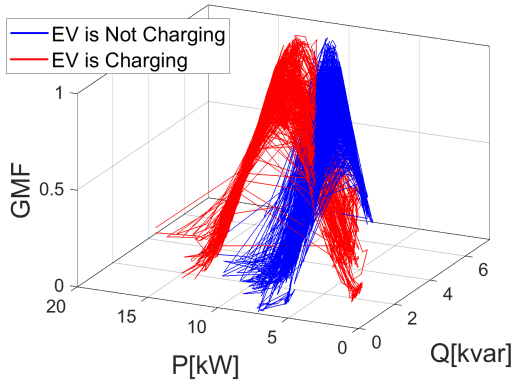

Detection and Localization of Electric Vehicles in Low Voltage Network

Manuel Francisco Pires de Sousa Nunes

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor(s): Prof. Paulo José da Costa Branco
Prof. João Filipe Pereira Fernandes

Examination Committee

Chairperson: Prof. Célia Maria Santos Cardoso de Jesus
Supervisor: Prof. Paulo José da Costa Branco
Member of the Committee: Dr. José Oliveira

November 2019

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Dedicated to the Universe
and its expression in my parents,
Helena and Manuel.

Acknowledgments

Foremost, I have to thank my thesis supervisor, Prof. Paulo Branco, not only for all the patience whenever I was feeling distressed and knowledge when I lacked it but also for believing in me and giving me the opportunity to conduct this work, which I'm passionate about. My next thanks must go to Prof. João Fernandes for showing incredible synthesis capacity and provide me with the most useful of insights. Unquestionably, it was a privilege to learn from both and I will dearly remember and cherish this chapter of my life. I also want to thank my laboratory colleagues, and in particular Francisco Silva, for their contribute in this thesis is far greater than what they can perceive. Moreover, I would like to thank ENEIDA and all the team involved in this research, namely Carlos Pina Teixeira, José Oliveira and Flávio Cordeiro. They have proved to be tireless whenever I was in need of some sort of information, and for that I am deeply grateful.

On a personal note, I have to mention my family. From the very start of my journey in IST, they have been given me nothing but full support in every step of the process, and this thesis is as much as mine, as it is theirs. To my father, thank you for teaching me the value of righteous work and to my mother, thank you for the unconditional love. My final thank you goes to Marta, for not letting me lose hope and always help me see the bigger picture in life.

Resumo

Nos próximos anos, é esperado um aumento de veículos elétricos (VEs) nas estradas, o que acarreta vários desafios em termos do seu carregamento, devendo por isso, haver uma cooperação direta entre as indústrias automotiva e de energia. Para permitir uma integração em larga escala de VEs na rede, os operadores de sistemas de distribuição devem garantir o fornecimento de energia elétrica nas áreas onde há uma grande proliferação de VEs. Esta tese propõe um algoritmo de classificação para detectar, em tempo real, o carregamento de VEs usando o *smart sensor* ENEIDA® EWS DTVI-g instalado no secundário da subestação de distribuição. Extraindo os valores de potência ativa (P) e reativa (Q) e cruzando essa informação com a da entidade responsável pela gestão de carregadores de VEs - MOBI.E® - foram realizadas técnicas de reconhecimento de padrões, usando regressões lineares e funções Gaussianas, resultando no desenvolvimento de um algoritmo de aprendizagem automática supervisionado que tem como intuíto a deteção do carregamento de VEs. Este algoritmo tem em consideração a relação entre P e Q, bem como a duração do carregamento dos VEs. Os limites de viabilidade do classificador proposto foram testados em termos da magnitude da carga na linha do carregador, bem como do modo de potência de carregamento. Além disso, foi realizado um estudo estatístico sobre o comportamento dos utilizadores do carregador de VEs em estudo, utilizando *Gaussian Mixture Models*, para obter um modelo realista que permita a criação de um perfil de carregamento de VEs.

Palavras-chave: Veículos elétricos, Deteção de carregamento, Algoritmo de aprendizagem automática, Rede de baixa tensão.

Abstract

In the next few years, an increase of electric vehicles (EVs) on the road is expected. This increase comes with several challenges in terms of the charging of EVs and there must be a straight cooperation between the automotive and the energy industries. To allow for a large-scale EV integration into the power grid, Distribution System Operators must ensure that the demand in the areas where there is a large proliferation of EVs is met. This thesis proposes a classification algorithm to detect, in real-time, EV charging events using ENEIDA® EWS DTVI-g smart sensor installed in the secondary of the distribution substation. By sampling the active (P) and reactive (Q) power values and crossing this information with the one from the EV charging station network - MOBI.E® - some pattern recognition was conducted, using linear regressions and Gaussian membership functions, resulting in the development of a supervised machine learning algorithm to detect EV charging events. This algorithm takes into consideration the P and Q relationship as well as the EV charging duration. The feasibility limits of the proposed classifier were tested in terms of the load magnitude of the EV charger line as well as the charging power mode. In addition, a statistical study on the behaviour of the EV charger users was conducted, using Gaussian Mixture Model, to obtain a realistic model that allows the creation of an EV charging profile.

Keywords: Electric vehicles, Detection of charging, Machine learning algorithm, Low Voltage network.

Contents

List of Tables

List of Figures

Nomenclature

Greek symbols

Δ Difference.

ϕ Apparent power complementary angle.

Roman symbols

arg Argument.

log Log-likelihood.

r Residual.

T Threshold.

f Frequency.

K Multiplication factor.

P Active Power.

I Current.

Q Reactive Power.

S Apparent Power.

U Voltage.

Subscripts

GMM Gaussian Mixture Model.

GMS Gaussian Membership Function.

i, j, k Computational indexes.

RMS Root Mean Square.

t Time.

x, y, z Cartesian components.

Superscripts

*	Adjoint.
®	Registered trademark.
BK	Background load.
C	Class ”EV is Charging”.
CH	EV charging load.
NC	Class ”EV is Not Charging”.
T	Transpose.

Glossary

AC	Alternate Current
ANN	Artificial Neural Network
API	Application Programming Interface
CM	Confusion Matrix
DC	Direct Current
DSO	Distribution System Operator
EM	Expected Maximization
ESS	Energy Storage Systems
EV	Electric Vehicle
GHG	Greenhouse Gas
GMM	Gaussian Mixture Model
ICT	Information and Communication Technologies
IoT	Internet of Things
LV	Low Voltage
MLE	Maximum Likelihood Estimation
PDF	Probability Density Function
PF	Power Factor
R&D	Research and Development
SDS	Secondary of the Distribution Substation
SG	Smart Grid
SoC	State of Charge
THD	Total Harmonic Disorder
V2G	Vehicle to Grid

Chapter 1

Introduction

This first chapter serves as an introduction for this dissertation and it contains the underlying motivation for conducting this research, in section 1.1, alongside the proposed objectives, section 1.2, as well as this thesis general outline, in section 1.3.

1.1 Motivation

According to the International Energy Agency, in 2018 the global electric vehicle (EV) fleet exceeded 5.1 million units, almost doubling the number of new EV registrations from the previous year [1]. The total number of light-duty vehicle chargers amounted to 5.2 million, 540 000 of which are publicly accessible and 157 000 are fast chargers for buses. Also in 2018, EVs on the road consumed about 58 terawatt-hours (TWh) of electricity and emitted 41 million tones of carbon-dioxide while saving 36 million tones compared to an equivalent Internal Combustion Engine (ICE) fleet [1].

To illustrate the consequences of already announced policy ambitions (*New Policies Scenario*), a projection was conducted by International Energy Agency - Mobility Model, for the year 2030, regarding the global EV stock and it is presented in figure 1.1. These policies were conducted by various global policy makers and it is expected a cut in the demand for oil products by 127 million tonnes (Mtoe) in 2030 [1].

This increase in the number of EVs on the road (figure 1.1) comes with several challenges in which there must be a straight cooperation between the automotive industry and the energy one. In general terms and with the objective of allowing a large-scale EV integration, it is necessary to mitigate resulting power grid malfunctioning and, at the same time, improve efficiency. To accomplish that, Distribution System Operators (DSOs) must:

- Identify areas where there is proliferation of EVs;
- Manage the network demand of those areas;
- Plan network reinforcement, accordingly.

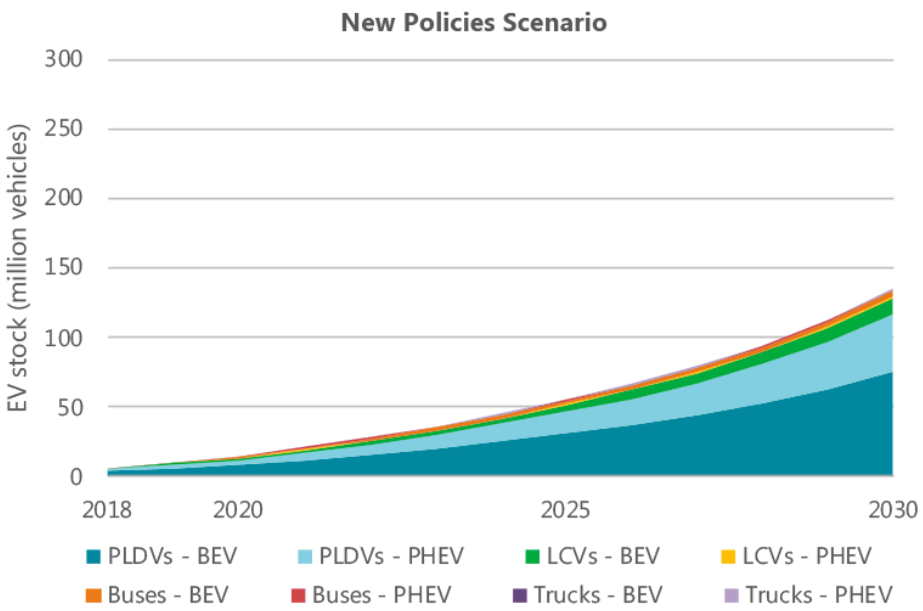

Figure 1.1: Projection of global EV Stock in which PLDVs are passenger light-duty vehicles, LCVs are light-commercial vehicles, BEVs are battery electric vehicles and PHEVs are plug-in hybrid vehicles. Adapted from *International Energy Agency* [1].

To accomplish such tasks, extensive research has been conducted for the past decades (this thesis included) and an overview of some relevant topics is presented in the Chapter *Background* (2).

In Portugal, a partnership was established with the European Commission, named *Portugal2020*, in which several objectives were traced and they are aligned with the key targets for Europe. For 2020, these targets are: 20% cut in *Greenhouse Gas* (GHG) emissions compared to 1990, 20% of total energy consumption from renewable energy and 20% increase in energy efficiency. Fast forwarding 10 years, for the year 2030, the targets are now: 40% cut in GHG emissions compared to 1990, 32% of total energy consumption from renewable energy and 32.5% increase in energy efficiency.

The budget assigned to *Portugal2020*, for the years between 2014 and 2020, is of €25 billion and the investments are done in four fundamental areas: competitiveness and internationalization; jobs and social inclusion; human capital; sustainability and efficiency in resource usage. Two specific measures that comprise the last investment area are: *"foster a low carbon economy"* and *"invest in the renewable energy sources utilization, energy efficiency and smart grids"* [2].

This thesis is then the product of a R&D project with a collaboration between Instituto Superior Técnico and Eneida Wireless and Sensors, S.A. in the ambit of *Portugal2020*, and the proposed objectives are thoroughly described and explained in the following section.

1.2 Objectives

More EV's are being manufactured and deployed, leading to an unprecedented electricity demand that will require proper grid management. To ensure that the grid power quality standards are being met, a coordination of EV charging is crucial and for that it is imperative to have the ability to detect EV charging

occurrences in the Low Voltage (LV) network. Moreover, stochastic models that assess the consumers behaviour of EV charger users are, furthermore, needed to forecast this events. Concerning EV charging, the combination of detection algorithms alongside forecasting models not only adds value to the on-going research on this topic, but has also practical application in the industry, particularly by DSOs, for the reasons stated above.

In this manner, this thesis objective is to build a binary classifier to predict, in real time, EV charging occurrences using LV network parameters, sampled using a smart meter in the Secondary of the Distribution Substation (SDS). This classifier is a supervised machine learning algorithm that labels each sample as belonging to class "*EV is Charging*" or "*EV is Not Charging*", respectively indicating whether an EV is connected to the EV charging station or not.

To build the classifier, not only data from the LV network is needed, but also a validation from the EV charging station network. Having both these datasets, extended pattern recognition is conducted, resulting in an EV charging classifier that takes into consideration the analytical relationship between sampled parameters as well consumers behaviour of the EV charging station users, particularly EV charging time duration. The creation of models that realistically characterize the consumers behaviour of the used EV charging station is also an objective, in the sense that they allow the creation of "fictitious" EV charges.

Lastly, to employ this classifier, the use of "fictitious" EV charging samples realistically modelled and extracted, together with historical data from the SDS, can "train" the classifier so that the need for a validation dataset from the EV charging station network is not needed.

1.3 Thesis Outline

As previously introduced, this section comprehends an overall thesis outline in which a brief explanation of the contents of the following chapters is presented.

In this matter, the chapter 2, comprehends the current *Background* concerning this thesis research. It encompasses the concept of smart grid, its definition, goals and applications, the current EV panorama with its charging modes, European standards and characteristic features of charging as well as the conduct research on forecasting its power demand.

In the following chapter (3) - "Data Acquisition and Pre-Processing" - the data acquisition process is thoroughly explained - how the data sampling procedure was conducted and where. In addition, a characterization of the used EV charging station is performed, based on its location and sampled data, and an overlapping of both datasets (LV network parameters and EV charging station validation) is conducted to understand the extent in which an EV charging influences the sampled LV network parameters. A pre-processing of the data takes place, in which, LV network samples extracted when an EV was connected to the charger are labelled as "EV is Charging", and samples extracted when EV was not connected to the charger are labelled as "EV is Not Charging".

In chapter 4 - "P-Q Data Analysis - Pattern Recognition" - another data processing is conducted, this

time with the objective of obtaining a data aggregation that best characterizes events from class ”EV is Charging” and ”EV is Not Charging”. For that, analyses were performed to understand whether it is preferred to aggregate data by weekday (grouping data samples by *Mondays*, *Tuesdays*, *Wednesdays*, *Thursdays*, *Fridays*, *Saturdays* and *Sundays* separately) or by working days (grouping data samples by *workdays* and *weekends*, separately). In addition, some different approaches were taken to assess whether it is preferable to use all samples, the average value in each time of the day or differentiating according to the time frame in which the data was sampled. Moreover, a study on the evolution of the samples derivative with respect to time was done in this chapter.

Chapter 5 - ”Power Factor Signature Analysis” - contains a description of the proposed classification algorithm. By plotting the active and reactive power values into a P-Q plot and making a distinction between samples that belong to class ”EV is Charging” and ”EV is Not Charging”, it is possible to obtain the likelihood of belonging to each of these classes. The proposed classifier is obtained by comparison of both these likelihoods (using *Linear Regressions* and *Gaussian Membership Functions*), alongside some algorithmic optimizations that comprehend the EV charging time duration and weighting filters. Additionally, some models that represent the consumers behaviour of the EV charging station users are presented, alongside a different algorithm for the creation of fictitious EV charges, from similar models. Furthermore, the metric for performance evaluation - *confusion matrix* - is presented.

Now, in chapter 6, the results concerning the EV detection using sampled LV network parameters, from the SDS, are presented, using the proposed classifier and its optimizations. To ensure impartiality, a 5-fold cross-validation of the data is undertaken and presented. Moreover, two different tests assess the feasibility limits of the proposed classifier under different scenarios: when the line adjacent load varies, simulating different power load magnitude levels on the LV network; when the charging station power rating mode is different than the used one. Also, the results concerning the fictitious EV charging are presented. In the end, an explanation of how this classifier can be used in the industry is conducted.

Finally, in chapter 7, the conclusions taken from this work are presented, namely the achievements and some future work.

Chapter 2

Background

In this sub-section, a comprehensive study of the concept of Smart Grid (SG) is conducted alongside the some previous research on the Electric vehicle panorama, namely its charging modes and European standards, statistical analysis of its charging and the main features for detection and forecasting.

2.1 Smart Grid

As previously introduced, a comprehensive analysis on the concept of *Smart Grid* is performed in this sub-section, as well as its goals and applications.

Quoting Shabanzadeh, M and Moghaddam, M. P. [3]:

> *"The electric power system delivery has often been cited as the greatest and most complex machine ever built. It consists of wires, cables, towers, transformers and circuit breakers — all bolted together in some fashion."*

The traditional power grid is then composed of a connection between the electricity generation, its transportation and finally, the distribution to the consumers. It can be simplistically represented as a one direction process (figure 2.1) and it was developed and deployed in the late nineteenth century.

Figure 2.1: Simplified representation of a traditional power grid. Adapted from *Siemens* [4].

The production of electricity is conducted in power stations that convert other forms of energy into electrical energy and it is market driven. The power stations are associated with transmission substations that increase the voltage so that electric power can be transported efficiently through transmission lines. As electricity has to reach the final consumer, a distribution operator must effectively reduce the

voltage, using distribution substations, and transfer power over its feeders, otherwise known as distribution lines [5].

In the mid twentieth century, some communication network technologies were introduced, namely Supervisory Control and Data Acquisition (SCADA) systems. These systems use dedicated software to monitor, supervise, and control the generation, transmission and distribution equipments, enhancing operational efficiency [5].

In the last decade, a large scale deployment of renewable energy solutions, combined with decentralized energy production driven by both environmental as well as economic reasons, has led to an inevitable upgrade in the electric grid [5, 6]. This upgrade is otherwise known as the *Energy Transition* and it is taking place worldwide resorting to *Smart Grids*.

The definition of *Smart Grid* varies according to different companies and organizations [5, 7]. However, it can be seen as a *smarter* version of the traditional power grid, in the sense that it interconnects all the active participants, from generators to consumers, so that the energy supply can be accomplished, ensuring greater reliability, security and efficiency of the grid [8]. Figure 2.3 is a representation of this interconnectivity.

Figure 2.2: Representation of a smart grid. Adapted from *Siemens* [4].

In another words, it takes advantage of *Information and Communication Technologies* (ICT) and integrate them into the current electricity grid, namely at the generation, distribution and consumption levels, resulting in wholesome systems, safer, more robust, flexible and efficient [6].

Not only does the *Smart Grid* definition varies according to organization and author, but so does its perspectives and ultimate goals. For that matter, some general *Smart Grid* goals are [6]:

- **Improvement of power generation and distribution systems** - combination of both the electric infrastructure with the ICT one, supporting new, more efficient and better managed power systems;
- **Increase use of renewable energy sources** - alternative use of energy sources, such as wind, solar and previous storage, while providing an integration into the electric infrastructure;
- **Better management of energy usage** - use of *Smart Meters* and *Demand Response Systems* to

reduce and balance energy usage. As an example, the bi-directional use of EV charging stations and the vehicle-to-grid (V2G) power interface solution [9].

Now, to ensure that the previously described *Smart Grid* goals are effectively met, the applications are [6]:

- **Electric mobility** : Enabling of large-scale integration of plug-in electric vehicles (EV's), among others.
- **Energy storage facilities** : Provide the means to store energy and easily access it across the grid.
- **Renewable energy sources integration** : Enable intermittent power generation sources, such as photovoltaic energy.
- **Monitoring of grid components** : Monitoring and display of power system components over large geographical areas, in near real time, to optimize its management and performance as well as to detect and prevent future faults.

To understand the level of commitment of the European Commission regarding the *Smart Grids*, quoting the *European Comission* [10]:

> *By 2020, in the EU, it is expected that almost 72% of European consumers will have a smart meter for electricity (close to 200 million units). About 40% will have one for gas (45 million units). This represents a potential investment of €45 billion.*

This dissertation's research is focused on the first of the previously stated *Smart Grid* applications: *Enabling of large-scale integration of plug-in EV's*. The next sub-sections provide an overview of previously conducted research on EV charging, as well as on power quality requirements.

2.2 Electric Vehicle Panorama

The integration of large-scale electric vehicle (EV) charging in the current power grid, represents both a challenge as well as a decisive resource in the context of *Smart Grids*. It can be seen as an unification between the automotive industry and the energy one and requires both these player to come up with innovative, state-of-the-art solutions.

On the one hand, the power grid needs to be upgraded so that there is a coordinated charging of EVs to mitigate an eventual negative impact on the grid, in terms of power outages, voltage fluctuations, harmonic pollution, among others [11]. On the other hand, the fact that EVs are connected to the grid represents an advantage in terms of *Energy Storage Systems* (ESSs), at the level of the consumer, providing peak shaving as well as power quality functions, while at the same time, making the charge time shorter [9, 11]. In this subsection, an overview of the electric vehicle charging, in Europe, is presented as well as some research developed on that matter.

By definition, EVs are any type of vehicle that use electrical drives with the objective of transporting or driving people, objects or physical load. They can use batteries, electrical motors, drive systems and, finally, velocity and torque control and are part of the group of vehicles labelled as ”zero emissions” . [12]

There are several different types of EV’s, with different characteristics and functioning modes. Not only an interest in this type of vehicles is increasing but also in other unconventional ones and some of the most used combine the electric engine with the internal combustion one or even the hydrogen fuel cell. This way, different vehicles can be divided into four categories: Electric vehicles (EVs) ou Battery electric vehicles(BEVs); Hybrid electric vehicles (HEVs); Plug-in hybrid electric vehicles (PHEVs); Fuel cell electric vehicles (FCEVs) [12].

2.2.1 EV Charging Modes and European Standards

Provided that this thesis research aims at detecting EV charging from LV network parameters, the EV charging process and features are considered in this subsection as well as its charging modes and international standards. In addition, to understand the EV drivers charging behaviour, the results of a statistical analysis of the charging behaviour of 221 real residential EV users, which have been monitored for a year during ”My Electric Avenue” project, are further presented [11, 13].

In Portugal, as well as in Europe, voltage supply frequency is $f = 50\,Hz$ and nominal/rated voltage is $230\,V$. In addition, the European standard EN 50160 defines norms regarding the frequency, amplitude, waveform, symmetry of the three phase voltage and the permitted harmonic levels [14]. The EV charging equipment is, obviously, crucial for the grid integration of EVs and daily usage. Its configuration varies according to grid frequency, voltage, output connection and compliance with the defined norms and standards. Succinctly, a charging station includes a charge stand, charge cord, attachment plug, power outlet, vehicle connection and protection system [15].

In general terms, EV chargers can have unidirectional or bidirectional power flow (enabling *Vehicle To Grid* (V2G) technology) and can either be categorized into on-board or off-board types. To ensure similar EV charging solutions across the entire Europe, the European Comission issued a standardization mandate CEN, CENELEC and ETSI(M/468), to promote an internal market for EV. In this matter, standards IEC 61851 [15] are available and in use throughout Europe, and they deal with the charging system, plugs and sockets.

There are four different EV charging modes, according to IEC 61851-1 Committee on ”Electric vehicle conductive charging system”, and they are described below, as in [15]:

- Mode 1: Slow charging from a household-type socket-outlet in AC;
- Mode 2: Slow charging from a household-type socket-outlet with an in-cable protection device in AC;
- Mode 3: Slow or fast Charging using a specific EV socket-outlet with control and protection function installed in AC;

- Mode 4: Fast charging using an external charger in DC.

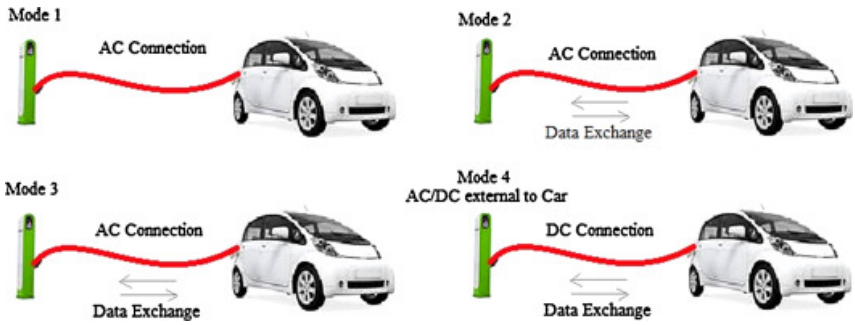

Figure 2.3: Standardized charging modes. From *IEC 61851-1* [15].

Moreover, in table 2.1 it is presented the various types of EV charging ratings, according to rated power [kW] (and consequently time of charge), connection type, maximum current [A] and mode, in Europe.

Table 2.1: Electrical ratings of different EV charger methods in Europe, and respective charging modes. Adapted from *IEC 61851-1* [15].

Charge Method	Connection	Power[kW]	Max Current[A]	Location	Mode
Normal power (slow charging)	1-phase AC connection	3.7	10-16	Domestic	1,2
Medium power (quick charging)	1- or 3-Phase AC connection	3.7-22	16-32	Semi-Public	2,3
High power	3-Phase AC connection	>22	>32	Public	2,3
(fast charging)	DC connection		>3.225		4

In addition to these different ratings, the European committee IEC 62196-2 has defined three types of socket-outlets [15]:

1. Type 1: Single phase vehicle coupler - SAE J1772/2009 automotive plug - Yazaki;
2. Type 2: Single and three phase vehicle coupler - VDE-AR-E 2623-2-2 plug specifications - Meneekes;
3. Type 3: Single and three phase vehicle coupler with shutters - EV Plug Alliance proposal - SCAME

Also according to European standards EN 61851-1 (2011) and NP 61851-22 (2013), voltage levels must not exceed 690 V and frequency $50\,Hz \pm 1\%$. The station must be able to operate with temperatures between -30^oC and 50^oC and a relative humidity between 5% and 95% [14].

When it comes to high power charging, promoting the recharge of an EV within a limited time frame, both DC off-board as well as AC on-board solution exist, being the first the most common. Furthermore, the European Automotive Industry is promoting the combined charging system with the Combo connector. This connector incorporates a single inlet for AC as well as DC charging, on-board, guaranteeing

up to 100 kW high-power charging in the future [15]. In the case of on-board, an AC to DC converter is used and can either be a diode bridge rectifier to charge the battery or a switch-mode converter which not only controls the charging of the battery, but also provides a means for injecting power back into the grid [14].

2.2.2 EV Charging Behaviour - Statistical Analysis and Features

To truly understand the impact on EV charging at the secondary of the distribution substation level, the EV charger characteristics alone are not enough. The forecast, presented in section 2.2.3, stresses the need for extended research of EV charging users behaviours to provide for coordinated EV charging solutions.

Various studies have been conducted on this matter, using data from small-scale trials as well as adapted travel surveys to respectively model EV demand and users behaviour. Some of these studies are in [16–18] and these models are limited to the particular set of typical users used to create them. On the other hand, a large scale study was conducted in the UK in which 200 Nissan LEAFs, used by 221 real residential users, were monitored during the "My Electric Avenue" project, led by *EA technology* and the UK Distribution Network Operator (DNO) *Scottish and Southern Energy Network* [13]. The used vehicles have $24\,kWh$ battery capacity and $3.6\,kW$ demand and more than 68000 charging occurrences have been monitored, over the period of one year (Dec 2013 to Dec 2014).

In this matter, for each EV charging event, the start time, end time, initial state of charge (SoC) and final SoC were recorded. By computing the probability density functions (PDFs) of these metrics, it is possible to create stochastic, realistic and detailed profiles that characterize the EV charging demand in a truthful way, as it was conducted in study [11]. A possible way of obtaining the PDFs of the above metrics is by using Gaussian Mixture Model (GMM), as it was done it [19].

With the objective of creating daily time-series EV profiles, it is also important to know the EV charging demand as well as its power factor ($PF\,or\,cos(\phi)$), which represents the ratio of the real power flowing to the load, to the apparent power. In the same study [11], the active (P), reactive (Q)and apparent (S) power of a specific EV were monitored during a total of 78 days, 1 minute average samples, representing a total of 112,320 samples. The results from that project demonstrate that the EV charging typically demand $3.6kW$ although lower demand values exist when the battery is reaching full charge ($SoC > 85\%$), represented in figures 2.4 and 2.5. Regarding the power factor (PF or $cos(\phi)$), the typical value is of 0.98, inductive, as it possible to observe in figure 2.6.

With the above considerations, the study in [11] proposes an algorithm to create stochastic, realistic and detailed EV profiles. Moreover, there was no significant difference found in the charging behaviour across seasons and the performed analysis divides the dataset into working days and weekends. It was concluded that 70% of EVs are connected once a day, the first connection usually happens with SoC between 25% and 75% and 65% of the EVs finish their first connection with full battery. In addition, it was showed that multiple daily connections do not impact evening peak, but this it does affect morning peak as well as the overall energy consumption [11].

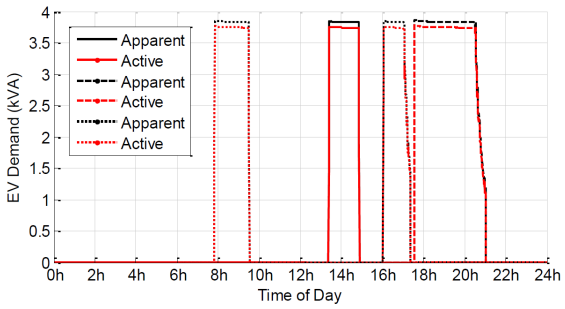

Figure 2.4: Three examples of monitored active and apparent power demand of EV charging. From [11].

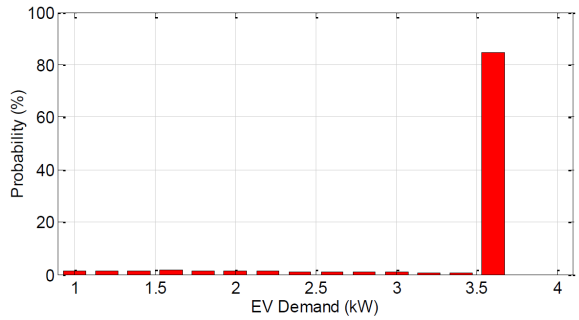

Figure 2.5: PDF of the real EV charging active power demand. From [11].

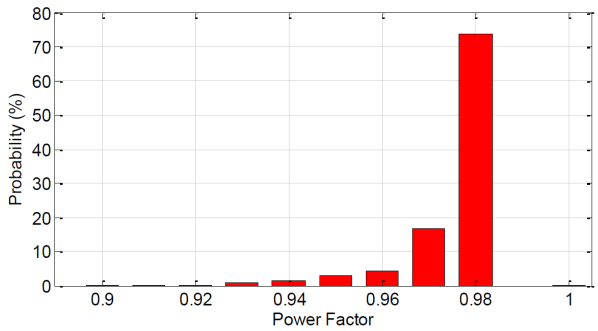

Figure 2.6: PDF of the real EV charging power factor monitored. From [11].

2.2.3 EV Charging Detection and Forecast

The next topic covered in this chapter is the EV charging detection and forecasting. When it comes to EV charging detection, the main features that allow an effective detection, from the LV network perspective, are the active power and power factor (previously described), as well as the injection of harmonics on the AC network. The EV charging demand forecast, as previously stated, is a stochastic phenomenon and to accurately produce an estimation of future EV charging load, advanced forecasting methods are required.

In order to effectively detect an EV charging event, its main features have to be assessed so that they can be monitored. Regarding slow charging, it was already seen that a characteristic feature of EV charging is active power constant and equal to $3.6kW$ until $SoC = 85\%$ and decaying until battery has full charge. Also, that the typical power factor is $FP = 0.98$ inductive.

