80 km/h.

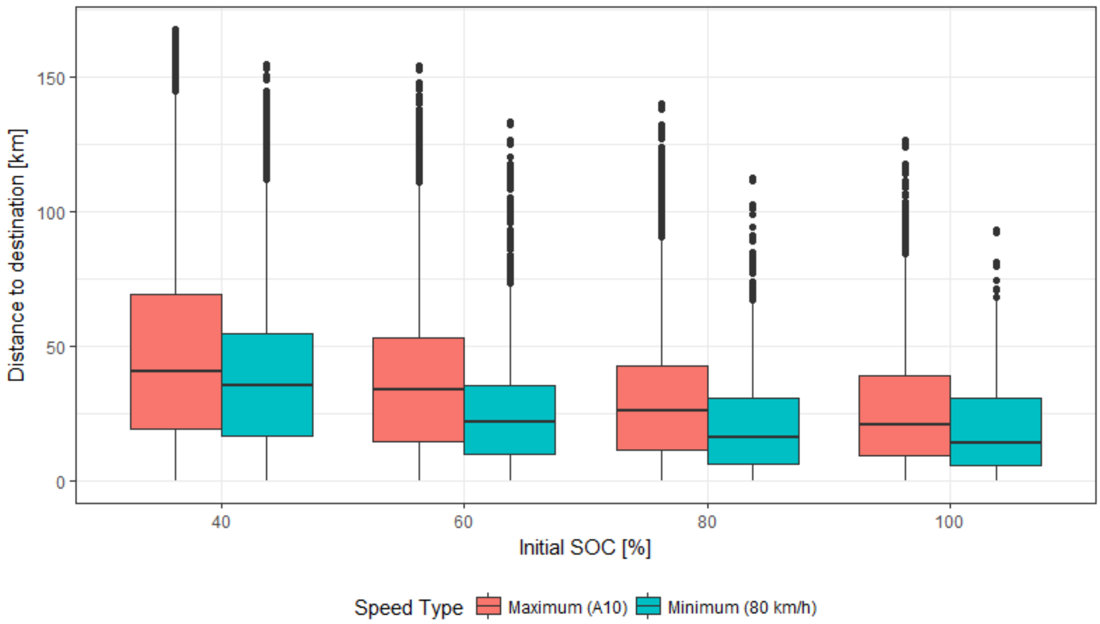

Figure 19: Boxplot of distance to destination for failure cases by SOC and Speed Type levels. Source: Own

This data representation is useful in order to see the current limits of distance from which the electric vehicles can not reach its destination, around 150 km. Although the effect of the initial SOC does not seem very significant, we have to take into account that some of the trips include a charge between origin and destination. Driving with the minimum speed has a positive effect on the total range of the vehicle, and the distance to destination is significantly decreased. Users may be aware of this characteristic, in order to increment the available range in case they need it.

When the vehicle initiates the trip with full battery (SOC = 100%), the success rate is increased to 89% in the scenario of zero degradation, whereas in the worst case, 30% of degradation, the rate decreases to approximately 66%. What is more, if we consider a best scenario with SOC = 100%, degradation = 0% and mean speed = 80 km/h, the success rate is increased to 98.54 %.

