

École polytechnique de Louvain

Influence of temperature on consumption of battery electric buses and consolidation of a prediction model

Analysis of the case with midibuses in Brussels

Author: **Régis DE WILDE**

Supervisors: **Emmanuel DE JAEGER, Youssef OUALMAKRAN**

Readers: **Francesco CONTINO, Pierre-André RULMONT**

Academic year 2021–2022

Master [120] in Electro-mechanical Engineering

Acknowledgements

I would like to express my deepest gratitude to Youssef Oualmakran, E-Mobility expert and project leader, for his welcome at Laborelec, his mentoring and regular follow-ups. His consideration and kindness during my recovery from the Covid touched and encouraged me.

I wish to acknowledge the help provided by prof. dr. Emmanuel De Jaeger, Professor of electrical energy systems at UCLouvain, Vice-Dean of the Louvain School of Engineering in charge of relations with industry, for the opportunity he offered me to collaborate with Laborelec. His follow-up, critical mind and helicopter view on my work helped me all along the way.

I'm extremely grateful to prof. dr. Francesco Contino, Professor of Thermodynamics at UCLouvain, who accepted to read my thesis and be part of the committee members.

I would like to extend my sincere thanks to Pierre-André Rulmont, CIO at STIB-MIVB, for taking part in the committee, and for putting me in contact with Nicolas Mascaux, Senior Manager Transport Solutions, who provided me with the necessary data for my thesis.

I am also grateful to dr. Laurent De Vroey, former Technical Manager Electric Mobility Lab at Laborelec, now Director Smart Mobility at Engie, for his proofreading and contributions to the article presented at the 2022 EVS35 Oslo Conference, the International Electric Vehicle Symposium & Exhibition.

Thanks should also go to Tommaso Difonzo, Business Development - SMATCH Smart Charging for EVs, who attended and presented the paper at the 2022 EVS35 Oslo Conference.

Contents

1	Introduction	1
1.1	Context	1
1.2	Scope of this work	3
1.3	Organization of this work	4
2	Theoretical approach and literature review	6
2.1	Main parameters of energy demand	6
2.1.1	Traction energy	6
2.1.2	Heating, ventilation and air conditioning	8
2.2	Lithium ion batteries	10
2.2.1	Operating principle of a battery	11
2.2.2	Temperature effect on the cell battery	12
2.2.3	Battery Thermal Management Systems	14
2.3	Measurement from the literature	15
3	Analysis of the data from STIB/MIVB	17
3.1	Bus line 33	17
3.2	Presentation of the raw data	19
3.2.1	Files from the bus computer	20
3.2.2	Files from the passenger count system	21
3.3	Data Preprocessing	22
3.3.1	Addition of elevation	22
3.3.2	Linking the raw data	23
3.3.3	Choice of time reference	23
3.3.4	Classification by trip	24
3.3.5	Validation Classification by trip	26
3.4	Analysis on a micro level	28
3.4.1	By shift	28
3.4.2	By trip	29
3.5	Aggregation by trip	29
3.5.1	Graphical analysis	29

3.6	Visualization of the influence of temperature on consumption	31
3.6.1	Energy consumption by temperature range	32
3.6.2	Selection of trips with similar conditions except temperature	33
3.6.3	Headers of the dataset	33
4	Machine Learning	35
4.1	Machine learning basics	36
4.1.1	Evaluation of the model	36
4.1.2	Scaling of the data	38
4.2	Presentation of the different algorithms	38
4.2.1	Linear regression	38
4.2.2	Polynomial regression	39
4.2.3	K-Nearest Neighbors (KNN) regression algorithm	40
4.2.4	Gradient Boosting regression algorithm	41
4.2.5	Multi-Layer Perceptron	42
4.3	Comparison of the performance	43
4.4	Results from the model	44
4.5	Conclusion and best model	49
5	Application in different scenarios	50
5.1	Shifts of three different temperature conditions	50
5.2	Validation of the energy demand	52
5.3	Conclusion and improvement of the model	53
6	Conclusion	54
A	Appendix	56
A.1	Heating output requirements as a factor of the outside temperature	56
A.2	Battery performance	57
A.3	OpenData STIB	58
A.4	Overfitting	58
A.5	Main results from previous work	59

Chapter 1

Introduction

1.1 Context

Global warming and its consequences are rapid, widespread and intensifying. As the news headlines illustrate everywhere in the world: glacier collapse, more extreme heats, heavy rainfalls, droughts and fire weather, ...

Although climate change is not the subject of this work, it is nevertheless an integral part of the context since its main cause is the rise of the greenhouse effect linked to the increase in carbon dioxide emissions (CO₂). Carbon dioxide levels are now the highest in at least 2 million years. They result from burning fossil fuels (sources of emission) and deforestation (sinks of emissions)[1].

The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report [2] reaffirms the very important finding that every emission of carbon dioxide into the atmosphere adds to global warming. This warming is irreversible for many centuries if not longer. This implies that stabilizing global warming requires emissions of carbon dioxide to reach net zero.

One of the European Green Deal (EGD) [3] goals is to transform the EU into a net emissions zone of greenhouse gases by 2050. The European Commission adopted a set of proposals to make the EU's climate, energy, transport and taxation policies fit for reducing net greenhouse gas emissions by at least 55 % by 2030, compared to 1990 levels. More specifically in relation to the scope of this work, the Sustainable and Smart Mobility Strategy [4] and the Clean Vehicles Directive [5] are leading to a gradual increase in electric buses.

The upcoming increase in demand for electric buses is confirmed by the fact-finding study carried out for the European Commission in 2021 on status and future needs

regarding low- and zero-emission urban mobility [6]: In the majority (57 %) of cities, clean public transport buses are in operation. The number of clean public transport buses is relatively low compared to conventional buses (the average share of clean buses of the total number of public transport buses is 16,1 percent). This is still far from the 38,5 % clean heavy-duty vehicle in the aggregate public procurement across the main Member State imposed by the Clean Vehicles Directives . The same study shows that Belgium with France and Italy are frontrunners on the level of support in relation to urban mobility (SUMP Sustainable Urban Mobility Plans, UVAR Urban Vehicle Access Regulations, LEZ Low Emission Zone).

In 2022, the Commission will propose a Recommendation to EU Member States for the development of national plans to assist cities in developing their mobility plans[7]. The main focus will be on public transport, walking and cycling.

The imminent arrival of many new electric buses on the market is visible in a more concrete and tangible way on the Cleanbus platform [8]. On the interactive map for an overview of open tenders, bus orders and deployed buses, we see for example:

- for Brussels: following a reduction of 43 % on the diesel-only share, the current bus fleet counts 48 % hybrid buses and 4.5 % battery electric buses. STIB's fleet will be fully electric by the end of 2035. The next step is the acquisition of about 100 electric buses in 2024.
- for Antwerp: Currently there is one battery electric bus pilot. In 2022, the city of Antwerp plans to deploy another 30 e-buses. No more diesel buses will be incorporated in the public transport fleet.

It is not only the low emission zones and the decarbonization of society that accelerate the increase of electric buses, but also decreasing cost and density improvement of batteries. Technical innovations help to reduce total cost of ownership and operational costs of electric buses.

All those buses will need electricity to accomplish their trip and transport passengers. Electric buses cannot operate if their batteries are not charged. Therefore, converting depots for e-buses is also a part of the context.

During the Bonn Climate Change Conference of June 2022 (The 56th session of the Subsidiary Body for Scientific and Technological Advice SBSTA), the "Mitigation of Climate Change" presentation in collaboration with IPCC [9] shows the importance of coordinated cross-sectoral policies and planning which can maximise synergies to obtain the net zero emission, e.g. in the transport sector: electrification by renewables combined with encouraging the use of public transport.

The Commission To The European Parliament [10] calls for public and private investment in local renewable energy production. ASSURED [11] a project started

in 2017 with funding from the European Union's Horizon 2020 research and innovation program, is aimed at boosting the deployment of urban commercial electric fleets (buses, trucks and vans) across Europe and ensure their integration with high power fast charging infrastructure.

De Lijn, active in public transport in Flanders, has started in May 2022 three awarding procedures for work on its first electric bus depots. Therefore, De Lijn can count on 15 million euros through European revitalization aid (in the context of NextGenerationEU and the Recovery and Resilience Facility)[12].

The knowledge to manage the demand of electricity to charge the batteries is also part of the context. In fact, in order to avoid negative grid impact, grid connection and battery of the bus have to be properly sized[13]. It is therefore essential to know the energy consumption of electric buses in different driving conditions including temperature.

The influence of temperature on consumption of battery electric buses is the purpose of this work which, as we have just seen, is set in a very broad context.

1.2 Scope of this work

This work is made in collaboration with Laborelec, a leading expertise and research centre in electrical power technology. The expertise of its 335 engineers and technicians extends from the production to the distribution of energy, including its storage and end use. The focus is always on energy transition and zero carbon. Laborelec SA/NV is part of ENGIE Research and Innovation, with ENGIE and independent grid operators as shareholders [14].

The division "Electric Mobility Lab" consolidates knowledge of recharging needs of electrical vehicles and their consequences on the electrical network. Services are based on measurement, laboratory testing and running simulations, in order to offer expert advice on reducing energy consumption and the feasibility and potential of emerging technologies.

To do this, it is important to know and be able to predict accurately the energy consumption of electric buses in different conditions. However, the bus manufacturers' documentation only contains averages or track tests which do not reflect the real conditions of bus traffic in the city.

In 2020-2021, a first master thesis [15] and internship made an analysis of the most relevant parameters on the energy demand of the battery electric bus, retrieved energy consumption for >150 cases, and built a regression model with Scikit-learn.

Building further on that master thesis, the purpose of this current work is :

- To deepen the technology analysis with focus on readiness and impact of preheating / precooling of batteries.
- To consolidate the prediction model. In particular by using real operating data from an electric bus line

1.3 Organization of this work

There are several chapters in this report, structured as follows:

- Chapter 2 describes theoretically the main parameters influencing the energy consumption, namely traction energy and HVAC (Heating Ventilation and Air Conditioning). Then, it presents the Li-ion batteries, their operating principle, the effects of temperature on their cells and the BTMS (Battery Thermal Management Systems) used to mitigate these effects. Finally, it reviews the results of the literature on the influence of temperature on energy demand.
- Chapter 3 presents how the real data of a STIB bus line was used. After a presentation of the characteristics of the routes followed and the buses, the raw data available in the 2 types of files received is reviewed. Next, a data preprocessing section explains what data has been added, how and above all why. This section also includes the choice to classify the data by trip. The analysis of a limited sample of data also allows the quality of the data to be tested. After a first graphical analysis of the aggregated data per trip, the last point of this chapter shows a visualization of the influence of temperature on consumption. This is shown both by presenting the data with a temperature slice and after selecting trips with similar conditions in all aspects except temperature.
- Chapter 4 presents the theoretical foundations of Machine Learning by making the link with the data of the previous chapter, how the model is evaluated and the scaling of the data. Then, after a brief summary of the previous year's work and the algorithms used then, two new algorithms used this year are presented. After having implemented these algorithms using the dataset of chapter 3, a comparison of their performances and results are presented.
- Chapter 5, uses the preferred algorithm on different temperature scenarios (cold day, normal day and hot day). The aim is to compute the consumption

of the bus line for a whole day. Impact of the extreme temperatures on the battery performance is highlighted. Different ways of use and improvement are discussed.

- Chapter 6 presents a conclusion and the different steps followed to carry out this work.

Chapter 2

Theoretical approach and literature review

This chapter begins by describing the energy demand of battery electric buses and the main parameters influencing it. Then, a summary of lithium-ion batteries and their operating principle is presented. The impact of temperature is discussed as Li-ion batteries are very sensitive to temperature [16]. The different Battery Thermal Management Systems are explained. Finally, the literature studying the influence of the temperature on energy demand is presented.

2.1 Main parameters of energy demand

There are two components of energy demand for the buses. First, the traction energy, E_{tr} , that is necessary to move and accelerate the vehicle. Second, the auxiliary energy, E_{aux} , that is needed for the Heating, Ventilation and Air-Conditioning (HVAC) and the electronics onboard. A physical model is a model based on physical laws in order to compute the energy demand of BEB's. This is a useful approach when no records of BEB's consumption are available. In this section, a physical model is derived in order to discuss the impact of different parameters on the energy demand rather than retrieving energy demands.

2.1.1 Traction energy

To derive the traction energy, a model of tractive force, F_{tr} , must be made. This model specifies all the force components that are considered and can be derived from physics laws learned in the Young and Freedman textbook [17] as it was done in this paper [18]. Information is added concerning the change of certain

coefficients due to temperature.

$$F_{tr} = F_{drag} + F_{climb} + F_{roll} + F_{inertia} [N] \quad (2.1)$$

- $F_{drag} = K_d v(t)^2$; $K_d = \frac{1}{2} \rho(T) C_D A$, is the aerodynamic drag and is equal to the drag coefficient K_d times the square of the velocity. The coefficient of drag is proportional to the air density in kg/m^3 , the dimensionless coefficient of drag C_D that depends on the shape of the vehicle and can be obtained in a wind tunnel and the frontal area of the vehicle in m^2 .

The air density ρ , depends on the temperature and can be derived from the Ideal Gas Law. $\rho = \frac{p}{R^* T}$. The air density when the temperature is 0, 15 or 30 °C is respectively 1.29, 1.22 or 1.16 kg/m^3 . Therefore, higher temperatures reduce aerodynamic drag for the same velocity.

- $F_{climb} = Mg \sin(\alpha)$ is the force to overcome the gravity when climbing. M is the total mass of the bus in kg , $g = 9.81 m^2/s$ is the gravitational acceleration and α is the angle of inclination of the road.
- $F_{roll} = Mg C_R \cos(\alpha)$ is the rolling resistance that the wheels have to overcome and is equal to the force normal to the ground (Mg) times the rolling coefficient C_R . The road surfaces has an influence on the rolling coefficient, the smoother the road, the lower the coefficient.

Ambient temperature has also an influence on the rolling coefficient according to [19]. The ISO 28580 standards suggest the following correlation $C_R = C_{R,25} / [1 + K_t (T_{amb} - 25)]$. At 0 °C the rolling coefficient increases by 22% compared to the reference $C_{R,25}$ at 25 °C. At 40 °C, C_R decreases by 17% compared to the reference.

- $F_{inertia} = \delta M a(t)$ is the force related to the change of kinetic energy. If the speed of the vehicle increases, the acceleration a is positive and $F_{inertia}$ is positive and this requires more traction force to gain kinetic energy. The term δ is a factor to take into account the inertia of all rotating components in the drive train like the wheels, drive shaft,...

The traction energy power as expressed in (2.4) is the the traction force times the velocity multiplied by the overall efficiency η . There are two types of efficiencies, one when the tractive force is positive corresponding to an energy demand and the other when the tractive force is negative corresponding to regenerative braking. The regeneration factor, r_{regen} , is added to take into account the limitation due to

the maximum charging power of the batteries.

$$F_{tr}(t) = K_d v(t)^2 + Mgf \cos(\alpha) + Mg \sin(\alpha) + \delta Ma(t) \quad (2.2)$$

$$P_{tr}(t) = \frac{1}{\eta_{traction}} F_{tr}(t)v(t); \text{ when } F_{tr}(t) \geq 0 \quad (2.3)$$

$$P_{tr}(t) = \eta_{regeneration} F_{tr}(t)v(t); \text{ when } F_{tr}(t) < 0 \quad (2.4)$$

$$\eta_{traction} = \eta_{transmission} \eta_{motor} \eta_{inverter} \quad (2.5)$$

$$\eta_{regeneration} = r_{regen} \eta_{transmission} \eta_{motor} \eta_{inverter} \quad (2.6)$$

$$E_{tr} = \int P_{tr}(t) dt \quad (2.7)$$

To summarize, the traction energy demand, E_{tr} , is dependent on the mass of the vehicle, the velocity, the slope of the road and the rolling coefficient.

2.1.2 Heating, ventilation and air conditioning

The main auxiliary energy consumption is the Heating, ventilation and air conditioning (HVAC) systems. Other energy consumption is the electronics onboard such as the screen and the motors to open the doors.

The cabin temperature, humidity and CO2 concentration conditions have an impact on the comfort of the passengers. The specific term for this comfort is called thermal comfort. According to ASHRAE and ISO 7730 standards, thermal comfort is defined as “*that condition of mind that expresses satisfaction with the thermal environment*”. The conditions of thermal comfort are different for each individual as it is explained in greater detail in [20].

HVAC systems are present in the bus in order to control the temperature, humidity and CO2 conditions. The functioning mode of the HVAC for the STIB vehicles depends on the external temperature and is detailed in TABLE 2.1.

External temperature [°C]	Function	Set temperature [°C]
<16	Heating	16
17-23	Ventilation	/
24-26	Cooling	22
27	Cooling	23
>27	Cooling	$T_{out} - 5$

Table 2.1: HVAC function depending on the external temperature [21]

The heating and cooling requirement of a BEB is a complex problem and is the main subject of the following master "study of Heating Solutions for Electric Buses" thesis written in 2019 [21]. The aim here is to list the equations of the heat transfer

rates and understand their origins. Increase or decrease of temperature in the cabin is dependent of the total heat transfer rate of the bus, $\dot{Q}_{tot}$, which is the sum of all the different heat transfer rates in (2.8).

$$\dot{Q}_{tot} = \dot{Q}_{passengers} + \dot{Q}_{solar} + \dot{Q}_{drivetrain_losses} + \dot{Q}_{conduction\&convection} + \dot{Q}_{doors} + \dot{Q}_{HVAC} \quad (2.8)$$

The detailed equations of all the heat transfer rate are developed in the appendix A.1 but here is a brief explanation.

- $\dot{Q}_{passengers}$: every passenger body generates heat around 100 W. There could be an important difference between 0 and 40 passengers onboard.
- $\dot{Q}_{solar}$: the solar irradiation received by the bus heats the cabin. Depends on the intensity of the sun's rays.
- $\dot{Q}_{drivetrain_losses}$: the energy lost in the drivetrain is dissipated in the form of heat and a part of this heat goes to the cabin.
- $\dot{Q}_{conduction\&convection}$: this is the heat transfer rate due to the difference of temperature between the inside of the cabin and the external temperature.
- $\dot{Q}_{doors}$: each time the bus doors open, there is an airflow between the air of the cabin and the outside air
- $\dot{Q}_{HVAC}$: this is the only heat transfer rate that can be controlled.

In summary, the HVAC power consumption depends on many parameters and the required temperature. FIGURE 2.1 is an estimation of the HVAC power consumption from the literature. Other heating requirements are shown in the appendix FIGURE A.1 for different percentages of passengers.