When considering the power electronic converters of EV charging systems, whether on-board or off-board, it is important to take into account the resulting harmonics being injected into the AC network, and the impact in terms of harmonic distortion. Harmonics are then defined as the sinusoidal component of a periodic waveform having a frequency that is an integer multiple of the fundamental power frequency and a distortion occurs when several of these harmonics are combined. The Total Harmonic Distortion (THD) is the ratio of the sum of the power of all harmonic components to the power of the fundamental frequency. In this manner, some studies have been conducted to determined the level of repercussion on the LV network.

In [20], a simulation case study of the entire low voltage network from a single utility in New Zealand was conducted and its results are presented and discussed. In this research, 10558, $11\,kV$ $415\,V$ transformers and their associated low voltage distribution feeders are used, as well as standard $10A$ chargers (*slow charging*). The purpose of it was to determine the constraining factors for wide-spread EV deployment.

It was demonstrated that, in urban networks, with 40% of EV penetration level (at the same time), the transformers started becoming overloaded, and a possible solution is to use coordinated EV charging. Moreover, it was concluded that the harmonic levels and the THD are below the imposed limits, regulatory levels will not be breached and that the 3rd and 9th harmonics are the most dominant. Moreover, the EV hosting capacity is more limited by the fundamental frequency than by harmonic distortion [20]. In these cases, where **slow charging** occurs, harmonic filtering is not required [20].

When it comes to higher power EV charging, or **fast charging**, other studies were conducted to understand the main features, so it would make it possible to detect an occurrence [14]. With this type of charging, different characteristics are manifested.

According to [14], up until 85% of SoC, the active power drawn is constant and the power factor is $PF = 0.98$, this time its capacitive. From 85% to 100% of SoC, the reactive power increases in steps while the active power decreases, decreasing the power factor. It was concluded that the behaviour of the charger has a much resistive character in the beginning than by the end of the charging, when power factor is very low, displaying a capacitive behaviour. The decrease in the active power flow makes the charging slower to minimize battery deterioration [14].

When it comes to voltage harmonics (figure 2.7), the THD is less than 3% when the standard imposed maximum value is 8%. Regarding individual harmonics, all the values are under the standard maximum value as well. Moving on to current harmonics (figure 2.8), the THD registers high values, near 20%, and the individual harmonics 5 and 7 are in the range of significant harmonic pollution, being the end of charge ($SoC > 85\%$) the most influential time period [14].

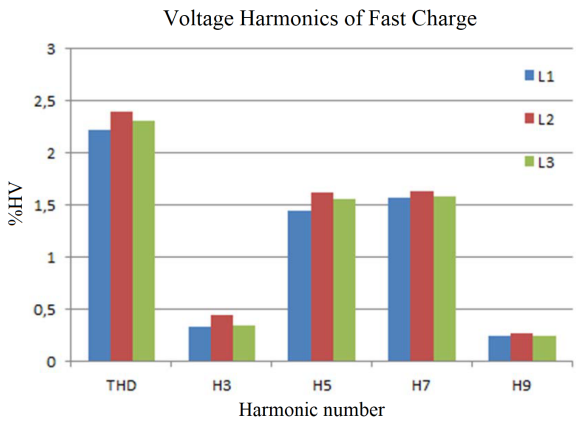

Figure 2.7: Voltage harmonics for fast EV charging. From [14].

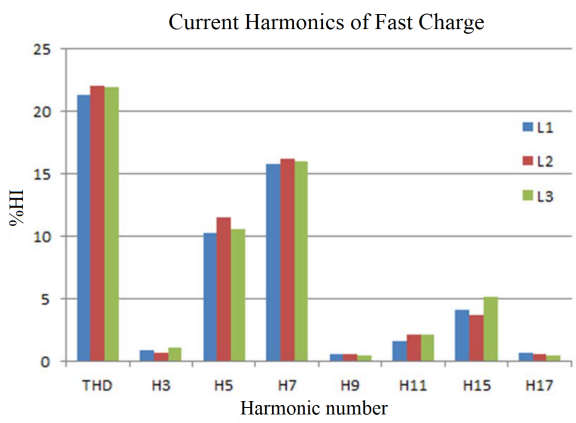

Figure 2.8: Current harmonics for fast EV charging. From [14].

In sum, concerning slow EV charging, the active and reactive power are the main LV network features that allow for a successful detection. When it comes to fast charging, the active and reactive power as

well as the current harmonics can be used for prediction.

Forecasting

In general terms, load forecasting is divided in three categories according to time frame, namely: short-term, medium term and long term. Each one of them plays a vital role in what concerns the power grid operation. Short-Term Load Forecasting comprehends load demand predictions from one hour to one week and it is important for power system stability and energy cost minimization. Medium-Term Load Forecasting ranges its predictions from one week to one year and it is used in power system planing. Lastly, Long-Term Load Forecasting is used for predictions over one year and it is very useful for investments on infrastructure and new generation technology [21].

Several different approaches have been taken when it comes to create power load forecasting models. Some of these include statistical and artificial intelligence models such as: time series method, autoregressive integrated moving average, regression analysis, Kalman filtering, artificial neural networks (ANNs), support vector machines and deep learning methods [22]. According to [23], ANNs for load forecasting have been proved accurate and efficient. Recently, deep learning techniques, that are based on ANNs, have been used, due to advances in the computation hardware as well as developed techniques providing strong learning and accurate models.

Furthermore and concerning EV charging demand, not only research on overall time-based forecasting is being conducted but also spatial load forecasting to obtain spatio-temporal clustering of charging demand [24]. Indeed, knowledge of both the location and time of EV charging events is needed to guarantee continuous supply of energy and install sufficient power capacity within each service area of the distribution grid.

Chapter 3

Data Acquisition and Pre-Processing

In this chapter, an explanation of the data acquisition process is conducted. With the objective of build an EV charging classifier - to detect beginning and ending of EV charging processes - it is necessary to have the LV Network parameters sampled in real time as well as some EV charging validation metric to ensure that an EV charging is occurring. To accomplish the first, data was sampled using Eneida's smart-meter and the second, from the MOBI.E® website [25].

By merging the data from these two distinct sources, it is possible to start conducting a pattern recognition analysis and eventually build an EV charging events classifier.

ENEIDA®

As it was previously introduced, Eneida's smart-meter EWS DTVI-g is installed in the Secondary Distribution Substation (SDS) and is used to monitor the LV Network. This monitoring consists of real time data sampling and a computation of it is conducted using an IoT solution. Then, using the web-based platform **eneida DeepGrid®**, it is possible to visualize and extract this data [26]. Figure 3.1 presents the smart-meter EWS DTVI-g.

This device measures, at maximum, 6 SDS feeders and 3 phases per feeder. By feeder and by phase, in each 5-minutes window, 4 measurements are made (voltage (v) and current (i)) of 1 second each, at a frequency $f = 4096\,Hz$, spaced equidistantly in time [26].

For each measurement, the following parameters are calculated: active energy, active energy direction, active power (P), apparent power (S), consumed energy, voltage frequency (f_v), RMS current (I_{RMS}), reactive energy, reactive energy direction, reactive power (Q), RMS voltage (U_{RMS}), power factor $(PF\ or\ cos(\phi))$, power factor direction, current phase angle, voltage phase angle, current frequency (f_V) and neutral current (I_n) [26].

For each set of 4 measurements of the parameters, the average is calculated and sent to the server, with U_{max}, U_{min}, I_{max} and I_{min} obtained as well. For some of these parameters, when there is a signal inversion, sending the average might not be the most indicated. In this matter, all physical phenomenon whose time duration occurs at a scale smaller than $75\,s\ (= 5min/4)$ are out of reach [26].

In this research, the DTVI used is located in Avenida Calouste Gulbenkian, Sto António Olivais,

Figure 3.1: Eneida's EWS DTVI-g smart-sensor installed in the SDS. Photo from *ENEIDA* [26]

Coimbra, Portugal, SDS 556 (coordinates 40.12507, -8.24485). The data of feeder 6, line 2 was used in this work provided that it was connected to some non-disclosed load that includes an EV charger, described further.

MOBI.E®

The MOBI.E® network is composed by EV charging stations for electric vehicles mostly situated in public access spaces [25].

In order to sample EV charging validation data, the MOBI.E® website [25] was used because it has publicly accessible real time data of the EV chargers that belong to its network. The state of an EV charger was monitored and sampled, with a sampling time of roughly 5 minutes.

The chosen EV charging station is a public one, located in Alameda Armando Gonçalves, Coimbra Portugal (coordinates 40.215254, -8.413305) and it is manufactured by Efacec. The charging socket is CBR-00004-02 and uses a Type 2 - 62196-2 Charging Station Socket of 3.7 kW (slow charging).

This EV charging station is located in a parking lot, near an industrial area, with some stores nearby as well as a couple of residential buildings.

3.1 Characteristic features of EV charging - Coimbra Case Study

In order to make sense out of the available data, sampled both from ENEIDA as well as from MOBI.E, a comprehensive study of it is accomplished in this chapter.

A characterization of the selected EV charger is presented in sub-section 3.1.1. A temporal analysis of ENEIDA's data together MOBI.E's one is performed to understand which LV network parameters are more influenced whenever an EV charging is occurring, in sub-section 3.1.2.

Furthermore, the days that comprise our *data lake* are: 5-12 and 14-31 of March, 1-23 and 27-30 of

April, 1-31 of May, 1-30 of June, 1-26 of July and 25 of September. This work will be conducted using data sampled in these 141 days exclusively. The days in between are not available provided that EV charging validation data could not be sampled.

3.1.1 EV charging characterization

As previously stated, the number of days used to conduct this study was 141 and out of those, 101 were working days and 40 were weekends. It is important to note that, all of the characterization analyses performed in this section were done using the entire *data lake* above mentioned, unless otherwise specified.

In this matter, the total number of EV charging events summed up to 365. In figure 3.2, a histogram of the number of EV charges per weekday is presented.

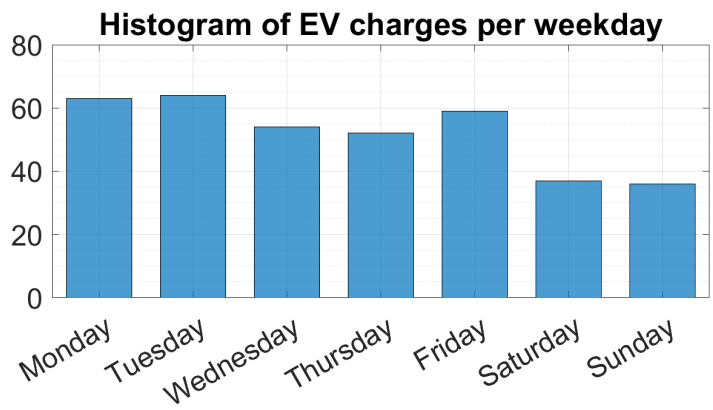

Figure 3.2: Histogram of EV charges per weekday.

In what concerns the *prevalence* of EV charging - percentage of time the EV charger was connected to an EV - was of 31.9% using all days, 34.9% during working days and 24.5% during weekends.

Another useful characteristic of the users of the EV charger in study is the number of EV charges per day. Figure 3.3 represents an histogram of that feature.

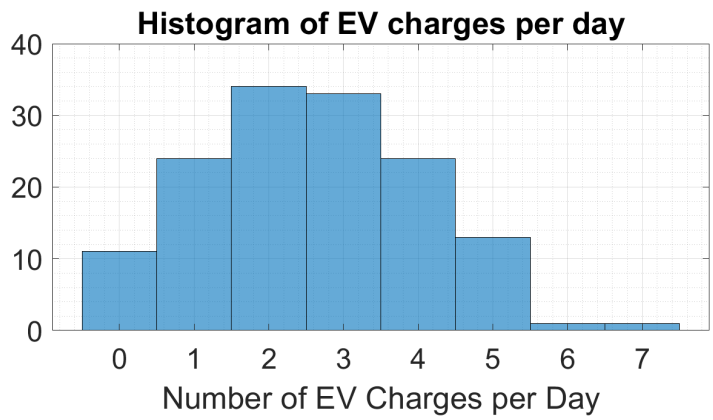

Figure 3.3: Histogram of EV charges per day.

Moving on to the time of the day in which EV charging begins and ends, a histogram was computed

for each, using the entire *data lake*. These histograms have a 1 hour bin width and are represented in figures 3.4 a) and b), respectively.

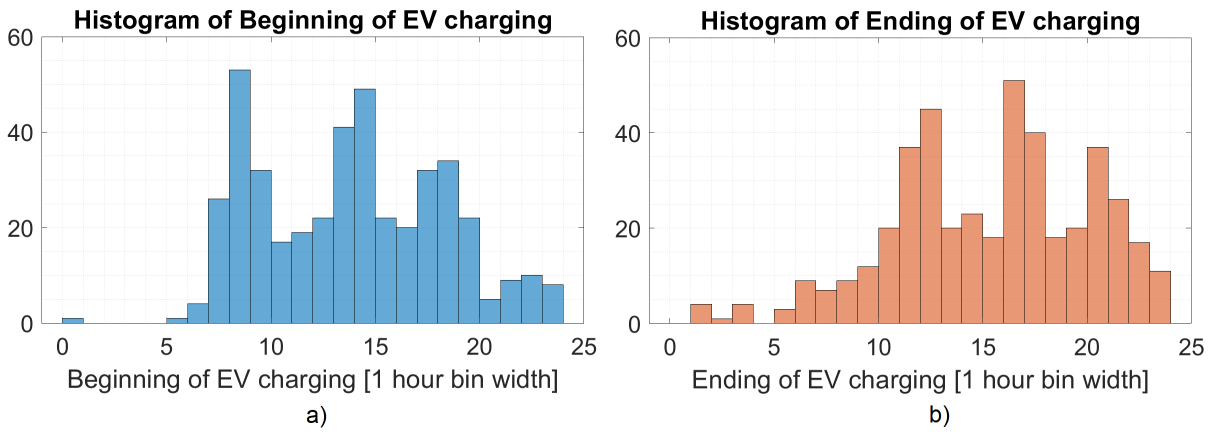

Figure 3.4: Histogram of EV charging beginning a) and ending b) hours of the day of the entire dataset, with a 1 hour bin width.

Looking at figure 3.4 a), one can empirically observe that the periods with greater number of occurrences are between $8:00h-9:00h$ and $14:00h-15:00h$. Correlating with the location of the charger in study being near an industrial area, one can assume that these time periods correspond to EV charger users that put their vehicle to charge upon arrival at work, in the morning, and after lunch. Elseways, looking at figure 3.4 b), the time periods in which more EVs disconnect are between $12:00h-13:00h$ and $16:00h-17:00h$, possibly when EV drivers leave work.

Another useful characteristic regarding the drivers behaviour of the EV charger in study is the duration of charging. Figure 3.5 represents the histogram of EV charging duration, of the charger in study.

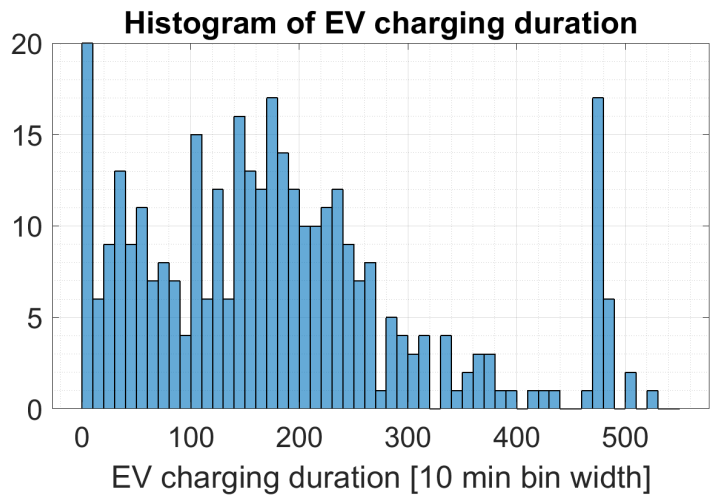

Figure 3.5: Histogram of EV charging duration with a 10 minutes bin width.

The most frequent time duration of an EV charging is between $0-10$ minutes and the second most is between $170-180$ minutes and $470-480$ minutes.

3.1.2 Temporal Analysis of Extracted Parameters

As it was previously mentioned in this sections introduction, an overlapping of data from both ENEIDA as well as MOBI.E sources is accomplished. The goal of it is to realize which sampled LV Network parameters are the most influenced whenever an EV charging occurs.

A slow EV charging is characterized mostly by an increase in the active power of $3.6\,kW$. The power factor is 0.98 inductive and the harmonic content is despicable. With this in mind, and taking into consideration the obtained parameters (from the point of view of the SDS), a temporal evolution of the samples P, Q, I_{RMS} and U_{RMS} is compared with the EV charging occurrences. In addition, the time derivative of each parameter is computed and presented as well.

In these matter, figures (3.6, 3.7, 3.8 and 3.9) correspond to an overlapping, respectively between the aforementioned parameters temporal evolution (blue), and the Mobi.e data (red) that validates the occurrence of an EV charging.

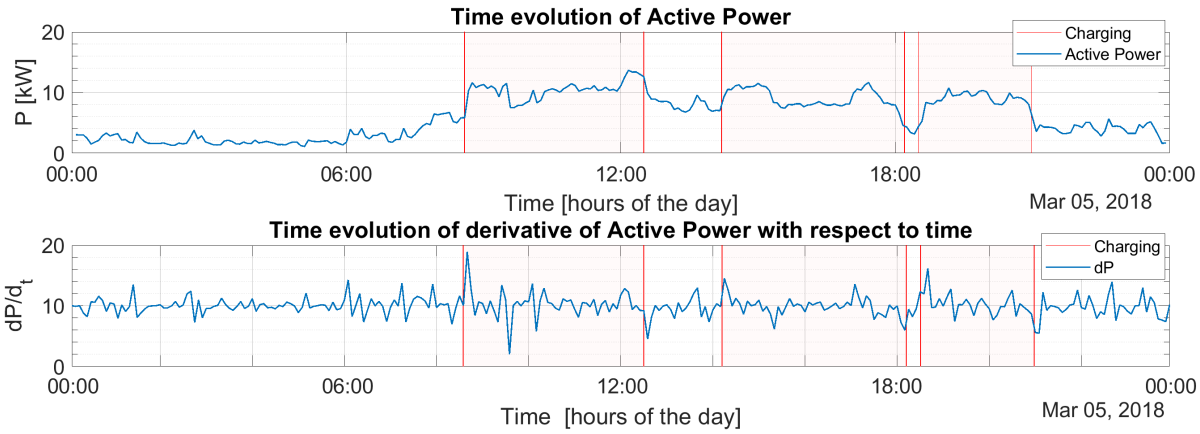

Figure 3.6: Time Basis plot of active power and corresponding derivative with respect to time (blue) and EV charging period (red) in 05/03/2018.

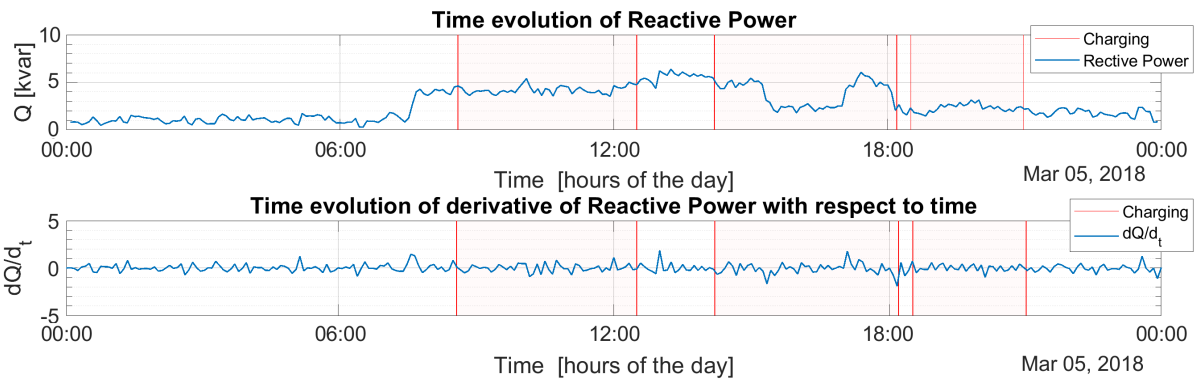

Figure 3.7: Time Basis plot of reactive power and corresponding derivative with respect to time (blue) and EV charging period (red) in 05/03/2018.

By inspection it is possible to correlate the charging of an EV with an increase in the active power - high derivative of P value in the beginning of an EV charging. The reactive power (Q) does not change significantly when an EV charging is occurring. Nevertheless, Q varies according to adjacent load.

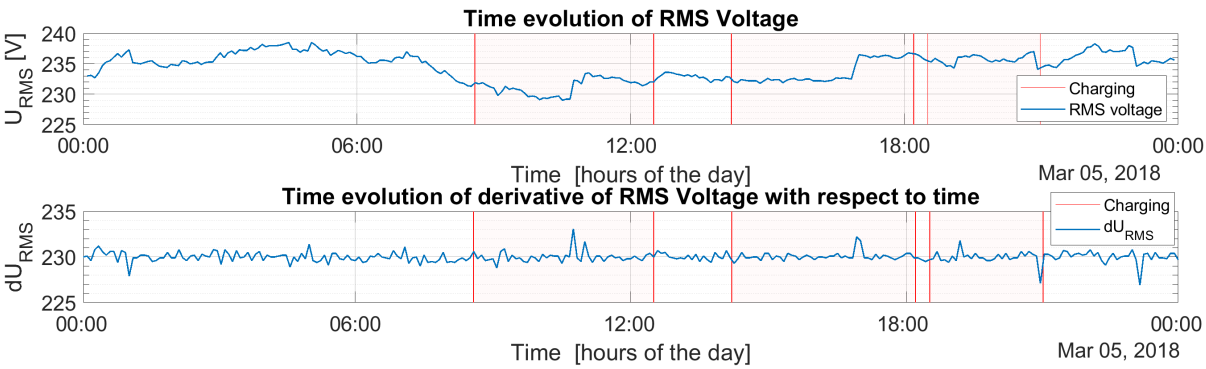

Figure 3.8: Time Basis plot of RMS voltage and corresponding derivative with respect to time (blue) and EV charging period (red) in 05/03/2018.

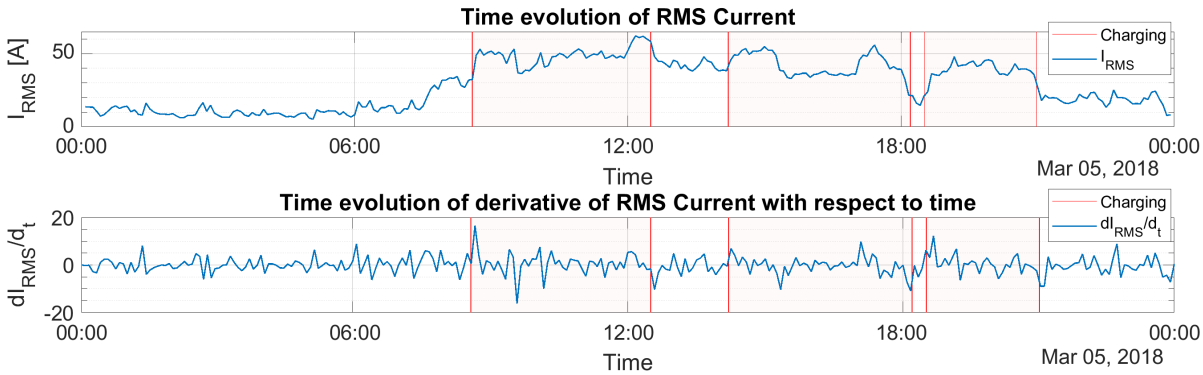

Figure 3.9: Time Basis plot of RMS current and corresponding derivative with respect to time (blue) and EV charging period (red) in 05/03/2018.

These empirical observations support the existing literature about the active power profile of the charging of an EV as well as the MOBI.E EV charging demand. Looking at figure 3.8 - temporal evolution of U_{RMS} - it is not possible to correlate directly the charging of an EV with a variation in the RMS voltage at the substation level. On the other hand, the RMS current, present in figure 3.9, is sensible to an EV charging.

This way, in what concerns the best features that characterize an EV charging, among the ones obtained by the smart meter, the active and reactive power were considered the most relevant.

Chapter 4

P-Q Data Analysis - Pattern Recognition

In this chapter, the study of some patterns in what concerns the EV charging active and reactive power (P and Q) is performed. The goal of this assessment is to obtain the two, as distinct as possible, data aggregations characterized by P-Q samples in which EV charging was occurring, class "*EV is Charging*", and EV charging was not occurring, class "*EV is Not Charging*". A P-Q sample (or P-Q point) indicates a pair of active and reactive power quantities, sampled at the same time instant by the smart-meter.

Moreover, a hypothesis was made in which, if the given data was manipulated in a way that these two classes samples characterization is as differentiated as possible, a better estimation in what concerns EV detection and classification is produced. This characterization will encompass the behaviour of the consumers of that particular EV charger.

To obtain data patterns, a P-Q scatter plot is used. Different approaches were applied to extract patterns concerning EV charging. They are:

- Weekday aggregation of P-Q samples, grouped in *Mondays*, *Tuesdays*, *Wednesdays*, *Thursdays*, *Fridays*, *Saturdays*, *Sundays*;
  - The average value at each time of the day, separating between weekdays;
  - All samples plotted in P-Q graph, separating between weekdays;
  - All samples plotted in P-Q graph, differentiated by time frames, separating between weekdays;
- Workday aggregation of P-Q samples, grouping working days in one set and weekends in another set;
  - All samples plotted in P-Q graph, separating between working days/weekends;
  - All samples plotted in P-Q graph, differentiated by time frames, separating between working days/weekends;

The same methodology is applied, using now a dP-dQ scatter plot. In this analysis, the change of active and reactive power is plotted with a differentiation for when an EV charging is beginning, further trying to obtain a disaggregation between the classes "*EV is Charging*" and "*EV is Not Charging*".

4.1 P-Q Plot with Linear Data Fitting

In this section, a comprehensive clarification of the P-Q graphic in the context of this thesis is performed. The terminology *P-Q graphic* refers to a scatter plot of aggregated active and reactive power (the y-axis corresponds to the active power (P) in kW and the x-axis to the reactive power (Q) in kvar) composed of multiple P-Q samples, sampled each one in different time instants. The scatter plot do not comprehend the temporal evolution of the samples and the obtained patterns reveal P and Q typical arrangements (patterns) when an EV is charging and when it is not.

By comparing the data extracted from the Eneida's smart-meter, in the secondary of the distribution transformer, with the data obtained from Mobi.e® website, it was possible to identify the values of P and Q when an EV charging was occurring and when it was not, providing a useful label and two distinct classes, respectively "*EV is Charging*" and "*EV is Not Charging*". It is, however, important to note that, the Mobi.e validation data indicates if an EV is connected to the EV charging station and not if the EV is extracting power from the grid [25].

As seen in the sub-section 3.1.2, when an EV charging occurs, there is an increase in P demand of 3.7 kW, while Q variation is less significant, characteristic of normal power charging method (slow charging). This represents a general behaviour in terms of EV charging power and occurs throughout the data set. Hence, having the two data sets - one for each class - it is expected that the previously stated EV charging behaviour can be represented in these P-Q scatter plots, thus with some patterns emerging.

Moreover, in this chapter, a linear fitting of both "*EV is Charging*" and "*EV is Not Charging*" classes P-Q samples is computed to assess the extent in which these two classes are disaggregated. This data fitting is constrained to the origin and, together with other methods, is used to make a comparative analysis between different data aggregations techniques conducted in this chapter. In addition, an unconstrained fitting is also performed provided that the data fitting has a lesser error.

4.2 Weekday Aggregation of P-Q Samples: Understanding the Behaviour of the EV Charger Users

In this section, an aggregation of P-Q samples is performed with the objective of understanding the behaviour of the users of the EV charger in study, and establish some patterns using the P-Q plot. It was further assumed that the EV charger in study has a somewhat fixed number of EV users with similar utilization patterns in the same days of the week, given that it is located near an industrial area.

Furthermore, in figure 3.2, it is possible to observe that there is a significant difference of EV charging

occurrences when it comes to working days compared with the weekends, reinforcing the need for a tailored solution.

In the following subsections, some analyses will be conducted and the randomly chosen weekday, which this work will focus, is on Wednesday. Moreover, with the purpose of demonstrating result consistency, the resulting constrained linear fitting of all weekdays will be presented and explained in subsection 4.2.2.

4.2.1 Average Value of P and Q

The first approach when trying to establish EV charging patterns was to conduct an average estimation of P and Q quantities whenever an EV charging was and was not occurring. In order to accomplish that, a plot of these quantities was done over the time period of a day and several days were aggregated, according to the weekday aggregation. Furthermore, the mean value in each sampled instant was computed and plotted, as well.

Figure 4.1 represents the individual load in the several Wednesday's, among the days presented in section 3.1, summing up to 20 days in total. Each P and Q day load is represented in blue, from the charge line in study, and the average value in each sampled instant is in red color. Note that, the P and Q quantities in each day correspond to the load attached to the line in study and represents a combination of both the EV charging load, as well as other adjacent load that is not possible to identify.

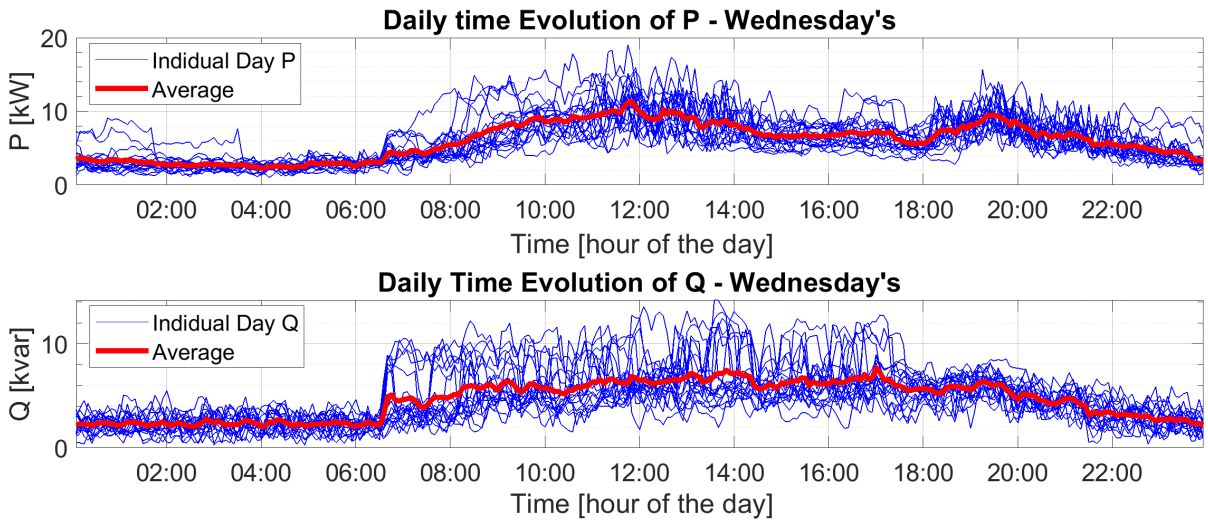

Figure 4.1: Wednesday's time evolution of P and Q (in blue) and the correspondent tendency (in red), in March, April, May, June and July.