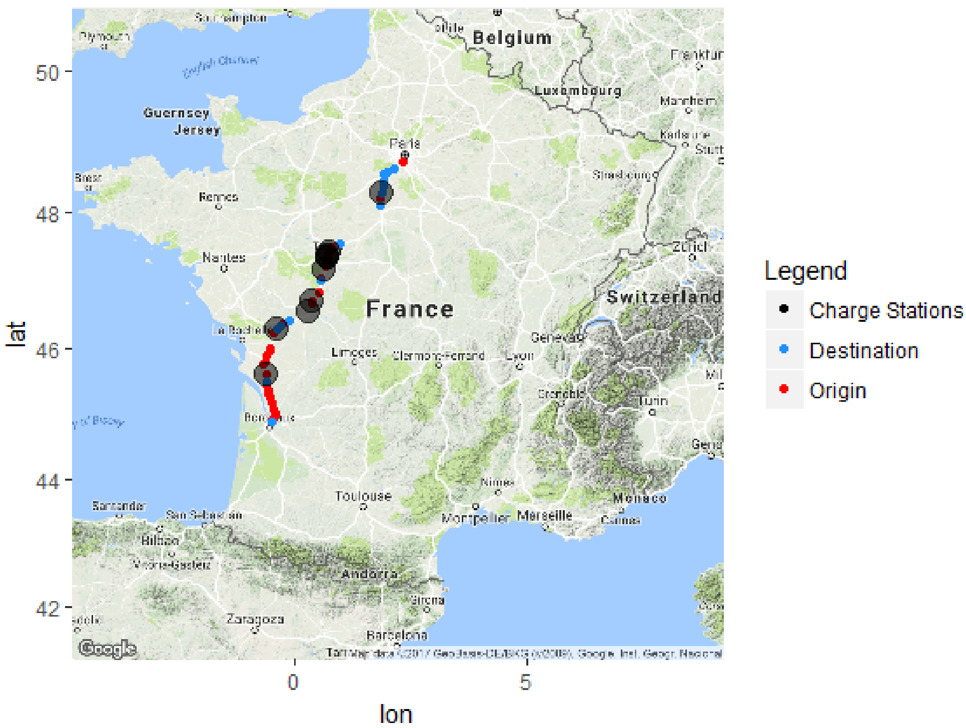

Figure 20: Integration of charging points and Autoroute POIs in the A10. Source: Own

Figure 20 shows the origin and destination points for this best scenario, as well as the charging stations, for all the trips that do not reach its destination. It is interesting to see that there is a long distance between two charging stations in the north of the A10 Autoroute, which is the cause of failure for these trips. Specifically, the distance is 150.67 km.

Chapter 6

Conclusions

Several studies for contributing in the integration of electric vehicle and charging station in highways have been presented in this project. This chapter details the main findings throughout the thesis and includes the future work that may be addressed.

6.1 Contributions

The contributions are detailed per chapter.

- In Chapter 3 the different data sources used for conducting this thesis are presented, including data transformation and data integration. Bearing in mind that this project was started from zero, data collection can be considered the critical phase: There is no public available data for electric vehicle range estimation, and charging infrastructure geospatial data is very limited in comparison the the real number of points.

- In Chapter 4, a model to predict the actual range of a vehicle given certain trip and vehicle features is trained, including vehicle features such as available SOC at the beginning of the trip and degradation level of the battery; user features including the mean driving speed; and trip features such as elevation between origin and destination and temperature. Even though the obtained test error is exorbitantly low, the lack of a extensive real world trip data sentences this model to be very vehicle and user dependent, and therefore can not be extrapolated to other vehicle or user types.

- In Chapter 5, a methodology for studying the availability of charging infrastructure for electric vehicles in highways is detailed. The method takes into account highway and near-highway charging points, as well as the different Points Of Interest (POIs) to build a grid of possible trips and simulate their success given a series of feature combinations, related with the range estimation explanatory variables. The algorithm is applied to the French Autoroute A10, and several results and conclusions are presented to allocate the critical points where electric vehicles could not finish the journeys.

6.2 Future work

Collecting or having access to data is a crucial part in order to perform an accurate analysis of the situation of EV charging stations in highways, as well as predicting the range of a vehicle given certain conditions for a trip. In this sense, future work can be done in the following directions:

- Finish the optimal location model for charging stations in highways and analyze the results with the proposed methodology for analyzing its availability.
- Compare the results obtained in the case study of Autoroute A10 for the availability analysis both with the current situation and with the optimal charging point scenario.
- Improving the range forecast using real data from a set of users in order to develop a framework that considers not only the type of vehicle and the discharge rate given speed or temperature, but also user behaviour on road.
- Refine the methodology for studying the availability of charging infrastructure in highways, building several cases of study to consider the whole Autoroute grid in France. Interpret the results in global terms to decide new strategies of charging infrastructure location.