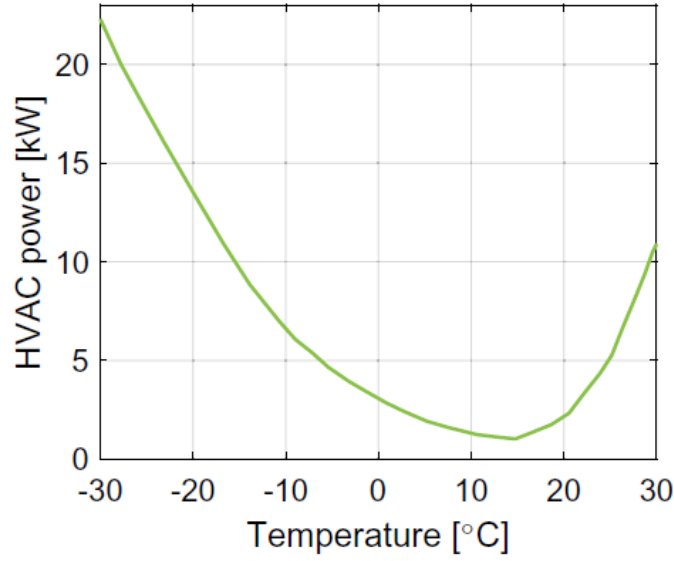


Figure 2.1: Estimation input HVAC power consumption as a function of ambient temperature [22]

It is useful to specify the minimum and maximum temperatures recorded in Brussels by the Royal Meteorological Institute of Belgium in order to know the temperature range. The coldest temperature recorded in winter 2022 is $-5\text{ }^{\circ}\text{C}$ [23] and the hottest temperature recorded in summer 2021 is $30\text{ }^{\circ}\text{C}$ [24]. The conclusion is that energy demand of HVAC for the Brussels temperature range in Brussels is more significant for high temperatures.

2.2 Lithium ion batteries

The BEB's will operate in a wide range of temperature. It is therefore important to evaluate the performance of the batteries at low and high temperatures. To do so, the battery technology is first presented and the operating principle explained. Then, the impact of low and high temperatures on the battery cell is explained. Finally, an overview of the battery temperature conditioning techniques is presented. Temperature conditioning is there in order to maintain the temperature of the battery in a good operating window. This temperature window is generally $[-20;60]\text{ }^{\circ}\text{C}$.

Lithium-ion batteries are the main type of batteries used for electric vehicles due to their interesting characteristics compared to other technologies. The batteries have to be lightweight, low footprint with a high energy storage and high power.

- Specific energy density [Wh/kg] : ranges between 90 and 200 [25]

- Volumetric energy density [Wh/L] : ranges between 150 and 600 [26]
- Power density [W/kg] :

FIGURE A.2 displays different performance compared to other types of technologies.

2.2.1 Operating principle of a battery

The goal of a battery is to store energy in the form of chemical potential and release this energy in form of electricity when needed. There are two operating conditions: the discharge when the battery provides energy and the charge when the battery is plugged in and restores the chemical potential energy.

A single battery cell is composed of two electrodes, the Anode and the Cathode dipped in an electrolyte and with a separator in between as shown in FIGURE 2.2. Electrolyte is conductive for the Lithium-ions, Li^+ , but not to the electrons, e^- . The operating principle relies on oxydation and reduction reaction defined in (2.9) in the electrodes.

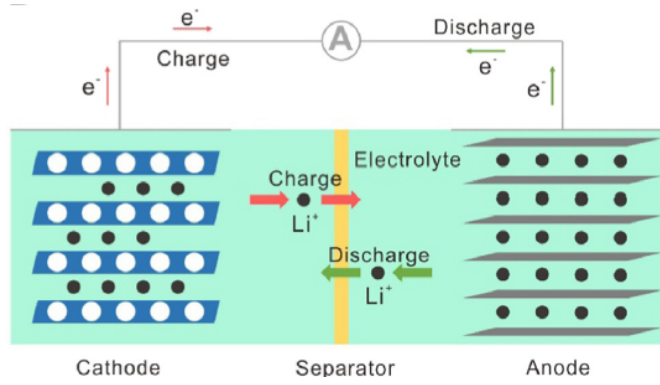
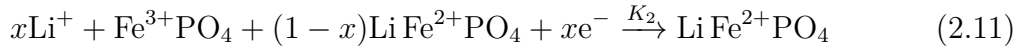
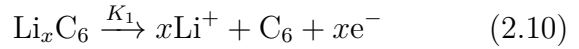
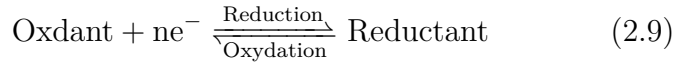


Figure 2.2: Schematic of the operating principle of a battery [27]



- Explanation of the discharge: the oxydation reaction (2.10) occurs in the anode and the free electron moves through the electric circuit whilst the lithium-ion moves in the electrolyte. In the cathode the reduction (2.11) reaction occurs and the the lithium-ion and electron are recombined in a more stable molecule. The more stable molecule is less likely to react, in other words it has lost chemical potential energy.

- Explanation of the charge: a voltage source is connected to the battery with the positive terminal at the cathode and the negative at the anode. The positively charged cathode attracts the electrons of $\text{LiFe}^{2+}\text{PO}_4$ and an oxidation takes place. The negatively charged anode attracts the positively charged lithium-ions through the electrolyte. Lithium-ions and electrons are stored in the loosely bounded structure of the graphene anode and ready to be used for the next discharge.

The voltage cell is the difference of electrochemical potential, E^0 , of the cathode and the anode. Electrochemical potential, E^0 , is the tendency of an oxydant to reduce and gain an electron.

$$V_{cell} = E_{cathode}^0 - E_{anode}^0 [V] \quad (2.12)$$

Cell voltage depends on the state of charge of the battery. When a battery is fully charged (100% state of charge), the cathode is at its highest potential for a reduction and absorbing electron whilst the anode is at its highest potential for a oxydation and provide an electron. The variation of cell voltage can be seen on the charge and discharge curve on FIGURE 2.3. The y-axis is the cell voltage and the x-axis is the electric charge capacity in Ah already consumed of the battery. The energy available in the cell is the area under the curve and is computed in (2.13).

$$E_{cell} = \int_0^{q_{discharge, max}} V_{cell} dq_{discharge} ; [V \cdot Ah] = [Wh] \quad (2.13)$$

2.2.2 Temperature effect on the cell battery

Low temperature effect

At lower temperature, the performance is decreased. An experiment tested 3.7 V 1500 mAh rechargeable Li-ion polymer cells in low temperature conditions [28]. The result is that the voltage and the discharge capacity is reduced as it can be seen in FIGURE 2.3. At $-15^\circ C$ the discharge capacity decreases by 23%.

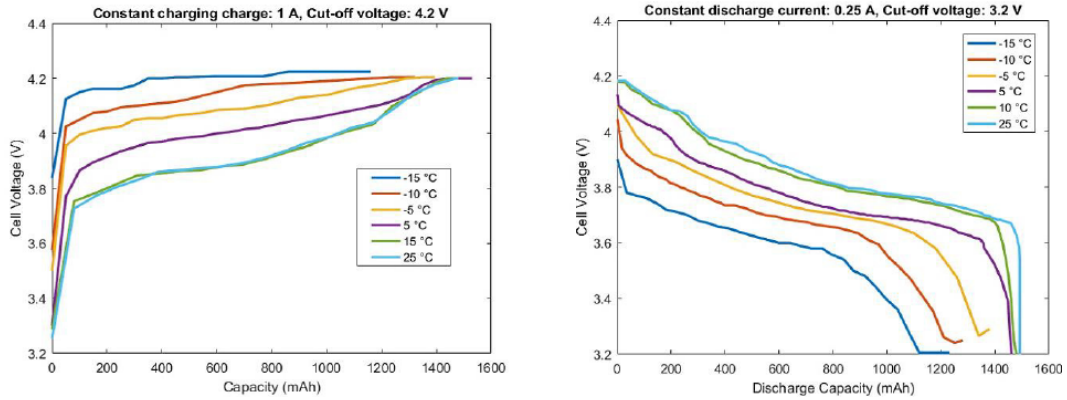


Figure 2.3: Charge and discharge curves for different temperatures [28]

In another experiment at $-15\text{ }^{\circ}\text{C}$, the cell discharge is 1.7 Ah instead of 2.2 Ah at $25\text{ }^{\circ}\text{C}$ [29]. Other studies from the literature have similar results. Performance degradation at low temperature is due to multiple different components that affect the functioning explained previously:

- slower chemical reaction rate: the oxydation and reduction of electrochemical reactions slow down according to the Arrhenius rate law [30]
- ionic conductivity decrease: at lower temperatures, the viscosity of the electrolyte increases and the lithium-ions, Li^+ , velocity is decreased. Lithium-ions take more time to arrive to the cathode and react.
- charge transfer resistance increase: the resistance to transfer an electron from 1 atom to another at the cathode is 3 times higher at $-20\text{ }^{\circ}\text{C}$ [31]

High temperature effect

High temperature increases the performance of the batteries at a given time but accelerates the aging process that will decrease the performance and the lifetime [32]. Here are some phenomenons that occur more at higher temperatures and why this decrease performance in the future:

- Irreversible intercalation covering the electrodes [33]: Electrodes have less reacting area remaining and the total number of reactions is reduced.
- Electrolyte is decomposed [34]: when less electrolyte is remaining, the ionic conductivity is decreased.

An experiment [35] tested the discharge capacity of two identical 2 Ah batteries at $80\text{ }^{\circ}\text{C}$ and $120\text{ }^{\circ}\text{C}$ respectively. After 25 charge-discharge cycles, the capacity of

the battery at 80 °C is 1.85 Ah and for the battery at 120 °C the capacity of the battery is 1.75 Ah. These temperatures were extreme in order to have an impact. Another study [36], tested the maximum storage capacity, Q_m , of batteries every fifty cycles. Four batteries were used at the test temperature of 25, 35, 45 and 50 °C. The percentage of capacity lost at higher temperature is more significant as shown in FIGURE 2.4

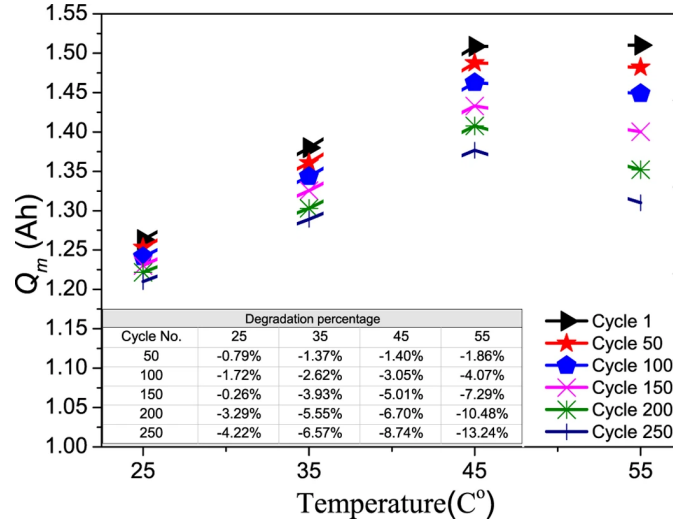


Figure 2.4: Maximum storage capacity depending of number of cycles and temperature (adapted [36])

2.2.3 Battery Thermal Management Systems

As explained previously, low temperature decreases the performance of batteries and high temperature increases their aging. To mitigate these effects, Battery Thermal Management Systems (BTMS) are used either to warm-up the batteries in cold climates or to cool them down during operation because they are generating heat [32]. Two categories of cooling techniques exist, techniques using Vapor Compression Cycle or not. The techniques are detailed in FIGURE 2.5.

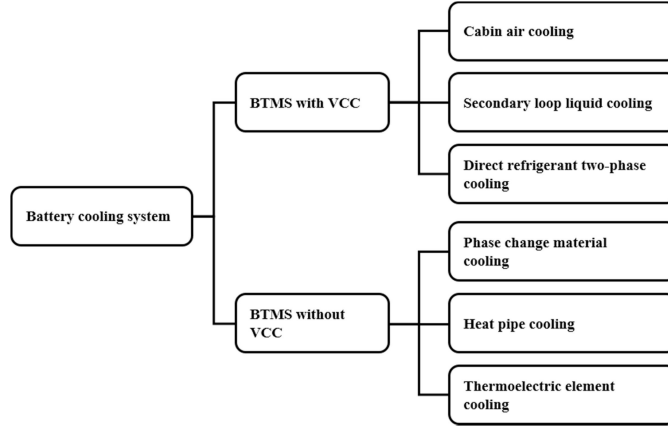


Figure 2.5: Different cooling techniques [37]

All the cooling techniques can be reversed in order to heat the battery. The important criterion before choosing a BTMS is the energy consumption, the cooling or heating duration, the temperature uniformity across the different cells of the battery, the cost and the additional weight.

2.3 Measurement from the literature

ViriCiti, a company providing data analytics for vehicles acquired by charging station manufacturer ChargePoint, investigated the impact of cold and hot temperatures on the consumption [38]. Data was gathered from 100+ electric buses in the Netherlands. Due to the flat landscape, the assumption was made that the average temperature during the shift is the only influencing parameter for the average energy demand of the shift. Shifts selected for the study are at least 40 km long to be representative. These shifts are divided in three temperature classes in function of the average temperature of the shift. The boundaries of the classes are the following, cold between $[-10 ; 14]$ °C, normal between $[15;19]$ °C and hot between $[20;29]$ °C. Cold temperatures increase the energy demand by 14% for 12m buses and 21% for 18m buses. Hot temperatures increase the energy demand by 9% for 12m buses and 12% for 18m buses.

Another study in collaboration with public sector transit agency SARTA used data from four transit agencies in the US 2.2. Similar to the previous study, the assumption was made that the average temperature during the day is the only influencing parameter for the average energy demand of the day. The four locations of the transit agencies are Washington DC (DDOT); Duluth Minnesota (DTA); Worcester Massachussets (WRTA) and Seneca South Carolina. The modelling of the impact of the temperature on the energy demand is done with reference to a

base temperature that is defined as the outdoors temperature where the lowest energy consumption for heating and cooling is registered. The function of the daily average energy demand depending on the daily average temperature is assumed to be in V-shape. The slopes and base temperature are obtained with a best-fit piecewise linear regression model and the results are displayed in Table 2.2.

Transit Agency	Data collection period	Daily Average Temp. Range [°C]	Base Temp [°C]	Increase in energy demand when below base temp [%/°C]	Increase in energy demand when above base temp [%/°C]
DDOT	MAR 2018 - JUN 2019	-10.3 ; 30.3	18	3.8	2.1
DTA	NOV 2018 - JUN 2019	-30.5 ; 20	12.2	1.5	Not enough hot data
Seneca	SEPT 2014 - JUL 2018	-9.4 ; 28.9	18.3	0.6	1.3
WRTA	SEPT 2013 - AUG 2017	-17.8 ; 27.2	13.7	2.7	2.6

Table 2.2: Summary of the main results [39] (adapted in °C from °F)

Chapter 3

Analysis of the data from STIB/MIVB

The STIB/MIVB is the main public transport operator in Brussels (Société des Transports Intercommunaux de Bruxelles, Maatschappij voor het Intercommunaal Vervoer te Brussel). The STIB is member of the International Association of Public Transport (UITP) a worldwide network bringing together all public transport stakeholders, sustainable transport modes, vehicle manufacturers and technology providers. UITP has more than 1,800 members in 100 countries and represents the key players needed for zero carbon urban mobility.

For many years, the STIB has attached great importance to the ecological performance of its vehicles. When purchasing its new buses, it has always taken care to anticipate the strictest European standards in terms of polluting emissions. In June 2018, after a test period, the STIB put the first fully electric bus line into service. The data from this line is analyzed in this chapter.

3.1 Bus line 33

The bus line that is studied is the line 33 of STIB/MIVB. It is a downtown local route between two terminals called "Louise" and "Dansaert". It connects the upper and lower parts of the city of Brussels by crossing the city centre.

- Louise, located on the Brussels inner ring road in the upper part of the city, is also a metro stop on lines 2 & 6.
- Dansaert is a district in the center of the Brussels city pentagon, close to the "Sainte-Catherine" metro stop on lines 1 & 5 and the Brussels Canal in the lower part of the city.

The buses of line 33 run every day at a time interval of 20 minutes. It is the first STIB bus line to be operated solely by battery electric buses with a capacity allowing them to manoeuvre easily in the narrow streets of the city centre.

The routes are shown on a map in FIGURE 3.1 with the details of each stops.

TABLE 3.1 summarize some additional information about the distance and elevation of the routes. The length difference between the two directions are due to one-way streets and choice of placement of the bus stops.

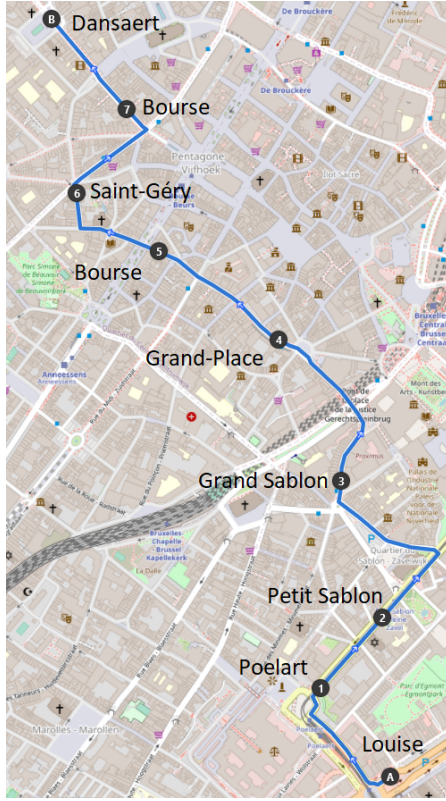
			elevation [m]		
	Distance [km]	stops [-]	gain	loss	net gain
from Louise to Dansaert	2.5	7	3	51	-48
from Dansaert to Louise	3.5	8	64	16	+48

Table 3.1: Information about the route of bus 33

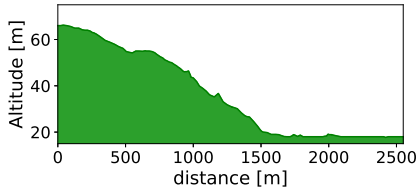


Figure 3.2: Illustration of the bus during a shift [40]

All buses are the same model, Solaris Urbino 8.9 LE electric fitted with 160 kWh battery [41]. The bus has a capacity of 22 seats and 20 standing places. The empty vehicle weights 10.7 tons.



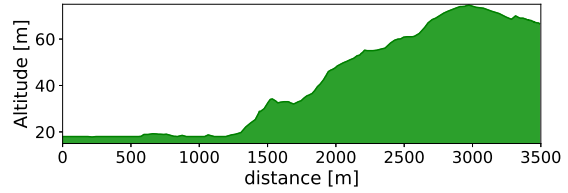
(a) from Louise to Dansaert



(c) from Louise to Dansaert



(b) from Dansaert to Louise



(d) from Dansaert to Louise

Figure 3.1: Route of the bus 33 on the map

3.2 Presentation of the raw data

The dataset at our disposal is the daily operation measurements logs of seven battery electric buses in Brussels. From the start at the depot to the return to the depot after a shift.