Thereafter, whenever an EV charging was occurring, the active and reactive power load of that line were not considered $(P = 0 \cup Q = 0)$. With the remaining P and Q load quantities, the average value, in each sampled instant, was computed, and it is present in figure 4.2. Then, both these average values are to be compared, as presented in figure 4.3 and plotted in a P-Q scatter plot, in figure 4.4.

In figure 4.3, it is possible to realize the daily average whenever EV charging occurrences are and are not considered. Looking at P evolution, majority of the time, the average encompassing EV charges

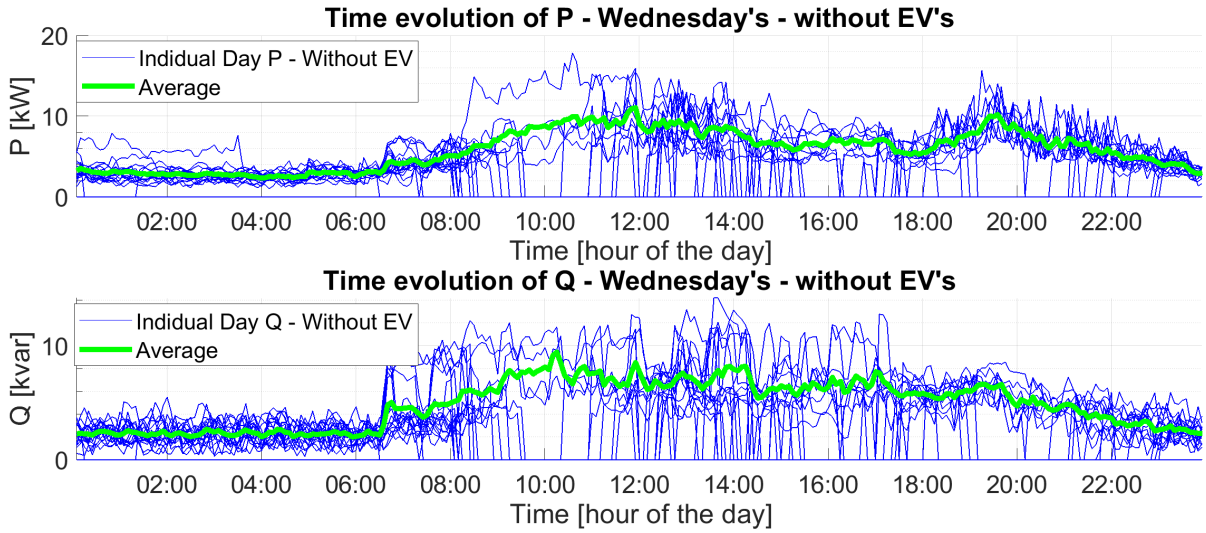

Figure 4.2: Wednesday's time evolution of P and Q (in blue) when an EV charging was not occurring, and the correspondent tendency (in green), in March, April, May, June and July.

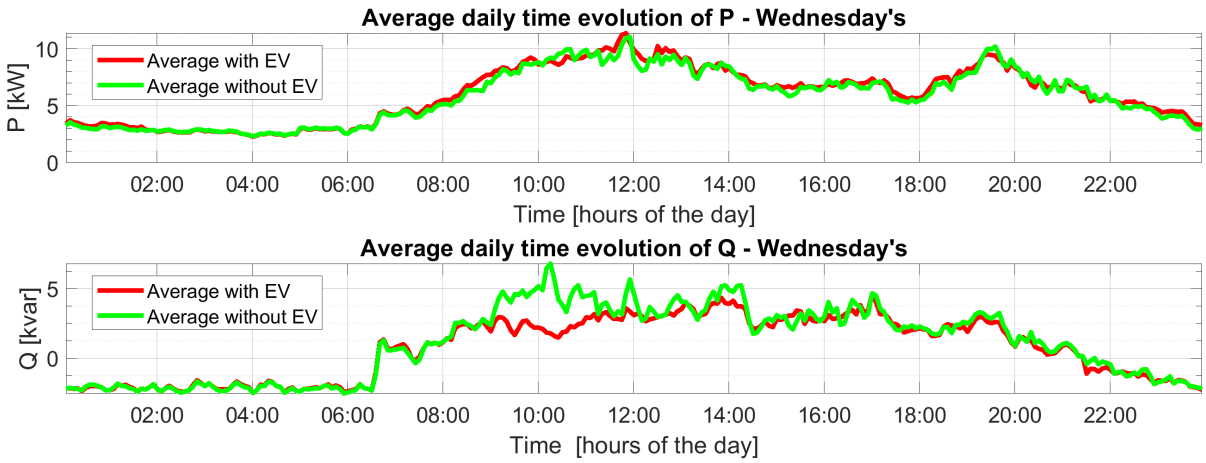

Figure 4.3: Tendency of P and Q when EV charging was happening (in red) and when EV charging was not happening (green).

is higher than the one excluding them. In this case, the sum of the residuals [27] (y-axis difference between "*Average with EV's*" and "*Average without EV's*") at each sampled instant, of 49.39. Moreover, and taking notice of the Q evolution also in the same figure, the average without EV's has higher values than the one with EV's during some periods of the day, with a sum of residuals of -38.90. These results reveal the lack of data in terms of line load without EV charging, making it not possible to extract a correct average value. According to some bibliography, the reactive power when *slow EV charging* is occurring is minimal, given that the PF associated with a slow EV charging is usually 0.98 inductive [11].

The correspondent P-Q scatter plot of these averages is presented in figure 4.4, and it is possible to infer that the two clusters are almost entirely overlapping each other, making very difficult any disaggregation.

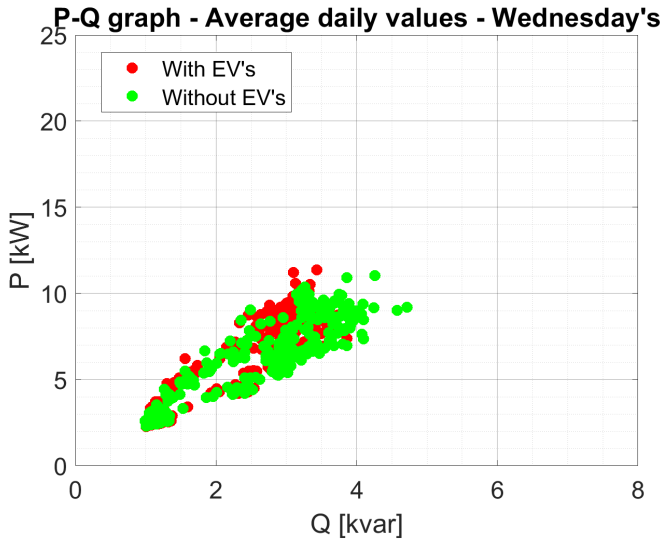

Figure 4.4: P-Q scatter plot of average daily values with load whenever EV was charging (in red) and without (in green).

4.2.2 All samples aggregated

Given the apparent difficulty in disaggregating the two classes by using the daily average with and without EV charging, in this section, one will consider all data points from the same weekday, and not just the average values.

As it was stated in this chapter's introduction, the remaining weekday's analysis was conducted to verify consistency in each day of the week and the results will be presented in the end of this section. However, for the sake of discussion and to understand the rationale behind these assessments and what patterns could be revealed, only Wednesday's figures will be presented in this section.

Figure 4.5 is representative of the P-Q scatter plot using all samples (no average values) from the previously plotted Wednesday's (figure 4.1). Moreover, given this plot results in an overly crowded P-Q graphic, possibly missing some patterns or data aggregations, roughly half of the dataset was used in figure 4.5 to conduct this assessment. In particular, the used days were the ones between 5 of March and 7 of May 2018.

By observation of figure 4.5, one can realize the existence of two classes: the one corresponding to P-Q samples when EV charging was occurring (in red) and was not (in blue). Although not completely disaggregated and distinct (there is quite some overlapping), it is possible to realize some degree of clustering in different areas. For instance, with roughly $P \in [6, 15]\, kW$ and $Q \in [0.5, 5.5]\, kvar$ there is a significant agglomeration of "EV is Charging", P-Q samples (red). Similarly, in the region $P \in [1, 9]$ and $Q \in [0.25, 5]$, there is a major aggregation of "EV is Not Charging" P-Q samples (blue). The y-axis displacement of these agglomerations further reinforces the active power increase when an EV is charging, not followed by a reactive power one.

Additionally, a linear fitting of the red and blue data samples was performed to have a representation of how disaggregated the classes could be. This linear fitting was computed using both an origin constrained linear fit of each class P-Q samples [28], in which the y-intersect is forced to zero, and an

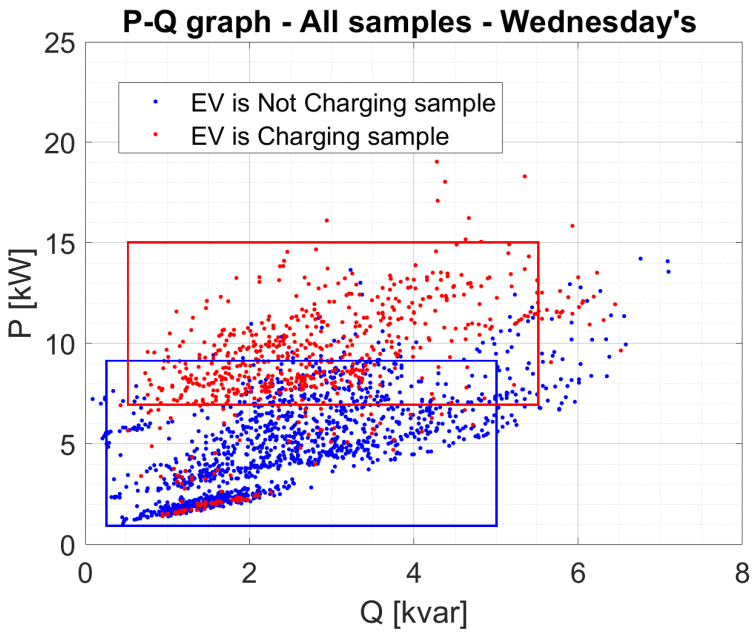

Figure 4.5: P-Q scatter plot of all samples, with (in red) and without (in blue) EV charging on Wednesday's [5 of March to 7 of May 2018].

unconstrained linear regression, with an explanation in section 5.1.2. Figure 4.6 is a representation of both an origin constrained linear fitting (figure 4.6 a)), as well as an unconstrained linear fitting (figure 4.6 b)) of each class P-Q samples.

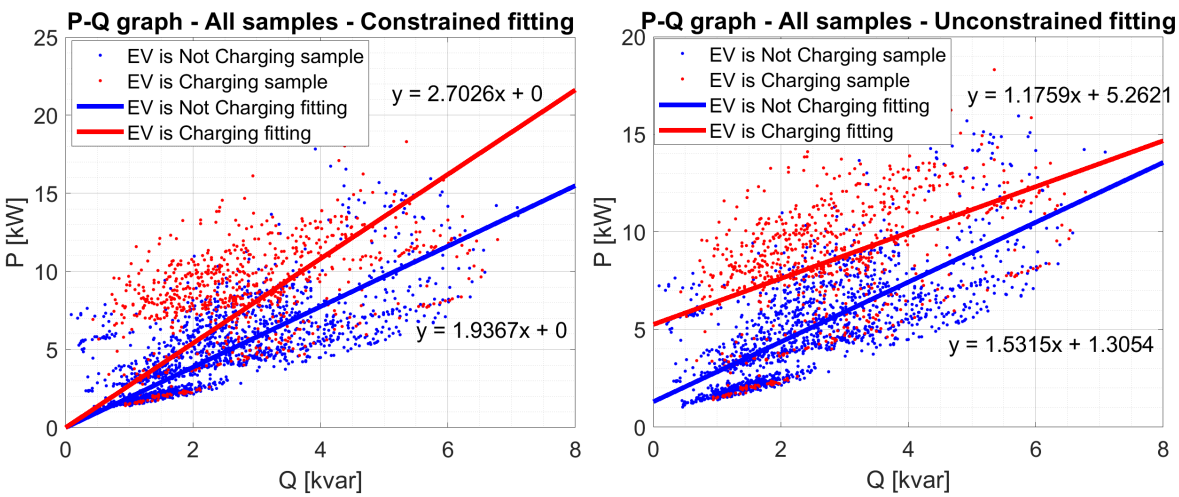

Figure 4.6: Constrained linear data fitting (a)) and unconstrained linear data fitting (b)) of all Wednesday's data samples of "EV is Charging" class and "EV is Not Charging" class.

In figure 4.6 a), a linear data fitting constrained to the origin is performed [28]. The "EV is Charging" data fitting function has an expression $y = 2.7026x + 0$ and the "EV is Not Charging" data fitting function has an expression $y = 1.9367x + 0$. The line angles are then respectively $\phi = 69.69^o$ and $\phi = 62.69^o$ with a difference of $\Delta\phi = 7.00^o$ between both.

In what concerns figure 4.6 b), the linear data fitting is unconstrained to the origin, meaning that a better characterization of these classes is done, in terms of the associated residuals. This way, the lines

have expressions $y = 1.1759x + 5.2621$ and $y = 1.5315x + 1.3054$, respectively "EV is Charging" and "EV is Not Charging" classes. The angles associated with these lines are not comparable in the sense that the y-axis intersection is different for both.

Moreover, the Euclidean norm of the residuals [27, 29] is also computed. The residual, in this case, is the y-axis difference between the actual value and the predicted fitting one [27], while the Euclidean norm of the residuals is the square root of the sum of squares of these differences. In the equation 4.1, $\|y\|_2$ is the residuals Euclidean norm, N is the total number of samples from each cluster and r is the residual value.

$$\|y\|_2 = \sqrt{\sum_{i=1}^{N} r_i^2} \tag{4.1}$$

Computing 4.1, using each line equation from the origin constrained data results in 97.52 and 80.34, and from the unconstrained one, in 75.23 and 76.51, respectively for "EV is Charging" and "EV is Not Charging" classes. As it was expected, the unconstrained data fitting has the least norm of residuals. Moreover, in order to ensure that there is consistency using different days of the week, this analysis was repeated for the remaining ones, listing the line angles ϕ, their difference $\Delta\phi$ and the respective norm of the residuals ($normres$) was obtained, from Mondays to Sundays. The results of this assessment are presented in Table 4.1.

Table 4.1: Linear fitting lines angle (ϕ) and their difference ($\Delta\phi$) of classes "EV is Charging" and "EV is Not Charging" for all weekdays using constrained linear fitting. Norm of the residuals ($normres$) for both constrained (**Cons.**) as well as unconstrained (**Uncons.**) linear fitting.

		MON	TUE	WED	THU	FRI	SAT	SUN
Cons.	$\phi(^o)$ "EV is Charging"	68.47	68.42	69.69	70.82	70.76	71.41	72.66
	$\phi(^o)$ "EV is Not Charging"	59.00	61.56	62.69	62.05	61.71	65.12	62.54
	$\Delta\phi(^o)$	9.47	6.86	7.00	8.77	9.05	6.29	10.12
	($normres$) w/ EV's	110.83	108.56	97.52	94.24	111.04	71.25	71.00
	($normres$) w/o EV's	52.41	53.89	80.34	58.70	79.99	77.37	48.01
Uncons.	($normres$) w/ EV's	62.11	78.91	75.23	64.45	74.26	62.93	45.65
	($normres$) w/o EV's	46.48	50.85	76.51	55.67	74.42	76.31	47.06

Taking table 4.1 in consideration, it is possible to verify that there is some consistency when it comes to the existing difference between the "EV is Charging" cluster and "EV is Not Charging" one, being the $\Delta\phi$ a good indicator, with mean $\mu = 8.22^o$ and standard deviation of $\sigma = 1.48^o$. Also, there groups of two consecutive days that present values of $\Delta\phi$ very similar, for instance, Tuesday and Wednesday, Thursday and Friday and Sunday and Monday.

Now, the **total** norm of residuals of all these weekdays is computed of both classes with constrained and unconstrained linear fitting is:

- $\|y\|_2^{Cons.\,w/\,EV's} = 254.77$
- $\|y\|_2^{Cons.\,w/o\,EV's} = 173.90$

- $\|y\|_2^{Uncons.\,w/\,EV's} = 177.36$
- $\|y\|_2^{Uncons.\,w/o\,EV's} = 165.15$

Naturally, the unconstrained linear fitting has least error and produces a better characterization of the data. These values will be further used for comparison purposes.

4.2.3 Time-frame analysis

Now, the P-Q plot will be used to understand the typical P-Q values, this time using a time-frame differentiation. This will be performed taking into consideration the dual-tariff and triple-tariff time schedules used by EDP, in Portugal [30] in which the electricity is charged differently according to the time of the day. As it was previously stated, the day of the week with which the study is conducted is on Wednesday's.

Dual-Tariff Time Schedule

As it was previously introduced, a time differentiation of samples is performed according to the dual-tariff practised by EDP [30]. In this manner and using the weekly cycle option, the used time slots are between 00h00-07h00 - *Off-Peak* hours - and 07h00-24h00 - *Peak* hours. Figure 4.7 shows a representation of the EDP dual-tariff.

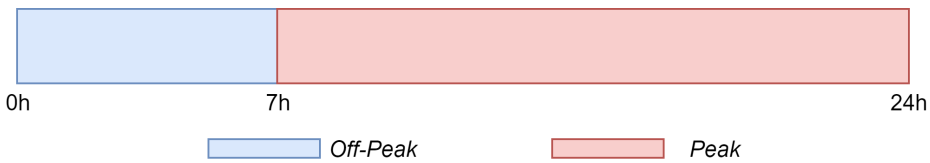

Figure 4.7: Dual-tariff time schedule representation.

In addition, figure 4.8 is the P-Q scatter plot distribution of both these time-schedules in different color, when EV's were charging.

In figure 4.8, there is a higher number of EV charging samples during *Peak* hours (88%) than during *Off-Peak* hours (12%). Moreover, during *Off-Peak* hours, the majority of the Q values are below 2 kvar (90% of samples), however with variable P values. Looking to the cluster with $Q \in [0.75, 2.5]\,kvar$ and with $P \in [1, 3]\,kW$, it is possible to realize that there is almost no cluster position distinction between P-Q points when an EV is charging and when it is not. This fact leads to the conclusion that the EV battery was already charged, no P-Q power was being extracted, and was still connected to the charging station. This means that there are samples in the used dataset that correspond to an EV charging scenario but with no P-Q power expression, henceforth denominated in this thesis as "*phantom charges*".

Triple-Tariff Time Schedule

In a similar way as it was done previously, a time differentiation was conducted when analysing the P-Q samples, but now using a triple-tariff time schedule such as the one used by EDP [30]. For simplicity

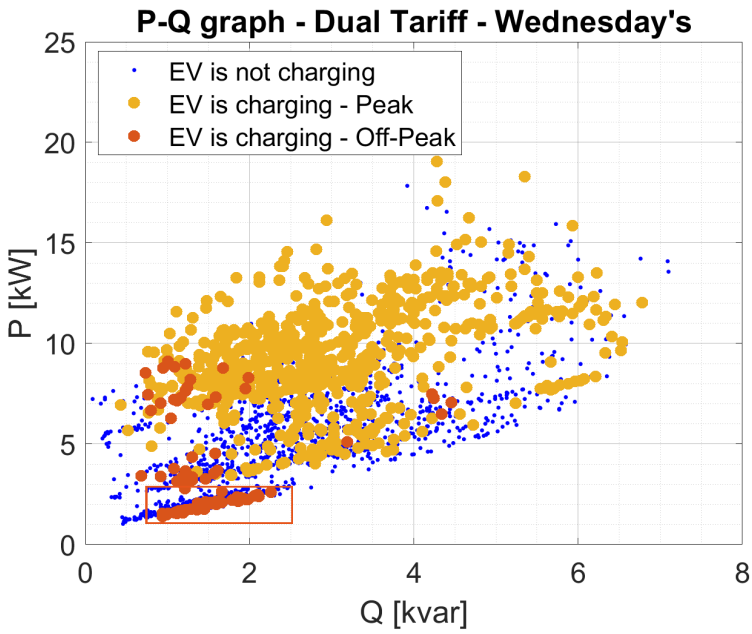

Figure 4.8: P-Q graphic using a dual-tariff time differentiation by color.

purposes, the weekly cycle option was considered, using the summer legal hour [30] in which the time differentiation indicated in 4.9 stays:

- 00h00-07h00 : *Off-Peak* hours;
- 07h00-09h15, 12h15-24h00 : *Peak* hours;
- 09h15-12h15 : *Half-Peak* hours.

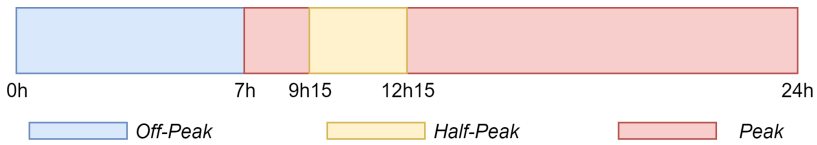

Figure 4.9: Triple-tariff time schedule representation.

Following the same reasoning as the one performed for figure 4.8, a time-frame color differentiation, using triple-tariff, was conducted using a P-Q scatter plot, presented in figure 4.10.

With the triple-tariff analysis, different levels of EV charging penetration, according to time periods, took place, as shown in figure 4.10. The *Off-Peak* period has a prevalence of 11%, *Half-Peak* has a prevalence of 22% and *Peak* has a prevalence of 67%.

Moreover, looking at 4.10, it is possible to realize that during *Off-Peak*, the majority of the samples have Q lower than 2 kvar, with three different P aggregations: $P \in [1, 2]\, kW$ (however *phantom charges*), $P \in [2, 5]\, kW$ and $P \in [6, 9]\, kW$.

The *Half-Peak* (marked in red), despite being the shortest period (9h15-12h15), has the second-highest incidence level (22%) with the majority of P being higher than 9 kW, which reinforces the assumption that during that period, users who arrived at work leave their car to charge.

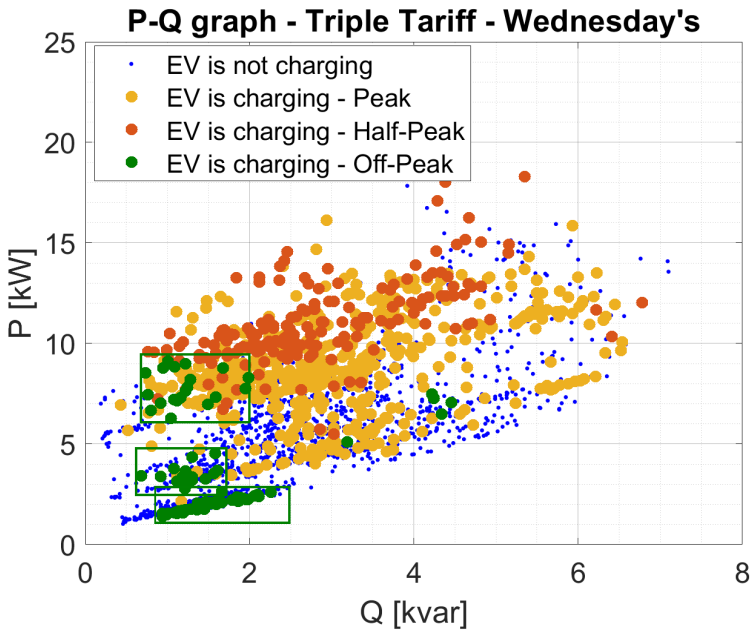

Figure 4.10: P-Q graphic using a triple-tariff time differentiation by color.

Lastly, *Peak* hours have both the largest incidence level, as well as the most dispersion in terms of P-Q samples. The Q values encompass almost the full spectrum of samples, from $Q \in [1, 7]\, kvar$ and $P \in [1, 14]\, kW$.

This analysis provides a good time reference for the existent data clusters, showing that a single day can be divided into three charging classes.

4.3 Workday and Weekend Aggregation of P-Q Samples

In this section, some of the previously conducted analyses are again taken into consideration, but now aggregating the data differently. Instead of making a weekday aggregation, as it was done in section 4.2, a working day aggregation is performed. Moreover, this analysis will test the hypothesis that during the working days of the week, the same EV charging P-Q characteristic occurs. The same analysis is applied to weekends.

By working day aggregation, a separation of working days and weekends is done, and for the same reasons as the ones explained in section 4.2.2, only data from 5 of March to 7 of May 2018 is considered.

The two different assessments performed in this section are the one including all samples and a second one with the dual-tariff time frame analysis.

4.3.1 All Samples

In this subsection, all samples regarding the working days as well as the weekends, are analysed distinctively. The objective for this analysis is the detection of patterns in the data that could be useful when building the EV charging classifier.

Working Days

Figure 4.11 is a representation of a P-Q plot containing all samples extracted during the working days of the week. The blue data marks the P-Q samples when an EV was not connected to the charger - ”EV is Not Charging” class. The red marks are the P-Q samples when an EV was connected to the charger - ”EV is Charging” class.

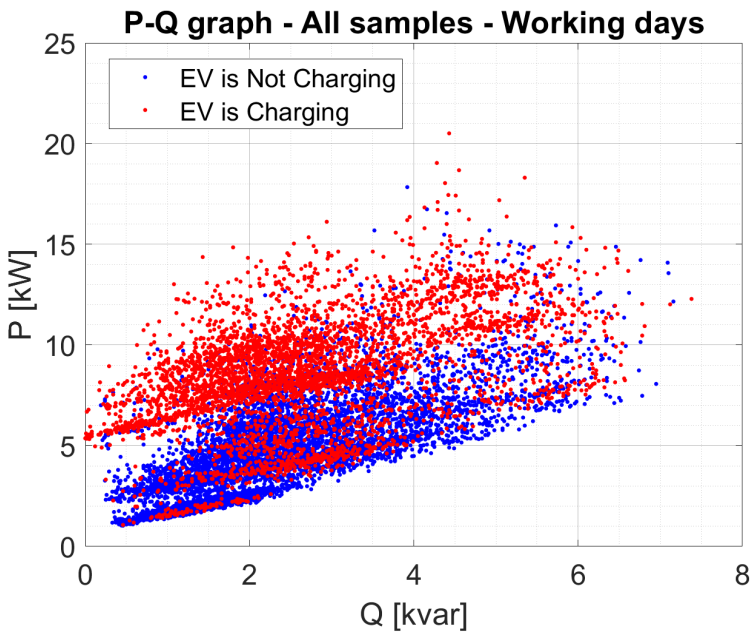

Figure 4.11: P-Q plot of all samples when an EV charging was connected to the charger (red), and when an EV was not connected to the charger (blue), during working days.

Looking at figure 4.11, it is possible to empirically observe different cluster’s of samples in both classes. In figure 4.12, the major, empirically observed, P-Q samples aggregations of each class are represented. Cluster’s 2, 4 and 6 belonging to class ”EV is Not Charging”, while cluster’s 1, 3 and 5 belonging to the class ”EV is Charging”.

In figure 4.12, cluster’s 4 and 5 represent the major P-Q samples aggregation, being expected a significant increase in active power once an EV charging begins. In other words, a P-Q sample belonging to cluster 4 (”EV is Not Charging” class) shifts to cluster 5 (”EV is Charging” class).

Concerning cluster’s 1 and 2, shown in figure 4.12, both have very low P values. These P-Q samples cluster’s correspond to EV charges that happened in *Off-Peak* hours, as figure 4.8 shows. Furthermore, cluster 1 samples belong to class ”EV is Charging” however, they are considered as being *phantom charges*.

Similarly to what was done in subsection 4.2.2, both a constrained as well as an unconstrained linear data fitting of each class’s samples was obtained for comparative purposes, respectively shown in figures 4.13 a) and b). Note that *phantom charges* are not removed.

In figure 4.13 a), the x-axis angle of ”EV is Charging” and ”EV is Not Charging” fitting lines is respectively $\phi = 69.52^o$ and $\phi = 61.38^o$ with a difference of $\Delta\phi = 8.14^0$. The Euclidean norm of the residuals is, respectively, 236.05 and 149.94. Now, in figure 4.13 b), the fitting lines of ”EV is Charging” and ”EV

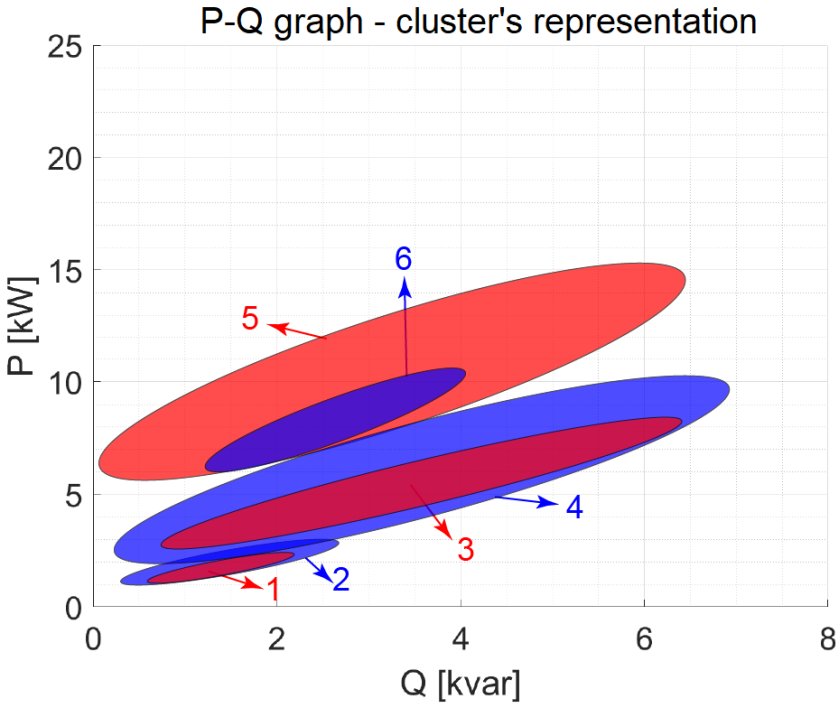

Figure 4.12: Empirical observation of major cluster's of P-Q samples when an EV charging is occurring (1,3,5) and when it is not (2,4,6). Cluster 1 P-Q samples are considered as *phantom charges*.

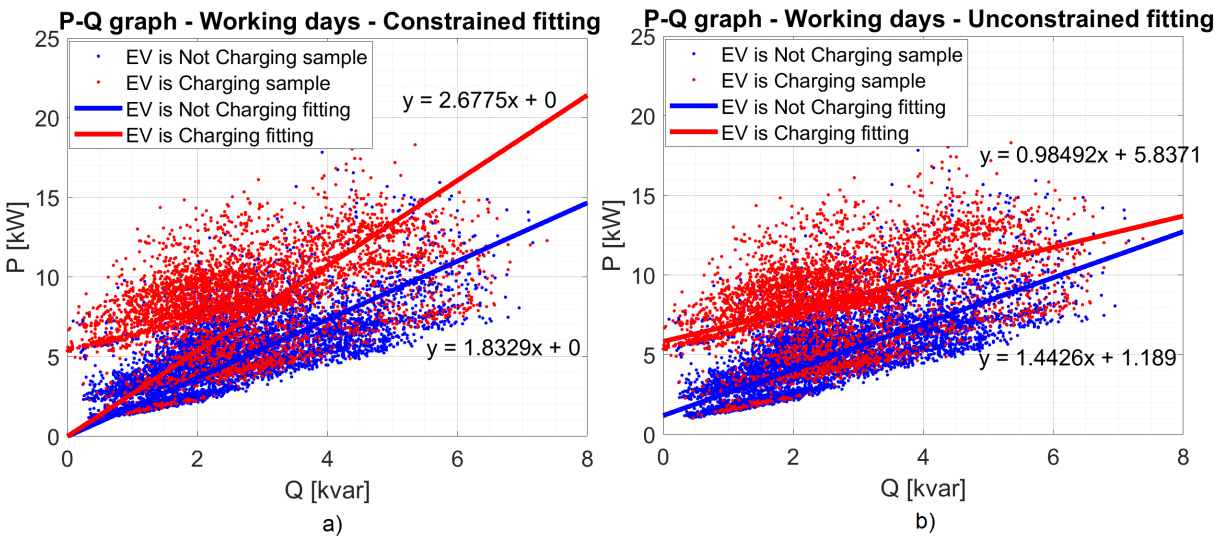

Figure 4.13: P-Q scatter plot of all samples when an EV charging was occurring (red) a when an EV charging was not occurring (blue) during working days with constrained (a)) and unconstrained (b)) linear fitting.

is Not Charging" classes have equations $y = 98492x + 5.8371$ and $y = 1.4426x + 1.189$, being the norm of the residuals respectively, 160.41 and 141.11.

Weekends

Now, similarly to the analysis made for working days, a weekend aggregation was taken into consideration to understand its feasibility. The P-Q scatter plot results are represented in figure 4.14.

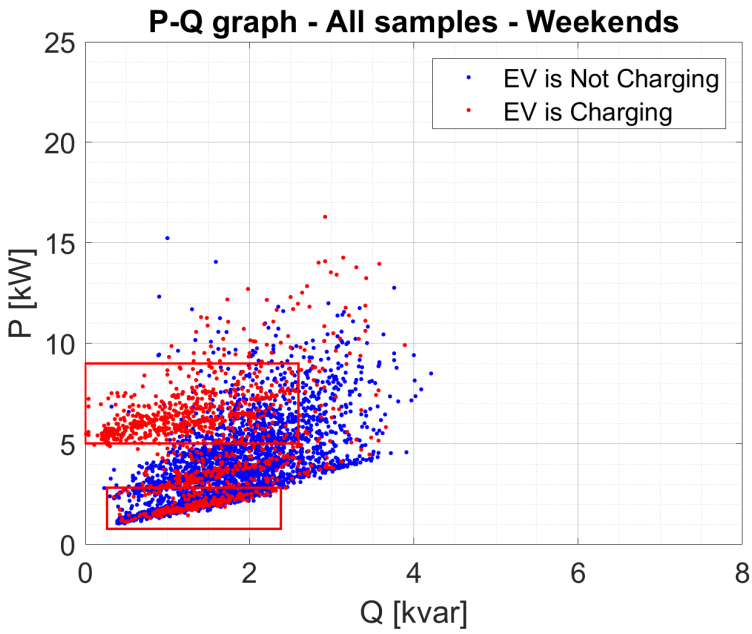

Figure 4.14: P-Q scatter plot of all samples when an EV charging was occurring (red) a when an EV charging was not occurring (blue) during weekends.

Results in figure 4.14 indicate that the Q values have a very strict range, $Q \in [0, 4]\, kvar$ when compared with the working days P-Q scatter plot, figure 4.11. Moreover, similarly to working days P-Q scatter plot, there is a P-Q data aggregation, with very low P values that, as it was previously described, no power is being withdrawn despite the EV being connected to the charger - *phantom charges*. Roughly with $P > 5kW$ another major P-Q "EV is charging" sample aggregation is displayed with low Q values, $Q < 3\, kvar$. This aggregation corresponds to actual EV charging occurrences, providing that the power withdrawn is concordant with the used EV charger, $P > 3.7\, kW$.

Samples from class "EV is Not Charging" have a much broader aggregation. However, noticing figure 4.14, these two classes samples are not completely distinct from one another. This way, to understand a general trend line, both a constrained as well as an unconstrained linear fitting is, again, realized in figure 4.15 a) and b). Note that *phantom charges* are not removed.

In figure 4.15 a), the x-axis angle of "EV is Charging" and "EV is Not Charging" fitting lines is respectively $\phi = 71.82$ and $\phi = 64.18$ with a difference of $\Delta\phi = 7.64$. The norm of the residuals is, respectively 100.76 and 91.94. Now in figure 4.13 b), the fitting lines of "EV is Charging" and "EV is Not Charging" classes have equations $y = 0.85612x + 3.9529$ and $y = 1.836x + 0.44391$ and the norm of the residuals is respectively, 80.86 and 91.36.

In this section, the $\Delta\phi$ of the working days aggregation and the weekends one is very similar, respectively 8.14^o and 7.64^o. However, in terms of dispersion, the working days cluster's present more dispersion than the weekends one, both in terms of P and of Q.

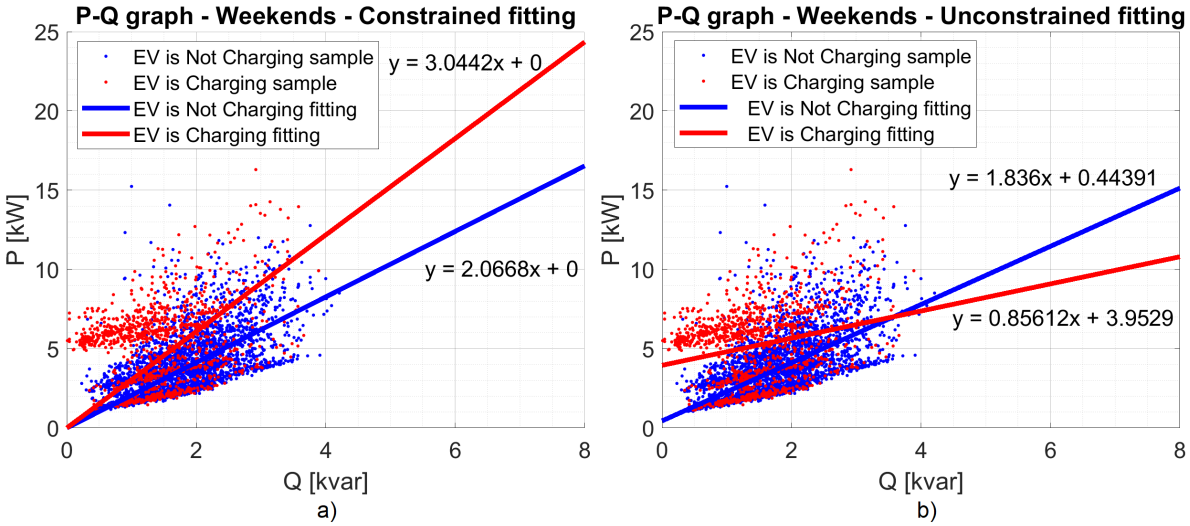

Figure 4.15: P-Q scatter plot of all samples when an EV charging was occurring (red) a when an EV charging was not occurring (blue) during weekends with constrained (a)) and unconstrained (b)) linear fitting.

4.3.2 Dual-Tariff Time Frame Analysis

Similarly to what was done in subsection 4.2.3, a dual-tariff time frame analysis is conducted again, but now having all the working day's data and the weekend's data aggregated.

Figure 4.16 represents the dual-tariff time frame analysis in both working days (a)) as well as weekends (b)), remembering that the time period between *00h00-07h00* is the *off-peak* hours (in orange) and between *07h00-24h00* is the *peak* hours (in yellow).

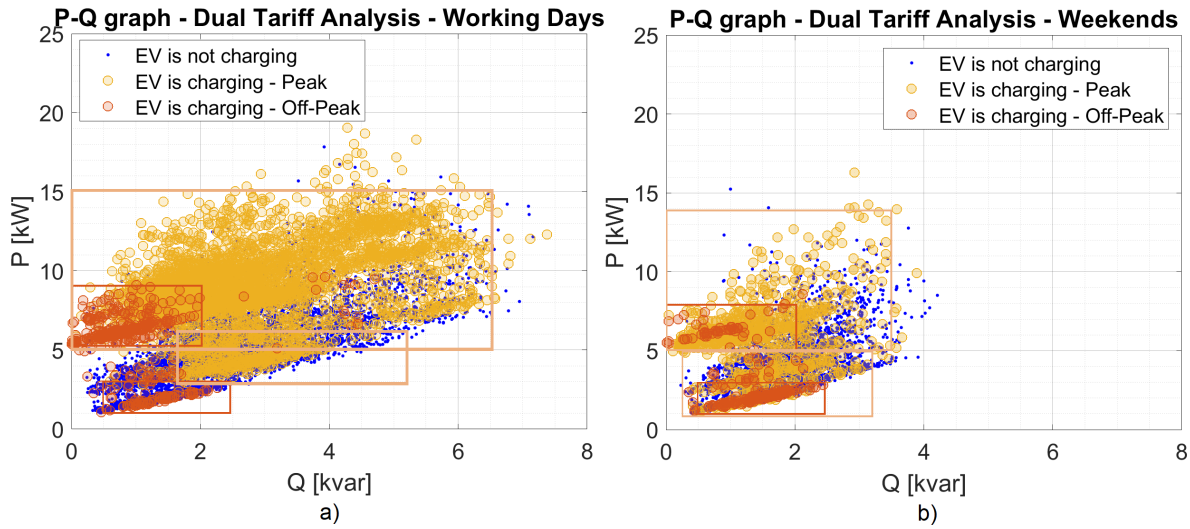

Figure 4.16: P-Q graphic using a dual-tariff time differentiation by color - (a)) working days aggregation and (b)) weekends aggregation.

Concerning the *Off-Peak* hours (in orange), the first thing to notice in both figures 4.16 a) and b), is the existence of *phantom charges*, with $P \in [1, 3]\, kW$ and $Q \in [0.5, 2.5]\, kvar$. Now considering figure 4.16 a) - working days, there is another considerable aggregation of P-Q samples with $P \in [5, 9]\, kW$

and $Q \in [0, 2]\, kvar$, corresponding to actual EV charges. Similarly, in figure 4.16 b) - weekends, the aggregation with P-Q samples between $P \in [5, 8]\, kW$ and $Q \in [0, 2]\, kvar$ also correspond to actual EV charges.

Concerning the *Peak* hours, P-Q samples (in yellow) are much scattered, going almost through the full Q spectrum. In figure 4.16 a) - working days, the P values have a major cluster with $P \in [5, 15]$ and a second one has $P \in [3, 6]$. These clusters have an elliptical shape and as the Q values increase, so the does the P ones, appearing to have a crescent trend line. In figure 4.16 b), - weekends, there is a significant difference when it comes to the magnitude and dispersion of P and Q, being lesser than during working days in both cases. In terms of active and reactive power, two cluster's with similar Q and different P values: one cluster is between $P \in [1, 5]$ and $Q \in [0.25, 3.25]$, and the other one is between $P \in [5, 14]$ and $Q \in [0, 3.5]$.

Working Days

Now, the unconstrained linear data fitting of both *Off-Peak* hours, as well as *Peak* hours, is done in order to understand the feasibility in making such a data aggregation. The constrained fitting is not shown in this section for simplicity purposes. The results are presented, respectively in figure 4.17 a) and b).

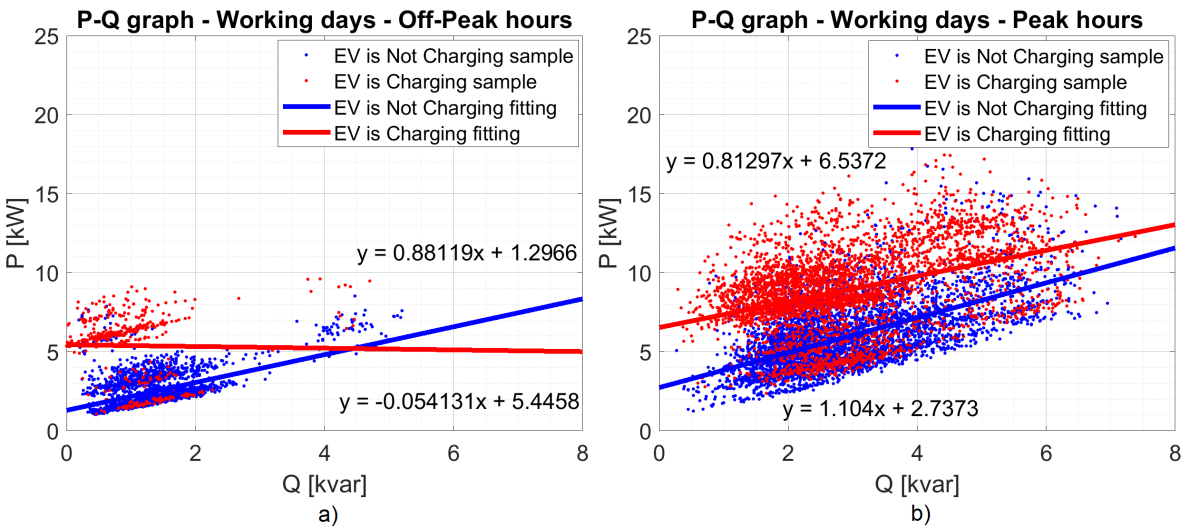

Figure 4.17: P-Q scatter plot of all samples and unconstrained linear fitting during working days in *Off-Peak* hours (a)) and *Peak* hours (b)) .

Looking at figure 4.17 a) - *Off-Peak* hours - it seems that the *phantom charges* have a considerable influence when computing the unconstrained fitting lines, providing a downward trend line as Q increases. This way, it was concluded that this fitting line does not represent the EV charging power behaviour. The lines have expressions $y = -0.05413x + 5.4458$ and $y = 0.88119x + 1.2966$ with a norm of the residuals of 44.37 and 49.62, respectively for classes "EV is Charging" and "EV is Not Charging".

Now, looking at figure 4.17 b) - *Peak* hours - it displays a different general trend, more aligned with an actual EV charge. It is possible to observe that both fitting lines follow the major data aggregations of each class, despite some cluster overlapping. Furthermore, the lines have expressions $y = 0.81297x +$

6.5372 and $y = 1.104x + 2.7373$ with a norm of the residuals of 147.35 and 113.54, respectively for classes "EV is Charging" and "EV is Not Charging".

Weekends

Now, for the weekends, the fitting lines of classes "EV is Charging" and "EV is Not Charging" during *Off-Peak* hours and *Peak* hours were computed with the results in figure 4.18, respectively a) and b).

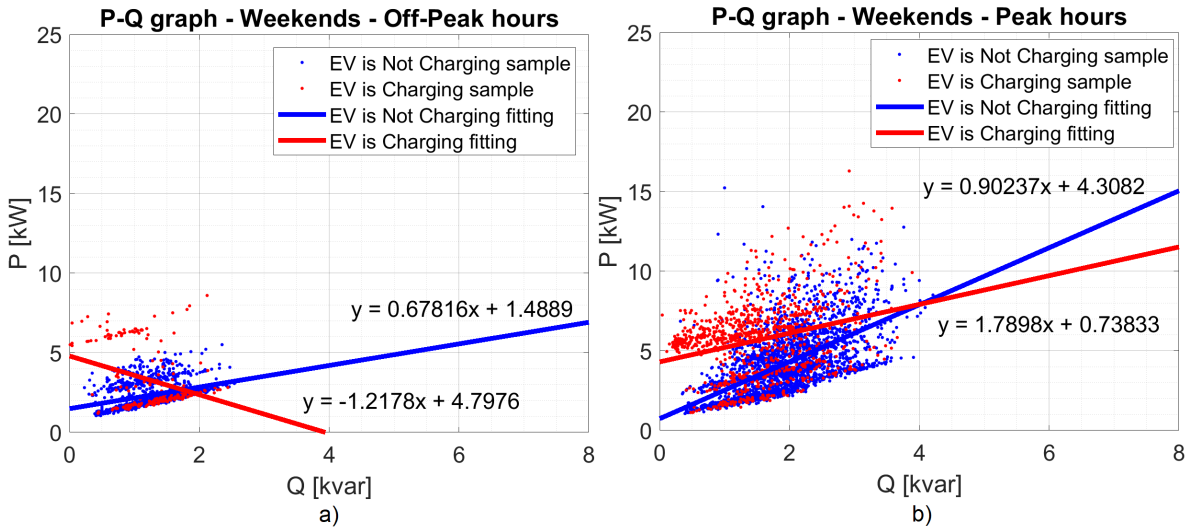

Figure 4.18: P-Q scatter plot of all samples and unconstrained linear fitting during weekends in *Off-Peak* hours (a)) and *Peak* hours (b)) .

It was previously stated that there is a very different power profile when comparing working days with weekends, namely the fact that the Q quantities are inferior. In figure 4.18 a) - *Off-Peak* hours - for the same reason as the one during working days, the fitting line of class "EV is Charging" does not fully represent its general behaviour. The line equations are $y = -1.2178x + 4.7976$ and $y = 0.67816x + 1.4889$ with the norm of residuals of 26.06 and 24.97, respectively for the classes "EV is Charging" and "EV is Not Charging".

In figure 4.18 b), the peak hours fitting lines represent better the EV charging power profile, the *phantom charges* are not that relevant. The line equations are $y = 1.7898x + 0.73833$ and $y = 0.90237x + 4.3082$ with norm of residuals of 68.37 and 83.49, respectively for classes "EV is Charging" and "EV is Not Charging".

4.4 Active and Reactive Power Time Derivative

Based on the temporal analysis of the time derivative of P and Q presented in section 3.1.1, and also to verify a third parameter to further disaggregate the two data *clusters*, a dP-dQ plot was assembled. This plot is represented in figure 4.19. The red points are all the dP-dQ points from the training set. The green ones correspond to a set of P-Q in which the beginning of charging was occurring.

Looking at figure 4.19, when an EV charging is starting (green), the derivative of P is higher (up to

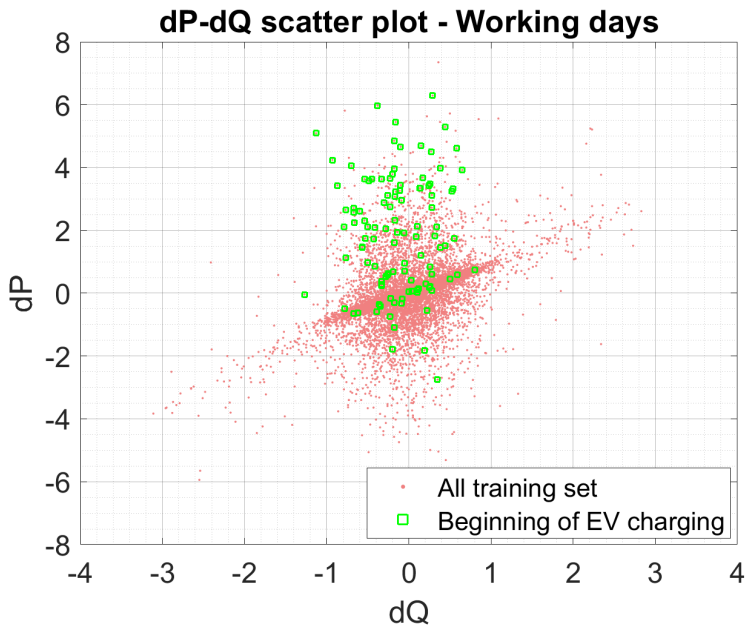

Figure 4.19: Representation of a dP-dQ scatter plot with all dP-dQ values in red and the derivative of the P-Q values of beginning of a charging in green, using the training set.

$\frac{dP}{dt} \approx 6$), while there is not much variation of the derivative of Q (up to $\frac{dQ}{dt} \approx 1$). It is possible to observe a semi-cluster. On the other hand, it is very difficult to disaggregate the samples corresponding to the beginning of an EV charging from the remaining. During the weekends, the results are similar but with less magnitude.

Moreover, there is uncertainty in what concerns the data extraction provided that, sometimes, the *timestamp* of sampled data from the Mobi.e® website is not properly synchronized with the DeepGrid® one in the sense that the sampling periods differ. This way, the P-Q pair corresponding to the beginning of an EV charging event might be one sample delayed or ahead when compared with the Mobie.e® one. With those uncertainties, this analysis will not be featured on the proposed EV charging classifier.

4.5 Chapter Conclusions

In this chapter, some data manipulation was conducted so that it was possible to obtain some patterns regarding the EV charging. The main tool used to realize these patterns was the P-Q scatter plot, described in section 4.1.

Several different data aggregations were tested, namely a weekday aggregation and a working day/weekend aggregation. In the weekday aggregation, the average value of P and Q (subsection 4.2.1), the aggregation of all P-Q samples (subsection 4.2.2) and a dual and triple time frame separation analyses (subsections 4.2.3) were conducted. In the working day/weekend aggregation, a grouping of all samples was conducted (subsection 4.3.1) as well as a dual time frame analysis (subsection 4.3.2).

Starting from the weekday aggregation, the assessment that presented more feasibility is the one in which all samples are used, without any time differentiation. When computing an origin constrained

fitting line, the average angle difference between both classes is 8.22^o with a standard deviation of 1.48^o. The P-Q average analysis was not feasible, providing that there was practically no differentiation between classes, as well as the time frame analysis in which during *Off-Peak* hours, there were more *phantom charges* than actual ones, making such differentiation unreliable.

In the working day/weekend analysis, the aggregation of all samples without time differentiation provided the best result. In fact, with such data aggregation, existent *phantom charges* are in less number, having less relevance when computing the fitting lines that will be used to build the classifier. The following phenomena occurs in both working days as well as weekends. The dual time frame analysis is not as effective for the same reasons as the ones having a weekday aggregation, namely the fact that phantom charges have a great impact when computing the fitting lines.

Furthermore, the total norm of the residuals using the unconstrained linear fitting is 177.36 and 165.15, respectively for classes ”EV is Charging” and ”EV is Not Charging” for the weekday aggregation For the working day/weekend aggregation, the norm of residual is 179.63 and 168.103 for each class. These results demonstrate that the weekday aggregation and the working day/weekend aggregation have similar error.

Moreover, the data necessary when constructing the weekday aggregation is 5 times greater during working days and 2 times greater during weekends than the one needed when using the working day aggregation. With more data, higher costs are associated, namely the storage of it and the down time of not predicting EV charges. In addition, given that the amount of time in which the sampling is occurring, training the classifier takes 5 times more time with the weekday aggregation, to have the same number of samples as the working day/weekend aggregation.

Knowing this, the chosen aggregation to build the baseline solution of the EV charging classifier was the one using all samples aggregated by working day/weekend.

In what concerns the dP-dQ scatter plot (in section 4.4) - a derivative of the active and reactive power - it was not considered feasible and will not be used in this work.

Chapter 5

Power Factor Signature Analysis

5.1 Theoretical Overview

In this section, a theoretical description of the proposed algorithm is made. This algorithm consists of binary classifier of EV charging events by disaggregation of the EV charging power load as seen from the secondary of the distribution substation. This classifier comprehends a **supervised learning** algorithm [31] and some of the tools used to construct it includes the use of a P-Q scatter plot (already explained in section 4.1), linear regression of the data samples (section 5.1.2) and Gaussian likelihood curves (section 5.1.3). The classes that can be predicted by the proposed classifier are then ”EV is Charging” and ”EV is Not Charging” as stated in the previous chapter.

Initially, the data will be plotted in a P-Q scatter plot with an aggregation by working day/weekend without a time frame separation, as it was concluded in the previous chapter. Then, an unconstrained linear regression is done in order to obtain the statistical relationship between P and Q of both classes. In addition, a Gaussian distribution model (*Gaussian membership function*) is applied [32], with the mean values resulting from the linear regression previously obtained, for both classes, resulting in a likelihood associated with each class.

The days used to train the algorithm are the working days: 5 to 9, 12 to 16, 19 to 23 and 26 to 30 of March, 2 to 6, 9 to 13, 16 to 20, 23, 27 and 30 of April, 1 to 4 and 7 of May, 2018. The days in between were not represented either because it was a weekend or because there was no data available from the Mobi.e® website.

5.1.1 P-Q analysis overview

In this section, a brief introduction of the P-Q scatter plot in the context of the algorithm making is done, provided that an explanation of this tool was already performed in section 4.1.

Since the chosen data aggregation was the one including a working day aggregation without time frame separation, (subsection 4.3.1), a representation of the P-Q scatter plot that was used in this chapter is the same as the one shown in figure 4.11. That figure represents the plot of values of $P\,(kW)$ and $Q\,(kvar)$ obtained from the smart-meter when an EV is connected to the charger (in red) and when

it was not (in blue).

5.1.2 Data Fitting

In order to best fit both classes (”EV is Charging” and ”EV is Not Charging”) of P-Q points and represent them using a function, a linear regression of these data points was performed for both [31]. This regression is a polynomial with order 1 given the sparsity, volume of data and theoretical assumption of almost unitary power factor when it comes to EV charging [33].

To implement the linear regressions, the *least squares method* [34] was applied, resulting in an unconstrained linear fitting of the data. This method is a tool to solve an overdetermined system of linear equations $\vec{V}\vec{P} = \vec{Y}$ (equation 5.1) with degree 1 resulting in the expression $p(x) = p_1\,x + p_2$. $\vec{V}$ is the Vandermonde matrix of $mx2$ with m being the number of points to fit and $x_{i=1}^{m}$ the reactive power values, $\vec{P}$ is the matrix $2x1$ with the requested coefficients and $\vec{Y}$ a matrix $mx1$ with the active power values.

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{5.1}$$

To obtain $\vec{P}$, one must solve $\vec{P} = \vec{V}(\vec{Y})^{-1}$ and to do that, the *QR decomposition* [34] was performed in which $\vec{V} = \vec{Q}\vec{R}$ (equation 5.2), being $\vec{Q}$ and orthogonal matrix and $\vec{R}$ being an upper triangular matrix.

$$\begin{bmatrix} \vdots & \vdots \\ v_1 & v_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ u_1 & u_2 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} r_1 & r_2 \\ r_3 & r_4 \end{bmatrix} \tag{5.2}$$

Furthermore, to obtain the $\vec{Q}$ matrix, the *Gram-Schmidt process* is used [34, 35] presented in equations 5.3 and having this matrix, the R is obtained through equation $\vec{R} = \vec{Q}^T\vec{V}$.

$$\begin{cases} u_1 = \frac{1}{\|v_1\|v_1} \\ u_2 = v_2 - (u_1\,v_2)\,u_1 \end{cases} \tag{5.3}$$

Having $\vec{Q}$ and $\vec{R}$, we are in the condition of obtaining matrix $\vec{P}$:

$$\vec{P} = \vec{R}\,(\vec{Q}^T\vec{Y})^{-1} \tag{5.4}$$

To further illustrate the line fitting in the scatter plot in the working days of the set designated above, figure 5.1 concerns the unconstrained linear fitting of P-Q samples of classes ”EV is Charging” and ”EV is Not Charging”. The linear regressions obtained were: $y = 0.98492x + 5.8371$ and $y = 1.4426x + 1.189$ with norm of residuals 160.41 and 141.11, respectively of classes ”EV is Charging” and ”EV is Not Charging”.

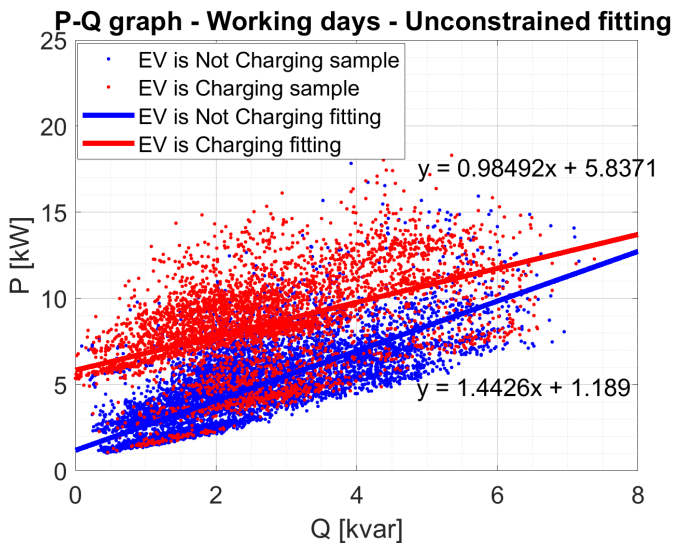

Figure 5.1: P-Q scatter plot of all samples when an EV charging was occurring (red) and when an EV charging was not occurring (blue) during working days with unconstrained linear fitting.

5.1.3 Gaussian Distribution Function

With the final objective of obtaining a likelihood of charging and of not charging, the next step of the algorithm computes the Gaussian distribution curve of each data class. From that, a comparison between both ("EV is charging" and "EV is not charging") can be established so that the classifier can make a prediction.

The mean value, μ, of each of these Gaussian distributions is given by the linear regression [31] of the data, with the standard deviation, σ, being given by the Cartesian distance between each data point (P-Q pair) and the correspondent cluster fitting lines.

To obtain the Gaussian function evaluated in each point over the fitting line, the *Gaussian membership function* [32] was used (equation 5.7). This function comes from the Gaussian probability density function (PDF), corresponding to the likely probabilistic state of belonging to a particular class. It has a maximum value of one, at the mean, and having the Gaussian computed from the data corresponding to "EV is Charging", and the one corresponding with the "EV is Not Charging", it will be possible to obtain a likelihood associated with each class, given a certain P-Q sample.

For each class, the distance between each sampled point (Q_i, P_i) and the fitting line of all data points $(ax + by + c = 0)$ is given by equation 5.5, with i being the index of the point.

$$d_i = \frac{\mid aQ_i + bP_i + c \mid}{\sqrt{(a^2 + b^2)}} \tag{5.5}$$

The standard deviation σ_{GMF} is obtained by equation 5.6, being n the number of total samples P-Q.

$$\sigma_{GMF} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{5.6}$$

Using (5.5) and (5.6), the *Gaussian membership function* is defined by equation 5.7, with d being the distance from a P-Q sample to the class line fitting. In this specific formula, the mean, μ, is always 0

because d is computed by correlating the distance to the linear data fitting - Gaussian mean.

$$f_{GMF}(d) = e^{\frac{-(d-\mu)^2}{2\sigma^2_{GMF}}} \tag{5.7}$$

Figure 5.2 is representative of the application of the Gauss membership function using the unconstrained data fitting (figure 5.1), with the previously mentioned data set. Now, we are in the condition of obtaining the likelihood of a certain P-Q sample belonging to the class "EV is Not Charging" (blue) and "EV is Charging" (red). The baseline solution to the problem addressed by this thesis - detection and classification of an EV charging occurrence - is then obtained by direct comparison of the likelihood of a given P-Q sample associated with the two Gaussians.

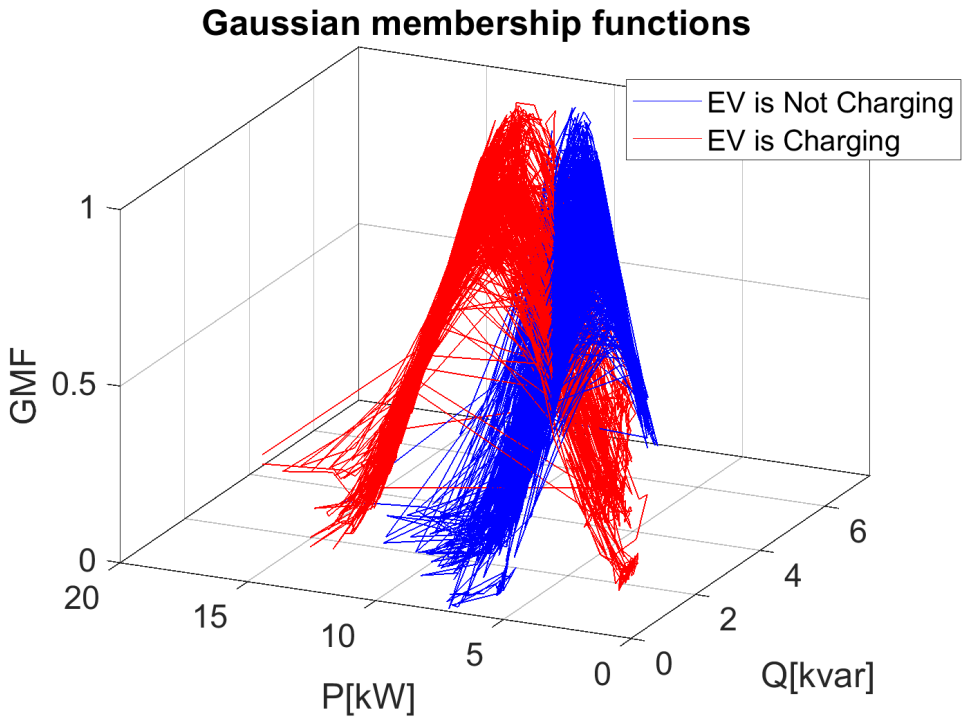

Figure 5.2: Representation of a Gaussian Membership Function for classes "EV is Not Charging" and "EV is Charging" using unconstrained data fitting as the mean value.

5.2 EV charging validation

In order to effectively predict an extracted sample as belonging to the class "EV is Charging" or "EV is Not Charging", different algorithmic optimizations were formulated and further applied to the dataset (both the training and testing). In the end, each of these optimizations will correspond a different classifier.

In this matter, having a sample $x = \{Q, P\} \in \mathbb{R}^2$ and wanting to predict its class $y = \Omega, \Omega = \{Y, N\}$, being Y the positive prediction - "EV is Charging" - and N the negative one - "EV is Not Charging" - each classifier optimization function will be denoted as $\hat{y}_i = f_i(x)$, with $i = \{A, B, C, D\}$. Moreover, the GMF likelihood associated to the class "EV is Charging" is denoted as L^C_{GMF} and the likelihood associated with the class "EV is Not Charging" is denoted as L^{NC}_{GMF}.

The optimizations conducted comprise the baseline solution - A: Without temporal filter, binary - and further derivations from there on, namely B: With temporal filter, binary; C: Without temporal filter, weighted and D: With temporal filter, weighted. Furthermore, some of these conjugate the correlation of P and Q with a time analysis or/and with a threshold establishment as additional tuning features explained in more detail in the following sub-sections.

5.2.1 A: Without time filter, binary

This optimization comprehend the baseline solution with which the classifier predicts whether an EV charging is occurring or not. In this matter, this optimization is indicative of the correlation degree existent between P and Q samples alongside its associated Gaussian curves.

For every sample of the test set (aggregated P-Q pair), the corresponding likelihood in each of the two Gaussians is obtained. If the L^{C}_{GMF} - GMF likelihood associated with the Gaussian "EV is Charging" - is greater or equal to the L^{NC}_{GMF} - GMF likelihood associated with the Gaussian "EV is Not Charging" - than the classifier will predict a positive output - "EV is Charging" class. On the contrary, when L^{C}_{GMF} its less than L^{NC}_{GMF}, the output will be negative - "EV is Not Charging" prediction.

$$\hat{y}_A = f_A(x) = \begin{cases} Y, & if\, L^{C}_{GMF} \geq L^{NC}_{GMF} \\ N, & otherwise \end{cases}$$

This way, this optimization is called "Without temporal filter, binary" because it does not feature a temporal analysis nor a threshold parameter. A prediction of EV charging occurrence is irrespective of the GMF likelihood difference between each of the Gaussians as long the likelihood of the class "EV is Charging" is greater or equal than the likelihood of the class "EV is Not Charging", constituting the baseline solution of the proposed classifier.

5.2.2 B: With temporal filter, binary

This algorithmic optimization relates the previous non-chronological analysis with the temporal evolution of the P-Q samples with respect to the GMF likelihood of the two classes - "EV is Charging" and "EV is Not Charging". Explicitly, this optimization will validate a set of consecutive "spikes" of power active, in the prediction, as well as its possible "deeps" with respect to its likelihood of charging.

The set of "spikes" will be accepted by the classifier as belonging to class "EV is Charging" only when the total time duration is greater than a pre-established window time threshold. In our case, it has been considered 20 minutes, which corresponds to 4 consecutive samples, given that the sampling time is 5 minutes.

To exemplify, in figure 5.3 a), given that the number of consecutive "spikes" is greater than 4, those samples are validated as belonging to class "EV is Charging". In figure 5.3 b), the number of consecutive "spikes" is less or equal than 4, validating those samples as belonging to class "EV is Not Charging".

Similarly, "deeps" represent a prediction made by the classifier of class "EV is Not Charging" whose total time duration is equal to 5 minutes (1 sample). In figure 5.4 a), given that the number of "deeps"

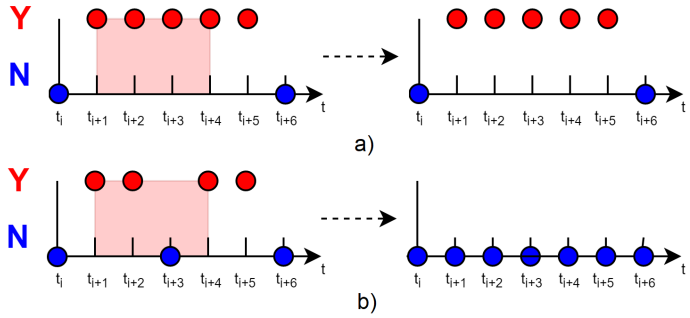

Figure 5.3: Representation of the validation of ”spikes” in two different scenarios.In a), samples validated as ”EV is Charging” and in b), samples validated as ”EV is Not Charging”.

samples is only 1, that sample is validated as belonging to class ”EV is Charging”. In figure 5.4 b), the number of consecutive ”deeps” is greater than 1, validating those samples as belonging to class ”EV is Not Charging”.

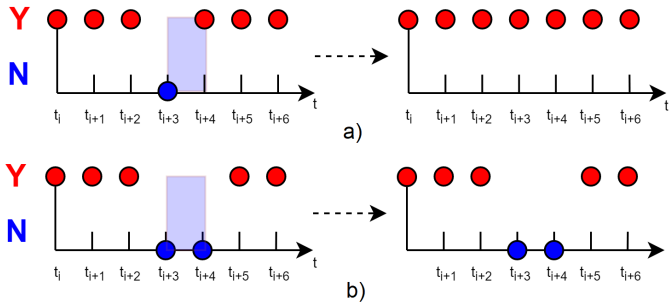

Figure 5.4: Representation of the validation of ”deeps” in two different scenarios. In a), sample validated as ”EV is Charging” and b), samples validated as ”EV is Not Charging”.

The ”deeps” validation proceeds the ”spikes” one.

Using this procedure, uncertainty in the prediction is suppressed, with the initial set of assumptions being that there are few EV charges whose time duration is below or equal to 20 minutes and the minimum time between EV charges is greater than 5 minutes. In order to decide upon this threshold value to use in this optimization, the histogram of EV charging duration during week-days (figure 5.5) was used.

Looking at the histogram in figure 5.5, it is possible to verify that most EV charges occurred with a time duration between 0 and 10 minutes, corresponding to roughly 5% of the total number of charges. Surprisingly, this bin is the one with most occurrences.

Now, to understand the impact of these occurrences in the power values, an individual inspection of the P and Q variation was performed. Some examples of these variations are presented in figure 5.6 as well as in appendix, figure A.1.

From these example figures, it is possible to visualize that there is practically no influence in P and Q when there is an indication of charges with a time duration below 10 minutes. This can be because the EV battery was already charged. From the same figures, it is possible to observe that around the beginning of an EV charging (other than the ones with 10-minute time duration), P increases further demonstrating the infeasibility of the previous results. Also in appendix , figure A.2, one can visualize

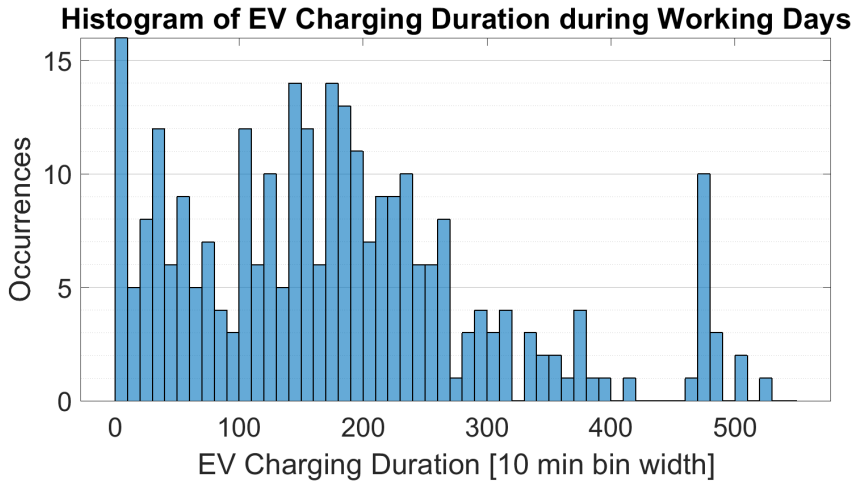

Figure 5.5: Histogram EV charging duration during working days with 10 minutes bin width.

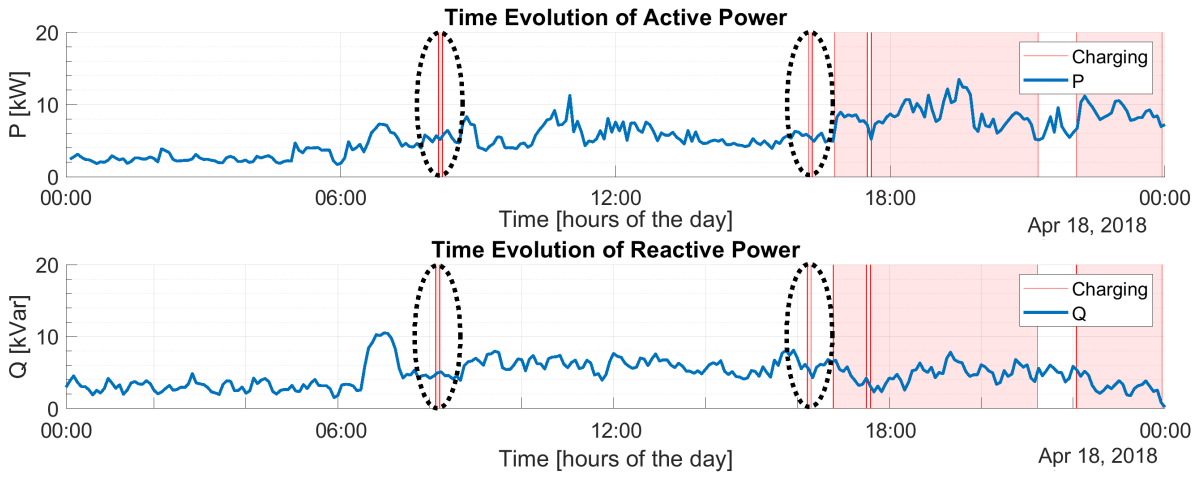

Figure 5.6: P and Q variation in 18/04/2018 with EV charging validation. EV charges with less than 10 minutes highlighted.

EV charging with time duration above 10 minutes and below 20 minutes, in which the variation of P and Q is very small. Following this reasoning, the previously stated assumption in which EV charging events with a duration below or equal to 20 minutes may not be accountable for prediction, is accepted.

The "spikes" expression ($\hat{y}_{B1}$) is represented in the 5.8 and proceeding from that, the "deeps" one ($\hat{y}_{B2}$) is expressed in 5.9.

$$\hat{y}_{B1} = f_{B1}(x) = \begin{cases} Y, & if\,[L^{C}_{GMF} \geq L^{NC}_{GMF}]\,AND\,[\Delta t_{(L^{C}_{GMF} \geq L^{NC}_{GMF})} \geq 20\,minutes] \\ N, & otherwise \end{cases} \tag{5.8}$$

$$\hat{y}_{B2} = f_{B2}(x) = \begin{cases} N, & if\,[L^{NC}_{GMF} > L^{C}_{GMF}]\,AND\,[\Delta t_{(L^{NC}_{GMF} > L^{C}_{GMF})} \geq 5\,minutes] \\ Y, & otherwise \end{cases} \tag{5.9}$$

This optimization - B: With temporal filter, binary - deepens the conducted research by correlating the purely analytical relationship between two variables P and Q when it comes to the charging of an EV

with the time variation of each charging, granting an additional tuning feature for this detection model, resulting in another classifier. It is also binary because it does not consider a threshold factor, when it comes to the GMF likelihood.

5.2.3 C - Without temporal filter, weighted

This optimization is again a purely relational analysis between P and Q, however, it consists in assigning a threshold value (other than 0), in terms of the difference between GMF likelihood associated with each Gaussian. The EV charging prediction is validated not only if $L^{C}_{GMF} \geq L^{NC}_{GMF}$ but if $L^{C}_{GMF} \geq T + L^{NC}_{GMF}$, T representing the chosen threshold. This threshold must be so that *true positive* and *true negative* predictions are still maximized and *false positive* and *false negative* predictions minimized.

To visually understand how the difference between the two GMF likelihoods is correlated with the *true positive* (TP) predictions [36], the graphic in figure 5.7 is presented. This graphic comprehends an analysis of the test data using the days 20, 23 to 27 and 30 of April, 1 to 4, 7 to 11, 14 to 18, 21 to 25, 28 to 31 of May, 1, 4 to 8, 11 to 15, 18 to 22, 25 to 29 of June, 2 to 6, 9 to 13, 16 to 20, 23 to 26 of July and 25 of September.

The number of *positive predictions* was determined by the classifier (in blue) - using the baseline solution: A - and out of those, the *true positive predictions* were obtained (in orange). This results are presented in an histogram in figure 5.7.

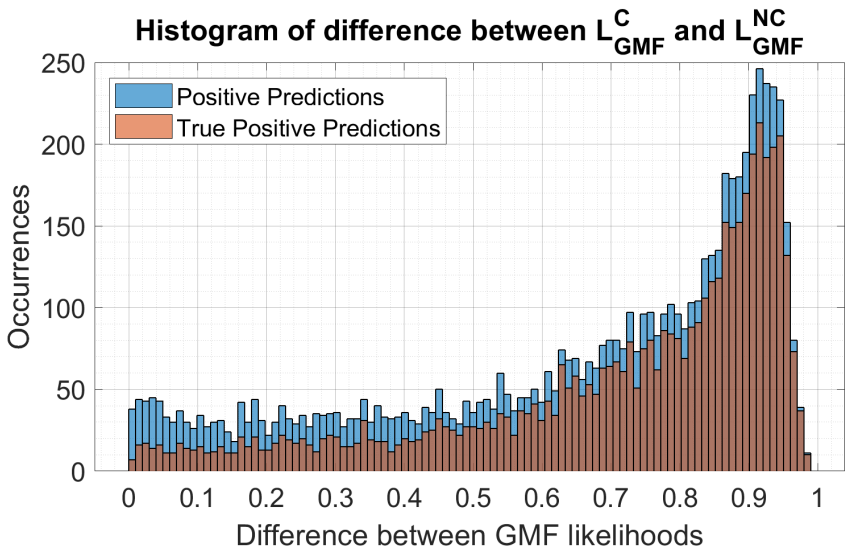

Figure 5.7: Histogram representing the number of positive and true positive occurrences with different likelihood difference between Gaussians "EV is Charging" and "EV is Not Charging".

Figure 5.7 indicates that, on average, with an increase in the likelihood difference between Gaussians, the number of TP samples increases relative to the *positive* predictions. When this difference is less than 0.1, the total percentage of *true positive* predictions out of the predicted as *positive* ones by the classifier, is only 36.9%. When the difference is greater than 0.1, the total percentage of *true positive* predictions out of the predicted as *positive* ones by the classifier, is 76.8%.

This way, the used threshold when making the classification will be 0.1, meaning that the likelihood of

belonging to the class ”EV is Charging” has to be at least 0.1 higher when compared with the likelihood of belonging to the ”EV is Not Charging” one. The optimization expression is presented in 5.10.

$$\hat{y}_C = f_C(x) = \begin{cases} Y, & if\, L^C_{GMF} \geq (0.1 + L^{NC}_{GMF}) \\ N, & otherwise \end{cases} \tag{5.10}$$

This optimization - C: Without temporal filter, weighted - does not use the temporal filter explained in sub-section 5.2.2 and makes use of a corresponding threshold validation factor of 0.1 between GMF probabilities in order to make a positive prediction.

5.2.4 D - With temporal filter, weighted

This sub-section develops the optimization - D: With temporal filter, weighted - merging the two previous ones B and C, using both the temporal filter as well as the threshold factor. Remembering, a number of consecutive samples, less or equal to 4, are considered ”spikes” and are labelled as ”EV is Not Charging” and ”deeps” as ”EV is Charging” and the relation 5.10 must hold also.

The ”spikes” expression is represented in 5.11 and proceeding that, the ”deeps” one is expressed in 5.12.

$$\hat{y}_{D1} = f_{D1}(x) = \begin{cases} Y, & if\, [L^C_{GMF} \geq (0.1 + L^{NC}_{GMF})]\, AND\, [\Delta t_{(L^C_{GMF} \geq 0.1 + L^{NC}_{GMF})} \geq 20\, minutes] \\ N, & otherwise \end{cases} \tag{5.11}$$

$$\hat{y}_{D2} = f_{D2}(x) = \begin{cases} N, & if\, [(L^{NC}_{GMF} + 0.1) > L^C_{GMF}]\, AND\, [\Delta t_{(L^{NC}_{GMF} + 0.1 > L^C_{GMF})} \geq 5\, minutes] \\ Y, & otherwise \end{cases} \tag{5.12}$$

Moreover, it is expected that the true positive rate of this optimization [36] be lower than the other ones, given the imposed constraints on the prediction. However, it is also expected that the specificity rate - samples correctly classified as negative - will increase.

Figure 5.8 resumes in a flowchart the proposed EV charging classification algorithm.

For inspection purposes, an example containing a day belonging to the test set (24/07/2018) with its respective Mobi.e® validation is presented in figure 5.9 and contains the GMF likelihood of belonging to each class alongside the classification of EV charging using every one of these four different optimizations.

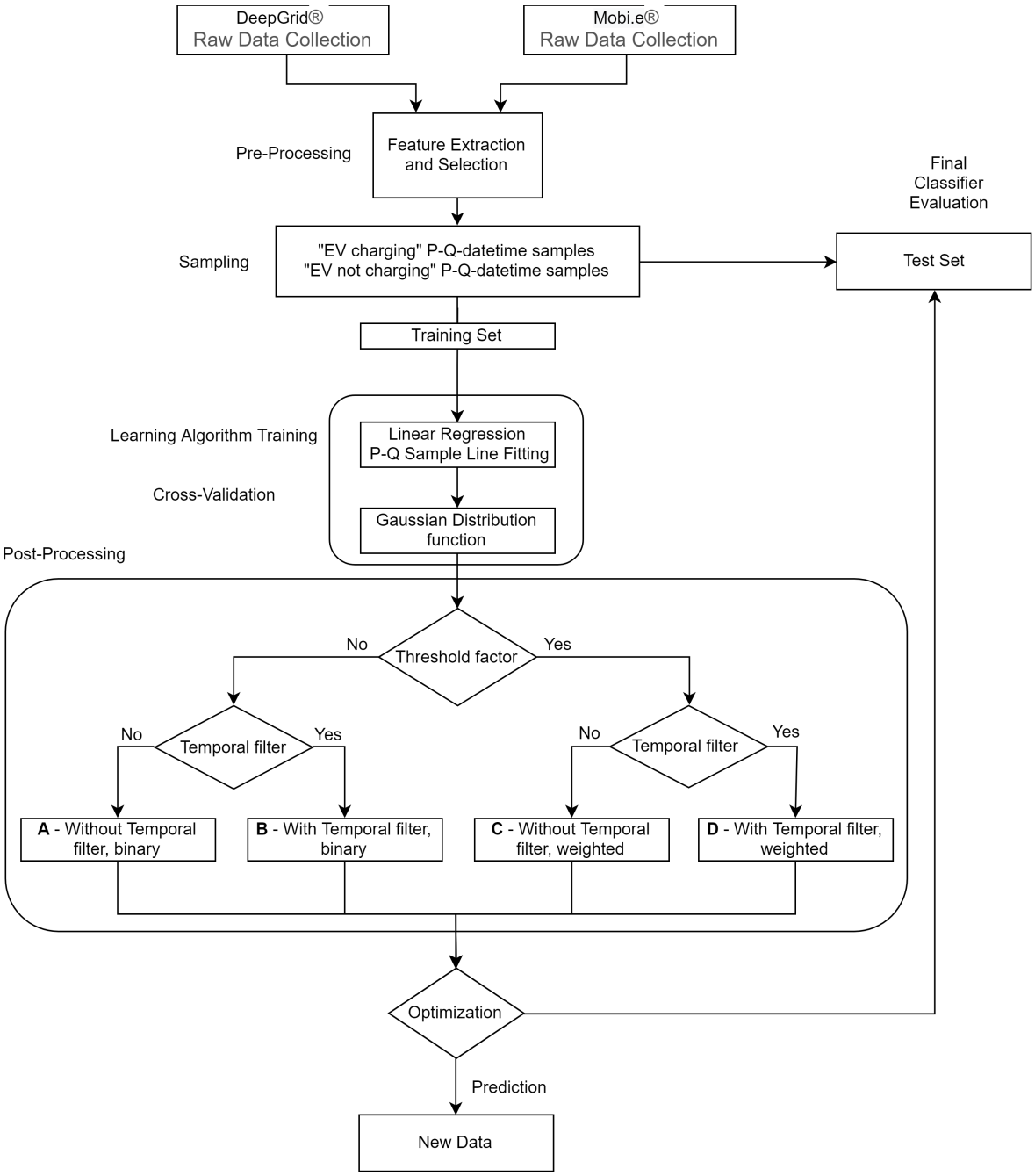

Figure 5.8: Flow chart of proposed EV charging classification algorithms.

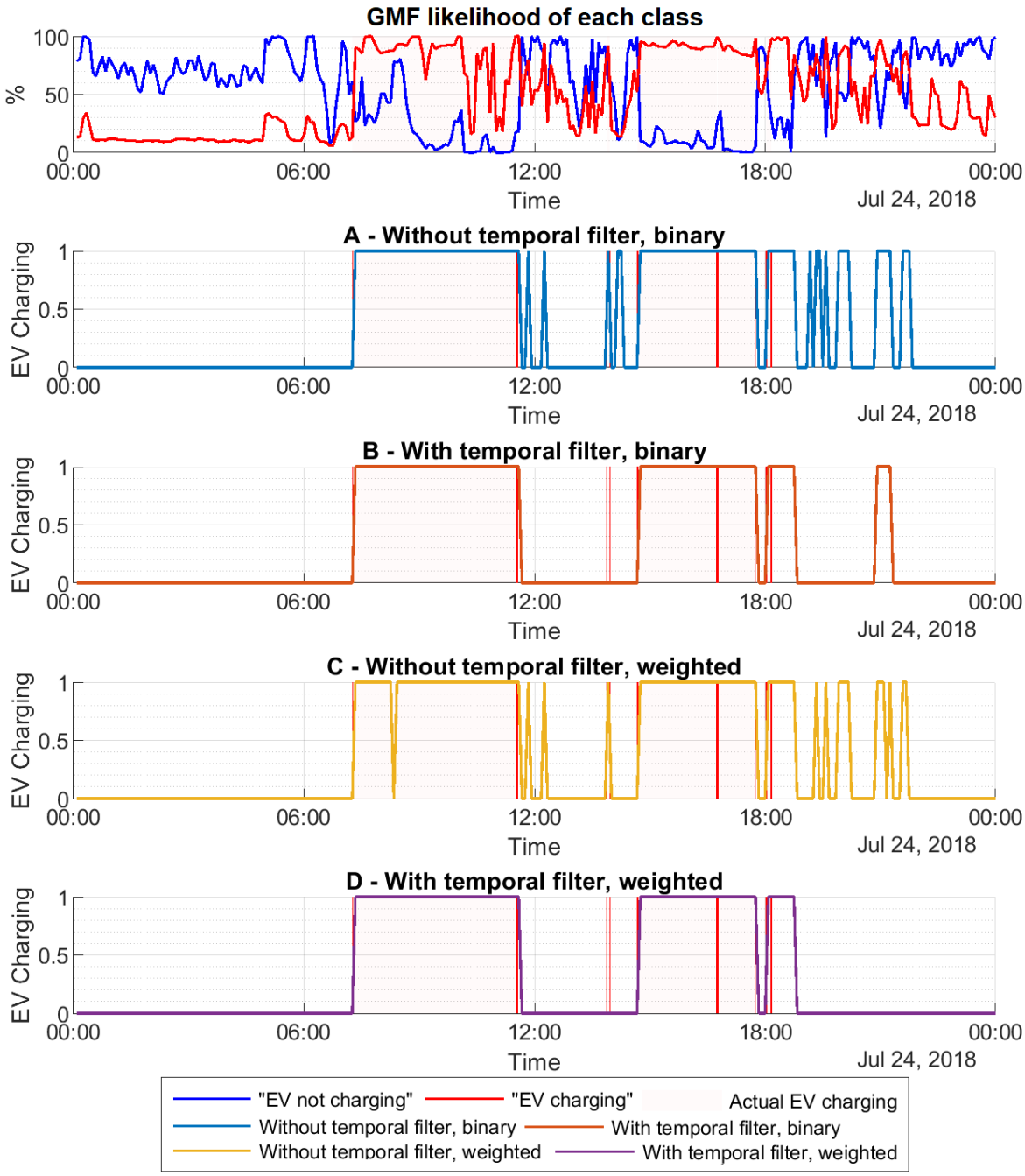

Figure 5.9: Visual representation of GMF Likelihoods in each Gaussian and corresponding optimizations of EV charging prediction in 24/07/2018 alongside actual EV charging validation.

5.3 EV Charging Profile

As previously stated in chapter 3.1, the total number of EV charges sampled and with which this classification model was built was 365, with data collected from over 5 months using both Eneida's smart-meter, in the secondary of the distribution substation, as well as the Mobi.e® website, with real-time information regarding the usage of EV public chargers. With enough data, a behaviour model of the car drivers using the EV charging station in study, can be obtained. Having then a more realistic approach, the proposed EV charging detection classifier can be optimized to increase its performance[31].

Taking into consideration the previous drawbacks, this section comprehends, a reverse-engineering method where additional data is created from previously analysed one. This will help to extend and reinforce the proposed research topic and extract more patterns in terms of this EV charger characterization.

The combination of historical power data with EV charging fictitious data, allows the training of the proposed classifier, requiring no validation of EV charging. In another words, it is possible to train the classifier having only historical data (where no EV charging occurs) and realistic, fictitious EV charging profile models. Moreover, these fictitious EV charges can also be used to select the "optimal" dataset with which one must train the proposed classifier if there is a data scarcity problem.

To test the fictitious profiles, days were no EV charges occurred are detached from their previous training or testing data set to ensure a certain degree of independence. Then, to these days, the fictitious EV charges are added, increasing the reliability of this assessment.

In this matter, the used days where no EV charges are recorded are 30 of March, 23, 27 and 30 of April, and they will be used as background load with which EV fictitious charges are going to be inserted. In order to create these fictitious EV charges, a statistical analysis on some features of actual EV charging from our dataset was performed and some probabilistic models were derived. The used statistical features, based on the histograms obtained, were the following:

- Number of EV charging events per day - section 5.3.2;
- Duration of EV charging - section 5.3.3;
- EV charging, starting time of the day - section 5.3.4.

Some of these features are not easily modelled and, as explained in section 5.3.1, the *Gaussian Mixture Model* [19, 37] algorithm was used.

Furthermore, a constraint had to be taken into consideration, given that it is not possible to extract the State of Charge (SoC) value of the EV being charged, and correlate it with the P demand - neither the Deepgrid® nor the Mobi.e ® website provide that information, with this being some possible future work. This way, the assumption taken into account is that the active power demand stays constant and does not decay with the SoC increase. In reality, the P demand is constant until 85% of SoC, and decays to zero, from 85% to 100% of SoC. This is possible to observe in figure 2.5, in the Background (chapter 2). Also, the power factor of an EV charge is assumed as unitary, no Q power expression of EV charging.

After computing the PDF models of the previously stated features, and in order to effectively create a profile characteristic of EV charging for the location studied, the following steps were taken, based on literature [19, 37]:

1. Random selection of the number of charges per day (section 5.3.2), constrained to be greater or equal to 0 charges. For each charging occurrence, follow steps 2 to 5:
2. Random selection of start time of the day (section 5.3.4) constrained to be greater than 00:00h and less than 24:00h;
3. Random selection of duration of charge (section 5.3.3) constrained to be greater than zero;
4. In the case where there is more than one charge per day, if after establishing the duration of a charge there is an overlap with another previously computed charge, a new duration random sample must be obtained. Repeat step 3 until there is none EV charging overlapping.
5. Having created the time periods in which EV charges are occurring, attribute to each charging time interval an active power of 3.7 kW.

As it is explained in detail in section 6.6, this study can be used as an industry application solution of the proposed classifier. In this matter, whenever there is no validation data to ensure whether an EV charging was occurring or not, it is still possible to train the model so that, when an actual EV charges occurs, predictions can be realized, using the proposed classifier.

Henceforth, the Gaussian models with which the random fictitious EV charges were computed can be used for future work whenever it is needed to have an extrapolation of the consumers behaviour profile of that EV charging station.

5.3.1 Gaussian Mixture Model

Some EV charging metrics have distributions with complex shapes (duration, start and end of EV charging) and the chosen approximation to deal with that level of complexity and fit a good distribution model is the *Gaussian mixture Model* (GMM) probability density function. This modelling algorithm was based on literature [19, 37].

This probabilistic model is composed of a finite sum of Gaussian PDFs [38], equal to Z, each with respective weight (ω_i), mean (μ_i) and variance (σ_i^2). The integral of the pdf over the sampled space has to be one, the mixture weights must be so that $0 \leq \omega_i \leq 1$ and their sum must be equal to one ($\sum_{i=1}^{Z} \omega_i = 1$) [37]. In this matter, $f(y)$ represents the GMM which is created by the sum of L individual weighted Gaussian components, whose random variable Y [37] is defined in equation 5.13.

$$f_Y(y) = \sum_{i=1}^{Z} \omega_i f_{N(\mu_i, \sigma_i^2)}(y) \tag{5.13}$$

The ith GMM component PDF is presented in equation 5.14, with mean and variance of the random variable Y represented in equations 5.15 and 5.16, respectively [19].

$$f_{N(x,\mu_i,\sigma_i^2)} = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}} \tag{5.14}$$

$$\mu_Y = \sum_{i=1}^{Z} \omega_i \mu_i \tag{5.15}$$

$$\sigma_Y^2 = \sum_{i=1}^{Z} \omega_i (\sigma_i^2 + (\mu_i - \mu_Y)^2) \tag{5.16}$$

It is also necessary to formulate an estimation problem in order to obtain the parameters used in the GMM, namely each ith individual component weight (w_i), mean (μ_i) and standard deviation (σ_i), having previously determined the number of components (Z) [19, 37].

To accomplish that, the Expected Maximization (EM) algorithm takes as input: $\Gamma = \{\gamma : \gamma = \{\omega_i, \mu_i, \sigma_i, \sum_i\}_{i=1}^{Z}\}$. This algorithm iteratively computes the expectation (E) of the log-likelihood of the complete data using the current estimate for the parameters Γ, followed by a computation of those same parameters, maximizing (M) the previous log-likelihood [19, 31]. This procedure is executed until convergence is achieved.

In this matter, the algorithm EM consists of [31]:

- **Initialization**:

  $\hat{\Gamma}^{(0)}$ - Choice of initial estimation Γ parameters when $t = 0$;

- **E step (Expectation):**

$$U(\Gamma, \hat{\Gamma}^{(t)}) = E\{log\, f_N(Y|\Gamma)|\hat{\Gamma}^{(t)}\} \tag{5.17}$$

- **M step (Maximization):**

$$\hat{\Gamma}^{(t+1)} = arg\, max_{\Gamma}\, U(\Gamma, \hat{\Gamma}^{(t)}) \tag{5.18}$$

In this matter, the EM cycle is repeated until the stop condition is achieved (the error is below a pre-determined threshold) and it is guaranteed that the *Maximum Likelihood Estimation* (MLE) function increases in each iteration. However, it is not possible to guarantee convergence to the global maximum of the MLE function.

To choose between the number of Gaussian components L in each analysed feature, firstly an empirical observation of the feature histograms shape was taken into account alongside the EV charger location characteristics.

5.3.2 Number of EV charges per day

Starting from the number of EV charges per day, figure 5.10 contains a histogram with the number of EV charging occurrences per day. A Normal distribution was fitted to this histogram distribution using the

Maximum Likelihood Estimation (MLE) for the estimated value of the mean and the standard deviation knowing a sequence of observations $X = (x_1, ..., x_n)$ and the normal likelihood function [31]. This PDF model, represented in orange in figure 5.10, was used to probabilistic recreate the number of EV charging events per day that would realistically resemble the realistic behaviour of the users of that EV charger.

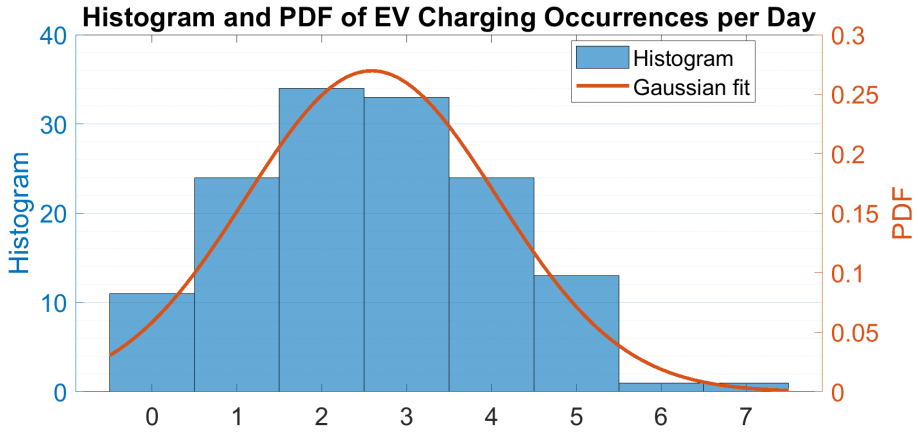

Figure 5.10: Histogram (in blue) and PDF (in orange) of EV charging occurrences per week day.

The estimated mean value for the number of charges per day was 2.58 with a standard deviation of 1.48. Naturally, when creating the EV charging profile, a round of each random sample extracted from this PDF must be conducted.

5.3.3 EV charging duration

A model that would realistically simulate the drivers behaviour regarding the EV charging duration was also constructed from the histogram of this feature using the extracted data. This histogram was shown before in figure 5.5 and as explained in section 5.2.2, the charging events with less than 10 minutes were not considered feasible. Hence, when probabilistic modelling this feature, the first bin (0-10 minutes) will be disregarded. However, the shape of this histograms curve is a rather complex one and does not follow any traditional distribution alone (normal, Weibull, Gamma, ...). Therefore, a mixture of Gaussian distributions appears to be a good choice given its simplicity (weighting, mean and variation for each), and knowing that the majority of computational tools can easily incorporate Gaussian PDFs [37]. In view of this and in order to have a realistic model of the drivers behaviour concerning EV charging duration, the Gaussian Mixture Model (GMM) [19] was used.

Figure 5.11 shows in blue the histogram of EV charging duration (excluding the first 10 minutes bin) as well as the PDF generated through the GMM and each individual Gaussian. For this simulation, the number of Gaussian components was $Z = 3$, and the initial Γ values, μ_i and σ_i, were randomly initialized. The weightings w_1, w_2 and w_3 are initialized with the same value between each of $1/3$. Moreover, the EM algorithm converged with error $\epsilon < 1\%$. The final Γ parameters - weighting, mean and variance of each GMM component - are presented in table 5.1.

This GMM fairly constitutes a model to statistically obtain EV charging duration random samples. It

Table 5.1: Γ parameters of EV charging duration GMM components.

$i^{th} component$	w_i	μ_i	σ_i
1	0.1341	39.3194	16.1568
2	0.7423	177.5260	69.1466
3	0.1236	427.7780	78.7451

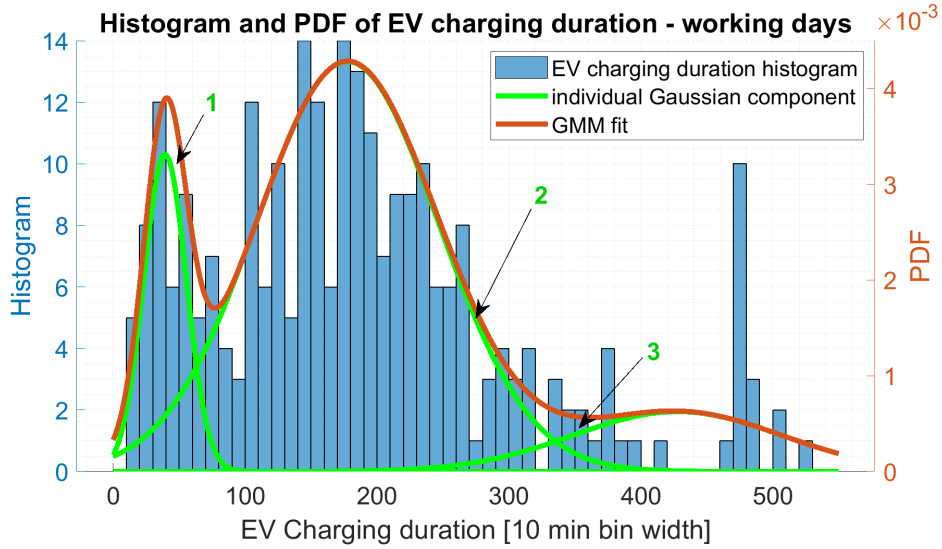

Figure 5.11: Histogram, in blue, and PDF, in orange, of EV charging duration occurrences. Individual Gaussians that compose the GMM in green.

is possible to establish three main charging duration periods, represented by the Gaussian components of the GMM, with mean around 40, 180 and 430 minutes, respectively.

5.3.4 Beginning of EV charging

Another feature considered to simulate EV charging events with respect to the active power demand was the beginning time of charging. Given the location of the EV charger in study, explained in section 3, as well as the histogram of daily beginning of charging, presented in figure 3.4, it was possible to verify that the charger belongs to an industrial area where most of the EV charges occur during the working day time. Furthermore, it is expected that pattern in the urban area. Here, the users will mostly start charging their EV when they arrive at work, possibly with a residential area nearby, given the rise in EV charging beginnings in the evening.

The histogram of these EV charging beginnings has a rather complex shape, forbidding the properly modelling with any of the traditional distribution functions. As in section 5.3.3, the GMM is used to provide a model that would realistically represent this feature. Figure 5.12 shows a representation of both the histogram of hourly EV charging beginnings in blue as well as the PDF of the computed GMM in orange. Again, the individual Gaussian components with which the GMM was computed are indicated in green. To construct the GMM, it was chosen a number of Gaussian components $Z = 2$. The initial Γ values μ_i and σ_i are, again, randomly initialized with the weightings w_1 and w_2 being initialized with the

same value between each of $1/2$. The EM algorithm converged with an error $\epsilon < 1\%$. Table 5.2 contains all the Γ parameters used in the GMM, for this feature.

Table 5.2: Γ parameters of beginning of EV charging GMM components.

$i^{th} component$	w_i	μ_i	σ_i
1	0.2330	8.3967	0.8271
2	0.7670	15.1401	3.7913

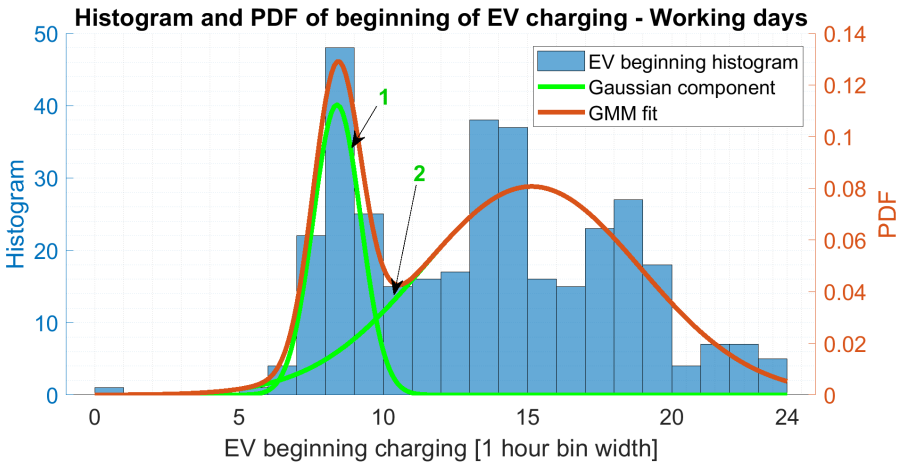

Figure 5.12: Histogram, in blue, and PDF, in orange, of beginning of EV charging. Individual Gaussians that compose the GMM in green.

Looking at figure 5.12 and by inspection, it is possible to understand the drivers behaviour regarding that EV charging station in the sense that the two most likely time periods are from 08:00h to 09:00h and 13:00h to 14:00h. These periods of time might as well represent the arrival at work in the morning and after lunch, respectively. Moreover, another time period with a great number of occurrences is from 18:00h to 19:00h, confirming that there are some residential areas nearby, indicating the time in which EV charger users arrive home.

5.4 Confusion matrix - Performance Assessment

In order to accurately assess the performance of this EV charging detection algorithm, it is imperative to have an appropriate classification metric. This way, the classification accuracy assessment will be made based on a *Confusion Matrix* (CM), otherwise known as the *error matrix*, which is the most commonly used tool [36]. Note that the purpose of this tool is to calculate the number of correct and incorrect predictions, summarize them in a matrix and extract relevant knowledge, namely the performance indices, from it.

The confusion matrix used has a dimension $2x2$ since our classification problem comprises two classes: "EV **is** Charging" and "EV **is Not** Charging". Furthermore, this matrix represents the existent relations between the output **predicted** by the classifier and the **actual/observable** output. In other

words, the confusion matrix columns correspond to the labels assigned by the classifier (prediction of EV charging event is positive, P', or negative, N') and the rows comprehend the test data (actual EV charging event is positive, P, or negative, N). The variable n corresponds to the total number of samples ($N' + P' = N + P = n$). This is illustrated in table 5.3 [36].

To each predicted sample, one of four different possible states can be assigned, corresponding to the matrix cells. The four states are:

- **True Positive** (TP): A charging was occurring, and the classifier predicted it was occurring;
- **True Negative** (TN): A charging was not occurring, and the classifier predicted it was not occurring;
- **False Positive** (FP): A charging was not occurring, and the classifier predicted it was occurring;
- **False Negative** (FN): A charging was occurring, and the classifier predicted it was not occurring;

Table 5.3: Representation of the confusion matrix.

		Prediction of EV charging event		
		Positive	Negative	Total
Actual EV Charging event	Positive	TN	FP	$TN + FP = N$
	Negative	FN	TP	$FN + TP = P$
	Total	$TN + FN = N'$	$FP + TP = P'$	n

Following the confusion matrix in table 5.3, several indices of classification accuracy can be derived [36], namely:

- *Accuracy* - overall, how often the EV charging detection is correctly classified: $\frac{TP+TN}{n}$;
- *Miss-Classification Rate* - overall, how often the EV charging detection is incorrectly classified: $\frac{FP+FN}{n}$;
- *True Positive Rate* (TPR) - how often does the classifier predict an EV charging event when it is actually occurring: $\frac{TP}{P}$.
- *False Positive Rate* (FPR) - how often does the classifier predict an EV charging is occurring when it is actually not occurring: $\frac{FP}{N}$.
- *True Negative Rate* / *Specificity* - how often does the classifier predict that an EV charging event was not occurring when in fact, it was not occurring: $\frac{TN}{N}$;
- *Precision* - when the classifier predicted an EV charging event, how often it was correct: $\frac{TP}{P'}$;
- *Prevalence* - how often there was an actual charging event in all the used data set: $\frac{P}{n}$.

With these 7 indexes, it is possible to have a more accurate depiction of the classifier performance. However, each of these indices alone could lead to misleading information. For example, the *accuracy* index alone is not a reliable one because the *prevalence* might be very different and having a high number of false positives in a set where the number of negatives is small can still result in a good overall accuracy.

Chapter 6

Results concerning the EV detection

In this chapter, the results of the proposed solutions for the EV charging classification will be presented and discussed, using the *confusion matrix* and its performance indices. To increase the impartially in what concerns the classifier performance results, a stratified 5-fold cross-validation [39] is done. Furthermore, the feasibility limits of the proposed classifier are tested in two distinct ways: with a variation of *background* load and with a variation of EV charging load.

As an additional feature, the results concerning the classifier performance with fictitious EV charges, statistically modelled from real data and explained in section 5.3, are presented and discussed. Moreover, section 6.6 features possible industrial applications of the proposed classification algorithm.

6.1 Problem Description

The data set (training and test set) with which the results presented in this section were obtained is composed of real data extracted both from the DeepGrid® IoT platform and from the MOBI.E® website. For the analysis, only working days of the week were considered, as further explained in Chapter 1.

The **training set** is composed of data from working days: 5 to 9, 12, 14 to 16, 19 to 23, 26 to 30 of March and from 2 to 6, 9 to 13 and 16 to 19 of April. This data is used to compute the GMF curves of classes ”EV is Charging” and ”EV is Not Charging”, explained in section 5.1.3. Having this Gaussian functions, each P-Q sample of the test set will have a correspondent GMF probability associated with each class and a prediction is then possible.

On the other hand, the **test set** is composed of data from working days: 20, 23 to 27 and 30 of April, 1 to 4, 7 to 11, 14 to 18, 21 to 25, 28 to 31 of May, 1, 4 to 8, 11 to 15, 18 to 22, 25 to 29 of June, 2 to 6, 9 to 13, 16 to 20, 23 to 26 of July and 25 of September.

The total ”useful” number of days between both data sets is 141 and the remaining ones are not a part of the training or test set because, either they correspond to weekends, or because data could not be extracted from the MOBI.E® website. Furthermore, the recorded EV charges are 292 (during working days) with this information in the data characterization section (3.1.1). Moreover, the *prevalence* level is of 30.8%, meaning the actual preponderance of EV charging events in the specific charger under

study. The *prevalence* is relatively small, hence it is then expected that the classifier will predict more samples from the class ”EV is Not Charging”, given the higher number of these samples. The Weekends results are presented in sub-section 6.3 and naturally comprehend a different set of days.

In the following section, the results concerning the EV charging detection are presented. Furthermore, multiple algorithmic optimizations (explained in section 5.2) are considered, specifically:

- A - Without temporal filter, binary;
- B - With temporal filter, binary;
- C - Without temporal filter, weighted;
- D - With temporal filter, weighted.

6.2 Confusion Matrix Results

The *confusion matrices* obtained for each one of the four different optimizations are presented, as well as the corresponding classification indices.

The final goal is to have the best possible EV charging event classifier. For that, a comparative analysis of these four different optimizations is presented, revealing its strengths as well as weaknesses in the context of the problematic addressed by this dissertation.

Optimization A - Without temporal filter, binary

For this optimization (A - Without temporal filter, binary), the associated confusion matrix is displayed in table 6.1, and the classification accuracy indices are displayed in table 6.2.

Table 6.1: Confusion matrix of EV charging events for optimization A - *Without temporal filter, binary*.

		Predicted		
		Negative	Positive	Total
Actual	Negative	12362	1689	14051
	Positive	1318	4938	6256
	Total	13680	6627	20307

Table 6.2: Confusion matrix classification indices for optimization A - *Without temporal filter, binary*.

Index	**value**
Accuracy	0.8519
Miss-Classification Rate	0.1481
True Positive Rate	0.7893
False Positive Rate	0.1202
Specificity	0.8798
Precision	0.7451

In this optimization and looking at tables 6.1 and 6.2, which corresponds to the baseline solution, the number of predicted samples N' and P' are, respectively, 13680 and 6627, confirming the hypothesis that there would be more predictions from class "EV is Not Charging". Following this reasoning, the *specificity* level is 88%, and the *true positive rate* is 79%, meaning that the classifier is more effective at predicting "EV is Not Charging" class.

Another important index is the *false positive rate*, indicating the *miss charges*, and it is in the range of 12%. The overall efficiency of this classifier is then given by the *accuracy* and *miss-classification rate*, respectively around 85.2% and 14.8%.

Optimization B - With temporal filter, binary

For this optimization (B - With temporal filter, binary), the associated confusion matrix is displayed in table 6.3, and the classification accuracy indices are displayed in table 6.4. Based on the explanation of the optimization in sub-section 5.2.2, and in short, a time filter is applied in which "spikes" - successive samples classified as "EV is Charging" for a period below or equal to 20 minutes - are further validated as "EV is not Charging". On top of that, existent "deep's" - successive samples classified as "EV is Not Charging" for a period below or equal to 5 minutes, when charging is occurring - are otherwise validated as "EV is Charging". This way, some of the previously classified samples will change its initial state, and the rectification of isolated predictions, with little meaning in the context of the EV charging panorama, is performed.

Table 6.3: Confusion matrix of EV charging event for optimization B - *With temporal filter, binary*.

		Predicted		
		Negative	Positive	Total
Actual	Negative	13263	788	14051
	Positive	1483	4773	6256
	Total	14746	5561	20307

Table 6.4: Confusion matrix classification indices for optimization B - *With temporal filter, binary*.

Index	**value**
Accuracy	0.8882
Miss-Classification Rate	0.1118
True Positive Rate	0.7629
False Positive Rate	0.0561
Specificity	0.9439
Precision	0.8583

In this optimization, table 6.4 indicates some significantly different results comparing with the baseline solution in 6.2, specifically in the *False Positive Rate*, which is now around 5.6%, and not 12%. Remember that this index is representative of the occurred *missed charges*. The *specificity* index is higher, around 94.4% (before was 88%), and the *miss-classification rate* is lower, around 11.2% confirming that

this classifier is good at predicting occurrences of the class ”EV is Not Charging”.

Again, contrasting with optimization A, the *true positive rate* is slightly lower, in the range of 76.3% (2.6% difference) indicating, as expected, that the classifier does not predict the occurrence of an EV charging event when it is in fact occurring, as good as in the previous optimization. This is due to the fact that there are more samples considered ”spikes” than ”deep’s”, thus decreasing the total number of positively predicted samples. This is possible to infer from the confusion matrix in Table 6.3 given that the number of negative prediction in optimization B is higher than in optimization A. The overall *accuracy* is around 88.8%.

Optimization C - Without temporal filter, weighted

Results of algorithmic optimization ”C - Without temporal filter, weighted”, explained in detail in sub-section 5.2.3, are now presented. Briefly contextualizing, the classification of a sample is predicted as positive - ”EV is Charging” - not only when $L^{C}_{gmf} \geq L^{NC}_{gmf}$, but also when the difference between each of these likelihoods is greater or equal than 0.1: $L^{C}_{gmf} \geq (0.1 + L^{NC}_{gmf})$. In other words, a threshold ($T = 0.1$) value is assigned to validate a prediction, and the temporal filter is not used.

This optimization’s confusion matrix is presented in table 6.5 and the proceeding performance indices in table 6.6.

Table 6.5: Confusion matrix of EV charging event for optimization C - *Without temporal filter, weighted*.

		Predicted		
		Negative	Positive	Total
Actual	Negative	12597	1454	14051
	Positive	1457	4799	6256
	Total	14054	6253	20307

Table 6.6: Confusion matrix classification indices for optimization C - *Without temporal filter, weighted*.

Index	**value**
Accuracy	0.8567
Miss-Classification Rate	0.1433
True Positive Rate	0.7671
False Positive Rate	0.1035
Specificity	0.8965
Precision	0.7675

Verifying the confusion matrix indices in table 6.6, the results are somewhat similar to the ones from optimization A (table 6.1). The overall classification *accuracy* is around 85.7% with a *miss-classification rate* of 14.3%. The *true positive rate* is the second higher (first is the one from optimization A - 78.9%), the *false positive rate* is around 10.4% and the *specificity* and *precision* are respectively 89.7% and 76.8%. Again, the TPR is not higher than optimization A, providing that the number of samples being predicted would always be less or equal, as it is possible to observe in figure 5.7 in sub-section 5.2.3.

Optimization D - With temporal filter, weighted

Lastly, the results of optimization D comprise the two previously formulated filters (temporal filter and threshold value). The obtained confusion matrix is displayed in table 6.7, and each of the derived classification indices are present in table 6.8.

Table 6.7: Confusion matrix of EV charging event for optimization D - *With temporal filter, weighted*.

		Predicted		
		Negative	Positive	Total
Actual	Negative	13427	624	14051
	Positive	1687	4569	6256
	Total	15114	5193	20307

Table 6.8: Confusion matrix classification indices for optimization D - *With temporal filter, weighted*.

Index	**value**
Accuracy	0.8862
Miss-Classification Rate	0.1138
True Positive Rate	0.7303
False Positive Rate	0.0444
Specificity	0.9556
Precision	0.8798
Prevalence	0.3081

These results are similar to the ones from optimization B, in part due to the fact that the same temporal filter is utilized to validate a prediction - validating ”spikes” as well as ”deeps” when it comes to the classification. The overall *accuracy* is around 89% and the *miss-classification rate* is 11%. Furthermore, the *true positive rate* is lower, around 73% but on the other hand, the *false positive rate* is the lowest from all of the optimizations, around 4%. Moreover, the specificity is around 96% and the precision is 88%.

6.2.1 Results Summary

Having all the confusion matrix indices from each of the proposed classifier optimizations, a comparative analysis is presented in this section. Moreover, an in-depth study of the advantages and disadvantages of each will be presented. Note that there are *phantom charges* in this simulation results.

Figure 6.1 is indicative of the CM indices evolution regarding each optimization.

Looking at the CM results in figure 6.1, it is possible to realize that, in fact, the *specificity* is fairly higher than the *true positive rate* (TPR), indicating that across optimizations (A to D), the proposed classifier has an apparent better performance at labelling samples from class ”EV is Not Charging” than from class ”EV is Charging”.

Now, looking at each optimization TPR score, with added filtering methods - temporal and threshold filter - the total number of TP samples decreases, indicating in the case of the temporal filter, that the

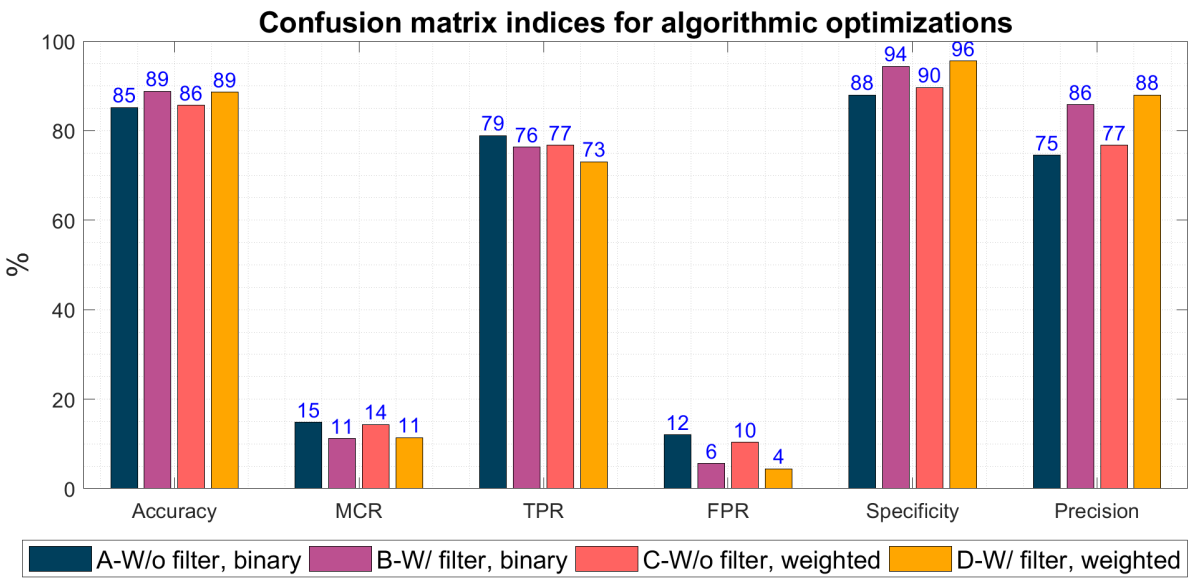

Figure 6.1: Overall display of the CM indices to evaluate the classifier performance of each of the optimizations, being MCR the *miss-classification rate*, TPR the *true positive rate* and FPR the *false positive rate*. Training and test set presented in section 6.1.

number of samples defined as ”spikes” is greater than the number of samples defined as ”deep’s”. Moreover, when the weighting filter (threshold) is applied, the TP samples are always less than when it is not, possible to see in figure 5.7 relating with the fact that the TPR of A is higher than C, and the TRP of B is higher than D, in figure 6.1.

On the other hand and with the addition of the previously mentioned filters, not only the TP samples decreased, as stated above, but also the FN ones leading to an improvement in the classification in what concerns the *precision*, also mirrored in both the FPR and in the *specificity*. Particularly, the FN predictions decrease substantially at the cost of less TP ones. Overall, the classifier accuracy improves, with the added filters.

With this information, the most advantageous optimization, if having the best TPR is more critical at the cost of relatively high FPR, is the one without temporal nor weighting filter - optimization A. In other words, an EV charging is more likely to be detected by the classifier at the expense of some inaccurate classifications, as represented by the FPR (miss-classifications), *specificity* and *precision*. Differently, when it is more important to have significantly higher *precision* and less *miss-classification*, followed by slightly lower TPR, both optimization B and D grant feasible results.

In the end, the B configuration is the one that best maximizes TPR and minimizes the FPR, even though the overall accuracy index is similar to D.

6.3 Cross-Validation

In this section, cross-validation [39] is used to evaluate and compare different algorithms more effectively but also to ensure that there is less bias in the used samples, assigning impartial training and test sets. Cross-validation can also be used to verify eventual *underfitting* or *overfitting* when building the classifier.

This way, the training and test set are crossed successively, so that each data sample can be validated against.

In what concerns this thesis, the used method was the stratified k fold cross-validation in which data is partitioned in k equally sized folds. Subsequently, in k successive iterations, the folds rotate in such a way that $k-1$ folders belong to the training set and the one remaining belongs to test set. Additionally, data is typically *stratified* prior to the fold attribution, meaning that there is a rearranging of the data in which the mean value is similar between folds, ensuring further neutrality.

Moreover, the k number used in this k fold cross-validation was of 5, meaning that the 101 days worth of data (during working days) will be almost equally split into $k = 5$ folds, as shown in figure 6.2, four of them with 20 days and the remaining one with 21 days. The number of iterations will be 5 and at each iteration there is a rotation with different training and test datasets. Figure 6.2 represents these rotations with $k = 5$. Note that there are *phantom charges* in this simulation results.

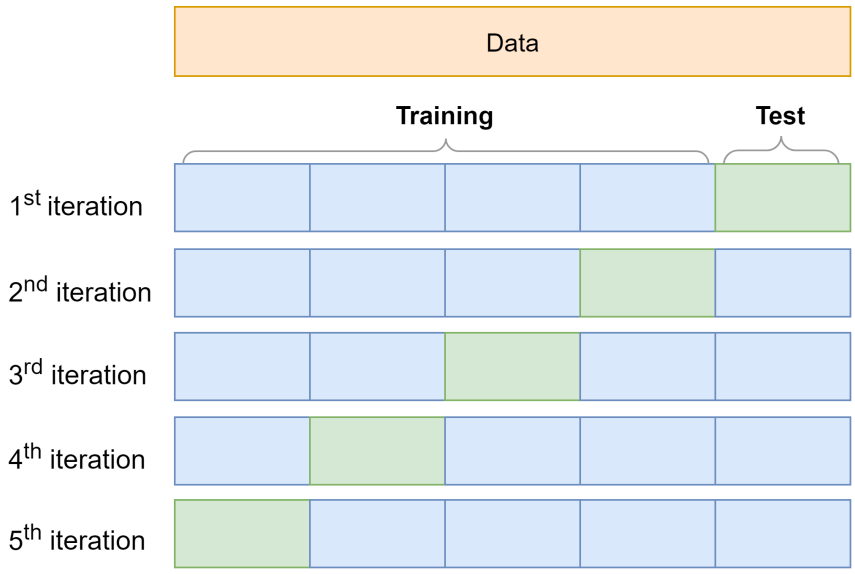

Figure 6.2: Representation of the k fold cross-validation algorithm with $k = 5$.

Cross-validation was then conducted consisting of 5 different simulations of the proposed classifier with different training and test sets. For each training-test verification, similarly to what was done in subsections 6.2.1, the CM was computed alongside its performance indices. After the 5th iteration, the average of all simulations was obtained, with its results presented in figure 6.3.

In this manner and looking at this results, by comparison with the ones presented in section 6.2.1, it is possible to perceive that they are similar in what concerns the overall *accuracy* and *Miss-Classification Rate*. The two of them are similar, respectively 2.5% on average between optimizations, in both indices. When looking at the *True Positive Rate*, the cross-validation displayed better results around 1.5% on average, between optimizations. Now, looking at the *False Positive Rate* (and subsequently the *specificity*) the cross-validation results demonstrate that the average score has a similar performance when compared with the results on subsection 6.2.1 with a difference of 3.5% on average between optimizations.

Furthermore, the precision has also a comparable score with a difference of 3.5%, in average across solutions.

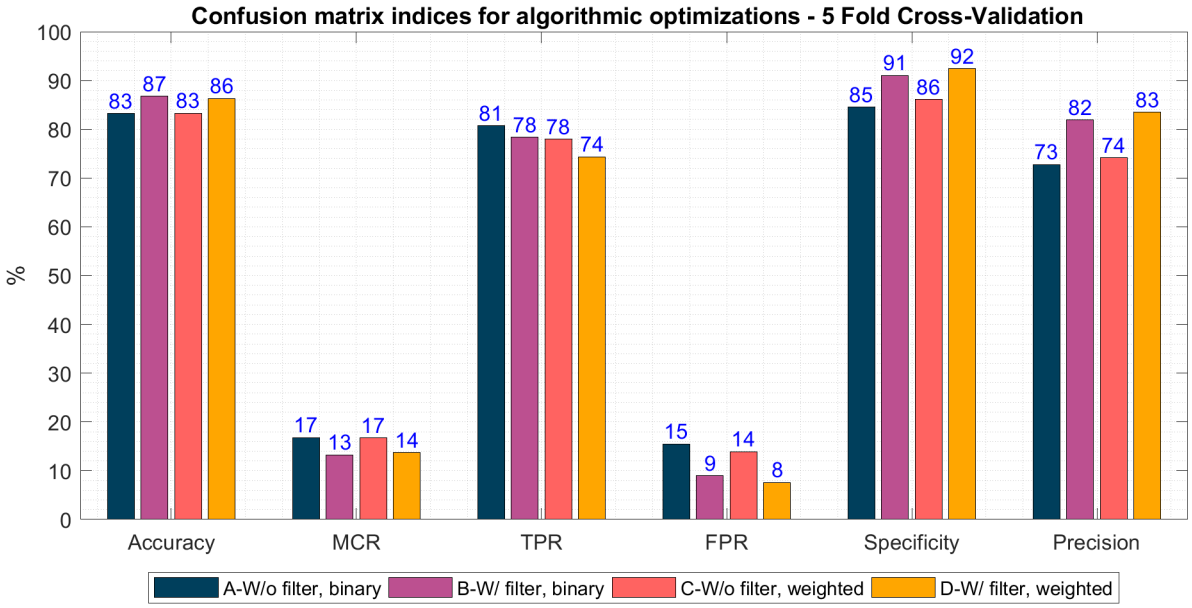

Figure 6.3: Confusion matrix indices results using the k fold cross-validation assessment.

Hereinafter, when addressing the performance of the proposed EV charging classifier, the results will be ones obtained trough cross-validation and disclosed in this subsection, given the unbiased generalisability it provides.

Weekends

The results of the weekends using the 5 fold cross-validation are also presented to extend the conducted research further. The used weekend days with which this analysis was performed are: 10, 11, 17, 18, 24, 25, 31 of March, 1, 7, 8, 14, 15, 21, 22, 28, 29 of April, 5, 6, 12, 13, 19, 20, 26, 27 of May, 2, 3, 9, 10, 16, 17, 23, 24, 30 of June and 1, 7, 8, 14, 15, 21, 22 of July of 2018.

Table 6.9 is representative of the results concerning the EV charging classification during weekends. In general and comparing with the analysis during weekdays, it is necessary to note that the *prevalence* of EV charging was of 25%, lower than during the working days.

Table 6.9: Confusion matrix performance indices EV charges during weekends.

Optimization	**CM performance indices(%)**					
	Accuracy	MCR	TPR	FPR	*Specificity*	*Precision*
A	82	18	76	16	84	61
B	87	13	73	9	91	74
C	83	17	72	14	86	64
D	87	13	68	7	93	76

Looking at table 6.9 and comparing these results with the working days, it is possible to conclude that they are similar in terms of overall *accuracy*. The *True Positive Rate* is greater during working days because the P-Q points have more dispersion (possible to see in figure 4.14) as well as because there

are more EV charging during working days. The *precision* level is also lower during weekends. This is because there are less samples being predicted, provided that there are less weekends than working days with the used *data lake*.

6.4 Feasibility limits of EV classifier

In this section, the limits of operation with which this classifier is able to effectively detect EV charges, are tested. In other words, how will the classifier perform when subjected to an increase in the line load as seen from the secondary of the transformer. Henceforth, the designation "*background load*" will be attributed to the load from the line in study, excluding the EV charging one (3.7 kW when an EV charging is occurring) - P_1^{BK} and Q_1^{BK}.

Furthermore, it is possible to divide this process into two different components thus testing two scenarios which are:

1. Scenario 1: multiplication factor K variation of *background* P *and* Q *load* a seen from the secondary of the distribution substation (P_k^{BK} and Q_k^{BK}), when the *background load* has different P and Q power magnitude. It is simulating the EV charging detection feasibility when, for example, load is added or removed to that line - $P_k^{BK} = (P_1^{BK} * K)\, kW$ and $Q_k^{BK} = (Q_1^{BK} * K)\, kvar$;

2. Scenario 2: multiplication factor K variation of P load of EV charging event, simulating different EV charging power modes (P_k^{CH} and Q_k^{CH}) with the same background load. For instance, instead of a *slow charge power mode*, as the one in study, how does the classifier behave when a *higher power mode* charger is installed.

The used constant multiplication factor K ranges from $K = [0.1,\, 5]$, both in scenario 1 and in scenario 2. To evaluate these scenarios the confusion matrix indices (section 5.4) were used, and similarly to the fictitious EV charging profile, some assumptions had to be established:

- EV charging event characterized by an unitary power factor ($PF = 1$);

- In scenario 1, the P in an EV charging event is constant and equal to $P_k^{CH} = 3.7\, kW$ - variation is in the background load and not in the EV charging load - $P_k^{BK} = (P_1^{BK} * K)\, kW$ and $Q_k^{BK} = (Q_1^{BK} * K)\, kvar$. In scenario 2, the EV charging load equals to $P_k^{CH} = (3.7 * K)\, kW$ - the variation is in the EV charging P load (hence the multiplication by K) with the same background load.

- The EV charging active power is constant throughout the charging, not decaying when the SoC is greater than 85%;

- Variation of P and Q of the background load is linear in scenario 1, i.e. both P and Q are multiplied by the same factor K;

6.4.1 Scenario 1 - Variation of Background P and Q Load in Transformer by Factor Multiplication

The load corresponding to the **EV charging station** in study (Alameda Armando Gonçalves 181, Coimbra, charger CBR-00004-02) does not correspond to the **total load associated with the line and feeder from the transformer** in which the data is being sampled by the smart-meter (SDS 556, feeder 6 and line 2). By looking at the sampled data (exemplified in figures 3.6, 3.7, 3.9 and 3.8), it is possible to observe the existence of *background load* and not just the load associated with the EV charger.

In this scenario, the performed assessment will aim to realize the limits with which this classifier is able to operate when the *background load* varies by a factor K - $P_k^{BK} = (P_1^{BK} * K)\,kW$ and $Q_k^{BK} = (Q_1^{BK} * K)\,kvar$. To exemplify a real-world case in which this phenomenon might occur, another load might be added or there might be an increase of the P and Q in existing load connected to the line and feeder in use.

Figure 6.4 represents the variation of the confusion matrix indices with a factor multiplication K of the background load, namely: *accuracy*, *miss-classification rate*, *true positive rate*, *false positive rate*, *specificity* and *precision*. In this simulation, both the training as well as the test set are computed having done the *background load* multiplication by K, and using a 5 fold cross-validation.

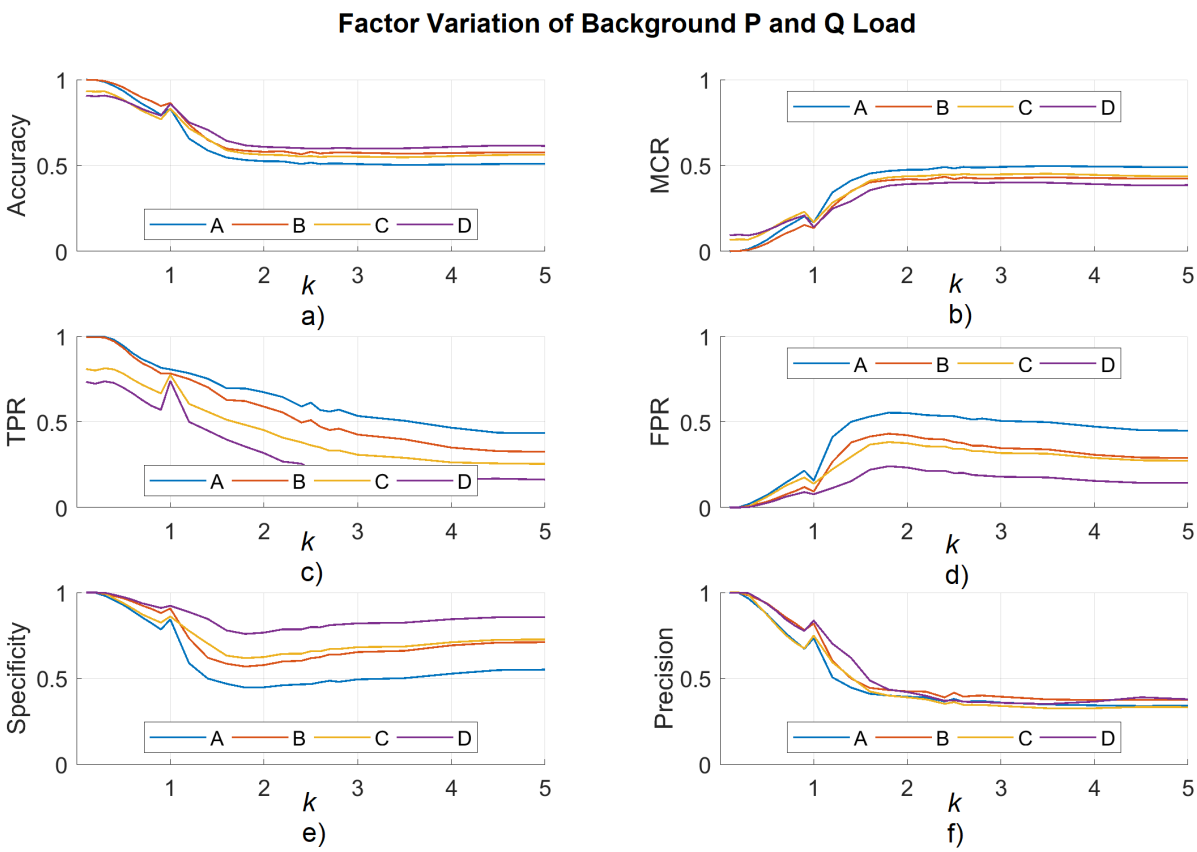

Figure 6.4: *Accuracy* (a), MCR (b), TPR (c), FPR (d), *specificity* (e) and *precision* (f) for *k* variation of background P and Q load from the secondary of the distribution substation, for each algorithm A to D.

From figure 6.4 and looking at each of the CM indices, it is clear how the classifier behaves when

there is both an increase as well as a decrease in the background load.

When K factor is above 1 and looking at the *accuracy* index, with an increase in the *line background power load*, classifier performance drops. Supporting this overall decrease in performance, the TPR decreases with a K increase, roughly stabilizing with the background load multiplication factor $K = 4$. Moreover, both the *specificity* and the *precision* drop significantly, stabilizing respectively around $K = 5$. The *miss-classification rate* increases with the factor multiplication, stabilizing when factor $K = 2$ and the FPR also increases stabilizing when the multiplication factor is around $K = 4$.

Contrarily, **when factor K is below 0.9**, the overall classifier performance is better. In fact, all of the confusion matrix indices report better results - *accuracy*, *true positive rate*, *specificity* and *precision* are higher while *false positive rate* and *miss-classification rate* decrease when the background multiplication factor decreases.

Between 0.9 and 1, regarding the multiplication factor, there is a slight inflexion in the performance of every CM index, indicating an adjustment in the model when using the real values. In fact, there is an overlapping of the data samples corresponding to the ”EV is Charging” with the ones corresponding to ”EV is Not Charging”.

In short and correlating with the physical reality, there is a logical reasoning behind the classifier behaviour when there is a variation of the background load. On the one hand, when the multiplication factor is below one, the EV charging P (constant and equal to 3.7 kW) becomes more relevant. As the multiplication factor tends to zero, so does the background load, making the EV charging load stand out and facilitate the disaggregation. On the other hand, when the multiplication factor is greater than 1 and as it increases, the EV charging P becomes less relevant, causing the EV active power variation to be insignificant in the remaining load, hardly disaggregating it.

In the appendix, figures B.1, B.2, B.3 and B.4 represent the PQ scatter plot of the training set with $K = 0.1$, $K = 0.5$, $K = 1$ and $K = 2$, respectively. They are illustrative of the existent disaggregation between the two clusters **with different background P and Q load multiplication factors *K***. Also in appendix, figure B.5 shows the average P and Q load throughout the day. This figure is presented so that there is a more accurate representation of the average P and Q power range when $k = 1$ (P_1^{BK} and Q_1^{BK}), facilitating the understanding when using multiplication factor K different than 1.

6.4.2 Scenario 2 - Variation of EV charging P Load by Factor Multiplication

Up until now, all results used the actual P of the EV charger - $P = 3.7\,kW$, multiplication factor $K = 1$ - corresponding to *Slow Charge* (P_1^{CH}). Now, the classifier performance with P variation of the EV charging by a factor K is tested $P_k^{CH} = (P_1^{CH} * K) = (3.7 * K)\,kW$. This test will infer the validity of this classifier within a range $K = [0.1,\, 5]$ evaluating if this algorithm is effective when it comes to classifying EV charging events with P lower and bigger than $3.7\,kW$. In fact, there are several power modes and corresponding procedures in what concerns EV charging established by the European Standards [15], described by table 2.1, in the background chapter.

In addition, with increased EV power charging, experimentally the power factor becomes different

than 1, due to some ineffectiveness in the boost PFC converter alongside different power factor correction electronic configuration specific for each EV charging power rating, as seen in the Chapter *Background* (section 2.2.3). For this experiment, an assumption was made in which the power factor remains unitary, varying only the P, for simplicity.

Figure 6.5 is representative of the *accuracy*, MCR, TPR, FPR, *specificity* and *precision* indices used to assess the classifier performance in this scenario - factor variation of EV charging P load - resorting to the 5 fold cross-validation.

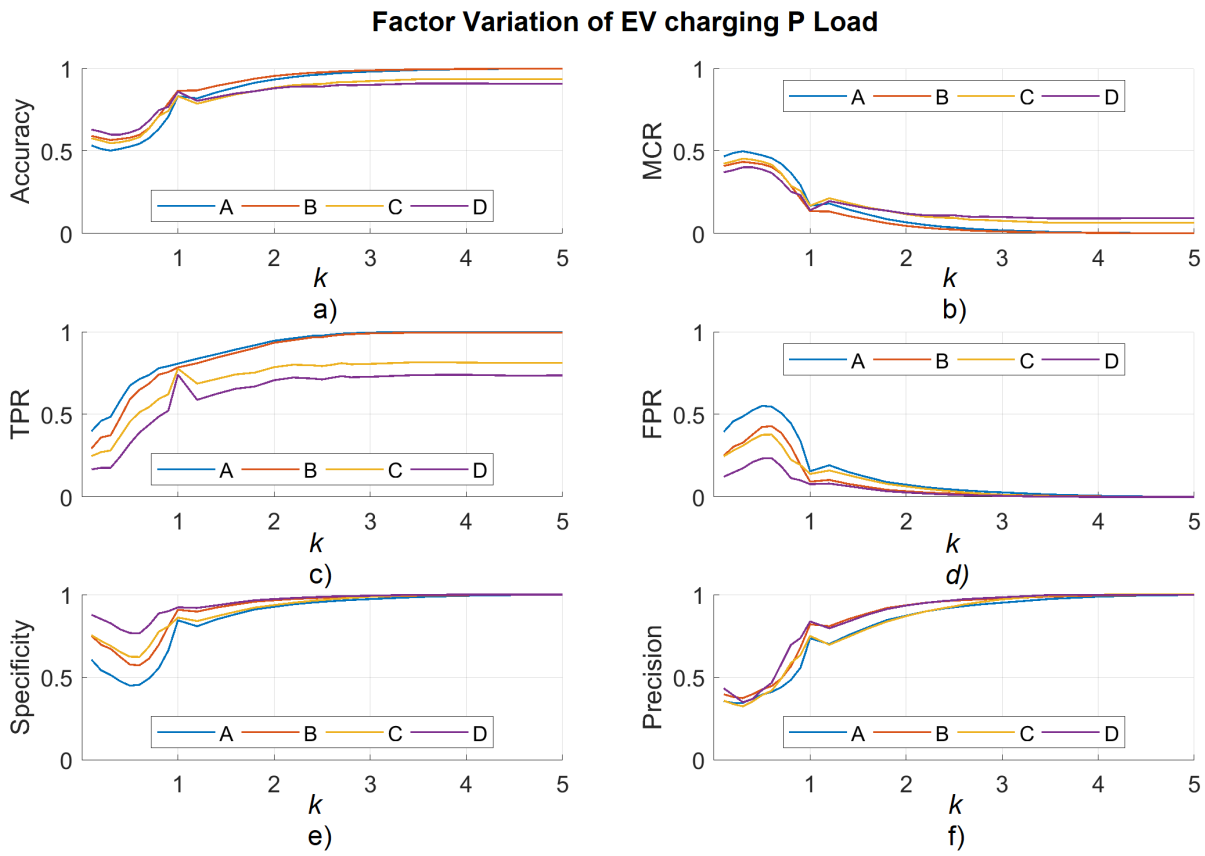

Figure 6.5: *Accuracy* (a), MCR (b), TPR (c), FPR (d), *specificity* (e) and *precision* (f) for *K* multiplication factor variation of EV charging P load in transformer.

Looking at the confusion matrix indices evolution with multiplication factor K in figure 6.5, some conclusions can be extracted regarding the classifier performance when using different EV charging power rates. When looking at the *accuracy* index, the overall performance of this classifier increases with an increase in the EV charging load indicating better performance with higher active power load. In fact and corroborating this statement, **when the multiplication factor is above 1**, the TPR, the *specificity* and the *precision* increase, while the FPR and the *miss-classification rate* decrease with an increase in the multiplication factor K.

Now, **with multiplication factor below 1**, firstly there is a decrease in the *accuracy* performance followed by an increase as K decreases. This phenomenon is due to the fact that there are two major group aggregations in the "EV is Charging" cluster. These aggregations are: when there are EV charges at night, background P is small (containing *phantom charges*); when there are EV charges during the

day, with background P higher. As K decreases and when $K < 0.5$, the clusters overlap, making more difficult any disaggregation.

On the one hand, with this background load, when there is an increase in the EV charging power rate, the span between the two data clusters increase making it easier to disaggregate and therefore for the classifier to detect it, largely improving the classifier performance.

On the other hand and with a decrease in the EV charging power rate, the P demand from the grid is more and more insignificant, until a point where the classifier has around 50% chances of detecting an EV charging, making it infeasible.

In appendix, figures C.1, C.2, C.3 and C.4 represent the PQ scatter plot of the training set with $K = 0.1$, $K = 0.5$, $K = 1$ and $K = 2$ respectively, and they are illustrative of the existent disaggregation between the two clusters **with different EV charging P multiplication factors *K***.

6.5 ”Fictitious” EV Charges

As explained in the implementation section regarding the design of fictitious EV charges (chapter 5.3), new data relative to EV charging events was created in order to try to have an isolated data setup, or otherwise named controlled dataset and to further evaluate the ideal conditions in which this classifier operates. However, this data set will not be completely independent given that the days that served as background load (where no EV charging events were reported) are only four and there is a repetition of those same days, obviously with different and random EV charging each time.

The data in the following simulation was obtained through statistical methods, and its purpose is not just to validate the classifier but also to understand if it is possible to extrapolate for other EV charger. These models can also be used in future work whenever it is necessary to understand the drivers behaviour of the users of that type EV charger and location.

In addition, by having these fictitious EV charges together with historical P and Q data with no EV charging load, it is possible to assess the impact on the LV network of introducing EV chargers.

6.5.1 Confusion Matrices Results Summary

To assess the feasibility of these fictitious EV charges, the CM of this simulation was calculated and displayed in table 6.10. Furthermore, the indices that allow its performance validation were computed and presented in figure 6.6. To perform this simulation, the training set is the same as the one used previously in subsection 6.2.1. The number of days in which fictitious EV charges were added was 71, which is the same number as the one from the sampled test set used in the *real data* simulation, comprehending the testing set.

The performed simulation shows a *prevalence* level of 27.7% with its CM parameters listed in table 6.10. The CM performance indices are displayed in figure 6.6.

Looking at figure 6.6, it is possible to infer some conclusions regarding the classifier performance with fictitious EV charging. The first thing to notice is the high values of TPR index in all 4 optimizations.

Table 6.10: *Confusion matrix* parameters for fictitious EV charges with n - total number of samples, P - number of actual positive samples, N - number of actual negative samples, P' - number of positively predicted samples, N' - number of negatively predicted samples, TN - number of true negative samples, FP - number of false positive samples, FN - number of false negative samples, TP - number of true positive samples.

Optimization	CM parameters								
	n	P	N	P'	N'	TN	FP	FN	TP
A	20413	5651	14762	8202	12211	12179	2583	32	5619
B				7399	13014	12960	1802	54	5597
C				7671	12742	12436	2326	306	5345
D				6655	13758	13247	1515	511	5140

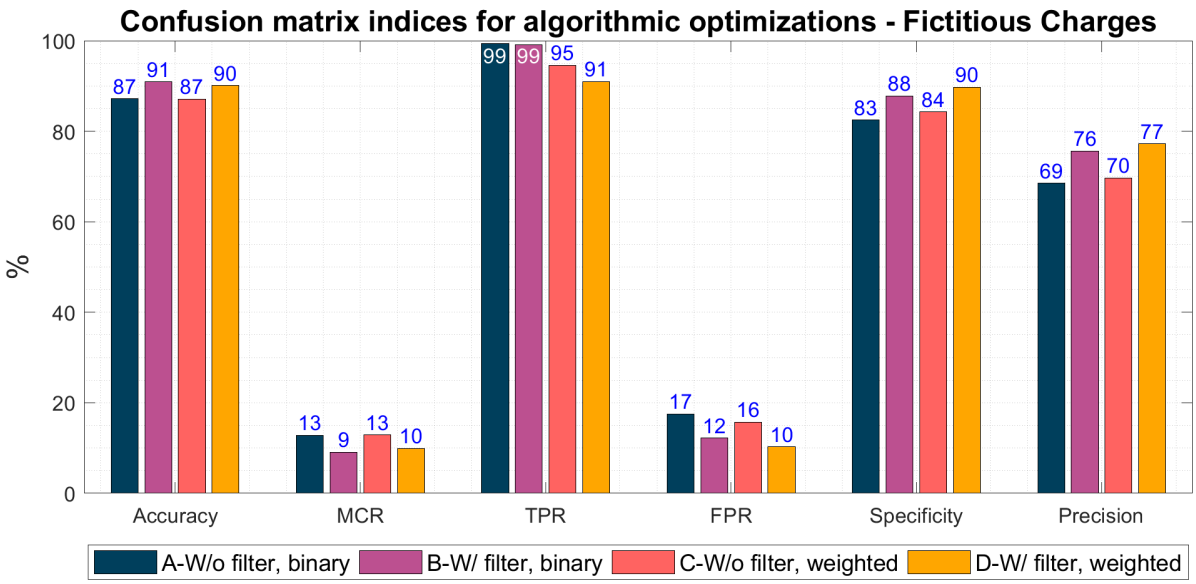

Figure 6.6: Overall display of the CM indices to evaluate the algorithmic performance of each of the optimizations for fictitious EV charges, being MCR the miss-classification rate, TPR the true positive rate and FPR the false positive rate.

This reveals that whenever an EV charge is occurring there is a high probability it would be positively classified. This means also that the fictitious EV charges do not comprehend the occurrence of *phantom charges* - an EV charge load is assumed to be constant and equal to $3.7\,kW$. This way, the classifier is optimized to detect EV charges that actually have power demand from the LV network.

The *specificity* is relatively high, between 83% and 90%, indicating the percentage of samples that are correctly classified as negative - "EV is Not Charging" - however, not as high as the TPR previously mentioned with the exception of optimization D with a 1% difference between each. With this, it would be expected than the *specificity* index percentage would be higher that the TPR one, provided that the *prevalence* of EV charging of this fictitious dataset is of 27.7%. One reason that explains this occurrence is the fact that there is more dispersion in the P-Q samples when an EV is charging, as it was displayed in the Pattern Recognition chapter in 4.

Looking now at the *precision* index, its values stays between 69% and 77%. They are representative

of the percentage of predicted positive charges that are correctly classified. Overall, the accuracy is between 87% and 91% and the MCR between 9% and 13%.

With this results, no *phantom charges* are represented, and no errors in the prediction contemplated, resulting in good predictions, in the sense that there are a lot of true positive ones, leading to high TPR. On the other hand, *precision* index indicates some error in the predictions made.

6.5.2 Comparison Between Real and Fictitious Datasets

Having extracted some conclusions regarding the fictitious EV charging using the CM, it is ever more important to compare its results with the real EV charges results, presented in section 6.2.1. This comparison will be executed in this section and the table represented in 6.11 contains the CM indices of both the real dataset - using the cross-validation results - as well as the fictitious one.

Table 6.11: Confusion matrix performance indices for real and fictitious EV charges

Training set	**Optimization**	**CM performance indices ($\%$)**					
		Accuracy	MCR	TPR	FPR	*Specificity*	*Precision*
Real	A	83	17	81	15	85	73
	B	87	13	78	9	91	82
	C	83	17	78	14	92	74
	D	86	14	74	8	92	83
Fictitious	A	87	13	99	17	83	69
	B	91	9	99	12	88	76
	C	87	13	95	16	84	70
	D	90	10	91	10	90	77

Looking at table 6.11, particularly the overall *accuracy* and MCR of the system, both the real as well as the fictitious present relative close values with a 4% difference between datasets, in each optimization.

At first glance, this results may seem rather similar, however, they are deceiving, in the sense that they alone do not provide the full picture regarding the feasibility of these fictitious EV charges as a substitute/complement of our real dataset. In order to have a more reasonable assessment, the comparison of the remaining indices, namely the TPR, FPR, *specificity* and *precision* has to be performed.

Concerning the TPR, it is now possible to realize a somewhat considerable difference between datasets. The fictitious values are higher than the real ones 18.25% on average, across optimizations. This higher value in the fictitious dataset reveals that, with the proposed assumptions, the classifier has a much better performance when it comes to classify a sample positively. This is because the presence of *phantom charges* largely influences the TPR index, as previously noted.

Moving on to the FPR, there is a difference between datasets of 2.25% on average across the four optimizations, with the lowest values belonging to the real dataset. This means that the negative predictions are similarly classified in the fictitious dataset. Now, looking at the *precision*, the real dataset present higher values with an average difference of 5%, reinforcing that the negative classification of samples is somewhat similar using the fictitious EV charges.

However, these results can be biased given that the background load used to ”create” the 71 days corresponding of this dataset consists of only 4 days, providing not enough randomness. It can also be due to the fact that this simulation does not consider the P decay after 85% of SoC. If that was to be considered, some of the P load corresponding to an EV charging would have been lesser than the used 3.7 kW, and the probability of being positively classified would have been lower. Moreover, a unitary PF constraint was assumed, when in reality PF is equal to 0.98 inductive, which could influence the results.

In conclusion, the fictitious data presents different results than the real data in the sense that it classifies not when an EV is connected to the EV charger (as it is done with the real data), but when the EV charger is demanding constant active power.

Comparison of P and Q Average Values Throughout the Day for both Datasets

To further understand the existent difference between both these datasets - real and fictitious - an average in terms of P and Q throughout the day was computed having all the 71 days. With this analysis, the mean absolute percentage error between datasets was obtained and by bridging it with the shape, it is possible to estimate whether there is a close match between both datasets.

Figure 6.7 represents an overlapping of P and Q averages of both datasets, with the real one in blue and the fictitious one in red.

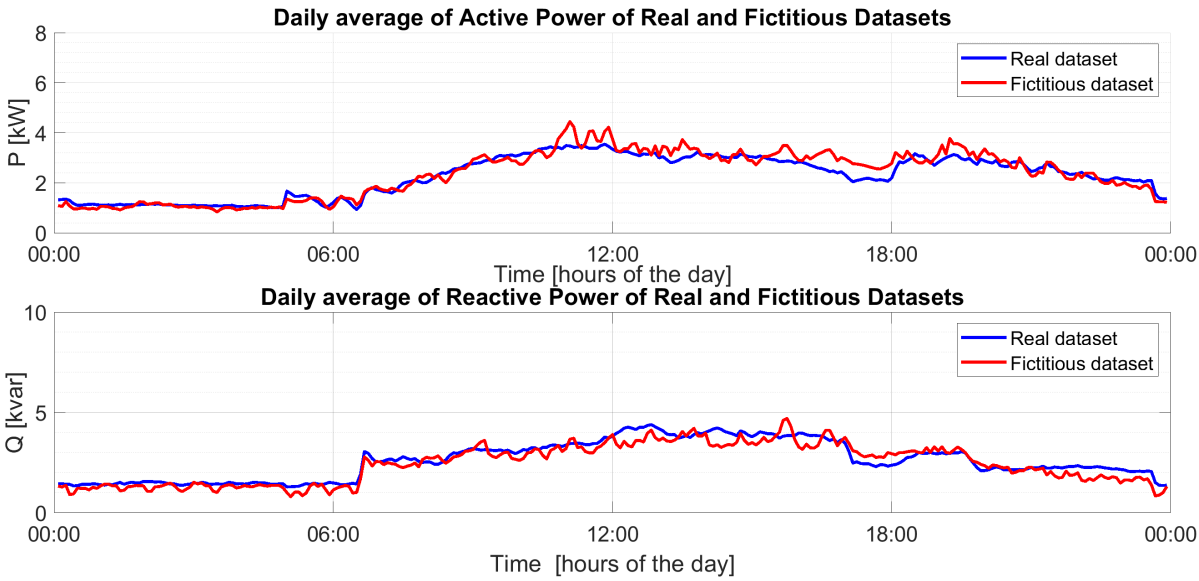

Figure 6.7: Daily P and Q average of both real and fictitious datasets.

In this simulation, the mean absolute percentage error in the active power graph is of 10.49% and in the reactive power is of 11.98%. Analysing this results, there is a relatively high value of Q error further that reinforces the infeasibility of assuming only 4 different days as *background load*. More days where no EV charges had occurred are needed in order to have enough randomness.

On the other hand the P error is relatively small and one might hastily say that, looking only at this simulation errors as well as the graphs shape, there could be some feasibility in the fictitious EV charges. However, as it was said before, different factors have to be taken into consideration when it comes to

realistically modelling an EV charging.

Overall Outcome

In short and looking at both the CM indices (real and fictitious) as well as the average results, the fictitious EV charges algorithm is characterized by an oversimplification. Following this reasoning, the use of only four days as background load with which the fictitious EV charges were added, could have biased the outcome providing not enough randomness.

The inability to extract data concerned with the initial and final SoC of the vehicle lead to the non-modelling of that feature and a further assumption of constant P injection in the EV, also constituted a simplistic approximation. Moreover, the unitary power factor assumption could also have been correlated with the present results.

After careful analysis, a reason for the TPR results being different than the real ones is the *phantom charges*. In this sense phantom charges are classified as ”EV is Not Charging” and the classifier labels those samples as *false negative* when in reality they should have been *true negative* given that no power is being demanded. This fictitious charges assess not the EV connection to the charger, but the actual power demand from there.

However, the models used to create fictitious EV charges that would resemble the drivers behaviour of that charger, in terms of duration, beginning time and number of EV charges during the day, were considered accurate and can be used in extrapolation to other EV chargers.

6.6 Industry Application

The proposed EV charging classifier has, not only academic validity but can also be used as an industry application solution. In this regard, it is possible to implement the proposed classifier using historical data, namely P and Q values in the secondary of the distribution substation, together with the fictitious EV charging model also presented.

In general terms, it was seen that a single EV charging is characterized by an increase in the active power by $3.7\,kW$. This increase displays a shift in P-Q samples, as it shown in figure 6.8. This is the general behaviour of an EV charging, when using the P-Q scatter plot.

Now, having historical P and Q data from the secondary of the distribution substation and using an algorithm to randomly (and realistically) obtain EV charging events, it is possible to train the proposed algorithm. The fictitious EV charging model proposed with the data from the used EV charger,presented in section 5.3.1, can be used as well as other models that realistically assess the EV charging behaviour of EV drivers [13, 19, 37].

Having the classifier trained - the Gaussian curves computed - it is possible to start conducting EV charging detection, using only data extracted using the smart-meter, from the secondary of the distribution substation. It is no more needed EV charging validation data, such as the one provided by MOBI.E®. Figure 6.9 provides a flowchart of the proposed algorithm.

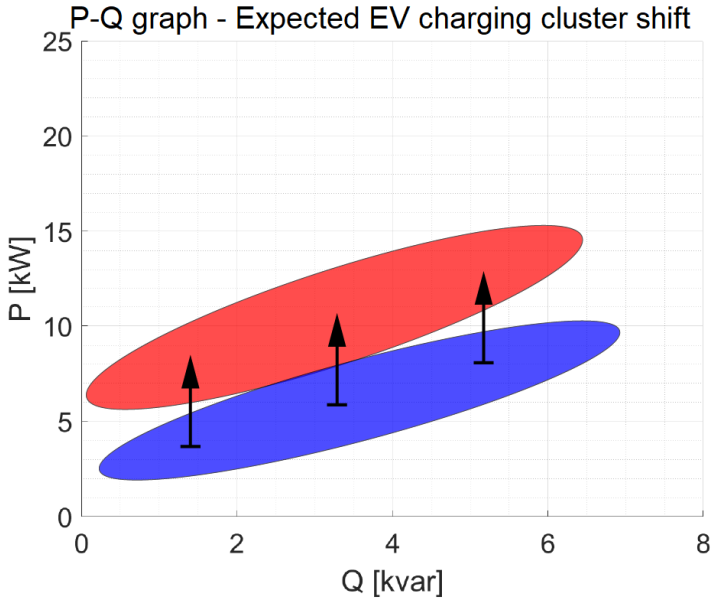

Figure 6.8: General behaviour of P-Q samples when an EV charging occurs, from ”EV is Not Charging” class to ”EV is Charging” class.

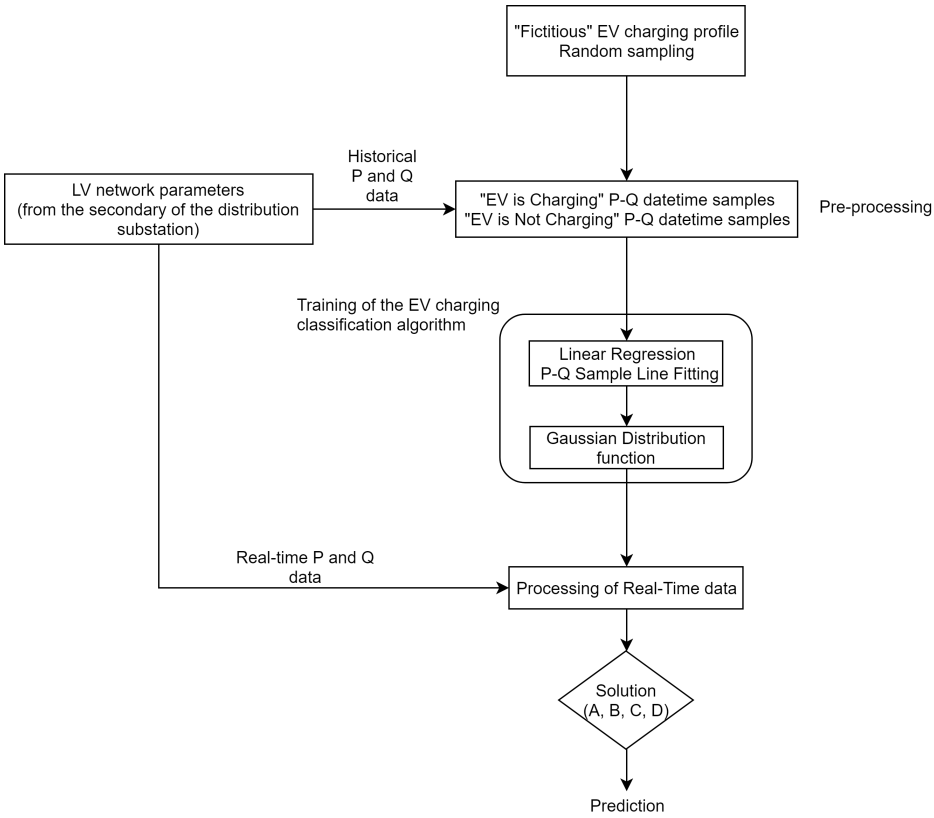

Figure 6.9: Flowchart of industry application EV charging detection algorithm

Chapter 7

Conclusions

In this final chapter, some conclusions are drawn, regarding the present thesis, namely the achievements and some future work.

7.1 Achievements

The work produced and presented in this thesis is part of the research that is being conducted on the large scale integration of EV's into the power grid. In this manner, the greatest achievement of this research work is the formulation of a classification algorithm for detection of EV charging occurrences, using sampled LV network parameters from the secondary side of the distribution substation. The smart meter ENEIDA® EWS DTVI-g was used for sampling.

This classification algorithm was extensively tested using a single EV charger with a *normal power* charging method (slow charging) from a public installation. The proposed algorithm makes use of linear regression of $P - Q$ samples alongside the *Gaussian Membership Function* and the prediction classes are "EV is Charging" and "EV is Not Charging". Several optimizations that included the use of both a time filter as well as a threshold filter were conducted and the classifier results ranged from $83 - 87\%$ overall accuracy, $74 - 81\%$ of *True Positive Rate* and $85 - 92\%$ of *Specificity*. To guarantee impartiality, a *5 fold cross-validation* was conducted. It was further concluded that the optimization that yields the best result is optimization B: *"With temporal filter, binary"* provided that it has the best ratio between the correctly and incorrectly classified samples, in both classes "EV is Charging" and "EV is Not Charging". The results obtained include "phantom charges" - $P - Q$ samples when an EV is connected to the charger but with full battery, *i. e.* no power demand - affecting the results in the sense that there are miss-labelled samples. In reality, it is expected that the results are better than the ones presented in this thesis, provided that the classifier objective is to detect an EV charging event with actual power demand.

Another achievement is related with the understanding that the best aggregation of active and reactive power data, concerning the detection of EV charging, is by using all available P-Q samples separated into working day/weekend datasets.

In addition, the feasibility limits in which the proposed classifier can operate are tested in two different

ways: when the line adjacent load varies, simulating different power load magnitude levels, on the LV network; when the charging station power rating mode is different than the used one (slow charging).

Moreover, the accurate modelling of consumers behaviour of the used EV charger was accomplished and it can be applied to realistically create fictitious EV charges. These models can then be used if an extrapolation of the characteristics of the EV charger in study is considered for other such EV chargers.

Finally, a solution for the implementation of the proposed classification algorithm in the industry is, furthermore, presented and explained in detail.

7.2 Future Work

There is some future work regarding what was developed in this thesis.

Foremost, all of this research was conducted taking into consideration a single EV charger. In the future, a large-scale study of multiple EV chargers can be performed to ensure greater representativeness as well as to understand the classifier performance when subjected to different circumstances, such as fast charging EV chargers. Moreover, an analysis on the classifier performance when simultaneous EV charges are occurring can be further analysed.

Moving on to the creation of fictitious EV charges, it is possible to complement the proposed algorithm by including a model of the vehicle *State of Charge*, which is linked with the active power extraction from the grid (when $SoC > 85\%$) as well as to include the power factor correction, according to the power mode of charging that is occurring.

Regarding the existence of ”phantom charges”, the correspondent *P-Q* samples were not mitigated in this thesis and that is something that could be done, in future work, to have an ever more realistic assessment of the classifier performance.

Lastly, several other types of classification algorithms can be used, and a possibility that was not tested in this thesis was the employment of deep learning algorithms that make use of artificial neural networks. This research did not use such algorithms due to time constraints and to the fact that, in every step of process of building the classifier, the physical aspect of the used variables was explicit.

References

[1] IEA(2019). Global EV Outlook 2019. Technical report, IEA, Paris. URL www.iea.org/publications/reports/globalevoutlook2019/.

[2] Portugal 2020. URL https://www.portugal2020.pt/.

[3] M. Shabanzadeh and M. P. Moghaddam. What is the Smart Grid ? Definitions, Perspectives, and Ultimate Goals. (February), 2014. doi: 10.13140/2.1.2826.7525.

[4] Smart grid – Desenvolvimento sustentável de megacidades – Siemens Brasil. URL https://www.siemens.com.br/desenvolvimento-sustentado-em-megacidades/smart-grid.html.

[5] M. Budka, Kenneth C., Deshpande, Jayant G., Thottan. *Communication Networks for Smart Grids*. Springer-Verlag London, 2014. ISBN 9781447163015. doi: 10.1007/978-1-4471-6302-2.

[6] R. Castro. *Uma Introdução às Energias Renováveis: Eólica, Fotovoltaica e Minihídrica*. IST Press, 2011. ISBN 9789898481016.

[7] M. Shabanzadeh and M. Moghaddam. What is the Smart Grid? Definitions, Perspectives, and Ultimate Goals. In *Power System Conference*, 2013. doi: 10.13140/2.1.2826.7525.

[8] SmartGrids Technology Plataform. URL http://www.e-control.at/market{_}players/electricity/{%}0Aspecific-topics/smart-grids.

[9] R. J. Ferreira, L. M. Miranda, and R. E. Ara. A New Bi-Directional Charger for Vehicle-to-Grid Integration. doi: 10.1109/ISGTEurope.2011.6162770.

[10] European Comission: Smart Grids and Meters. URL https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters/overview.

[11] J. Quirós-Tortós, L. F. Ochoa, and B. Lees. A statistical analysis of EV charging behavior in the UK. In *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, pages 445–449, oct 2015. doi: 10.1109/ISGT-LA.2015.7381196.

[12] M. Pereira. *Reconfiguração Dinâmica de Redes de Baixa-Tensão para Gestão de Pontas de Carregamento de Veículos Elétricos e Geração Distribuída*. PhD thesis, IST, 2015.

[13] My Electric Avenue. URL http://myelectricavenue.info/.

[14] R. J. C. Pinto, J. Pombo, M. R. A. Calado, and S. J. P. S. Mariano. An Electric Vehicle Charging Station : Monitoring and Analysis of Power Quality. In *2015 9th International Conference on Compatibility and Power Electronics (CPE)*, number Dc, pages 37–42, Costa da Caparica, Portugal, 2015. IEEE. ISBN 9781479963010. doi: 10.1109/CPE.2015.7231046.

[15] M. C. Falvo, D. Sbordone, I. S. Bayram, and M. Devetsikiotis. EV Charging Stations and Modes : International Standards. In *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, Ischia, Italy. IEEE. doi: 10.1109/SPEEDAM.2014.6872107.

[16] P. Richardson, M. Moran, J. Taylor, A. Maitra, and A. Keane. Impact of electric vehicle charging on residential distribution networks: An irish demonstration initiative. In *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, pages 1–4, jun 2013. doi: 10.1049/cp.2013.0873.

[17] K. Petrou, J. Quiros-Tortos, and L. F. Ochoa. Controlling electric vehicle charging points for congestion management of UK LV networks. In *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, feb 2015. doi: 10.1109/ISGT.2015.7131843.

[18] J. Quirós-Tortós, L. F. Ochoa, S. W. Alnaser, and T. Butler. Control of EV Charging Points for Thermal and Voltage Management of LV Networks. *IEEE Transactions on Power Systems*, 31(4): 3028–3039, jul 2016. doi: 10.1109/TPWRS.2015.2468062.

[19] J. Quirós-Tortós, A. N. Espinosa, L. F. Ochoa, and T. Butler. Statistical Representation of EV Charging: Real Data Analysis and Applications. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7, jun 2018. doi: 10.23919/PSCC.2018.8442988.

[20] J. D. Watson and N. R. Watson. Impact of electric vehicle chargers on harmonic levels in New Zealand. In *2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pages 1–6, dec 2017. doi: 10.1109/ISGT-Asia.2017.8378374.

[21] E. S. Xydas, C. E. Marmaras, L. M. Cipcigan, A. S. Hassan, and N. Jenkins. Forecasting Electric Vehicle charging demand using Support Vector Machines. In *2013 48th International Universities' Power Engineering Conference (UPEC)*, pages 1–6, 2013. doi: 10.1109/UPEC.2013.6714942.

[22] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang. Short-Term Load Forecasting for Electric Vehicle Charging Stations Based on Deep Learning Approaches. *Applied Sciences*, 9:1723, 2019. doi: 10.3390/app9091723.

[23] H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, feb 2001. doi: 10.1109/59.910780.

[24] F. Heyman, C. Pereira, V. Miranda, and F. J. Soares. Spatial load forecasting of electric vehicle charging using GIS and diffusion theory. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2017 - Proceedings*, volume 2018-Janua, pages 1–

6. Institute of Electrical and Electronics Engineers Inc., jul 2018. ISBN 9781538619537. doi: 10.1109/ISGTEurope.2017.8260172.

[25] MOBI.E - Mobilidade Elétrica (c). URL https://www.mobie.pt.

[26] ENEIDA.IO (c). URL https://eneida.io.

[27] J. Martin, D. Adana, and A. Garcia Asuero. *Fitting Models to Data: Residual Analysis, a Primer*. IntechOpen. ISBN 978-953-51-3279-0. doi: 10.5772/68049.

[28] Mark Mikofski. *Copyright (c) 2014*. 2014.

[29] R. W. Farebrother. *L 1 -Norm and L inf -Norm Estimation*. Springer, Berlin, Heidelberg, 2013. ISBN 9783642362996. doi: 10.1007/978-3-642-36300-9.

[30] Opções horárias. URL https://www.edp.pt/particulares/apoio-cliente/perguntas-frequentes/tarifarios/o-que-preciso-de-saber-para-contratar/o-que-e-a-opcao-horaria-e-qual-a-melhor-para-mim/faq-4823.

[31] J. S. Marques. *Reconhecimento de Padrões: Métodos Estatisticos e Neuronais*. IST Press, 2 edition, 2005. ISBN 978-972-8469-08-5.

[32] T. J. Ross. *Fuzzy Logic with Engineering Applications*. Copyright (c) 2010 John Wiley and Sons, Ltd, third edit edition, 2010. ISBN 9780470743768. doi: 10.1002/9781119994374.

[33] V. Monteiro, H. Gonçalves, and J. L. Afonso. Impact of Electric Vehicles on Power Quality in a Smart Grid Context. In *13th IEEE International Conference on Electrical Power Quality and Utilization (EPQU)*, pages 17–19, Lisbon,Portugal, 2011. IEEE. doi: 10.1109/EPQU.2011.6128861.

[34] N. Badowski. QR Decomposition in Least Squares Analysis, with an Application to the Reduction of Rotational Spectra to Molecular Constants. *Spectroscopy Letters*, 20:365–373, 1987. doi: 10.1080/00387018708081557.

[35] A. Bjorck. Solving least squares problems by Gram–Schmidt orthogonalization. *BIT*, 7:1–21, 1967. doi: 10.1007/BF01934122.

[36] K. M. Ting. *Confusion Matrix*, page 209. Springer US, Boston, MA, 2010. ISBN 978-0-387-30164-8. doi: 10.1007/978-0-387-30164-8_157. URL https://doi.org/10.1007/978-0-387-30164-8{_}157.

[37] R. Singh, B. C. Pal, and R. A. Jabr. Statistical Representation of Distribution System Loads Using Gaussian Mixture Model. *IEEE Transactions on Power Systems*, 25(1):29–37, feb 2010. doi: 10.1109/TPWRS.2009.2030271.

[38] D. Aljburi. Probability Density Function (PDF), 2016.

[39] H. Refaeilzadeh, Payam and Tang, Lei and Liu. *Cross-Validation*. 2009. ISBN 9781351605175. doi: 10.1007/978-0-387-39940-9_565.

Appendix A

Time Evolution of P and Q

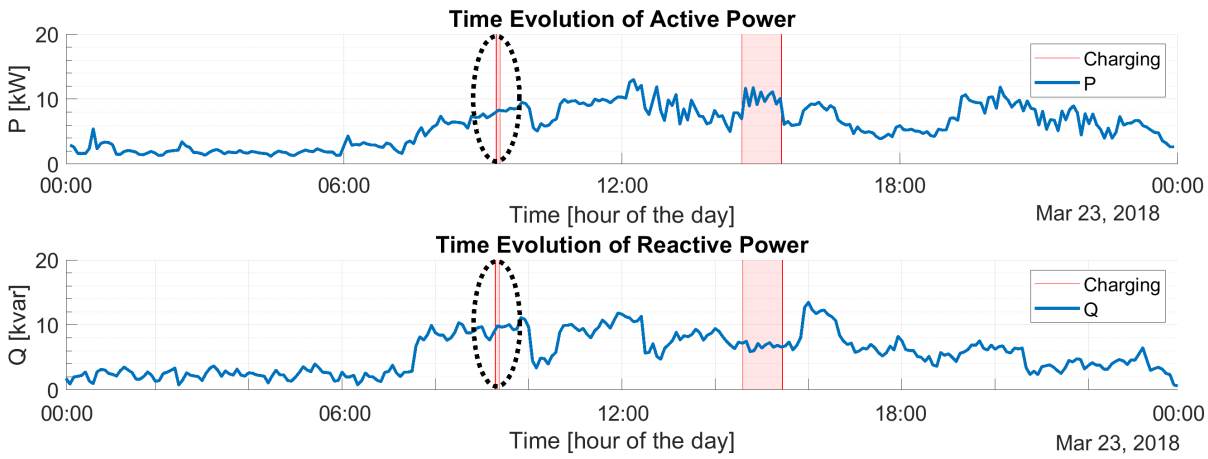

Figure A.1: P and Q variation in 23/03/2018 with EV charging validation. EV charges with less than 10 minutes highlighted.

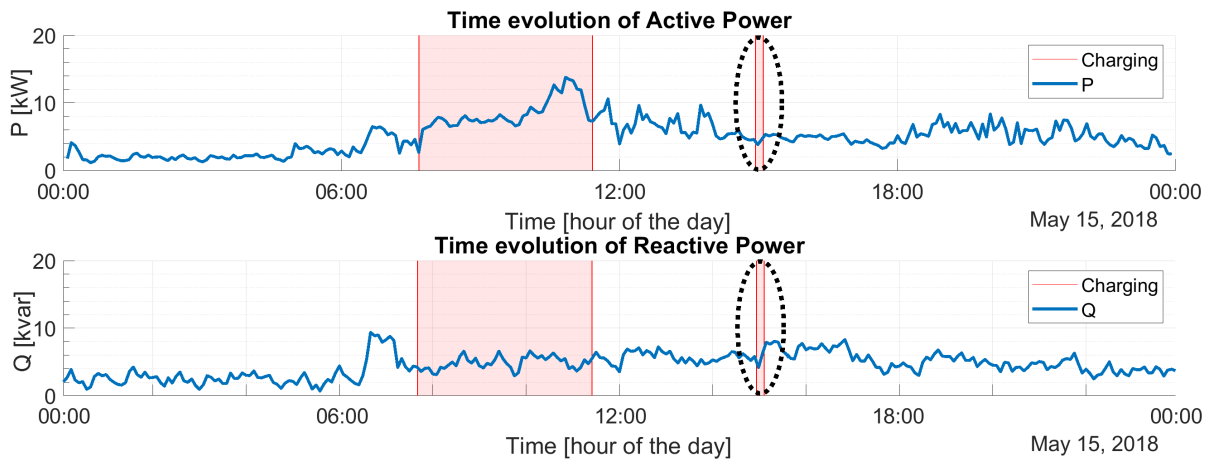

Figure A.2: P and Q variation in 15/03/2018 with EV charging validation. EV charges between 10 and 20 minutes highlighted.

Appendix B

Feasibility Limits - Scenario 1

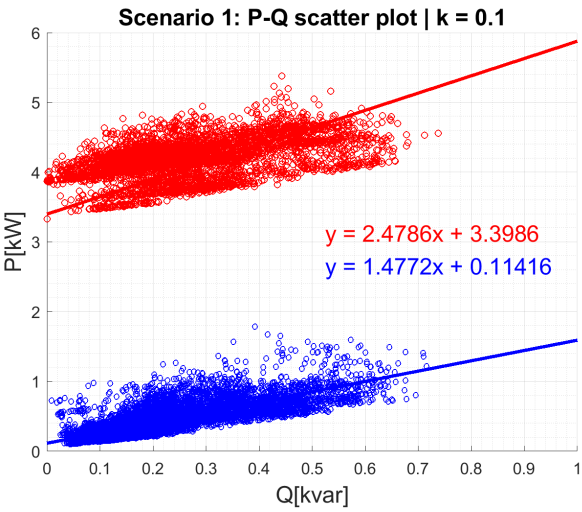

Figure B.1: Scenario 1 - PQ scatter plot with K = 0.1.

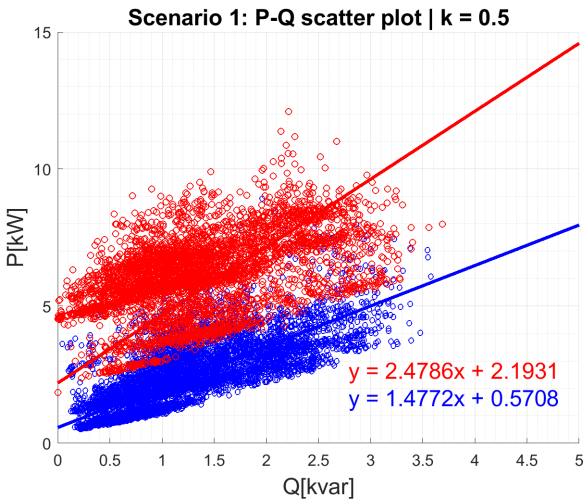

Figure B.2: Scenario 1 - PQ scatter plot with K = 0.5.

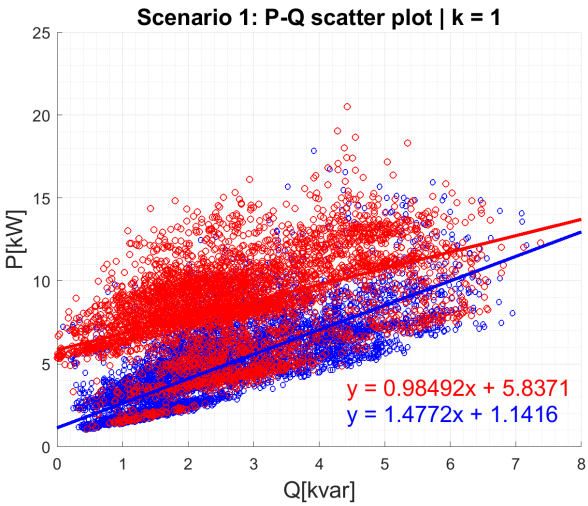

Figure B.3: Scenario 1 - PQ scatter plot with K = 1.

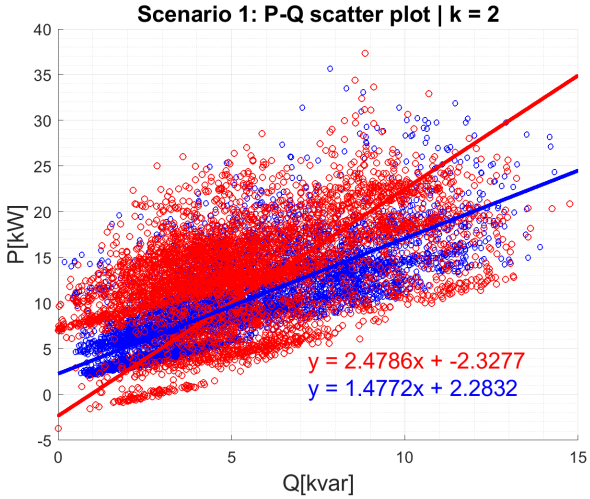

Figure B.4: Scenario 1 - PQ scatter plot with K = 2.

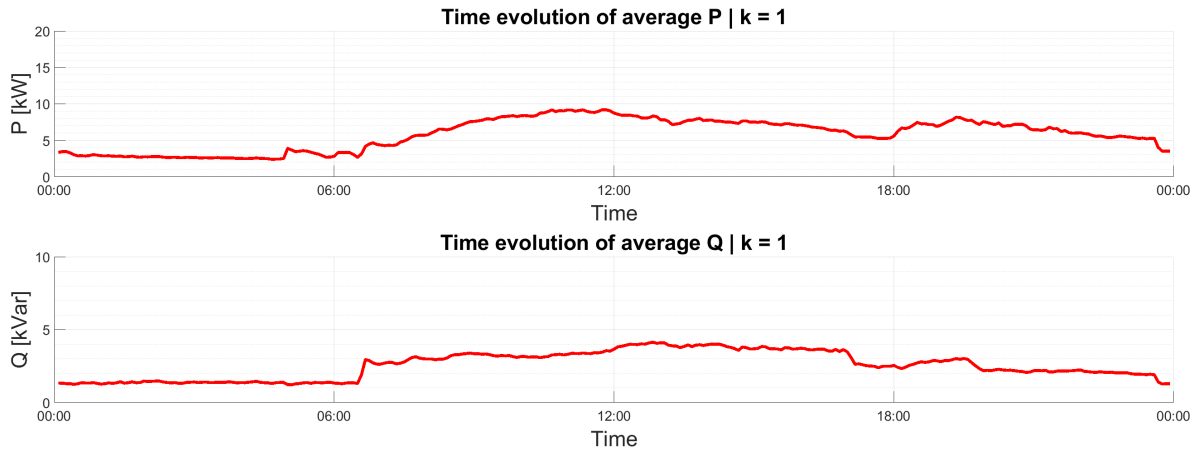

Figure B.5: Average P and Q power values throughout the day - visual representation of the average load of used line and feeder when K = 1

Appendix C

Feasibility Limits - Scenario 2

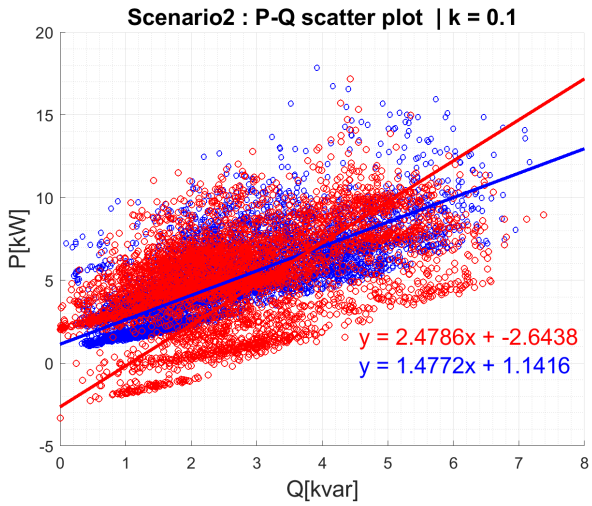

Figure C.1: Scenario 2 - PQ scatter plot with K = 0.1.

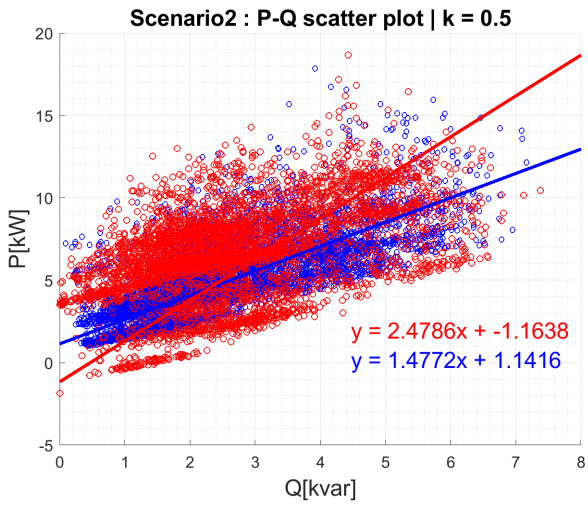

Figure C.2: Scenario 2 - PQ scatter plot with K = 0.5.

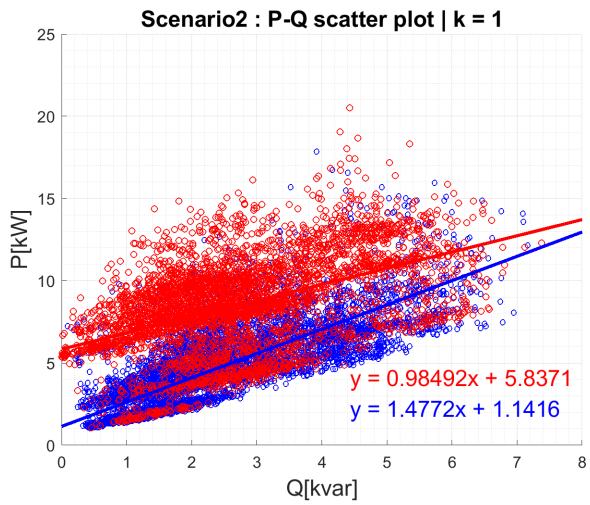

Figure C.3: Scenario 2 - PQ scatter plot with K = 1.

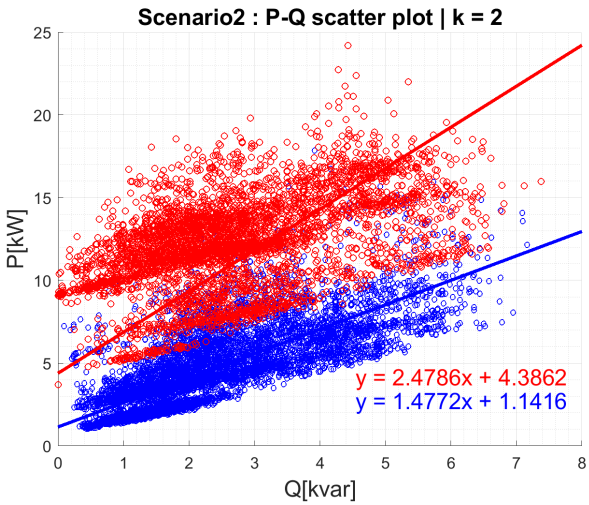

Figure C.4: Scenario 2 - PQ scatter plot with K = 2.