Bibliography

[1] Sjoerd Bakker (North Sea Region Electric Mobility Network) *Standardization of EV Recharging Infrastructures*, Interreg IVB project E-Mobility NSR, 2013. 1, 11

[2] European Alternative Fuels Observatory, *Electric Vehicle Charging Infrastructure* [online]. Available at: http://www.eafo.eu/electric-vehicle-charging-infrastructure [Accessed 10th November 2016] 2, 13, 25

[3] International Energy Agency (IEA). *Global EV Outlook 2016, Beyond one million electric cars*. IEA Publications, 2016. 2

[4] Austrian Energy Agency, et al. *State of the Art Electric Propulsion: Vehicles and Energy Supply*. COMPETT Project, 2013. 5, 9

[5] C. C. Chan. *The State of the Art of Electric, Hybrid, and Fuel Cell Vehicles*. Proceedings of the IEEE, Vol 95, Issue 4, 2007. 5

[6] Electric Vehicles News *Electric Vehicle History*[online]. Available at: http://www.electricvehiclesnews.com/History/historyearlyIII.htm [Accessed 08th November 2016] 5

[7] International Energy Agency. *Electric and Pug-in Hybrid Vehicle Roadmap*. 2013. 6

[8] Lukic, Srdjan M., Jian Cao, Ramesh C. Bansal, Fernando Rodriguez, and Ali Emadi. *Energy Storage Systems for Automotive Applications* IEEE, pages 2258-2267, 2008. 8

[9] Khaligh, Alireza. *Battery, Ultracapacitor, Fuel Cell, and Hybrid Storage Systems for Electric, Hybrid Electric, Fuel Cell, and Plug-In Hybrid Electric Vehicles: State of the Art*. IEEE, pages 2806-2814, 2010. 8

[10] Karden, Eckhard, Serve Ploumen, Birger Frickea, Ted Millerb, and Kent Snyderb. *Energy Storage Devices for Future Hybrid Electric Vehicles.* Journal of Power Sources, pages 2-11, 2006. 8

[11] Dixon, Juan, Ian Nakashima, Eduardo F. Arcos, and Micah Ortúzar. *Electric Vehicle Using a Combination of Ultracapacitors and ZEBRA Battery.* IEEE, 2010. 9

[12] Paul Wolfram and Nic Lutsey. *Electric vehicles: Literature review of technology costs and carbon emissions.* International Council on Clean Transportation Working Paper, 2016. 9

[13] The Car Guide *Tesla Model S 85 2016.*[online] Available at: http://www.guideautoweb.com/en/specifications/tesla/model-s/85/2016/ [Accessed 19 November 2016] 7

[14] MIT Technology Review *Machine Learning and Data Are Fueling a New Kind of Car.* [online] Available at: https://www.technologyreview.com/s/603853/machine-learning-and-data-are-fueling-a-new-kind-of-car/ [Accessed 18 December 2016] 9

[15] Volkswagen. *Volkswagen Car-net.* [online] Available at: https://www.volkswagen.es/es/modelos/passat-gte.html#item=1&powerLayer=passat/car-net_passat-gte-e-remote.display [Accessed 02 December 2016] 10

[16] TeslaRati. *Tesla API, Controlling your Model S and Model X.*[online] Available at: http://www.teslarati.com/tesla-api-controlling-model-s-model-x/ [Accessed 02 December 2016] 10

[17] Tim Dorr. *Tesla Model S JSON API.* [online] Available at: http://docs.timdorr.apiary.io/# [Accessed 02 December 2016] 10

[18] Damià Valero-Bover, Pol Olivella-Rosell, Roberto Villafàfila-Robles, Sivia Cestau-Cubero. *Performance analysis of an electric vehicle fleet for commercial purposes.* International Electric Vehicle Conference (IEVC), IEEE, 2014. 9