Data are available for January to mid-April 2021 and also for June 2021. Data are logged every 5 seconds. In total, there are approximately 1 million samples.

The dataset provided by the STIB is divided into two file types.

- Firstly, the files that come from the Controller Area Network (CAN) of each bus and are called mbse. There are 2560 files.
- Secondly, the files that come from the passenger count system called covo. There are 566 files.

3.2.1 Files from the bus computer

The sampling time is 5 seconds or 1 minute depending on the file. An example for a sample is shown in TABLE 3.2. The pattern for the files names is as follows:

`mbse_busnumber_6_yyyymmdd_hhmmss`,

with an example: `mbse_1001_6_20210601_163534.csv`.

mbse headers	mbse data example	mbse headers	mbse data example
build	12-10-20	charging	0
timestamp	2021-06-01T16:44:19	soc	99
vehicle	1001	autonomy_alarm	0
started	1	charging_completed	0
line	33	charge	62193
driver	72437	regen	609152
indice	706	discharge	82784
previous_stop	3257	traction_input_power	6
latitude	508.494.983	traction_battery_power	43,6
longitude	43.295.633	input_of_container	690
meter	59679720	battery_current	31
temp_ext	25,4	speed	28

Table 3.2: Example of a mbse file with all columns

FIGURE 3.3 is a screenshot from the mbse file received as is by STIB. It can be observed that the columns **line**, **driver** and **indice** are always empty at the beginning of a shift. This is the same at the end of a shift. These values depend of a manipulation of the driver. The assumption is that it manages the display on the information screen above the bus (see FIG 3.2). However, there is unfortunately no column for the destination.

Some columns are empty, constant nor bring information that can be exploited. The selected columns are colored in light green.

1	build	timestamp	vehicle	started	line	driver	indice	previous_latitude	longitude	meter	temp_ext	charging	soc	autonomy_alarm	charging_completed	charge	regen	discharge	traction_input_power	traction_battery_power	input_of_container	battery_current	speed	alarm
2	12-10-20	2021-06-01T16:35:35	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	-1.8.0		703	11		BOOTING
3	12-10-20	2021-06-01T16:35:40	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		702	11		BOOTING
4	12-10-20	2021-06-01T16:35:45	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	-1.8.1		702	12		BOOTING
5	12-10-20	2021-06-01T16:35:50	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.3		702	11		BOOTING
6	12-10-20	2021-06-01T16:35:55	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.1		703	11		BOOTING
7	12-10-20	2021-06-01T16:36:00	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.3		702	12		BOOTING
8	12-10-20	2021-06-01T16:36:05	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.2		701	11		BOOTING
9	12-10-20	2021-06-01T16:36:10	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		701	11		BOOTING
10	12-10-20	2021-06-01T16:36:15	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	-1.8.1		702	11		BOOTING
11	12-10-20	2021-06-01T16:36:20	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.1		702	11		BOOTING
12	12-10-20	2021-06-01T16:36:25	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		701	11		BOOTING
13	12-10-20	2021-06-01T16:36:30	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	-1.8.0		701	11		BOOTING
14	12-10-20	2021-06-01T16:36:35	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.1		701	11		
15	12-10-20	2021-06-01T16:36:40	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		701	11		
16	12-10-20	2021-06-01T16:36:45	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		701	11		
17	12-10-20	2021-06-01T16:36:50	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	11		
18	12-10-20	2021-06-01T16:36:55	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	11		
19	12-10-20	2021-06-01T16:37:00	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		701	11		
20	12-10-20	2021-06-01T16:37:05	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		701	11		
21	12-10-20	2021-06-01T16:37:10	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.8		700	11		
22	12-10-20	2021-06-01T16:37:15	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		700	11		
23	12-10-20	2021-06-01T16:37:20	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	11		
24	12-10-20	2021-06-01T16:37:25	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	11		
25	12-10-20	2021-06-01T16:37:30	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	10		
26	12-10-20	2021-06-01T16:37:35	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.8.0		699	11		
27	12-10-20	2021-06-01T16:37:40	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.8		699	11		
28	12-10-20	2021-06-01T16:37:45	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	10		
29	12-10-20	2021-06-01T16:37:50	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	-1.7.7		699	10		
30	12-10-20	2021-06-01T16:37:55	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		700	11		
31	12-10-20	2021-06-01T16:38:00	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.9		699	11		
32	12-10-20	2021-06-01T16:38:05	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.8		698	10		
33	12-10-20	2021-06-01T16:38:10	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.7		698	11		
34	12-10-20	2021-06-01T16:38:15	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.8		699	11		
35	12-10-20	2021-06-01T16:38:20	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.8		699	11		
36	12-10-20	2021-06-01T16:38:25	1001	1				508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.7.9		698	11		
37	12-10-20	2021-06-01T16:38:30	1001	1				508.463.150	43.201.150	59678460	25.1	0	100	0	0	62193	609152	82780	0.7.7		698	11		
38	12-10-20	2021-06-01T16:38:35	1001	1				508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.7.7		698	10		
39	12-10-20	2021-06-01T16:38:40	1001	1				508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.7.8		699	10		
40	12-10-20	2021-06-01T16:38:45	1001	1				508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	0.7.9		698	11		
41	12-10-20	2021-06-01T16:38:50	1001	1				508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.7.7		699	11		
42	12-10-20	2021-06-01T16:38:55	1001	1	33	72437	706	508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	0.7.6		698	10		
43	12-10-20	2021-06-01T16:39:00	1001	1	33	72437	706	508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.7.8		697	11		
44	12-10-20	2021-06-01T16:39:05	1001	1	33	72437	706	508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	-1.8.3		698	12		
45	12-10-20	2021-06-01T16:39:10	1001	1	33	72437	706	508.463.150	43.201.150	59678460	25.4	0	100	0	0	62193	609152	82780	6.16.7		698	12		
46	12-10-20	2021-06-01T16:39:15	1001	1	33	72437	706	508.463.150	43.201.150	59678465	25.4	0	100	0	0	62193	609152	82780	0.23.2		697	11	3	
47	12-10-20	2021-06-01T16:39:20	1001	1	33	72437	706	508.462.433	43.202.200	59678475	25.4	0	100	0	0	62193	609152	82780	27.10.3		697	29	7	
48	12-10-20	2021-06-01T16:39:25	1001	1	33	72437	706	508.461.050	43.201.735	59678500	25.4	0	100	0	0	62193	609152	82780	49.50.2		694	84	17	

Figure 3.3: Example of a mbse file received from STIB

3.2.2 Files from the passenger count system

The Brussels Intermunicipal Transport Company (STIB-MIVB) count their passengers to improve their services and offer a better performing, more comfortable network [42]. Passenger counting is achieved by entry/exit sensors at the doors. Each time the bus opens the doors it adds a row of measurement to the file as it is displayed on FIGURE 3.4. The selected columns are colored in light green. The **passengers** column is not given explicitly but calculated afterwards by computing the running total of the difference of people getting **in** minus the people getting **out**. Unfortunately, the measurements are not always reliable and the number of passengers can be negative. Another problem is that 40% of the data does not have information about the passengers. The patch for this problem will be explained later.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	veh	date	hour	duration	status	line	serv	dest	ppt	stop	lat	long	odo	in	out	devid	in1	out1	status1	in2	out2	status2
2	1001	01-06-21	08:22:50	0	0x00000000	0	0	904	0	0	0000.0000	0000.0000	0	0	0	4	0	0	OK	0	0	OK
3	1001	01-06-21	08:23:58	8	0x00000000	0	0	904	0	0	508.463	43.203	1529	0	0	4	0	0	OK	0	0	OK
4	1001	01-06-21	08:27:34	22	0x00000000	0	0	904	0	0	508.471	43.212	2080	1	1	4	1	1	OK	0	0	OK
5	1001	01-06-21	08:28:56	58	0x00000000	0	0	904	0	0	508.455	43.210	3490	1	1	4	1	1	OK	0	0	OK
6	1001	01-06-21	08:33:06	3598	0x00000000	0	0	904	0	0	508.455	43.208	3763	0	0	4	0	0	OK	0	0	OK
7	1001	01-06-21	09:49:02	24288	0x00000000	0	0	904	0	0	508.463	43.203	1949	0	0	4	0	0	OK	0	0	OK
8	1001	01-06-21	16:34:55	250	0x00000000	0	0	904	0	0	508.463	43.203	1949	6	6	4	6	6	OK	0	0	OK
9	1001	01-06-21	16:50:45	26	0x00000000	33	706	3301	0	2212	508.512	43.446	10709	3	0	4	0	0	OK	3	0	OK
10	1001	01-06-21	16:59:11	10	0x00000000	33	706	3301	0	7142	508.463	43.431	14419	1	0	4	0	0	OK	1	0	OK
11	1001	01-06-21	17:02:57	10	0x00000000	33	706	3301	0	7142	508.448	43.460	16954	3	0	4	0	0	OK	3	0	OK
12	1001	01-06-21	17:06:03	10	0x00000000	33	706	3301	0	7142	508.412	43.517	20198	0	0	4	0	0	OK	0	0	OK
13	1001	01-06-21	17:06:03	10	0x00000000	33	706	918	0	7142	508.412	43.517	20198	1	0	4	0	0	OK	1	0	OK
14	1001	01-06-21	17:06:47	22	0x00000000	33	706	3301	0	1128	508.415	43.533	20686	4	1	4	0	0	OK	4	1	OK
15	1001	01-06-21	17:08:21	20	0x00000000	33	706	3301	0	1128	508.399	43.559	21729	0	1	4	0	0	OK	0	1	OK
16	1001	01-06-21	17:11:43	20	0x00000000	33	706	3301	0	2177	508.392	43.613	24525	0	6	4	0	0	OK	0	6	OK
17	1001	01-06-21	17:14:07	724	0x00000000	33	706	3302	0	2177	508.362	43.554	26922	14	7	4	6	5	OK	8	2	OK
18	1001	01-06-21	17:28:13	12	0x00000000	33	706	3302	0	3630	508.393	43.554	28949	3	0	4	0	0	OK	3	0	OK
19	1001	01-06-21	17:32:05	12	0x00000000	33	706	3302	0	1781	508.447	43.524	32490	1	0	4	0	0	OK	1	0	OK

Figure 3.4: Example of a covo file received from STIB

3.3 Data Preprocessing

Before you can use the data in a machine learning model, it is necessary to link the 2 raw file types but also to add some important information, such as altitude, which can be deduced from GPS coordinates. This section explains the process illustrated on FIGURE 3.5 and how some data were enriched and others aggregated and why. All the data processing has been realized in Python unless specified otherwise.

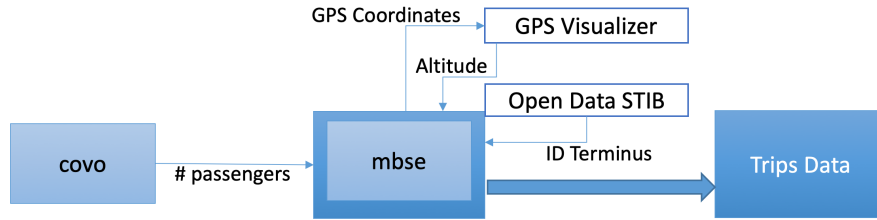


Figure 3.5: Flow chart of the Data Preprocessing

3.3.1 Addition of elevation

The elevation is an important metric regarding the energy consumption of vehicles but this information is not present in the files. Here, given the elevation profile of the bus line studied, it is essential to add the altitude. This was done, by using the GPS coordinates present in the mbse file and injecting them into the external GPS Visualiser site [43], which provides the elevation corresponding to the GPS coordinates. The gain or loss of elevation between 2 measurement points is also added. The result is visible in FIGURE 3.6. The Python code is not included in the appendices but is available to Laborelec.

3.3.2 Linking the raw data

Before continuing the analysis, passenger data must be added to the mbse file. First, the corresponding covo file is found for each mbse file by matching the dates. The dates are part of the file names and this is how the covo file can be found. Second, each in and out of passengers from the covo file is added in the newly created columns in the mbse file with an example at line 592 in FIGURE 3.6. The index where to add the data is chosen by comparing the timestamps of the mbse and covo files. Finally, a cumulative sum of the in-out is computed to obtain the numbers of **passengers** onboard and set in a column.

	timestamp	weekday	prev_stop	latitude	longitude	temp_ext	soc	P_in	P_batt	speed	altitude	altitude_gain	meter_travelled	travel	terminus	passengers	in	out
571	2021-06-01 17:26:32	1	3505	50.836	4.355	26.4	94	-2	28.2	7	66.0	0.0	10	1	0	11.0	NaN	NaN
572	2021-06-01 17:26:37	1	3505	50.836	4.355	26.4	93	0	45.0	7	66.0	0.0	10	1	0	11.0	NaN	NaN
573	2021-06-01 17:26:42	1	2177	50.836	4.355	26.4	93	49	79.4	21	66.1	0.1	30	2	1	11.0	NaN	NaN
574	2021-06-01 17:26:47	1	2177	50.836	4.354	26.4	93	4	39.0	25	66.0	-0.1	35	2	1	11.0	NaN	NaN
575	2021-06-01 17:26:52	1	2177	50.837	4.354	26.4	93	-3	-2.5	25	64.6	-1.4	35	2	1	11.0	NaN	NaN
576	2021-06-01 17:26:57	1	2177	50.837	4.354	26.4	93	-55	-8.3	25	65.0	0.4	35	2	1	11.0	NaN	NaN
577	2021-06-01 17:27:02	1	2177	50.837	4.354	26.4	93	-7	-31.4	17	64.1	-0.9	25	2	1	11.0	NaN	NaN
578	2021-06-01 17:27:07	1	2177	50.837	4.353	26.4	93	-29	-3.5	14	64.0	-0.1	20	2	1	11.0	NaN	NaN
579	2021-06-01 17:27:12	1	2177	50.837	4.353	26.4	93	-1	1.9	14	64.0	0.0	20	2	1	11.0	NaN	NaN
580	2021-06-01 17:27:17	1	2177	50.838	4.353	26.4	93	0	-16.2	14	63.5	-0.5	20	2	1	11.0	NaN	NaN
581	2021-06-01 17:27:22	1	2177	50.838	4.353	26.4	93	-28	-18.0	10	63.0	-0.5	15	2	1	11.0	NaN	NaN
582	2021-06-01 17:27:27	1	2177	50.838	4.353	26.4	93	36	25.7	7	62.8	-0.2	10	2	1	11.0	NaN	NaN
583	2021-06-01 17:27:32	1	2177	50.838	4.354	26.4	93	73	30.4	17	62.3	-0.5	25	2	1	11.0	NaN	NaN
584	2021-06-01 17:27:37	1	2177	50.838	4.354	26.4	93	-27	0.8	25	60.8	-1.5	35	2	1	11.0	NaN	NaN
585	2021-06-01 17:27:42	1	2177	50.838	4.354	26.4	93	-13	-23.2	25	59.7	-1.1	35	2	1	11.0	NaN	NaN
586	2021-06-01 17:27:47	1	3629	50.839	4.354	26.1	93	-52	-35.4	21	58.9	-0.8	30	2	1	11.0	NaN	NaN
587	2021-06-01 17:27:52	1	3629	50.839	4.355	26.1	93	-20	-47.5	28	58.2	-0.7	40	2	1	11.0	NaN	NaN
588	2021-06-01 17:27:57	1	3629	50.839	4.355	26.1	93	-85	-11.0	17	57.4	-0.8	25	2	1	11.0	NaN	NaN
589	2021-06-01 17:28:02	1	3629	50.839	4.355	26.1	94	-38	-18.0	14	56.8	-0.6	20	2	1	11.0	NaN	NaN
590	2021-06-01 17:28:07	1	3629	50.839	4.355	26.1	94	0	-1.6	7	56.5	-0.3	10	2	1	11.0	NaN	NaN
591	2021-06-01 17:28:12	1	3629	50.839	4.355	26.1	94	-1	2.2	0	56.5	0.0	0	2	1	11.0	NaN	NaN
592	2021-06-01 17:28:17	1	3629	50.839	4.355	26.1	94	0	4.3	0	56.5	0.0	0	2	1	14.0	3.0	0.0
593	2021-06-01 17:28:22	1	3629	50.839	4.355	26.1	94	0	5.4	0	56.5	0.0	0	2	1	14.0	NaN	NaN
594	2021-06-01 17:28:27	1	3629	50.839	4.355	26.1	94	0	6.5	0	56.5	0.0	0	2	1	14.0	NaN	NaN

Figure 3.6: Data after preprocessing

3.3.3 Choice of time reference

It is important to define the appropriate time reference in which the data will be averaged and the energy consumption will be predicted with the machine learning regression algorithms.

First, let's look at the data in raw form and show why taking the sampling time as

time reference is not ideal. Each point in FIGURE 3.7 represents a measurement that is taken every 5 seconds. It can be observed that the data is mainly independent because the data points are mostly evenly distributed. In other words, it is not possible to tell some change in one variable when changing the value of the other. The data seems noisy and according to [44] the occurrences of noisy data in data set can significantly impact prediction of any meaningful information. Averaging the data will smooth the data and reduce the noise.

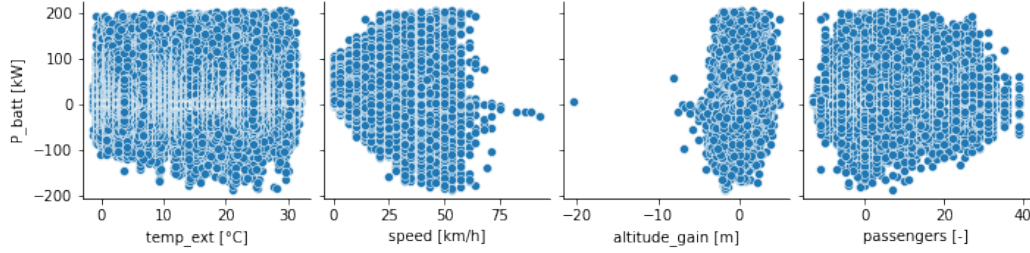


Figure 3.7: Scatter plots of the power of the battery

An averaging time must be decided. Instead of arbitrarily setting an average time of 10 minutes, 20 minutes or one hour, it seemed more appropriate to take an average of data per trip. This will allow to make comparisons on similar trips with similar elevation profiles. In addition, average temperatures over the duration of a 20-30 minute trip will also be more representative than over a full 8 hours and 30 minutes shift where temperature variations can be significant.

3.3.4 Classification by trip

The separation of the data between trips is achieved using the column `previous_stop` in the mbse file FIGURE 3.7. Thanks to a table A.3 (available on the opendata platform of STIB [45]) that matches the bus stops ID and bus stops names. We know that the stop ID **3633B** or **3633C** corresponds to Dansaert and the stop ID **2177** corresponds to Louise. The aim is to fill the columns:

- **travel**: the account of trips made during the shift (starting at zero)
- **terminus**: information about the type of trip. (-1: 'from depot', 0: 'from Louise to Dansaert', 1: 'from Dansaert to Louise', 2: 'to Depot')

Here is a simplified summary of the steps followed in the code to create the distinction between the different trips.

1. At the beginning of each msbe shift file the samples where **line**, **driver** and **index** are null are discarded. Those are generally the few samples at the bus depot when the bus is turned on.

2. Create empty columns for the values `travel` and `terminus`.
3. For the first samples `travel` is set to zero because it is the first of the day and `terminus` is set to -1 because the bus is leaving the depot.
4. When the bus leaves a terminal, the column `previous_stop` becomes either **2177** when it leaves Louise or **3633C/3633B** when it leaves Dansaert. This is the signal that another trip should be created.
5. For the following samples `travel` is set to the incremented value because it is an additional trip during the shift. `terminus` is set to '0' if the bus has left Louise to go to Dansaert and '1' if the bus has left Dansaert to go to Louise
6. Steps 4 and 5 are repeated until the columns `line`, `driver` and `indice` become empty at the end of the mbse file. This occurs generally when the bus reaches the depot and the few remaining data are discarded.
7. All the values of `terminus` for the current trip are set to '2' because it is the trip returning to the depot.

The following points describe the trips in more detail and mention areas for improvement in the trip classification.

- from depot: are the trips at the beginning of the shifts until the bus has arrived to one or another terminus. Average time of all these trips is 25 minutes. In order to increase precision, it would be possible in a later phase, by adding a function in the code, to separate this classification to specify whether the bus reaches Dansaert or Louise.
- from Louise to Dansaert, the bottom of the city: Average time of all these trips is 20 minutes for 2,5 km with 7 stops.
- from Dansaert to Louise, the top of the city: Average time of all these trips is 35 minutes for 3,5 km with 8 stops.
- to Depot: these trips start when the bus leaves for the last time of the shift a terminal and ends when the columns `line`, `driver` and `indice` become empty. As a reminder, this occurs generally when the bus reaches the depot. Unfortunately, in rare occasions the columns `line`, `driver` and `indice` become empty when the bus reaches the depot for the last time of the day. The last trip with passengers is then falsely categorized as a trip 'to depot'. In a later phase, a more robust approach could be to check that the last trip matches with the return to the depot with a comparison of the GPS coordinates. In this way, it would be ensured that no Dansaert or Louise trip

is put in the To Depot category when there is a reboot of the system, which may happen during an incomplete shift.

The data analysis presented in this work mainly focused on the trips with passengers between Dansaert and Louise. This is why these improvements to separate the classification "from Depot" and to clarify the trip "to depot" have not yet been added to the code.

3.3.5 Validation Classification by trip

Now that the trips are separated from each other, the energy consumption can be computed with the equation (3.1).

$$\text{Energy Consumption [kWh/km]} = \frac{\text{Energy [kWh]}}{\text{distance [km]}} \quad (3.1)$$

Energy consumption is a useful measure for comparing efficiency on routes of different lengths, which is the case in this study. In addition to energy consumption, other data has been added to the trip data: the mean and standard deviation of the speed to give an idea of the smoothness of the ride. FIGURE 3.9 are all the different trips saved during the complete shift of june first 2021.

FIGURE 3.8 is an overview of the observations that can be made when aggregating the data by trip. It can be seen that the trips to Louise have a higher energy consumption than the others and this can be explained by the fact that the trip is mainly uphill as explained previously. This supports the intuition that a classification of data by trip is a good solution.

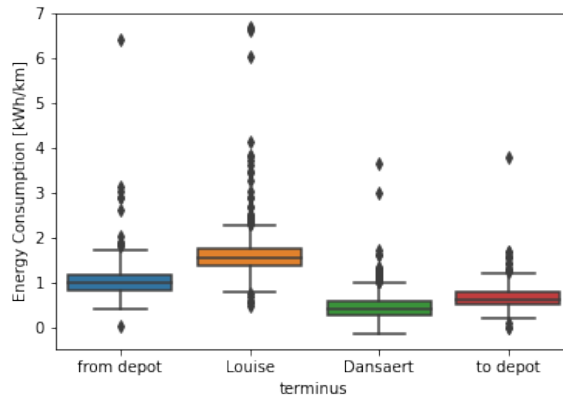


Figure 3.8: Boxplot of the energy consumption

Before further analysis of the data, some outliers were removed. This concerns 45 trips, or only one percent of the 3.659 trips studied in the following sections. The

energy outliers were mainly "false starts", where the bus moves forward 10 meters and then there is consumption at standstill for 1 hour. As explained previously, another problem is that 40% of the data does not contain information about the passengers. The patch to make the program work is to set the empty values with the mean of the `passenger_mean` of all the other trips.

	start time	weekday [0;6]	travel time [s]	vehicle	terminus	v_mean [km/h]	v_std	temp_ext [°C]	P_in_mean [kW]	P_batt_mean [kW]	distance [km]	Energy [kWh]	Energy Consumption [kWh/km]	elevation gain [m]	elevation loss [m]	passengers mean [m]
0	2021-06-01 16:38:55	1	750	1001	from depot	12.583	12.525	25.646	4.007	10.287	2.710	2.143	0.791	9.1	-27.3	0.179
1	2021-06-01 16:51:30	1	2107	1001	Louise	6.685	9.309	25.817	10.917	12.695	4.065	7.430	1.828	65.5	-17.5	7.517
2	2021-06-01 17:26:42	1	1366	1001	Dansaert	6.299	7.746	25.673	-0.624	2.466	2.505	0.936	0.374	4.0	-52.0	8.091
3	2021-06-01 17:49:33	1	2333	1001	Louise	6.017	8.828	25.414	8.473	8.541	4.065	5.535	1.362	65.8	-17.8	2.829
4	2021-06-01 18:28:31	1	1286	1001	Dansaert	6.655	7.874	25.169	-0.833	3.081	2.500	1.101	0.440	4.7	-52.7	3.651
5	2021-06-01 18:50:02	1	2456	1001	Louise	5.740	9.094	25.293	9.496	9.607	4.070	6.554	1.610	65.0	-16.9	6.732
6	2021-06-01 19:31:03	1	1107	1001	Dansaert	7.752	9.686	24.917	-1.626	7.502	2.485	2.307	0.928	3.9	-52.0	4.815
7	2021-06-01 19:49:35	1	2398	1001	Louise	5.981	9.971	24.391	9.438	9.895	4.125	6.591	1.598	65.2	-17.1	4.308
8	2021-06-01 20:29:38	1	1126	1001	Dansaert	7.646	9.429	24.120	-1.199	2.902	2.495	0.908	0.364	3.7	-51.8	7.111
9	2021-06-01 20:48:29	1	1451	1001	Louise	9.773	10.856	23.625	15.247	14.269	4.080	5.751	1.410	64.9	-16.8	6.357
10	2021-06-01 21:12:45	1	1011	1001	Dansaert	8.419	9.775	23.280	-1.399	4.378	2.470	1.230	0.498	4.7	-52.8	6.443
11	2021-06-01 21:29:41	1	1371	1001	Louise	10.269	11.673	22.892	17.495	17.449	4.060	6.645	1.637	65.3	-17.3	8.676
12	2021-06-01 21:52:37	1	1066	1001	Dansaert	8.150	10.119	22.618	3.089	7.469	2.515	2.212	0.879	3.1	-51.1	7.136
13	2021-06-01 22:10:28	1	1281	1001	Louise	11.012	11.871	22.134	16.930	16.621	4.060	5.914	1.457	64.1	-16.1	8.249
14	2021-06-01 22:31:54	1	1126	1001	Dansaert	7.748	10.091	21.801	-0.553	3.224	2.515	1.008	0.401	4.7	-52.7	6.588
15	2021-06-01 22:50:45	1	1261	1001	Louise	11.281	12.493	21.577	18.545	15.593	4.085	5.462	1.337	64.0	-15.9	7.209
16	2021-06-01 23:11:51	1	1051	1001	Dansaert	8.232	10.130	21.439	-0.749	2.272	2.500	0.663	0.265	4.4	-52.5	7.645
17	2021-06-01 23:29:27	1	796	1001	Louise	17.956	15.333	21.110	28.212	20.371	4.095	4.504	1.100	64.0	-16.8	NaN
18	2021-06-01 23:42:48	1	510	1001	to depot	27.010	16.741	20.689	21.107	17.117	3.930	2.425	0.617	23.5	-50.2	NaN

Figure 3.9: Example of data saved by trip

3.4 Analysis on a micro level

This section aims to check the sanity of the data by observing plots of the logs before continuing the analysis. It presents the data from a shift and then the data from a trip. The conclusion is that the data can be considered valid and exploitable.

3.4.1 By shift

Here are the data for a full shift of 8 hours and 30 minutes (here from 15:00 to 23:30) in FIGURE 3.10.

As expected, the state of charge of the battery decreases during the shift while the travelled distance increases.

Looking at the altitude, and knowing the complete route of line 33 and the timetable (the bus remains at the terminals between two travels), we can see that during a shift the bus makes the route several times. We can also see that the time to complete the route is shorter at the end of the day, when there are fewer people and the traffic is more fluid. In terms of passenger numbers, there is an anomaly with negative figures as explained in the section 3.2.2. However, the following trend can be observed: more passengers between 4pm and 5pm when leaving the office, then less in favour of the metro because of traffic jams, a little more before 9pm for parties and a peak just before the end of the day for the late evening in the Dansaert district with its restaurants and bars.

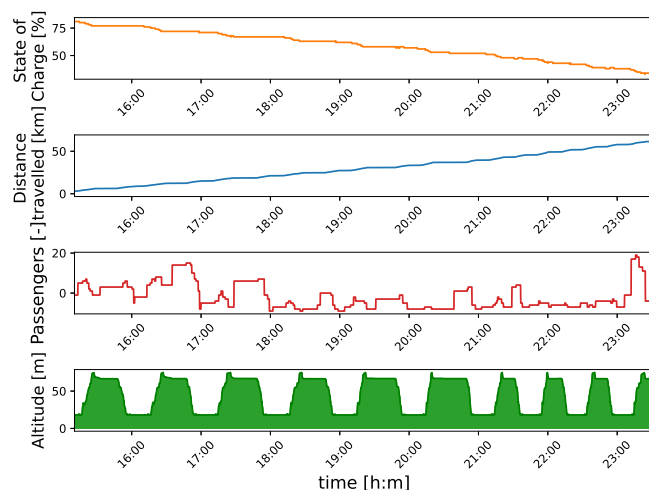


Figure 3.10: Log of a shift, Thursday January 14th 2021

3.4.2 By trip

If we look in detail at a trip from 15:00 to 15:30, we also see consistent things. Bus stops are represented with black dots on the speed graph in FIGURE 3.11. The number of passengers varies clearly near stops and when the speed is zero. Power is negative when the bus is decelerating. This can be explained by regenerative braking. Regenerative braking is a means of recovering electrical energy from the braking mechanism, typically implemented in electric vehicles [46].

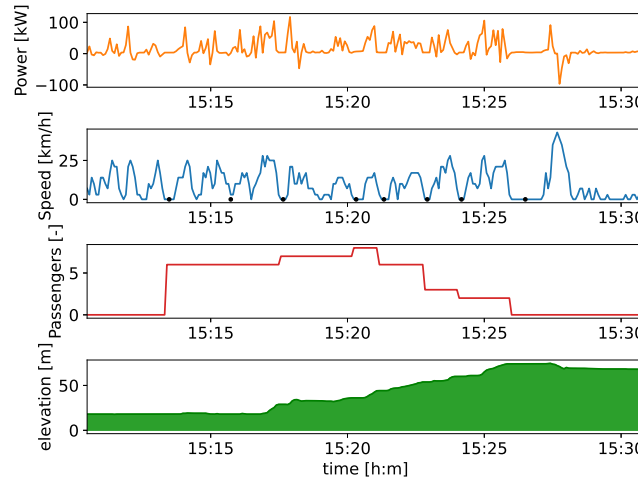


Figure 3.11: Log of a trip, Thursday January 14th 2021

3.5 Aggregation by trip

3.5.1 Graphical analysis

In this subsection, scatter plots of the energy consumption depending on certain parameters like the external temperature, average speed and the standard deviation of the speed can be found.

In FIGURE 3.12 FIGURE 3.13 and FIGURE 3.14, we can observe the following:

- The higher energy consumption for the trip to Louise because it is uphill.
- At higher temperatures, more energy is consumed due to the use of air conditioning.
- Higher consumption per km at lower average speed because even at a standstill the bus consumes energy (and even more if air conditioning). Moreover, a

lower average speed also implies more frequent stops and therefore the energy needed to restart a bus of 10.7 tons.

- The standard deviation taken separately is not very relevant and no trend can be identified.
- The same outlier in consumption can be identified everywhere. They are not related to temperature. They are the same outliers as those found in FIGURE 3.8

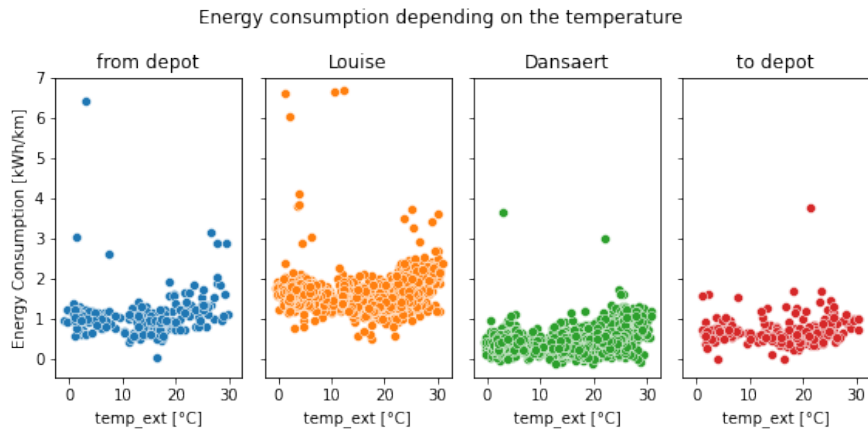


Figure 3.12: Energy Consumption depending on the temperature

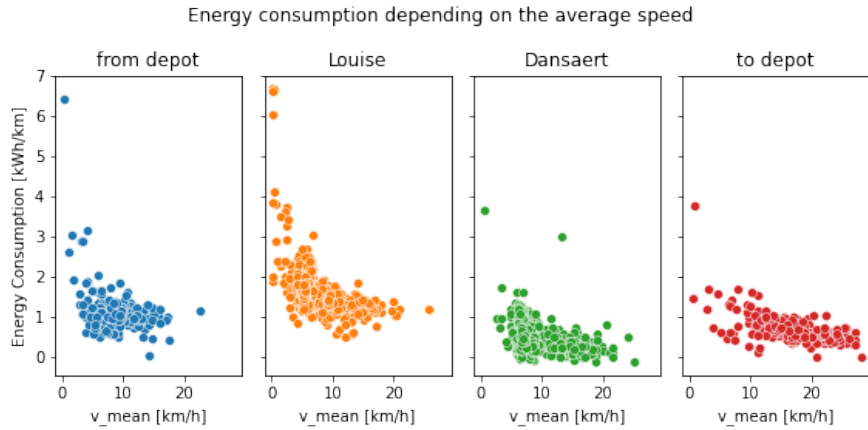


Figure 3.13: Energy Consumption depending on the mean speed

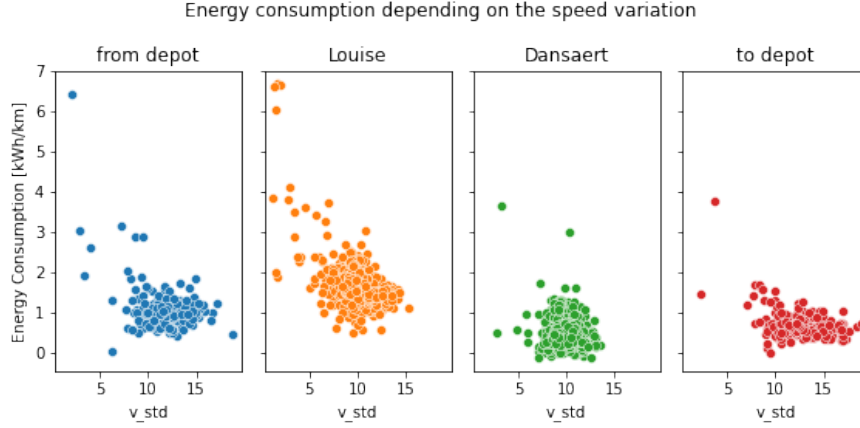


Figure 3.14: Energy Consumption depending on the standard deviation of the speed

3.6 Visualization of the influence of temperature on consumption

The goal of this section is to visualize the influence of the temperature on the energy consumption. To do so, different graphs are plotted in order to get a 'feel' of the data before starting the modelization with machine learning algorithms. The dataset at our disposal contains trips at different average temperatures. The diversity of these temperatures can be seen on FIGURE 3.15. The positive point temperature range is wide going from freezing temperatures up to 30 $^{\circ}C$. Ideally, the distribution would be uniform meaning that we have the same number of samples for every temperature in order to have a good representation of all the temperatures.

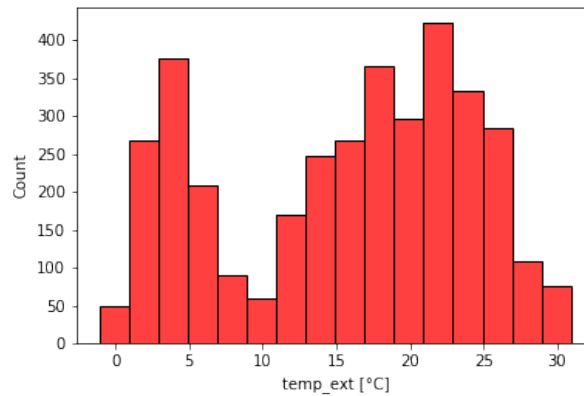


Figure 3.15: Histogram of the average temperature of the trips

3.6.1 Energy consumption by temperature range

It has been seen previously that the type of trip has a significant impact on the energy consumption, therefore the plots are separated by type of trips. FIGURE 3.16 displays boxplots of energy consumption of the subdatasets divided by temperature range of 3 $^{\circ}C$. Energy consumption is the highest at 30 $^{\circ}C$ and a minimum can be observed in the temperature ranges between 9 $^{\circ}C$ and 12 $^{\circ}C$. This consolidates the hypothesis of a base temperature where the energy consumption for heating or cooling is minimal. A slightly polynomial trend can be discerned when looking at the different plots. This has to be investigated with machine learning algorithms.