[19] Pavel Brandstetter, Jan Vanek, Tomas Verner. *Electric Vehicle Energy Consumption Monitoring.* International Scientific Conference on Electric Power Engineering (EPE), IEEE, 2014. 9

[20] European Environment Agency. *Electric vehicles in Europe.* EEA Report, Nº 20, Copenhagen, 2016 11, 12

[21] Genovese, A., Ortenzi, F. and Villante, C. *On the energy efficiency of quick DC vehicle battery charging.* 28th International Electric Vehicle Symposium and Exhibition, Korea, 2015. 11

[22] International Energy Agency *French Charging Infrastructure Strategic Roadmap* 2014 12

[23] Export.gov (U.S. Department of Commerce) *Market Brief - E-Mobility in France (2016)* 2016 13

[24] ChargeMap *French ChargeMap* [online] Available at: https://fr.chargemap.com/ [Accessed 21 January 2017] 14

[25] Open Charge Map [Online] Available at: https://www.openchargemap.org/site/ [Accessed 03 November 2016] 13, 14, 25

[26] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati and G. P. Hancke. *A survey on smart grid potential applications and communication requirements* IEEE Transactions on Industrial Informatics, vol. 9, pp. 28-42, 2013. 14

[27] J. Ferreira, V. Monteiro, and J. Afonso. *Vehicle-to-anything application (V2Anything App) for electric vehicles.* IEEE Transactions on Industrial Informatics, vol. 10, pp. 1927-1937, 2013. 14

[28] M. Ehsani, M. Falahi, and S. Lotfifard. *Vehicle to grid services: Potential and applications* Energies, vol. 5, pp. 4076-4090, 2012. 14

[29] Chung-Hong Lee and Chih-Hung Wu. *Collecting and Mining Big Data for Electric Vehicle Systems Using Battery Modeling Data* 12th International Conference on Information Technology, IEEE, 2015 14, 16

[30] Earth911 *3 Must-Have Apps for Locating Electric Vehicle Charging Stations* [Online] Available at: http://earth911.com/eco-tech/transportation/3-must-have-apps-for-locating-electric-vehicle-charging-stations/ [Accessed 23 January 2017] 15

[31] EV World *Six Apps for Electric Cars* [Online] Available at: http://evworld.com/news.cfm?newsid=31328 [Accessed 23 January 2017] 16

[32] EcoCars4Sale *Electric Car Smartphone Apps* [Online] Available at: https://www.ecocars4sale.com/smartphone-apps [Accessed 23 January 2017] 15

[33] S. Mehar, S. M. Senouci, and G. Rémy. *EV-planning: Electric vehicle itinerary planning*. International Conference on Smart Communications in Network Technologies, vol. 1, pages 17-19, 2013. 16

[34] Bohan Zheng, Peter He, Lian Zha and Hongwei Li. *A Hybrid Machine Learning Model for Range Estimation of Electric Vehicles*. IEEE, 2016. 17

[35] Joonki Hong, Sangjun Park and Naehyuck Chang. *Accurate Remaining Range Estimation for Electric Vehicles*. IEEE, 2016 18

[36] Daniela Chrenko, Alexandre Ravey, Robin Roche and David Bouquain. *Autonomy Estimation for EV based on Road Planning Software*. IEEE, 2016 18

[37] H. Rahimi-Eichi and M.-Y. Chow. *Big-data framework for electric vehicle range estimation*. Proceedings of 40th Annual Conference IEEE in Ind. Electronics Society (IECON), 2014. 17, 18

[38] C. H. Lee and C. H. Wu, *A novel big data modeling method for improving driving range estimation of EVs*. IEEE, 2015 18

[39] Y. Zhang, W. Wang, Y. Kobayashi, and K. Shirai. *Remaining driving range estimation of electric vehicle*. Proceedings of IEEE International Electric Vehicle Conference (IEVC), 2012. 18