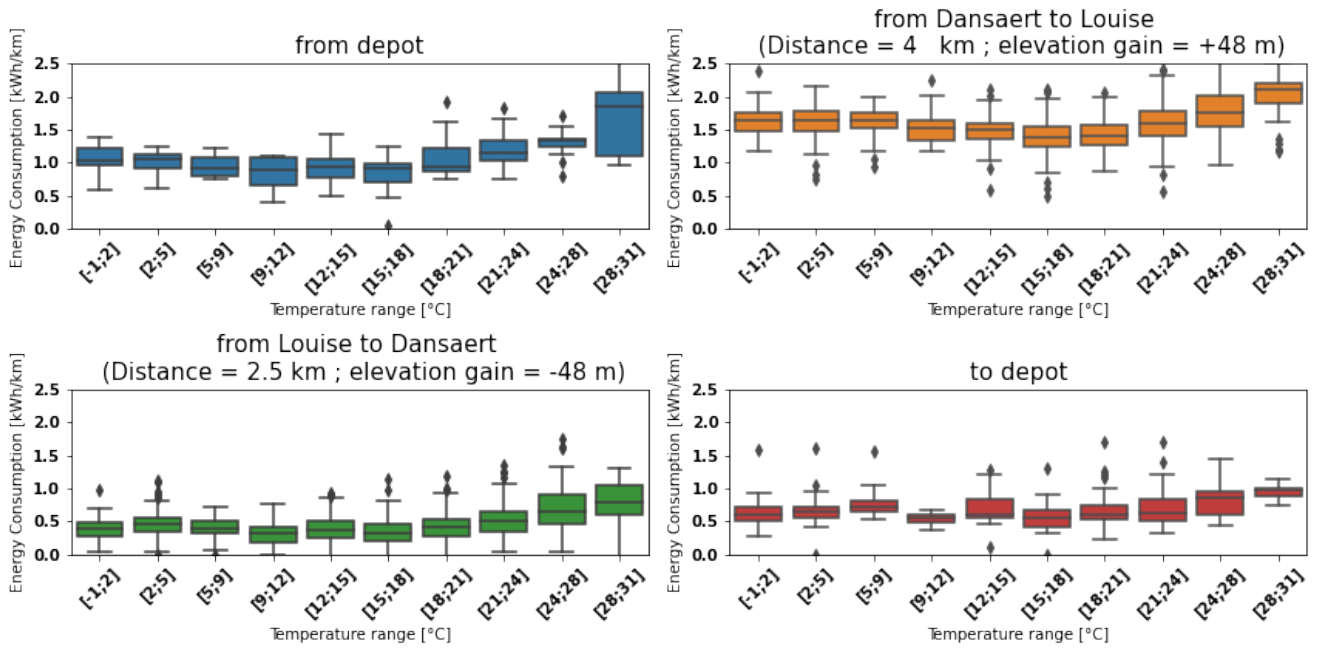


Figure 3.16: Boxplot of the EC by temperature range

3.6.2 Selection of trips with similar conditions except temperature

For this approach, data from trips with similar characteristics in terms of traffic conditions (mean and standard deviation of speed) and number of passengers were assembled in subdatasets. The strong assumption is that the only influencing parameter on the energy consumption in these subdatasets is temperature. Once again, a polynomial trend of energy consumption with respect to temperature can be guessed.

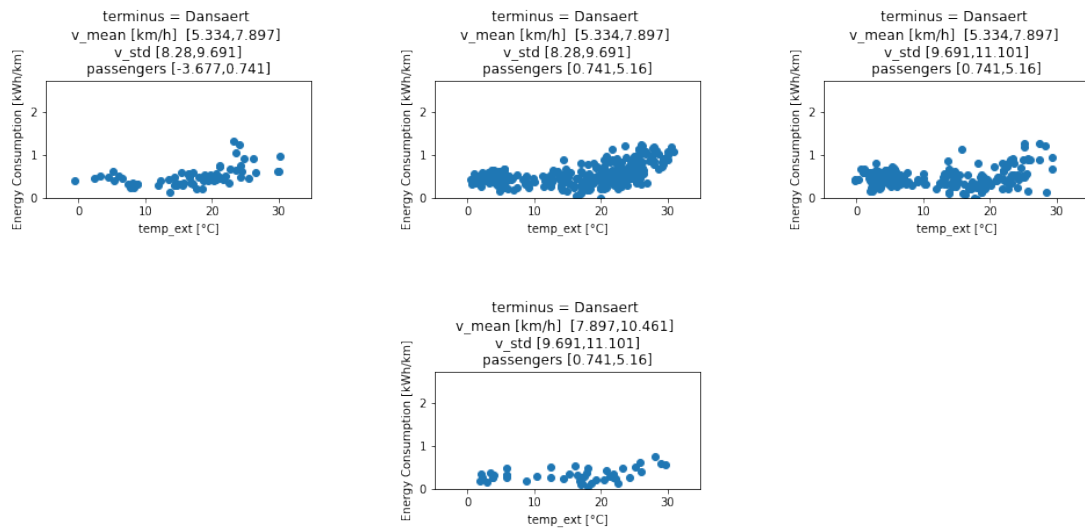


Figure 3.17: Trips with similar conditions

3.6.3 Headers of the dataset

Here are all the headers of the dataset which can be used as feature for the machine learning:

- start time
- weekday [0;6]
- travel time [s]
- vehicle
- driver

- line
- terminus
- v_mean [km/h]
- v_std
- temp_ext [°C]
- p_mean [kW]
- soc_max [%]
- soc_min [%]
- distance [km]
- Energy [kWh]
- Energy Consumption [kWh/km]
- elevation gain [m]
- elevation loss [m]
- passengers_mean
- passengers_max
- passengers_min

Chapter 4

Machine Learning

This chapter presents the theoretical foundations of Machine Learning by making the link with the data of the previous chapter, how the model is evaluated and the scaling of the data. Then, the used algorithms are presented. After having implemented these algorithms using the dataset of chapter 3, a comparison of their performances and results are presented.

As the subject of this work also aims to consolidate the regression model built with Scikit-learn last year by Frédéric Jadoul [15], here is a brief summary for ease of reading. His most relevant parameters for energy consumption are:

- The length of the bus: in [m]
- Temperature: the exterior temperature in [K]
- Passenger load: from 0 to 1
- Average speed: in [km/h]
- HVAC system: 0 or 1

The main results from his work are available in the appendix A.4 and it is recommended to read it before continuing. The best R^2 test set score is 0.61 and is obtained with a KNN regression algorithm and a MinMax scaler.

4.1 Machine learning basics

Machine Learning is 'a type of artificial intelligence in which computers use huge amounts of data to learn how to do tasks rather than being programmed to do them' according to Oxford Learner's Dictionaries [47]

In the context of this study, the task is to find the energy consumption of battery electric buses in $[kWh/km]$ for a trip with the data preprocessed previously.

In machine learning, the input data, (necessary to find the output value) are called the features and represented with an $\mathbf{X}$. The output value is called the target or prediction as shown in FIGURE 4.1 and represented with a $\mathbf{y}$. The box in the middle is a black-box that contains the rules governing the relation between the input and the output. It is called a regression algorithm because the target is a continuous value.

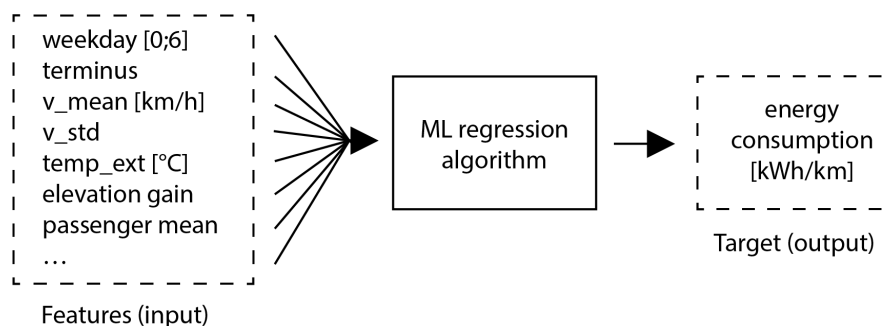


Figure 4.1: Diagram of use of a trained ML regression algorithm

For information purpose, the other type of algorithms in machine learning are the classification algorithms which classifies data into categories. Example: classifying fish in species (e.g 'salmon', 'sea bass', 'tuna',...) based on different features (e.g size, weight, color,..)

4.1.1 Evaluation of the model

A method to quantify the quality of a model is necessary. Different scoring functions are possible. The Mean Square Error is well known and expressed in (4.1). The error can also be called residuals because it is the remaining part when subtracting the predicted to the observed value.

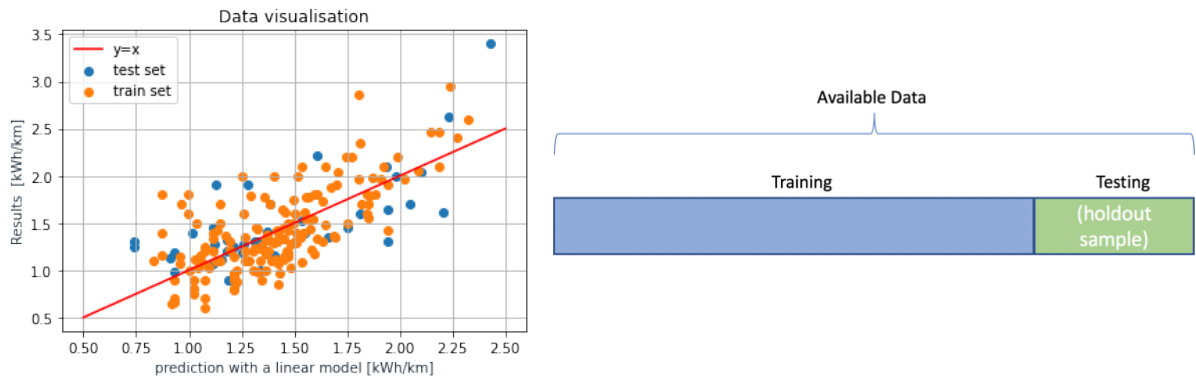
$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{Observation}_i - \text{Prediction}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.1)$$

Jadoul used the coefficient of determination, R^2 , to quantify the goodness of a fit.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{MSE_{mean} - MSE_{prediction}}{MSE_{mean}} \quad (4.2)$$

The R^2 score is easy to interpret depending on the 3 possible values

- $R^2 = 0$: the model does not fit the data better than taking the mean of all the observed values as prediction
- R^2 close to 1: the model is a good fit for the data because the $MSE_{prediction}$ is close to zero or much smaller than mean square error when taking the mean of all the observed values as prediction
- $R^2 < 1$: the model fits the data less well than the average of all observed values as a prediction



(a) Scatter plot of the predicted and real consumption [15] (b) Splitting of the dataset in a training set and testing set [48]

Figure 4.2: Evaluation of Machine Learning techniques

Another way to see the coefficient of determination, R^2 , is as the correlation between the predicted values and the actual values as they are plotted in FIGURE 4.2a.

Splitting the dataset

It is always important to hold a part of the dataset only for testing and not for training as shown in FIGURE 4.2b. This has to be done in order to detect overfitting. Overfitting is the production of a model that corresponds too closely to the training dataset and will have a bad scoring on the testing set. A case of overfitting is shown in FIGURE A.4. Splitting the whole dataset randomly in training and testing subdatasets is achieved with the `train_test_split` function of the `model_selection` class of **Scikit learn**. The first parameter is the `test_size` that sets the desired size of the test size, generally between 20 and 30 percent. The second parameter is `random_state` that controls the shuffling before splitting the data. If the value is the same, the result will be the same. This is useful in order to have reproducible training data to evaluate changes made in the regression algorithm [49].

4.1.2 Scaling of the data

Non-linear algorithms can be sensitive to the difference of scale of the different features. For instance, the feature `elevation gain [m]`, with values between 0 and 65 could have a greater impact on the prediction than the feature `weekday [0;6]` with values only between 0 and 6.

To solve this problem different scaling functions are tested in order to have a similar scale for all the features:

- MinMax scaler: $x_{MinMax} = \frac{X - X_{min}}{X_{max} - X_{min}}$
- Standard scaler: $x_{standar} = \frac{X - \mu}{\sigma}$
- Robust scaler: $x_{robust} = \frac{X - Q_1}{Q_3 - Q_1}$

In the case of this particular dataset, applying scaling function did not improve the performance even for non-linear functions. This is due to the removing of extreme outliers beforehand. More information about the scaling can be found in the work of Jadoul and in online documentation [50].

4.2 Presentation of the different algorithms

The approach is to train different models in order to compare them to gain more knowledge.

In this section, algorithms used by Jadoul are first briefly presented. Those are not very complex because of the small size of his dataset. This choice was made following the recommendation of prof. dr. Michel Verleysen (machine learning professor at UCLouvain) and dr. Mickaël Tits (data scientist at Laborelec).

Then new and more complex algorithms are explained. This year we had a bigger dataset which is a requirement in order to train properly complex algorithms.

4.2.1 Linear regression

The goal of this algorithm is to find the intercept a_0 and the coefficients for all the features $[a_1; a_{n_{feat}}]$ that fits the best the data. FIGURE 4.3 is a visual representation for two features.

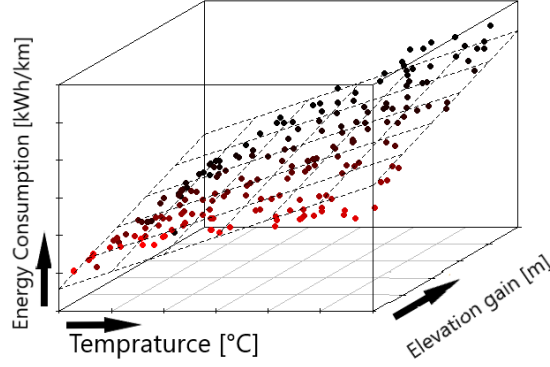


Figure 4.3: Example of a multiple linear regression with 2 features (adapted [51])

The expressions for the n predictions and errors are the following:

$$\begin{pmatrix} \hat{y}_1 \\ \dots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1n_{feat}} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nn_{feat}} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_{n_{feat}} \end{pmatrix} ; \quad \begin{pmatrix} e_1 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} - \begin{pmatrix} \hat{y}_1 \\ \dots \\ \hat{y}_n \end{pmatrix} \quad (4.3)$$

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{a} ; \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (4.4)$$

The intercept a_0 and the coefficients $[a_1; a_{n_{feat}}]$ are computed with the least square method that minimizes the loss function (4.5).

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\mathbf{y} - \mathbf{X}\mathbf{a}\|^2 \quad (4.5)$$

A variation of the linear regression is the ridge regression. The only difference is the loss function. The coefficient α is called the penalty. The goal is to avoid overfitting by penalizing the model with big coefficients, $\mathbf{a}$, that can fit well a training set but not a test set [52].

$$L(\mathbf{y}, \hat{\mathbf{y}}, \alpha) = \|\mathbf{y} - \mathbf{X}\mathbf{a}\|^2 + \alpha \|\mathbf{a}\|^2 \quad (4.6)$$

This algorithm is implemented in the class *linear_model* of **Scikit learn** [53].

4.2.2 Polynomial regression

In the previous chapter, we could observe a polynomial trend between the energy consumption and the temperature 3.6.1. It is therefore important to test a polynomial regression algorithm. The basis of this algorithm is the same than for the linear regression. The only difference is that $\mathbf{X}$ are the values of the features for the linear regression and $\mathbf{X}$ is a matrix with all the polynomial interactions

of the values of the features [54]. The chosen maximum degree is 2 because the polynomial trend seems like a square function.

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{11}x_{12} & x_{11}^2 & \dots & x_{1(n_{feat}-1)}x_{1n_{feat}} & x_{1n_{feat}}^2 \\ \dots & & & & & & & & \\ 1 & x_{n1} & x_{n1} & \dots & x_{n1}x_{n2} & x_{n1}^2 & \dots & x_{n(n_{feat}-1)}x_{nn_{feat}} & x_{nn_{feat}}^2 \end{pmatrix} \quad (4.7)$$

This function is implemented in the class *preprocessing* of **Scikit learn** [54].

4.2.3 K-Nearest Neighbors (KNN) regression algorithm

This algorithm projects all the training dataset in a multidimensional space. The number of dimensions corresponds to the number of features. For the example in FIGURE 4.4, the two features are respectively the percentage of passenger load on the x-axis and the temperature on the y-axis.

All the blue dots are training data and the red in the center is the data for which energy consumption is to be predicted. The prediction is the average of the energy consumption of the k nearest neighbors.

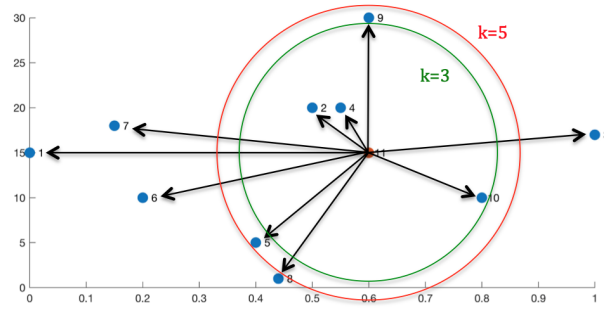


Figure 4.4: Example of KNN [15]

$$\hat{y}_{11} = \frac{y_4 + y_2 + y_{10}}{k} ; \text{ when } k = 3$$

There are two different ways to compute the distance:

- The euclidian distance: $d(a, b) = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$
- The Manhattan distance: $d(a, b) = |x_b - x_a| + |y_b - y_a|$

This algorithm is implemented in the class *neighbours* of **Scikit learn** [55].

4.2.4 Gradient Boosting regression algorithm

Gradient Boosting regression algorithm is based on regression trees that should first be explained.

Trees are generally used for classification algorithms but can also be used for regressions. The result of a regression tree is a piecewise defined function. Regression trees will be explained with FIGURE 4.5 that is a simplified example with only a few features. Trees are composed of different type of elements:

- Nodes: all the blocks with a condition inside
- Root: node at the top of the tree
- Branches: arrows that links the nodes together. Either true and goes to the left or false and goes to the right
- Leaves: final blocks without conditions. If a sample of the test dataset arrives in that leaf, the predicted value will be the average of all the training dataset in that leaf

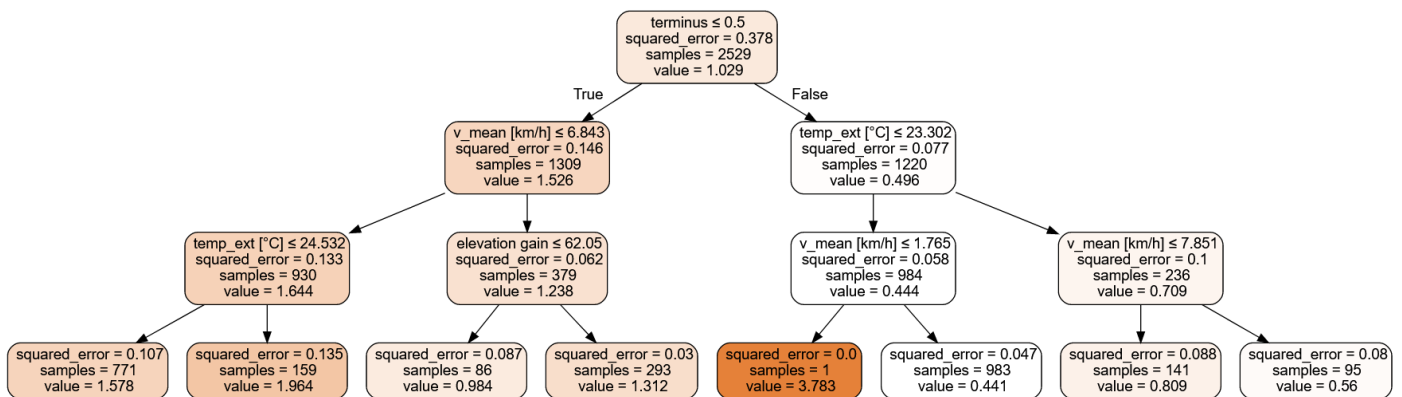


Figure 4.5: Simplified example of a regression tree

Here are the steps to create a tree:

1. Find the first condition for the root. It's the best condition to divide the data in two. In the example the condition is $\text{terminus} \leq 0.5$ and this separates the data between trip destination. The trips "from depot" and "Louise" go to the left of the tree and the trips "to depot" and "Dansaert" to the right of the tree. Trip destination has a significant impact on the energy consumption (cfr 3.3.5). It is therefore the best condition for the root.

2. Find the best conditions for the nodes of the left and right branches. The best condition minimizes the mean square error of the two newly created nodes.
3. Continue step 2 until the mean square error is below the threshold or the decided shape of the tree is obtained.

Generally, a tree regression is susceptible to overfitting the data and having an abrupt piecewise defined prediction function can be problematic. The solution to that is Gradient Boost. Gradient boost is an iterative learning process.

The first prediction is the average of all the targets. This will not be precise and there will be a big error or residual. A tree regression algorithm predicts these residuals. Then the predicted residuals, $\hat{\mathbf{e}}^{(i)}$, multiplied by a `learning_rate` (generally equal to 0.1) is added to the previous prediction in order to generate the new prediction. The training process stops after a given number of iterations or once a satisfactory score is achieved.

$$\hat{\mathbf{y}}^{(0)} = \bar{\mathbf{y}} \quad (4.8)$$

$$\mathbf{e}^{(i)} = \mathbf{y} - \hat{\mathbf{y}}^{(i)} \quad (4.9)$$

$$\hat{\mathbf{e}}^{(i)} = \text{tree_regressor}(\mathbf{X}, \mathbf{e}^{(i)}) \quad (4.10)$$

$$\hat{\mathbf{y}}^{(i+1)} = \hat{\mathbf{y}}^{(i)} + \text{learning_rate} \times \hat{\mathbf{e}}^{(i)} \quad (4.11)$$

According to Jerome Friedman, the inventor of this algorithm, empirical evidence shows that taking lots of small steps in the right direction results in better predictions with a testing dataset [56].

This algorithm is implemented in the class *tree* of **Scikit learn** [57].

4.2.5 Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) is a complex machine learning algorithm that can develop complex non-linear regression functions. In fact, it is so complex that it is impossible to fully explain it in this work. MLP is also called Neural Networks because historically people thought that the functioning was quite similar to the brain [58]. Vocabulary still remains, the nodes can be called neurons and the arrows between synapses.

The features enter the network in the input layer on the left of the schematic in FIGURE 4.6 and the information only goes from left to right. The internal computations of the neurons are represented at the bottom where the activation function can be a Rectified Linear Unit (ReLU)¹ or other function.

¹ReLU: $f(x) = x^+ = \max(0, x)$

The MLP is trained by backpropagation [59] in order to find the different values of weights, $\mathbf{w}$ and biases $\mathbf{b}$.

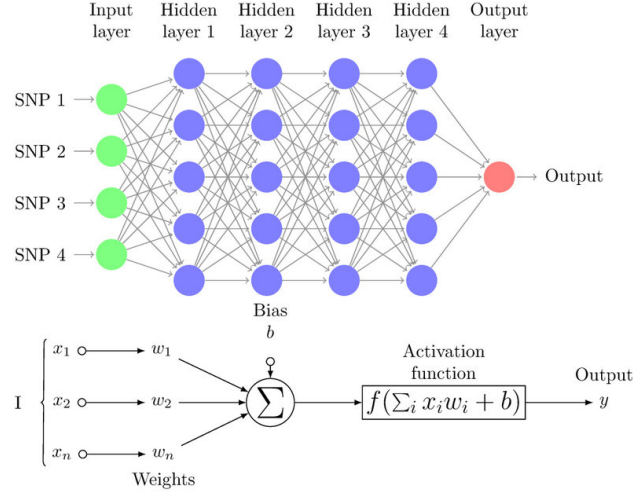


Figure 4.6: Example of a Multi Layer Perceptron [60]

To understand the basics of neural networks and other machine learning algorithm the following illustrated guide to machine learning is recommended [61]. A video series explain step by step the backpropagation computation [62]. This algorithm is implemented in the class *neural_network* of **Scikit learn** [63].

4.3 Comparison of the performance

The training and testing R^2 scores of the different regression algorithms are compared and summarized in TABLE 4.1. Different test sizes are compared. For the first column 30% of the dataset is put aside for testing. For the middle column, the test size is 20% and with the same random state as the first column in order to have the same base of training dataset and add 10% (cfr 4.1.1). For the last column, the training and testing set are obtained with a new shuffle. This is done to confirm the scores with different training datasets.

	test size = 30% split 1		test size = 20 % split 1		test size 20% split 2	
Type of regression	Train Score	Test Score	Train Score	Test Score	Train Score	Test Score
linear	0.767	0.8	0.794	0.773	0.792	0.772
linear temperature normalized	0.796	0.771	0.792	0.776	0.792	0.77
K-NN	1	0.868	1	0.87	1	0.857
Polynomial	0.877	0.865	0.879	0.855	0.879	0.853
Gradient Boost	0.929	0.842	0.927	0.842	0.929	0.858
Multi Layer Perceptron	0.87	0.867	0.874	0.854	0.875	0.846

Table 4.1: R^2 scores for training and testing for different algorithm and train&test splitting

Here are a few observations regarding the scores.

- The scores are similar between the different training datasets. This is a positive point and a validation of the robustness of the models
- Linear regression models are under-performing in comparison with the other models. The model is underfitting because it is not complex enough.
- The training score of the KNN is perfect. This is due to the design of the algorithm.
- The gradient boost descent has a significant difference between the training score and the test score. However, it performs well and has a high test score.

All the models are selected except the linear regressions.

4.4 Results from the model

Now that the models are trained and that their performance were assessed, the models can be plotted with more details. This will allow a better understanding of the algorithms that were used and to compare them beyond a simple score.

The aim of the prediction model was to analyze the influence of the temperature on the consumption. Rather than setting arbitrary values for the other features and plotting the predictions for the energy consumption depending on the temperature (as in FIGURE A.5) another approach is chosen. Each sample of the dataset will plot the energy consumption depending on a range of temperature with the other

features constant.

The function from the **Scikit learn** library that is used is called *PartialDependenceDisplay.from_estimator* [64]. The principal parameters are:

- *estimator*: this will be the regression model that we want to represent
- *X*: This a dataframe (n_samples,n_features) of the data we want to plot
- *features_names*: features that are selected for the plot

FIGURE 4.7 will be taken as an example to better understand the function. The estimator is the polynomial regression model, all the trips are taken to be plotted and the selected feature is *temp_ext [°C]*. As a reminder, TABLE 4.2 is an example of 1 line of the dataframe *X*.

weekday [0;6]	terminus	v_mean [km/h]	v_std	temp_ext [°C]
1	Louise	6,684834	9,308526	25,81659
elevation gain	elevation loss	passengers_mean	passengers_max	passengers_min
65,5	-17,5	7,516588	11	3

Table 4.2: Example of a sample of the dataframe (1 trip) for Partial dependence plot

Each blue line represents a different sample and one line is drawn as follows.

Multiple energy consumption predictions are computed with the polynomial regression model based on the input same input features of a sample. The only difference between all these predictions is that the value of the *temp_ext [°C]* feature changes. All the temperature range from 2 [°C] to 27 [°C] is covered in order to draw one blue line.

The process is repeated for the remaining samples and all the blue lines are displayed. The red line is the average between all the blue lines.

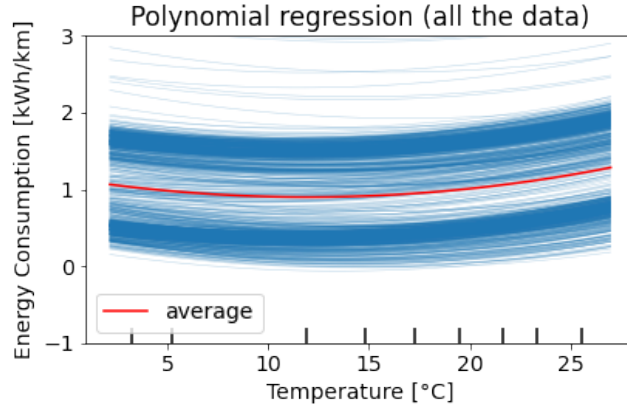


Figure 4.7: Example of Partial Dependence Display of temperature

All the blue lines have a similar shape but there are two more dense zones. In FIGURE 4.8, different data are entered as parameter. For 4.8a parameter X contains all the trips, 4.8b parameter X contains all the trips trips to 'Louise' and 4.8c all the trips to 'Dansaert'. It can be observed that the average energy consumption curve is higher for the trips to 'Louise' than for the trips to 'Dansaert'. This is the expected result because those trips are respectively uphill and downhill.

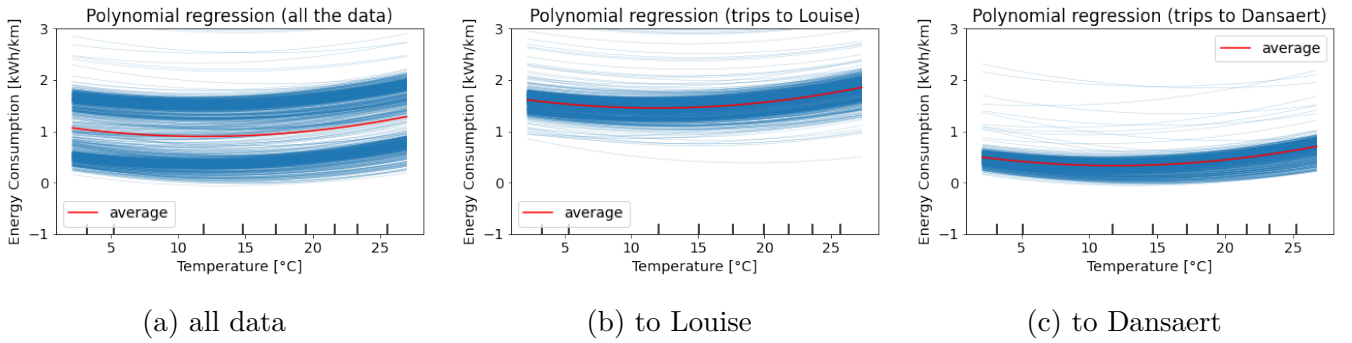


Figure 4.8: Partial Dependence Display Polynomial regression

Other features of interest can also be represented as is shown in FIGURE 4.9. It is a good way to identify errors with specific features. In this case, the `passengers_mean` is not a good feature for the model. This can be explained by the fact that the measurements are not always reliable as explained in subsection 3.2.2. Another observation is that the standard deviation, `v_std`, and the day of the week, `weekday` [0;6] does not have an impact on the energy consumption according to the model. This can be concluded because the partial dependence displays are nearly flat for

these features.

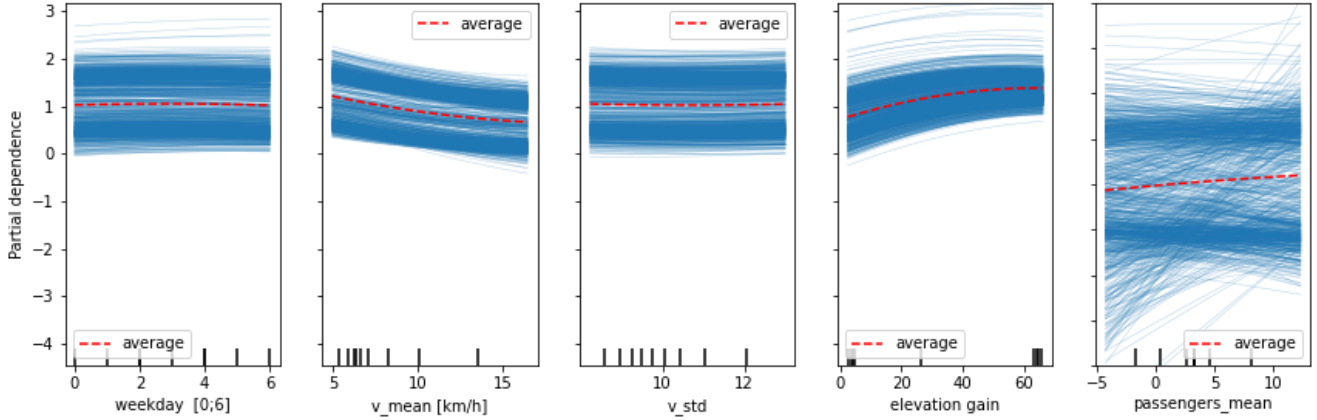
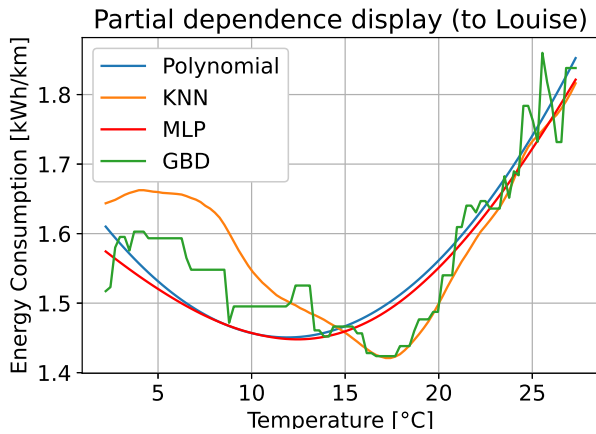
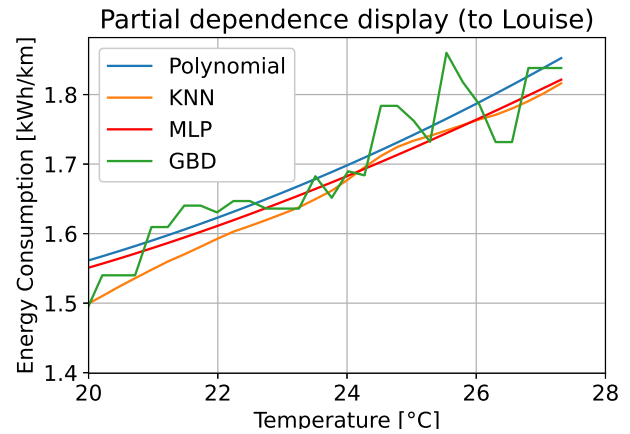


Figure 4.9: Partial Dependence Display Polynomial of different features (all the data)



(a) all temperature range



(b) zoom on high temperatures

Figure 4.10: Partial Dependence Display of 4 algorithms with the trips to Louise

The influence of the temperature on the energy consumption according to the four algorithms is displayed in FIGURE 4.10a and 4.11 for the trips to 'Louise' and 'Dansaert' respectively.

It is a satisfying result because the lines have the same general shape even though they were obtained with different algorithms. Two sub-groups can be identified. Firstly, the polynomial and multi layer perceptron that have a minimum energy

consumption around 12 °. Secondly, the KNN and the Gradient Boosting Descent that have a minimum energy consumption around 17.5 °. These values match with the HVAC function depending of the external temperature seen in the theoretical approach (TABLE 2.1). The KNN algorithm is the smoothest and is considered as the best model.

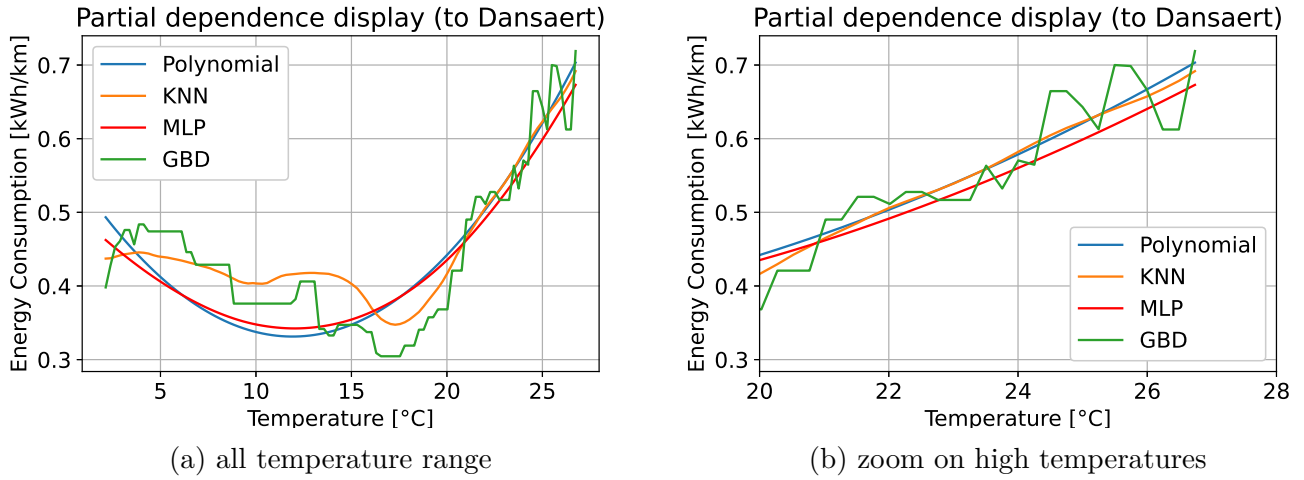


Figure 4.11: Partial Dependence Display of 4 algorithms with the trips to Dansaert

4.5 Conclusion and best model

The best model is the K-Nearest Neighbour regression model with the following parameters:

- Formula to compute the distance: Euclidian
- Number of neighbors taken into account: 16
- Weights: Distance, this means that the neighbors close to the the prediction node has a greater weight

Here are the only few features that are retained because they have a great impact on the prediction:

- **v_mean [km/h]**: when the mean speed is higher, the ride is more smooth and the energy consumption decreases
- **temperature_ext [°C]**: has an important impact on the consumption due to the HVAC
- **elevation gain** and **elevation losses**: are the most relevant features because it reflects the difficulty of the terrain. **terminus** was helpfull in order to divide the problem but is not necessary to make a good prediction.