[40] A. Bolovinou, I. Bakas, A. Amditis, F. Mastrandrea, and W. Vinciotti. *Online prediction of an electric vehicle remaining range based on regression analysis*. Proceedings of IEEE International Electric Vehicle Conference (IEVC), 2014. 18

[41] Eugene Kim, Jinkyu Lee, and Kang G. Shin. *Real-Time Prediction of Battery Power Requirements for Electric Vehicles.* ICCPS, ACM, 2013. 18

[42] Chung-Hong Lee and Chih-Hung Wu *An incremental learning technique for detecting driving behaviors using collected EV big data.* ACM, 2015 18

[43] SARATLAS. *France, Base de Données des Autoroutes et Voies Rapides* [online]. Available at: http://saratlas.free.fr/ [Accessed 28th November 2016] 23

[44] Ignacio J.Pérez-Arriaga *Regulation of the Power Sector* Springer, 2013

[45] U.S. Government's open data. *Data.gov* [online]. Available at: http://www.data.gov/ [Accessed 11 April 2016]. 24

[46] Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R* Springer, 2013. 30

[47] Discussion Forum for the Nissan LEAF EV. *My Nissan Leaf Forum* [online]. Available at: http://www.data.gov/ [Accessed 11 February 2017]. 24

[48] Electric Range for the Nissan Leaf & Chevrolet Volt in Cold Weather. *Fleet Carma* [online]. Available at: http://www.fleetcarma.com/nissan-leaf-chevrolet-volt-cold-weather-range-loss-electric-vehicle/ [Accessed 15 February 2017]. 24, 30

[49] Google Maps API. *Google Elevation API* [online]. Available at: https://developers.google.com/maps/documentation/elevation/intro [Accessed 15 January 2017]. 26

[50] Google Maps API. *Google Distance Matrix API* [online]. Available at: https://developers.google.com/maps/documentation/distance-matrix/ [Accessed 18 January 2017].

[51] Association of French Motorway Companies (ASFA). *Autoroutes.fr* [online]. Available at: http://www.Autoroutes.fr/index.htm?lang=en [Accessed 21 December 2016] 33

[52] Wikipedia. *List of elevation extremes by country* [online]. Available at: https://en.wikipedia.org/wiki/List_of_elevation_extremes_by_country#cite_note-Blanc-20 [Accessed 22 December 2016] 31

[53] Nissan. *Nissan Leaf* [online]. Available at: https://www.nissan.es/vehiculos/nuevos-vehiculos/leaf.html [Accessed 09 January 2017] 34

[54] U.S. department of Energy. *Compare Electric Cars Side-by-Side* [online]. Available at: https://www.fueleconomy.gov/feg/evsbs.shtml [Accessed 13 January 2017] 34

[55] Ludwig Fahrmeir, Thomas Kneib, Stefan Lang and Brian D. Marx. *Regression Models, Methods and Applications*. Springer, 2013. 34, 36

[56] Kaggle *Home Depot Product Search Relevance* [online] Available at: https://www.kaggle.com/c/home-depot-product-search-relevance [Accessed 09 February 2017] 34

[57] Trevor Hastie, Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* Springer, 2008. 36

[58] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and other Kernel Based Learning Methods*. Cambridge University Press, 1999. 36

[59] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Suport Vector Machines, Regularization, Optimization and Beyond*. The MIT Press, 2002. 36

[60] Ron Kohavi. *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. International Joint Conference on Artificial Intelligence (IJCAI), 1995 38

[61] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. *A Practical Guide to Support Vector Classification*. Department of Computer Science, National Taiwan University, 2003. 39

[62] Wikipedia. *A10 autoroute*. [online] Available at: https://en.wikipedia.org/wiki/A10_autoroute [Accessed 24 March 2017] 51

[63] Mark A. Hall *Correlation-based Feature Selection for Machine Learning* (Doctoral disseration) University of Waikato, Department of Computer Science, 1999.