The test R^2 score is 0.859. A more intuitive score is the mean absolute error: $mean(\sum_i |Prediction_i - Target_i|)$ of the energy consumption. The MAE for the whole testing set, only the trips to Louise and only the trips to Dansaert are respectively 0.155, 0.164 and 0.148 kWh/km . This is a satisfying result considering that the medians are respectively 1.04, 1.56 and 0.44.

Chapter 5

Application in different scenarios

In this chapter, three shifts for three different temperature conditions (cold, normal and hot) are selected. Observation are made regarding the data. Then, the energy demand is computed with the regression algorithm. Finally, possible improvements are discussed.

5.1 Shifts of three different temperature conditions

Three shifts of different temperature conditions, cold, normal and hot are selected. TABLE 5.1 and FIGURE 5.1 displays the external temperature measured by the bus.

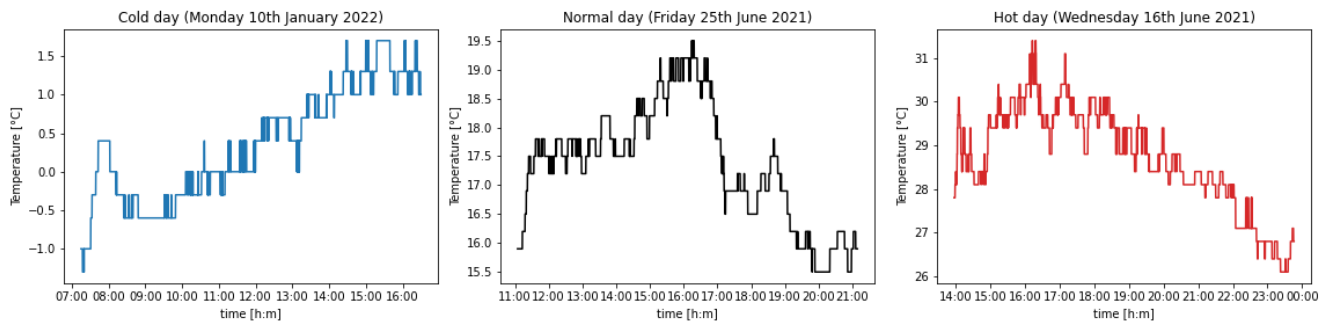


Figure 5.1: Temperature recordings for cold, normal and hot conditions

	cold shift	normal shift	hot shift
start time	2022-01-10 07:14:59	2021-06-25 11:04:38	2021-06-16 13:57:18
end time	2022-01-10 16:29:46	2021-06-25 21:08:58	2021-06-16 23:45:47
vehicle	1001	1001	1006
v_mean [km/h]	6,23	6,71	7,76
temp_ext mean[°C]	0,36	17,37	28,73
temp_ext min[°C]	-1,3	15,5	26,1
temp_ext max[°C]	1,7	19,5	31,4
P_in mean [kW]	6,74	7,67	6,42
P_batt mean [kW]	7,32	7,88	9,93
soc_max [%]	99	100	100
soc_min [%]	44	53	42
distance [km]	59,69	70,11	78,725
Energy [kWh]	67,66	79,41	97,43
elevation gain [m]	836	1005	1023
elevation loss [m]	888	1078	1081
Energy Consumption [kWh/km]	1,134	1,133	1,238

Table 5.1: Data of three shift for different temperature conditions

The first observation is that the available energy from the battery computed in (5.1) decreases in cold conditions. This is expected from the literature (cfr 2.2.2). The available energies in batteries when fully charged in cold, normal and hot conditions are respectively 121, 170 and 167 *kWh*. According to the datasheet the battery is a 160 *kWh* model. The available energy is slightly overestimated for the normal and hot conditions and decreases 24% for cold conditions

$$\text{Available energy @100\%SOC} = \frac{\text{Energy}_{shift}}{SOC_{max} - SOC_{min}} * 100\% \quad (5.1)$$

Second observation, 'P_in mean' is the mean power that goes to the traction of the bus and 'P_batt mean' is the mean power provided by the battery. The percentage of the battery power that goes to the auxiliary (mainly HVAC) is computed for the different temperature conditions. In normal conditions it is only 2.66% for supplying the electronics and the ventilation. When the heating is on, the ratio is 7.92%. For extreme heats, the air conditioning consumes 35.35% of the power coming out of the batteries.

$$\% \text{ to auxiliary} = 1 - \frac{P_{in,mean}}{P_{batt,mean}} \quad (5.2)$$

5.2 Validation of the energy demand

The best KNN model was not trained with the data of the three selected trips. This guarantees that a good result is not obtained with overfitting. The total energy demand in kWh for the shift is computed by adding the energy demand of each trip. Energy demand are obtained by multiplying the energy consumption resulting from the regression algorithm by the distance of the trip.

$$EC_i = \text{regression_algorithm}(X_i) \quad [kWh/km] \quad (5.3)$$

$$ED_{tot} = EC_{fromdepot} * d_{fromdepot} \quad (5.4)$$

$$+ \sum_{i=1}^{trips_Dansaert} EC_{i,Dansaert} * d_{Dansaert}$$

$$+ \sum_{i=1}^{trips_Louise} EC_{i,Louise} * d_{Louise}$$

$$+ EC_{todepot} * d_{todepot} \quad [kWh]$$

The prediction of the energy demand for the shift and the errors interpreted in different ways are in TABLE 5.2. The performance is good and the errors are within the safety margins.

	cold shift	normal shift	hot shift
Target [kWh]	67,66	79,41	97,4
Prediction [kWh]	67,53	70,84	107,96
Difference [kWh]	0,13	8,57	-10,56
Difference [% of target]	0,2	10,8	-10,8
Average Energy consumption [kWh/km]	1,15	0,91	1,39
Estimated range remaining at the end of the shift [km]	-0,11	-9,42	7,6

Table 5.2: Prediction of energy demand for a shift

5.3 Conclusion and improvement of the model

The model implementing the K-Nearest Neighbour regression algorithm performs well. High R^2 scores of around 0.85 were obtained and the model predicts energy demand for Battery Electric Buses within + or - % 10. Here are some areas where the model can be developed further:

- Create a dataset based on the weather forecast to obtain the outdoor temperature and the traffic conditions or the opendata platform [45] to obtain the average speed
- Further automate the predictions and learning process in order to suggest charging strategies
- Correct the problems with the number a passengers to increase the performance of the model
- Enrich the model with other routes and BEB's. Adding other bus types with new features like **length**, **mass** or HVAC **system** will allow to make predictions for other types of BEB's that are not in the database
- Make suggestion for the electrification of other bus lines by computing the energy demand for this line based on the data gather from the internal combustion engine buses

Chapter 6

Conclusion

The main objective of this work is to validate an energy consumption prediction model for electric buses. This model had been built with a limited set of data.

One of the first steps of the work was therefore to collect as much data as possible in order to train the model. The raw data of an electric bus operated line of the STIB was obtained.

A lot of pre-processing work was done to understand the data, enrich it, make links between them and classify them. The Python code used is available at [Laborelec](#). This made it possible to obtain a large dataset with parameters that seemed significant to estimate the power consumption. The classification by trip as time reference was validated by micro analysis of the data, both at the level of a whole day and at the level of a specific trip to a terminus. The quality of the data could be demonstrated in this way as well, although there were some doubt regarding the reliability of the number of passengers.

After eliminating the few outliers, a set of 3600 trips was available with all the useful data such as the external temperature, average speed and energy consumption.

I then adapted the model in Scikit learn. Indeed, compared to my predecessor, the type of feature available is not the same. Compared to last year, the dataset obtained from the STIB contains in addition: the gain and loss of altitude, much more precise outdoor temperature data, and the type of journey. However, the data only relates to the same type of bus, so there is no way to take into account the length of the bus.

I was then able to use the dataset with the different machine learning algorithms. First to train it with part of the data, then to test it.

As the dataset was much larger than the previous year, 2 new and more complex algorithms could be added.

A comparison of the performances shows that only the linear regression should be discarded because of its low score. The 4 other algorithms, although using very different functions, show very similar trends in the impact of temperature on energy consumption, which confirms the quality of the models. A preference is given to the KNN model which best reflects reality.

The best model is then applied in different temperature scenarios.

3 shifts are selected (cold - normal - hot) to observe the impact of cold on the battery performance and the increase in energy consumption in hot weather. The total energy spent during the 3 shifts was simulated by the ML model and then compared to the samples. The results were very similar which again corroborates the relevance of the model.

Finally, the possibilities of improving the model are discussed so that it can be used to predict energy consumption in other situations.

Appendix A

Appendix

A.1 Heating output requirements as a factor of the outside temperature

$$\dot{Q}_{tot} = \dot{Q}_{passengers} + \dot{Q}_{solar} + \dot{Q}_{drivetrain_losses} + \dot{Q}_{conduction\&convection} + \dot{Q}_{doors} + \dot{Q}_{HVAC} \quad (A.1)$$

$$\dot{Q}_{passengers} = \sum_{passengers} M \cdot A_{body} \quad (A.2)$$

$$\dot{Q}_{solar} = \dot{Q}_{solar,window} + \dot{Q}_{solar,wall} \quad (A.3)$$

$$\dot{Q}_{solar,window} = \frac{1}{2} \cdot SHGC \cdot I_S \cdot A_{window} \quad (A.4)$$

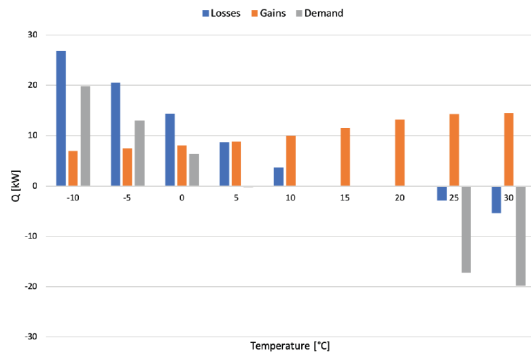
$$\dot{Q}_{solar,wall} = a \cdot I_S \cdot \frac{U_{wall}}{h_o} \cdot A_{wall} \quad (A.5)$$

$$\dot{Q}_{drivetrain_losses} = r_{drivetrain \rightarrow cabin} \cdot P_{tr} \cdot (1 - \eta_{traction}) \text{when in traction} \quad (A.6)$$

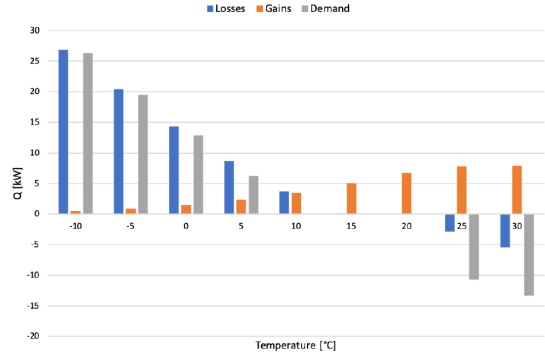
$$\dot{Q}_{drivetrain_losses} = r_{drivetrain \rightarrow cabin} \cdot P_{tr} \cdot \frac{1 - \eta_{regen}}{\eta_{regen}} \text{when Regenerative braking} \quad (A.7)$$

$$\dot{Q}_{conduction\&convection} = U \cdot A \cdot (T_i - T_{out}); U = \frac{1}{R_{th}} \Rightarrow R = \frac{1}{h_o} + \frac{t}{k} + \frac{1}{h_i} \quad (A.8)$$

$$\dot{Q}_{doors} = 0.04 W H^{3/2} \sqrt{\Delta T} \cdot (h_{in} - h_{out}) \quad (A.9)$$



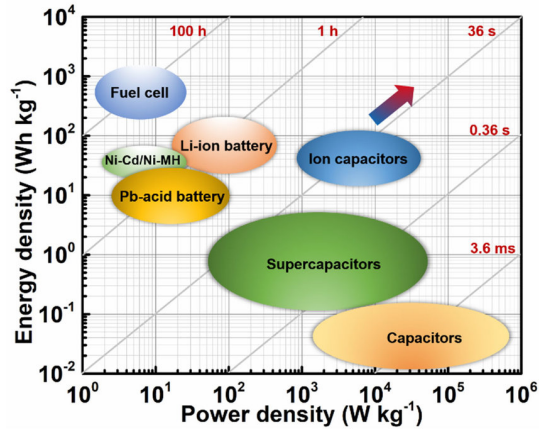
(a) 100% passengers



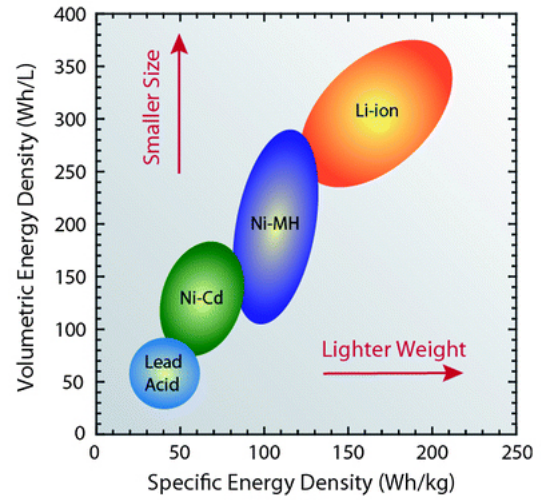
(b) 0% passengers

Figure A.1: Heating output requirements as a factor of the outside temperature [21]

A.2 Battery performance



(a) Specific Energy and power density [65]



(b) Volumetric and Specific Energy density [25]

Figure A.2: Performance of different types of batteries and storage technology

A.3 OpenData STIB

line	direction	id	order	name_fr	name_nl	latitude	longitude
33	City	2177	1	LOUISE	LOUIZA	50.83.615	4.355.323
33	City	3629	2	POELAERT	POELAERT	50.838.003	4.353.594
33	City	6353	3	PETIT SABLON	KLEINE ZAVEL	50.839.354	4.355.459
33	City	3630	4	GRAND SABLON	GROTE ZAVEL	50.842.189	4.354.148
33	City	1781	5	GRAND-PLACE	GROTE MARKT	50.845.003	4.351.728
33	City	2392	6	BOURSE	BEURS	50.846.598	4.348.295
33	City	1707	7	SAINT-GERY	SINT-GORIK	50.84.775	4.345.878
33	City	33208	8	BOURSE	BEURS	50.849.372	4.347.381
33	City	3221	9	DANSAERT	DANSAERT	50.851.241	4.344.952
33	Suburb	3633C	1	DANSAERT	DANSAERT	50.851.142	434.437
33	Suburb	2212	2	BOURSE	BEURS	50.848.383	43.467
33	Suburb	2578	3	ANNEESSENS	ANNEESSENS	50.845.686	4.346.176
33	Suburb	1063	4	CESAR DE PAEPE	CESAR DE PAEPE	50.842.882	4.350.791
33	Suburb	1127	5	CHAPELLE	KAPELLEKERK	50.841.246	4.351.814
33	Suburb	13768	6	GRAND SABLON	GROTE ZAVEL	50.841.413	4.353.383
33	Suburb	1128	7	PETIT SABLON	KLEINE ZAVEL	50.839.898	4.355.875
33	Suburb	6354B	8	ROYALE	KONING	50.841.545	4.358.548
33	Suburb	3505	9	PORTE DE NAMUR	NAAMSEPOORT	50.838.945	4.361.604
33	Suburb	2177	10	LOUISE	LOUIZA	50.83.615	4.355.323

Figure A.3: ID stop number of bus line 33 in the order of the route

A.4 Overfitting

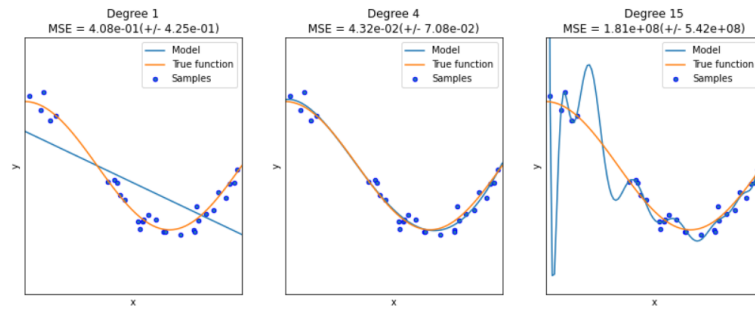


Figure A.4: Underfitting and Overfitting in Machine Learning [66]

A.5 Main results from previous work

	Ridge regression with normalized temperature using a Standard scaler	KNN regression using a MinMax scaler	Polynomial regression using a Robust scaler and a Ridge regression algorithm to resolve the matrix coefficients
α	2.5	/	1.3
Number of neighbors	/	2	/
Metric	/	manhattan	/
Weights	/	uniform	/
Train set score	0.57	0.86	0.68
Cross-validation score (k=5)	0.43	0.61	0.53
Test set score	0.51	0.61	0.52

Table A.1: best combination of algorithms and best score for each algorithm type

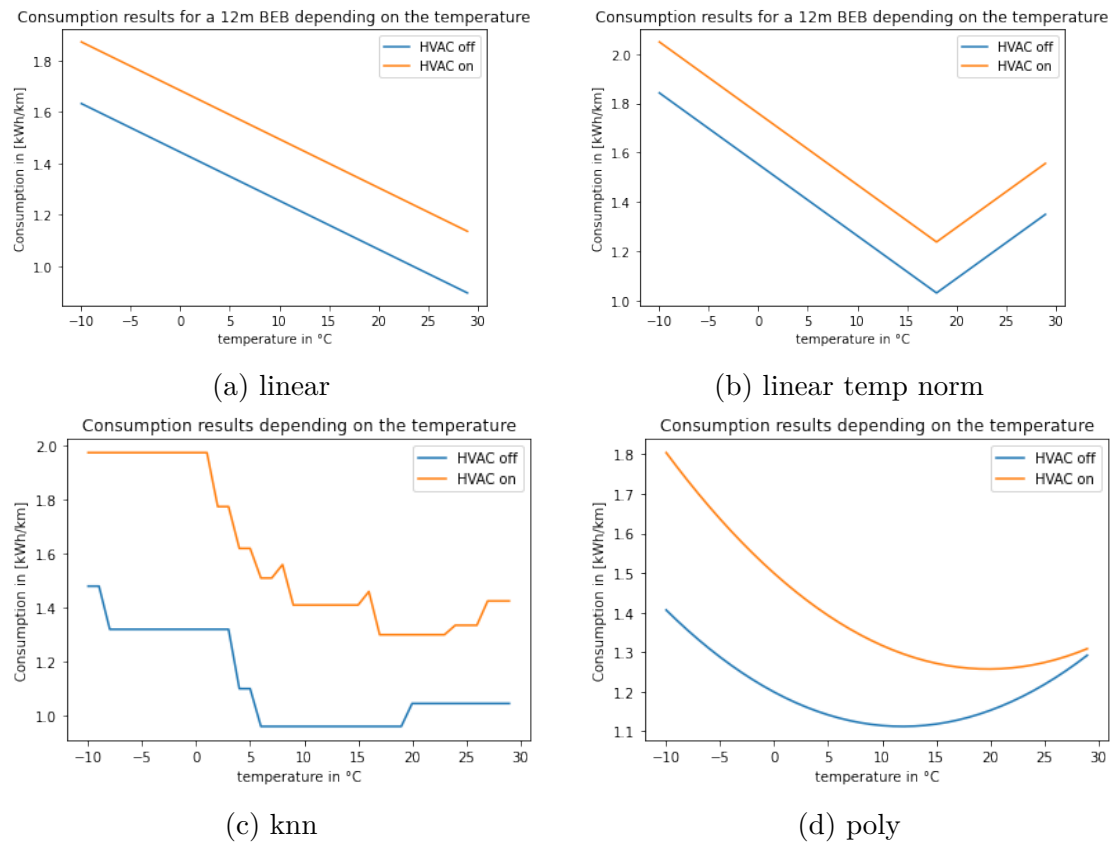


Figure A.5: Plot of the consumption depending on the exterior temperature for different algorithms

Bibliography

- [1] *Causes of Climate Change*. URL: https://ec.europa.eu/clima/climate-change/causes-climate-change_en (visited on 08/17/2022).
- [2] Valérie Masson-Delmotte et al., eds. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021. DOI: 10.1017/9781009157896.
- [3] *Delivering the European Green Deal | European Commission*. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en (visited on 08/19/2022).
- [4] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Sustainable and Smart Mobility Strategy – Putting European Transport on Track for the Future*. 2020. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789> (visited on 08/19/2022).
- [5] *Directive 2009/33/EC of the European Parliament and of the Council of 23 April 2009 on the Promotion of Clean Road Transport Vehicles in Support of Low-Emission Mobility (Text with EEA Relevance)Text with EEA Relevance*. Aug. 1, 2019. URL: <http://data.europa.eu/eli/dir/2009/33/2019-08-01/eng> (visited on 08/19/2022).
- [6] *Studies*. URL: https://transport.ec.europa.eu/transport-themes/clean-transport-urban-transport/studies_en (visited on 08/19/2022).
- [7] *Efficient and Green Mobility*. URL: https://transport.ec.europa.eu/news/efficient-and-green-mobility-2021-12-14_en (visited on 08/19/2022).
- [8] *Market Monitoring*. Clean Bus Europe Platform. URL: <https://cleanbusplatform.eu/toolkit/market-monitoring> (visited on 08/19/2022).

- [9] *Climate Change 2022: Mitigation of Climate Change*. URL: <https://www.ipcc.ch/report/ar6/wg3/> (visited on 08/19/2022).
- [10] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Sustainable and Smart Mobility Strategy – Putting European Transport on Track for the Future*. 2020. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789> (visited on 08/19/2022).
- [11] *Home – ASSURED Project*. URL: <https://assured-project.eu/> (visited on 08/19/2022).
- [12] *De Lijn Ready to Tender Work on Three Electric Bus Depots*. Sustainable Bus. May 16, 2022. URL: <https://www.sustainable-bus.com/electric-bus/de-lijn-ready-to-tender-work-on-three-electric-bus-depots/> (visited on 08/19/2022).
- [13] Youssef Oualmakran et al. “Mitigating the Impact of High Power Charging of Electric Buses: Perspective of European Distribution Grid Operators”. In: *EVS33 International Electric Vehicle Symposium 133nd Electric Vehicle Symposium (EVS33)* (Sept. 2020), pp. 1–10.
- [14] *Company Profile*. ENGIE Laborelec. URL: <https://www.laborelec.com/company-profile/> (visited on 08/19/2022).
- [15] Frédéric Jadoul. “Modelling of Energy Consumption of Electric Buses, Sizing of the Battery for Selected Routes, and Impact of Their Recharge on the Grid”. 2021.
- [16] Todd M. Bandhauer, Srinivas Garimella, and Thomas F. Fuller. “A Critical Review of Thermal Issues in Lithium-Ion Batteries”. In: *Journal of The Electrochemical Society* 158.3 (Jan. 25, 2011), R1. ISSN: 1945-7111. DOI: 10.1149/1.3515880. URL: <https://iopscience.iop.org/article/10.1149/1.3515880/meta> (visited on 08/15/2022).
- [17] Hugh D. Young and Roger A. Freedman. *University Physics*. 14th edition. Harlow: Pearson, Jan. 15, 2015. 1328 pp. ISBN: 978-0-13-396929-0.
- [18] Marc Gallet, Tobias Massier, and Thomas Hamacher. “Estimation of the Energy Demand of Electric Buses Based on Real-World Data for Large-Scale Public Transport Networks”. In: *Applied Energy* 230 (Nov. 15, 2018), pp. 344–356. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2018.08.086. URL: <https://www.sciencedirect.com/science/article/pii/S030626191831256X> (visited on 08/09/2022).

- [19] Jerzy Ejsmont et al. “Influence of Temperature on the Tyre Rolling Resistance”. In: *International Journal of Automotive Technology* 19.1 (Feb. 1, 2018), pp. 45–54. ISSN: 1976-3832. DOI: 10.1007/s12239-018-0005-4. URL: <https://doi.org/10.1007/s12239-018-0005-4> (visited on 08/15/2022).
- [20] K. Charles. “Fanger’s Thermal Comfort and Draught Models”. In: (Jan. 1, 2003). DOI: 10.4224/20378865.
- [21] Jérôme Goffart. “Study of Heating Solutions for Electric Buses”. 2019.
- [22] Jari Vepsäläinen et al. “Computationally Efficient Model for Energy Demand Prediction of Electric City Bus in Varying Operating Conditions”. In: *Energy* 169 (Feb. 15, 2019), pp. 433–443. ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.12.064. URL: <https://www.sciencedirect.com/science/article/pii/S0360544218324307> (visited on 08/13/2022).
- [23] *Bilan climatique saisonnier - hiver 2022*. Institut Royal Météorologique de Belgique, Mar. 7, 2022, p. 9. URL: <https://www.meteo.be/fr/climat/climat-de-la-belgique/bilans-climatologiques/2022/hiver> (visited on 08/13/2022).
- [24] *Bilan climatique saisonnier - été 2021*. Institut Royal Météorologique de Belgique, Jan. 31, 2022. URL: <https://www.meteo.be/fr/climat/climat-de-la-belgique/bilans-climatologiques/2021/ete> (visited on 08/13/2022).
- [25] *Battery Comparison of Energy Density - Cylindrical and Prismatic Cells*. URL: <https://www.epectec.com/batteries/cell-comparison.html> (visited on 08/14/2022).
- [26] *Comparison of Commercial Battery Types*. In: *Wikipedia*. Aug. 14, 2022. URL: https://en.wikipedia.org/w/index.php?title=Comparison_of_commercial_battery_types&oldid=1104343151 (visited on 08/14/2022).
- [27] Biao Li and Dingguo Xia. “Anionic Redox in Rechargeable Lithium Batteries”. In: *Advanced Materials* 29.48 (2017), p. 1701054. ISSN: 1521-4095. DOI: 10.1002/adma.201701054. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201701054> (visited on 08/14/2022).
- [28] Asma Mohamad Aris and Bahman Shabani. “An Experimental Study of a Lithium Ion Cell Operation at Low Temperature Conditions”. In: *Energy Procedia*. 1st International Conference on Energy and Power, ICEP2016, 14-16 December 2016, RMIT University, Melbourne, Australia 110 (Mar. 1, 2017), pp. 128–135. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.03.117. URL: <https://www.sciencedirect.com/science/article/pii/S1876610217301479> (visited on 08/14/2022).

- [29] Yan Ji, Yancheng Zhang, and Chao-Yang Wang. “Li-Ion Cell Operation at Low Temperatures”. In: *Journal of The Electrochemical Society* 160.4 (Feb. 13, 2013), A636. ISSN: 1945-7111. DOI: 10.1149/2.047304jes. URL: <https://iopscience.iop.org/article/10.1149/2.047304jes/meta> (visited on 08/15/2022).
- [30] Miltiadis Papalexandris. *Combustion and Fuels*. Presses Universitaires de Louvain - UCL, Aug. 29, 2020. 188 pp. ISBN: 978-2-87558-979-8.
- [31] Fei Gao and Zhiyuan Tang. “Kinetic Behavior of LiFePO₄/C Cathode Material for Lithium-Ion Batteries”. In: *Electrochimica Acta* 53.15 (June 1, 2008), pp. 5071–5075. ISSN: 0013-4686. DOI: 10.1016/j.electacta.2007.10.069. URL: <https://www.sciencedirect.com/science/article/pii/S0013468607013576> (visited on 08/15/2022).
- [32] Shuai Ma et al. “Temperature Effect and Thermal Impact in Lithium-Ion Batteries: A Review”. In: *Progress in Natural Science: Materials International* 28.6 (Dec. 1, 2018), pp. 653–666. ISSN: 1002-0071. DOI: 10.1016/j.pnsc.2018.11.002. URL: <https://www.sciencedirect.com/science/article/pii/S1002007118307536> (visited on 08/15/2022).
- [33] Heike Gabrisch, Yasunori Ozawa, and Rachid Yazami. “Crystal Structure Studies of Thermally Aged LiCoO₂ and LiMn₂O₄ Cathodes”. In: *Electrochimica Acta*. ELECTROCHEMICAL ENERGY CONVERSION AND STORAGE - ISE MEETING 2005 52.4 (Dec. 1, 2006), pp. 1499–1506. ISSN: 0013-4686. DOI: 10.1016/j.electacta.2006.02.050. URL: <https://www.sciencedirect.com/science/article/pii/S0013468606004130> (visited on 08/15/2022).
- [34] Patricia Handel et al. “Thermal Aging of Electrolytes Used in Lithium-Ion Batteries – An Investigation of the Impact of Protic Impurities and Different Housing Materials”. In: *Journal of Power Sources* 267 (Dec. 1, 2014), pp. 255–259. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2014.05.080. URL: <https://www.sciencedirect.com/science/article/pii/S0378775314007721> (visited on 08/15/2022).
- [35] Lucille Bodenes et al. “Lithium Secondary Batteries Working at Very High Temperature: Capacity Fade and Understanding of Aging Mechanisms”. In: *Journal of Power Sources* 236 (Aug. 15, 2013), pp. 265–275. ISSN: 0378-7753. DOI: 10.1016/j.jpowsour.2013.02.067. URL: <https://www.sciencedirect.com/science/article/pii/S0378775313003522> (visited on 08/15/2022).

- [36] Feng Leng, Cher Ming Tan, and Michael Pecht. “Effect of Temperature on the Aging Rate of Li Ion Battery Operating above Room Temperature”. In: *Scientific Reports* 5.1 (1 Aug. 6, 2015), p. 12967. ISSN: 2045-2322. DOI: 10.1038/srep12967. URL: <https://www.nature.com/articles/srep12967> (visited on 08/15/2022).
- [37] Jaewan Kim, Jinwoo Oh, and Hoseong Lee. “Review on Battery Thermal Management System for Electric Vehicles”. In: *Applied Thermal Engineering* 149 (Feb. 25, 2019), pp. 192–212. ISSN: 1359-4311. DOI: 10.1016/j.applthermaleng.2018.12.020. URL: <https://www.sciencedirect.com/science/article/pii/S135943111835614X> (visited on 08/15/2022).
- [38] *ViriCiti Releases First-Of-Its-Kind Report on E-Bus Performance*. ViriCiti. URL: <https://viriciti.com/press-releases/viriciti-electric-bus-performance-consumption-report/> (visited on 08/10/2022).
- [39] Mark Henning, Alison Smyth, and Andrew Thomas. *An Analysis of the Association between Changes in Ambient Temperature, Fuel Economy, and Vehicle Range for Battery Electric and Fuel Cell Electric Buses*. Nov. 15, 2019.
- [40] Trougnouf. *English: Solaris Electric Bus 1005 Operated by STIB on Line 33 in Brussels*. 22 June 2018, 15:08:35. URL: [https://commons.wikimedia.org/wiki/File:Solaris_electric_bus_operated_by_STIB_on_line_33_in_Brussels_\(DSCF6911\).jpg](https://commons.wikimedia.org/wiki/File:Solaris_electric_bus_operated_by_STIB_on_line_33_in_Brussels_(DSCF6911).jpg) (visited on 07/25/2022).
- [41] STIB/MIVB. *Une flotte moderne au service des clients*. 2018. URL: http://www.stib-mivb.be/irj/go/km/docs/WEBSITE_RES/Attachments/%20Vehicules/18-2884-Brochure_NouveauxVehicules_2018_FR_LR.pdf (visited on 10/20/2021).
- [42] STIB/MIVB. *Pourquoi Compter Nos Voyageurs Améliore Le Confort de Nos Lignes ?* June 2, 2022. URL: <https://stibstories.be/2022/06/02/pourquoi-compter-nos-voyageurs-ameliore-le-confort-de-nos-lignes/> (visited on 07/25/2022).
- [43] Adam Schneider. *GPS Visualizer*. URL: <https://www.gpsvisualizer.com/> (visited on 07/31/2022).
- [44] Shivani Gupta and Atul Gupta. “Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review”. In: *Procedia Computer Science* 161 (2019), pp. 466–474. ISSN: 18770509. DOI: 10.1016/j.procs.2019.11.146. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1877050919318575> (visited on 08/01/2022).
- [45] *GET INVOLVED AND HELP US KEEP BRUSSELS MOVING*. URL: <https://opendata.stib-mivb.be?lang=en> (visited on 08/02/2022).

- [46] M. Islameka, E. Leksono, and B. Yulianto. “Modelling of Regenerative Braking System for Electric Bus”. In: *Journal of Physics: Conference Series* 1402.4 (Dec. 2019), p. 044054. ISSN: 1742-6596. DOI: 10.1088/1742-6596/1402/4/044054. URL: <https://doi.org/10.1088/1742-6596/1402/4/044054> (visited on 08/04/2022).
- [47] *Machine-Learning Noun - Definition, Pictures, Pronunciation and Usage Notes / Oxford Advanced Learner’s Dictionary at OxfordLearnersDictionaries.Com*. URL: <https://www.oxfordlearnersdictionaries.com/definition/english/machine-learning> (visited on 08/09/2022).
- [48] Ben. *Clarification on Train, Test and Val and How to Use/Implement It*. Data Science Stack Exchange. Oct. 9, 2019. URL: <https://datascience.stackexchange.com/q/61467> (visited on 08/17/2022).
- [49] *Sklearn.Model_selection.Train_test_split*. scikit-learn. URL: https://scikit-learn/stable/modules/generated/sklearn.model_selection.train_test_split.html (visited on 08/18/2022).
- [50] *Compare the Effect of Different Scalers on Data with Outliers*. scikit-learn. URL: https://scikit-learn/stable/auto_examples/preprocessing/plot_all_scaling.html (visited on 08/18/2022).
- [51] ivanmp. *Plot Linear Model in 3d with Matplotlib*. Stack Overflow. Oct. 17, 2014. URL: <https://stackoverflow.com/q/26431800> (visited on 08/18/2022).
- [52] Qshick. *Ridge Regression for Better Usage*. Medium. Jan. 3, 2019. URL: <https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db> (visited on 08/18/2022).
- [53] *1.1. Linear Models*. scikit-learn. URL: https://scikit-learn/stable/modules/linear_model.html (visited on 08/18/2022).
- [54] *Sklearn.Preprocessing.PolynomialFeatures*. scikit-learn. URL: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html> (visited on 08/18/2022).
- [55] *1.6. Nearest Neighbors*. scikit-learn. URL: <https://scikit-learn/stable/modules/neighbors.html> (visited on 08/18/2022).
- [56] Jerome H. Friedman. “Greedy Function Approximation: A Gradient Boosting Machine.” In: *The Annals of Statistics* 29.5 (Oct. 2001), pp. 1189–1232. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013203451. URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full> (visited on 08/19/2022).

- [57] 1.10. *Decision Trees*. scikit-learn. URL: <https://scikit-learn/stable/modules/tree.html> (visited on 08/19/2022).
- [58] Timothy P. Lillicrap et al. “Backpropagation and the Brain”. In: *Nature Reviews Neuroscience* 21.6 (6 June 2020), pp. 335–346. ISSN: 1471-0048. DOI: 10.1038/s41583-020-0277-3. URL: <https://www.nature.com/articles/s41583-020-0277-3> (visited on 08/19/2022).
- [59] Yves Chauvin and David E. Rumelhart. *Backpropagation: Theory, Architectures, and Applications*. Psychology Press, Feb. 1, 2013. 576 pp. ISBN: 978-1-134-77581-1. Google Books: B71nu3LDpREC.
- [60] Pérez-Enciso and Laura Zingaretti. “A Guide for Using Deep Learning for Complex Trait Genomic Prediction”. In: *Genes* 10 (July 20, 2019), p. 553. DOI: 10.3390/genes10070553.
- [61] Josh Starmer PhD. *The StatQuest Illustrated Guide To Machine Learning*. Independently published, May 2, 2022. 305 pp. ISBN: 9798811583607.
- [62] 3Blue1Brown, director. *But What Is a Neural Network? / Chapter 1, Deep Learning*. Oct. 5, 2017. URL: <https://www.youtube.com/watch?v=aircAruvnKk> (visited on 08/19/2022).
- [63] 1.17. *Neural Network Models (Supervised)*. scikit-learn. URL: https://scikit-learn/stable/modules/neural_networks_supervised.html (visited on 08/19/2022).
- [64] 4.1. *Partial Dependence and Individual Conditional Expectation Plots*. scikit-learn. URL: https://scikit-learn/stable/modules/partial_dependence.html (visited on 08/18/2022).
- [65] Vanchiappan Aravindan et al. “Insertion-Type Electrodes for Nonaqueous Li-Ion Capacitors”. In: *Chemical Reviews* 114.23 (Dec. 10, 2014), pp. 11619–11635. ISSN: 0009-2665. DOI: 10.1021/cr5000915. URL: <https://doi.org/10.1021/cr5000915> (visited on 08/14/2022).
- [66] *Underfitting and Overfitting in Machine Learning*. URL: <https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning> (visited on 08/17/2022).

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl