UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR INGENIERÍA INDUSTRIAL VALENCIA

INDUSTRIAL ENGINEERING MASTER THESIS

DESIGN AND ANALYSIS OF ARTIFICIAL NEURAL NETWORKS FOR THE PREDICTION OF THE ELECTRICITY DEMAND OF PUBLIC ADMINISTRATIONS BASED ON TYPE OF CONSUMPTION DECOMPOSITION

AUTHOR: Sergio Mateo Barcos

SUPERVISOR: Manuel Alcázar Ortega

SUPERVISOR: David Ribó Pérez

Academic year: 2021-22

Acknowledgement

Gracias a mi familia por apoyarme en todo aquello que deseo hacer y a la Cátedra de Transición Energética y el Instituto de Ingeniería de la Energía por la continua formación, así como a David por su contínua atención e interés por el trabajo.

Resumen

La predicción de la demanda eléctrica es uno de los procesos más importantes en el sector energético. Desde la operación de los sistemas hasta la gestión de compras y ventas de energía en los distintos mercados existentes, predecir bien el consumo es vital para evitar problemas como congestiones en las redes que pueden llegar a causar desabastecimientos o para evitar importantes pérdidas económicas como consecuencia de planificar mal las compras de los agentes en los mercados mayoristas. En este TFM se plantea un modelo de predicción de la demanda eléctrica para administraciones públicas a partir de una descomposición en tipos de consumo y la aplicación de redes neuronales artificiales (ANN). Normalmente, todas las administraciones suelen contar con una cierta tipología de tipos de consumo; iluminación pública, oficinas, colegios, semáforos. . . etc, cuya desagregación puede ser útil para fragmentar el problema y observar patrones que sean más fácilmente predecibles. La metodología se aplica al caso de estudio del Ayuntamiento de Valencia. Se dispone de datos reales de los consumos municipales para el periodo de 2017 y 2018. Estos consumos se van a clasificar por tipos y se van a tratar para posteriormente facilitar el proceso de entrenamiento de las redes neuronales y que estas ofrezcan la predicción más acertada posible. La predicción de cada uno de los tipos de consumo se hará por separado, dado que cada uno depende de unos inputs diferentes. Posteriormente, se agregarán todas las predicciones individuales para obtener una predicción global del consumo del ayuntamiento de Valencia.

Palabras Clave: Demanda eléctrica, Redes Neuronales Artificiales, Administración pública, Valencia, Predicción, Eficiencia, Mercado eléctrico.

Resum

La predicció de la demanda elèctrica és un dels processos més importants en el sector energètic. Des de l'operació dels sistemes fins a la gestió de compra i venda d'energia en els diferents mercats existents, predir bé el consum és vital per a evitar problemes com congestions en les xarxes que poden arribar a causar desproveïments o per a evitar importants pèrdues econòmiques a conseqüència de planificar malament les compres dels agents en els mercats majoristes. En aquest TFM es planteja un model de predicció de la demanda elèctrica per a administracions públiques a partir d'una descomposició en tipus de consum i l'aplicació de xarxes neuronals artificials (ANN). Normalment, totes les administracions solen comptar amb una certa tipologia de tipus de consum; il · luminació pública, oficines, col · legis, semàfors. . . etc, la desagregació dels quals pot ser útil per a fragmentar el problema i observar patrons que siguen més fàcilment predictibles. La metodologia s'aplica al cas d'estudi de l'Ajuntament de València. Es disposa de dades reals dels consums municipals per al període de 2017 i 2018. Aquests consums es classificaran per tipus i es tractaran per a posteriorment facilitar el procés d'entrenament de les xarxes neuronals i que aquestes oferisquen la predicció més encertada possible. La predicció de cadascun dels tipus de consum es farà per separat, atés que cadascun depén d'uns inputs diferents. Posteriorment, s'agregaran totes les prediccions individuals per a obtindre una predicció global del consum de l'ajuntament de València.

Paraules clau: Demanda elèctrica, Xarxes Neuronals Artificials, Administració pública, València, Predicció, Eficiència, Mercat elèctric.

Abstract

Electricity demand forecasting is one of the most important processes in the energy sector. From the operation of the systems to the management of energy purchases and sales in the different existing markets, predicting consumption is vital to avoid problems such as grid congestion that can lead to shortages or to avoid significant economic losses as a result of poorly planned purchases by agents in wholesale markets. In this TFM we propose a model for predicting electricity demand for public administrations based on a decomposition into types of consumption and the application of artificial neural networks (ANN). Usually, all administrations have a certain typology of consumption types; public lighting, offices, schools, traffic lights, etc., whose disaggregation can be useful to lighten the problem and observe patterns that are more easily predictable. The methodology is applied to the case study of Valencia City Council. Real data on municipal consumption is available for the period 2017 and 2018. These consumptions are going to be classified by type and treated to subsequently facilitate the training process of the neural networks so that they offer the most accurate prediction possible. The prediction of each type of consumption will be done separately, given that each one depends on different inputs. Subsequently, all the individual predictions will be aggregated to obtain a global prediction of the consumption of Valencia City Council.

Keywords: Electricity demand, Artificial Neural Networks, Public administration, Valencia, Forecast, Efficiency, Electricity market

Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Networks
CM	Cemeteries
DOP	Day Of Prediction
DS	Doctoral student
EME	Energy Mean Error
JE	Junior Engineer
LM	Libraries and museums
MAPE	Mean Absolute Percentage Error
MLP	Multi-layer perceptron
MSP	Metered supply point
PA	Period of analysis
PG	Parks and gardens
PL	Public lightning
PM	Public markets
PO	Public offices and working buildings
REE	Red eléctrica de España
SC	Schools
SP	Specialist professor
TH	Threshold
TL	Traffic lights and tunnel ventilation systems

Part I

MEMORY

1 Introduction and objectives

1.1 Motivation

For decades now, cities have been the center of work and social activity, as well as the place where most of the resources necessary for life are located: hospitals, universities, administration centers, etc. This make cities places with a high population density and, by 2050, it is estimated that over 65% of the world's population will live in cities (Urban Climate Change Research Network (UCCRN), 2018). As a consequence, cities, which only occupy around 2% of the planet's surface, are energy sinks, consuming two thirds of the world's energy and being responsible for 70% of carbon dioxide emissions (C40 cities, 2021; Wei et al., 2021). That high energy consumption is also present for municipalities, which own facilities like schools, offices and health centres that have to offer services to a huge population. Municipalities also own public lightning, traffic lights and other infrastructures that in big cities can represent a considerable share of the energy consumption for the municipality. This consumption constitutes an important economic expense for municipalities and an important part of their budgets (Ajuntament de Barcelona, 2020; Ajuntament de València, 2021; Ayuntamiento de Madrid, 2020). For this reason, some municipalities have started to procure electricity in wholesale markets to eliminate intermediary costs (Cambranos, 2019; Diario de Rivas, 2018; Radiotelevisión del Principáu d'Asturies, 2014). Others had more ambitious plans and have created municipal electricity retail companies not to only purchase electricity for municipal loads but also to offer this service to residential consumers (Barcelona Energía, 2021; Bristol Energy, 2021; Eléctrica de Cádiz, 2021; Hamburg energie, 2021). However, buying electricity in the wholesale markets has a large set of associated risks (Bartelj et al., 2010; Boroumand & Zachmann, 2012; Ojanen, 2002). One of the most important are penalties due to unbalances between the electricity bought and the actually consumed. (Alcázar Ortega et al., 2019; Cabello García, 2020; Carbajo, 2007). These imbalances can lead to important economic losses, as will be discussed later on this document.

1.2 Background

One of the methods to forecast electricity demand that has gained most popularity in the last years is Artificial Neural Networks (ANN) (Kumar et al., 2013; Li et al., 2015; Zhang et al., 1998). ANN's are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, imitating the way that biological neurons signal to one another (IBM Cloud Education, 2020). ANN can be trained with past input and output data and then forecast future outputs given only the inputs. Also, one strategy followed by some authors is dividing the electricity consumption by end uses (Farinaccio & Zmeureanu, 1999; Ghisi et al., 2007; Murthy et al., 2001). In (Escrivá-Escrivá et al., 2014) and (Escrivá-Escrivá et al., 2011), authors propose a method that combine the ANN strategy with dividing consumption into end uses. Some of the end uses considered are Heat pumps, only chillers systems, public lightning and some others. Nevertheless, this end-use decomposition is very difficult to apply to big consumers like municipalities, since they count on a lot of consumption points whose end uses are wide. In practice, conducting an end-use decomposition to a whole municipality would result in a deep study of the end uses that every building or municipal facility has and a big expense on measurement devices and a good data acquisition system. Hence, the point is obtaining a methodology that can offer as good results as combining end-use decomposition with ANN that is applicable to facilities which can have a lot of consumption points of different nature. To achieve this, in this TFM, a methodology to forecast municipalities' electricity demand based on the classification of Metered Supply Points (MSP) and the application of Artificial Neural Networks (ANN) is proposed. MSP classification is a similar approach to end use decomposition but instead of splitting consumption into end uses, consumption is divided by groups of MSP that share a similar load curve and similar types of day and are affected by the same variables (temperature, sunset time...etc). The most common groups of MSP in municipalities are public lightning, public offices and working buildings and schools, among others. Once municipality consumption is divided into groups of MSP, ANN's are applied to each of the groups. By performing the classification before the ANN application, the learning process is easier and forecast results are improved.

The novelty of this work lies in the following aspects:

- The methodology is applied to an actual case of study and can be replicated to forecast other municipalities' electric consumption. That could lead to public administrations having a deeper knowledge of their future consumption, facilitating demand-side strategies or even buying their own electricity in the wholesale markets, getting the corresponding savings and avoiding possible dependence on third parties.
- Municipality MSP are classified according to their load curve shape, variables influencing consumption and the different types of day.
- This classification is easier to perform for municipalities, since it offers good results without spending resources on a more complex measurement systems that is able to measure the consumption of all the end uses. Only a smart meter is required for every MSP. Since in most of the cases smart meters are mandatory, the measurement system doesn't represent an extra investment for the municipality.
- The methodology offers a big potential for the future, in the big data era, when all consumption will be monitored. The highest the data availability, the better performance is offered by the method, since the classification process can be done more precisely, improving ANN's training process and results.

1.3 Objectives

The objective of this TFM is proposing an energy forecast method replicable for municipalities that offers high precision without spending a lot of resources on a complex data acquisition system. The methodology is suitable for medium and large municipalities, with a number of MSP high enough to have reproducible consumption patterns. If municipality is small, consumption tend to be very variable and can have random behaviours that can make a forecast methodology difficult to success. To reach this general objective, the following set of specific objectives has been fixed:

- Study the evolution of the energy forecast methods.

- Describe the most important working principles of the wholesale market of Spain
- Explain how ANN work
- Apply the strategy described above to a real case of a study and evaluate the performance
- Draw some conclusions and analyse further work that can be done.

The methodology is used to forecast the electricity demand of a whole year of the municipality of Valencia. To do so, 525 actual load curves provided by the Valencia Council are used. The load curves represent a whole year consumption from the 2017-2018 period.

1.4 Organization of the document

2 Electricity Market

2.1 The MIBEL

Since the methodology that is going to be developed is going to be applied to a case study in Valencia, Spain, in this section, the basics of the *Mercado Ibérico de Electricidad* (MIBEL) are going to be explained. If the methodology wants to be applied to other countries different to Spain and Portugal, other markets should be studied. The following figure illustrates the different market operators that run markets across Europe (Nord Pool, Epex Spot and OMIE). The little icons are the different system operators existing in Europe.

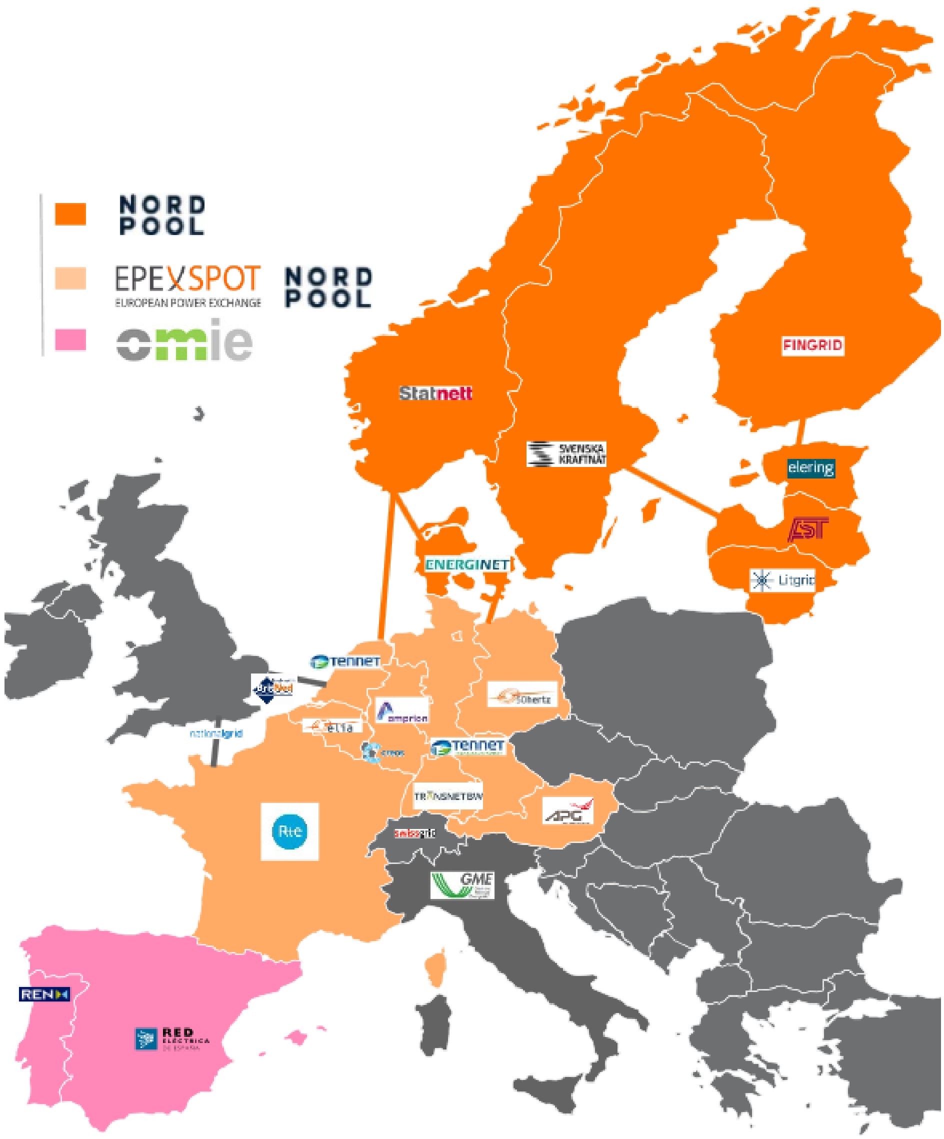


Figure 1: Different market and system operators across Europe (Magnus commodities, 2022).

Currently, the Spanish electricity markets are included in the Iberian Electricity Market, the MIBEL. The MIBEL is a set of organized and unorganized markets common for Spanish and Portuguese agents in which they have the possibility to carry out transactions or electricity contracts and trade different financial products (Alcázar Ortega et al., 2019).

The negotiation process for the constitution of MIBEL started in 1998 and subsequently, in 2001, the cooperation protocol was signed between the Spanish and Portuguese authorities for its creation. Three years later, the Santiago de Compostela agreement was signed, the terms of which were revised in 2006 at the XXII Luso-Spanish summit in Badajoz and in 2008 with the signing of the Braga agreement. MIBEL became fully operational on July 1, 2007 (MIBEL, 2021). With the operation of MIBEL, the aim is to equalize the electricity legislation in both countries, thus facilitating transactions, movements and actions of the companies and also to achieve more competitive prices for consumers.

In the future, the aim is to achieve a common European electricity market, and the European Commission is laying the foundations and setting the guidelines. Currently there is an alliance between several countries in Europe that are interconnected (PCR Alliance) and that use a unique algorithm called EUPHEMIA that calculates and assigns the daily market price in all the countries of the alliance maximizing social welfare and guaranteeing objectivity and transparency. The initiative started in 2009 but the cooperation and co-ownership agreement was signed by all RCP members in June 2012. The countries forming this alliance are: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Slovakia, Slovenia, Spain, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Romania, Sweden, Switzerland and the United Kingdom (OMIE, 2020).

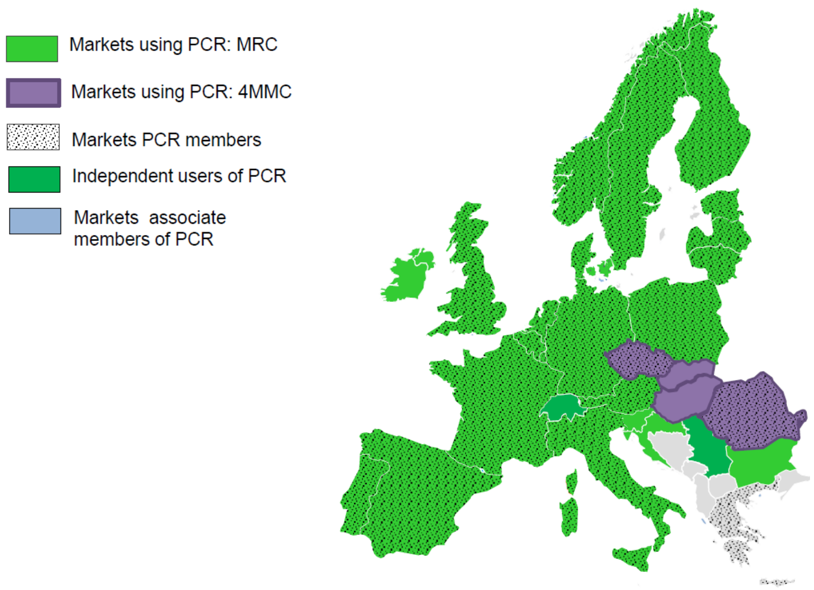


Figure 2: Countries joining the PCR alliance (Gestore mercati energetici, 2021)

In spain and in General in the electricity systems, two kind of markets exist: the retail market and the wholesale markets. On the one hand, the wholesale markets are those in which big amounts of energy are exchanged. These exchanges are done between big energy producers or generators and big consumers. Big consumers are usually retail companies who buy energy in the wholesale markets and then sell it to the final consumers. Actually, every consumer is able to buy energy in the wholesale markets. However, to do so they have to pay the so-called guarantees, a payment done to the Spanish market operator (OMIE), in order to ensure that the market participants will be able to respond economically to the undertaken obligations. That guarantees are high so, in practice, buying electricity in the wholesale markets is only an option for big consumers. These markets are common for Spain and Portugal. On the other hand, the retail market enables the exchange of little amounts of energy, usually between energy retailers and final consumers. This market is different for the different countries. Also, in this market consumers negotiate a price for the energy and they pay what they finally consume. Nevertheless, in the wholesale markets, buyers pay previously for a precise quantity of energy. When the energy is finally consumed, differences between the energy bought and the energy finally consumed (imbalances) are settled (Alcázar Ortega et al., 2019). Imbalances will be explained in Section 2.4.2.

2.2 The wholesale market

2.2.1 Forward or futures market

The forward market is based on a daily session in which operations can be negotiated from 4 years to 2 days before the exchange. The manager of these markets is the Portuguese market operator (OMIP), which in turn is divided into two entities: OMIP for the technical management of trading and OMICLEAR for the economic management (clearings).
Each futures market session starts at 8:00 and ends at 18:30 and consists of three phases (OMIE, 2021):

- Pre-trading phase (8:00 - 9:00): brokers can only make inquiries.
- Negotiation phase (9:00 - 17:00): Agents make offers to buy and sell.
- Pre-close phase (17:00 - 18:30): As in the first phase, brokers can only make inquiries.

In the forward market, sellers and buyers make public their purchases and sales offers on an electronic platform managed by OMIP. At the time of contracting a product on the forward market, the quantity of energy to be traded, the date of delivery and payment of the energy and the price to be paid for this energy are agreed upon. The products in the forward markets can be both physically or financially settled:

- Physical settlement: implies that on the day of exchange, the seller will deliver an amount of energy to the buyer, who in turn will consume that energy.
- Financial settlement: implies that on the day of delivery, the buyer and seller will not exchange any amount of energy, but will simply settle the differences between the price negotiated in the futures market and the actual price of the energy on the day of execution of the contract.

The better the agents' forecasts of the daily market price, the more efficient their operations in the futures market will be. For example, if subject A buys 1 MWh from subject B for €1/MWh on the futures market and at the time of delivery the daily market price is €2/MWh, subject A has saved €1 per MWh.
However, no payment or collection is made on the date of the agreement, but only once the energy has been delivered to the buyer. The futures contract, like any other contract, establishes obligations, so those who sign them incur penalties in the event of non-compliance, as commented above.

2.2.2 Daily market

The daily market, also known as pool, is managed by OMIE and offers to buy and sell energy are made the day before the exchange.
The purchase and sale bids are included in a matching procedure, which determines a different price for energy in each of the 24 hours of the following day, depending on the purchase and sale bids accepted in each hour. Once the bids for each of the hours of the following day have been submitted, OMIE orders the purchase bids from highest to lowest price and the sale bids from lowest to highest. With this process, the demand is covered first with the cheapest bids and then with the most expensive bids until the point where the supply and demand curves intersect.
Normally, at the lower part of the supply curve are producers that need to sell the energy they produce due to the impossibility of shutting down or storing it, as is the case of nuclear power plants. Plants such as pumped-storage hydroelectric plants would be located in a higher part of the supply curve, as they can store their resource to produce energy at another time that is more convenient for them.
The electricity market is a marginalist market; the point at which the curves intersect is the market price for each hour, which implies that the market price is always that of the last accepted bid, the most expensive generation one.

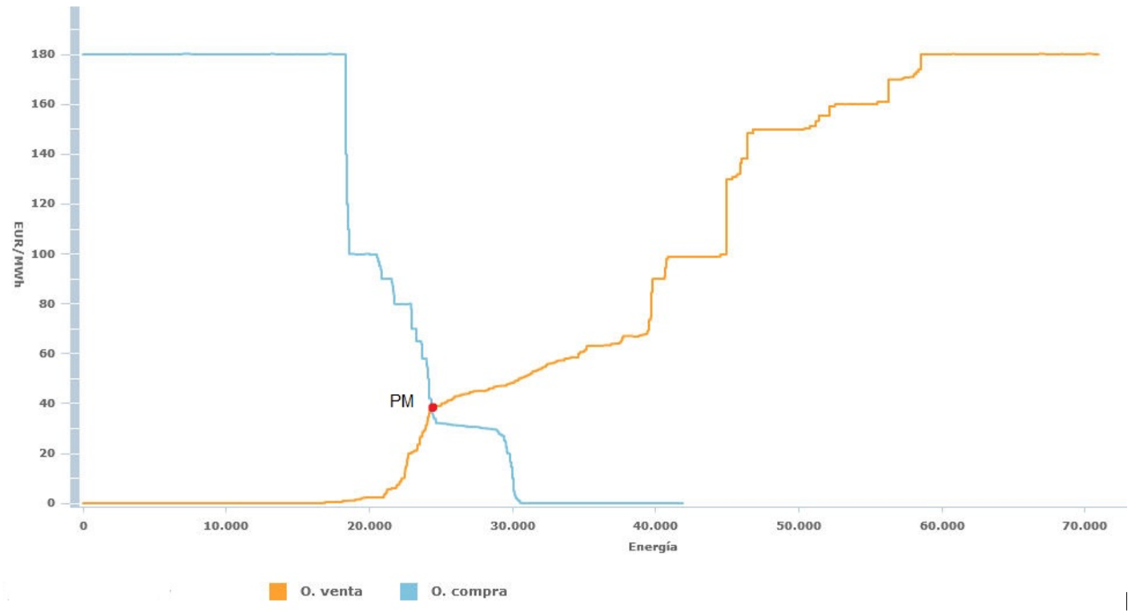


Figure 3: Supply-demand matching for any given hour.

Bids to the right of the matching point do not enter the market. Generators that have bid above the market price do not find a consumer to buy the power at that price. The same is true for consumers if they have bid their purchase at a price below the market price.

2.2.3 Intraday market

Once the quantities to be exchanged and the price of energy for each hour of the day D has been determined, buyers or sellers may notice changes in their energy predictions. In the intraday market, new purchase and sale offers can be made after the close of the daily market to balance the operations more accurately, thus avoiding deviations that can lead to significant costs. For example, if a trader had bought 1 MWh and after the close of the daily market his predictions show that 2 MWh will be consumed instead of the MWh he had previously calculated, the agent can make a purchase offer in the intraday market to balance his mismatch.
Bids may be placed up to a few hours before the exchange, with several sessions being held after the daily market: The following figure also shows the times for which trades are effective in

Table 1: Trading sequence in the intraday market (OMIE, 2021)

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Opening of the session	14:00	17:00	21:00	01:00	04:00	09:00
Closing of session	15:00	17:50	21:50	01:50	04:50	09:50
Matching and publication	15:07	17:57	21:57	01:57	04:57	09:57

each of the sessions. For example, trades submitted in session 1, which opens at 14:00, will be executed from 00:00 on day D until 23:00 on the same day D.

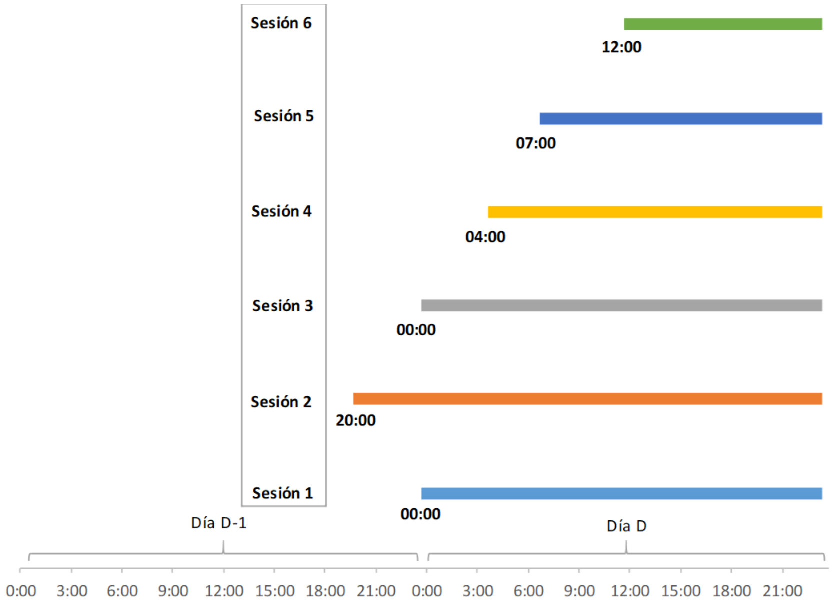


Figure 4: Intraday market session scheduling horizon. Prepared by the authors on the basis of (OMIE, 2021)

Prices in this market do not necessarily have to be more expensive than those of the daily market, as they obey a matching system equal to that of the latter so that the difference will depend exclusively on the purchase and sale bids made in one or the other market.
Apart from the intraday market by auction, there is a second option for adjusting bids, which is the continuous contracting or Single Intraday Coupling (SIDC) market, which allows energy to be traded continuously between different regions in Europe. This market currently includes 23 countries and allows adjustments to be made up to one hour before the delivery of energy. Both the intraday and day-ahead markets are managed by OMIE. The SIDC is managed by all market operators in its member countries.

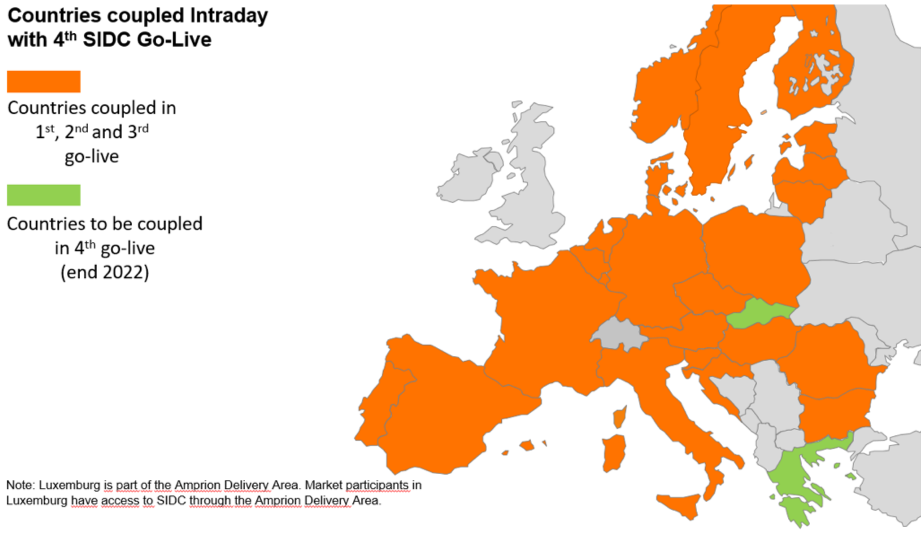


Figure 5: Current and expected members of the SDIC market.(Entso-e, 2021)

2.3 Wholesale market purchasers

In Spain, is considered that a qualified consumer is the one who is able to choose freely his electricity provider and negotiate with him the economic conditions of his contract. Since the year 2009, every consumer within the MIBEL is considered a qualified consumer. Hence, no matter the volume of the annual energy consumed, they can acquire electricity either through a electricity retailer or through the wholesale markets or even by signing a bilateral contract with a provider outside the regulated market. However, as explained on Section 2.1, these options are only profitable for large consumers.

In the last years, entities like municipalities have started to satisfy their own electricity demand by purchasing on the wholesale markets, especially on the daily market. In Spain, some examples of this are Aviles (Asturias) and Rivas Vaciamadrid (Madrid), with 79.000 and 86.000 inhabitants, respectively. The first city started operating in the day-ahead market in 2014 and the second did it in 2018. Both started acquiring the electricity for their most predictable consumption to not incur in imbalances and once they had enough experience, they extended it to most of their consumption points. Both cities have reported saving of around 20% a year from their initial bills (Cambranos, 2019; Diario de Rivas, 2018).

Barcelona and Cadiz had more ambitious projects and they started municipal retail companies. Barcelona, a 1,64 M inhabitants city is first city in Spain that created a new retail company, *Barcelona Energia*, that provides electricity to the municipal consumption and 20,000 customers through a public company, TERSA. The company buys the electricity and does the retail business but also serves as a seller of the municipal generating infrastructure and has a program to help its clients to install systems for self-generation. Barcelona does not offer cheaper prices but offers a more transparent service with counselling to their customers.

Cadiz, with 116.000 inhabitants, has a municipal retail company too. Since the mid-20th century, *Eléctrica de Cádiz* serves the municipal consumption and more than 60.000 customers. Its

experience allows the company to provide competitive prices and services, which have recently wined several prices regarding Energy Poverty and decentralised PV installations.
Other cities such as Pamplona and Palma de Mallorca have studied the possibility to create this similar companies (Som energia, 2016). However, both projects have not been implemented yet due to political and legal barriers.

2.4 Risks of the wholesale markets purchasers

Acquiring electricity in the wholesale markets is usually cheaper, since by doing it, no intermediaries like retailers are needed. Nevertheless, it has a set of associated risks that need to be known in order to avoid problems like having important extra costs or losing the right to operate in the market, which might lead to shortages.

2.4.1 Guarantees

In order to ensure the correct payment to the corresponding market agents, the Market Operator and the System Operator require to every agent participating in the wholesale market to pay a deposit called guarantees. Not fulfilling these payments, which are done weekly, can end up in fees to the buyer and even its disqualification from the market.

2.4.2 Imbalances

Deviations are the difference between the energy that is scheduled in the different markets and the energy that is actually produced or consumed in real time. There are always deviations, since it is physically impossible for the system to generate or consume exactly what was scheduled initially.
The system operator has to ensure that at all times the energy produced is equal to the energy consumed, otherwise there could be problems with the frequency values of the grid and another operation problems. To do this, market operator uses adjustment services offered in the operating markets, which have a cost that the system has to pay. The cost of these adjustment services is covered only by the agents that deviate in their predictions based on the system state. Deviations can go against or in favour of the system:

1. In favour:
   When the whole system is generating more energy than it consumes (surplus system) and purchasers consume more energy than they bought on the market. Purchasers can also deviate in favour of the system when it is consuming more energy than is generated (deficit system) and they consume less than expected. In both situations, a contribution is made to the generation-demand compensation in the system, so that the agents don't receive any penalty.

2. Against:
   When the system is in surplus state, consumers consume less than expected and when the system is in deficit state, consumers increase their consumption. In both cases, the generation-demand gap increases, so agents are penalised.

Table 2: Types of imbalances summarized. Own elaboration

System status	Consumer action (compared to what they initially bought)	Type of imbalance	Penalty
Surplus	They consume less	Against	Yes
Surplus	They consume more	In favour	no
Deficit	They consume more	Against	yes
Deficit	They consume less	In favour	no

3 Electricity load forecasting

3.1 Importance of electricity load forecasting

Forecasting models are widely used in different fields (Kuster et al., 2017a); For example, in the finance area to forecast stock exchange movements or different indices of the stock markets (Bianco et al., 2009), in business to schedule staff, manage inventory and predict demand (Intergovernmental Panel on Climate Change, 2014), in medicine to monitor the spread of diseases (Generous et al., 2014), and in meteorology for predicting weather.

In the electricity field, forecasting is esential and very used. Load forecasting can offer useful information for utilities that they can use to make important decisions like purchasing or generating electricity, switching loads or decisions on infrastructure development (Singh et al., 2012a). Is considered specially important in the operation of power systems. Since system operators need to schedule enough electricity to meet demand at every moment, accuracy of load forecasts has considerable effects on the operational economies of the system. Indeed, operation costs can be largely increased due to forecasting errors (either positive or negative)(Haida & Muto, 1994).

Load forecasting is critically important when talking about large consumers or purcharsers in the wholesale markets. As was discussed on Section 2.4.2, umbalances between the energy bought and the actually consumed can be penalised if that umbalance goes against the system. In (Mateo Barcos et al., 2020), a quantification of costs due to umbalances is done, assuming different percentages of hourly imbalances from a retailer that is purchasing electricity in the Spanish daily wholesale market. This calculation was done considering that every imbalance was against the system, which is the type of imbalance that has economic penalties in the Spanish electricity market (Alcázar Ortega et al., 2019).The study was done for the period from 2017 to 2018, so market prices were the ones from that period and therefore also cost overruns.

Table 3: Volume of the hourly imbalances assumed (Mateo Barcos et al., 2020)

Hourly imbalances (%)	Hourly imbalances (MWh)
5%	0,54
10%	1,09
15%	1,63
20%	2,18
25%	2,72
30%	3,26
35%	3,81

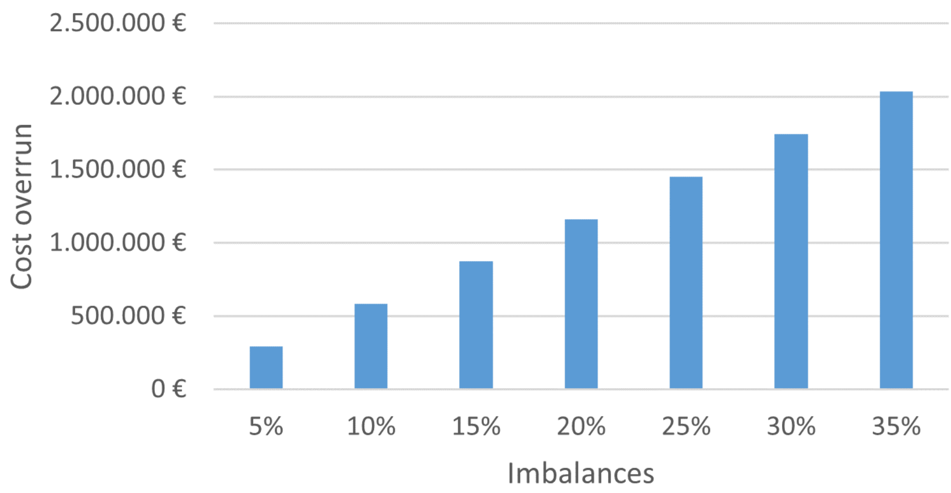


Figure 6: Effect of the assumed imbalances over costs (Mateo Barcos et al., 2020)

Figure 6 show that imbalances play an important role when managing energy acquisition in the wholesale markets. As it can be seen, cost overrun can reach 500.000€ with hourly deviations from the purchased energy of a 10%. Therefore, it is essential for municipalities to count on good tools to forecast their demand, not only to avoid important losses, but also to plan demand-side strategies (Moghaddam et al., 2011; Pina et al., 2012; Strbac, 2008).

3.2 Evolution of electricity load forecasting methods

Forecast methods have evolved a lot in the last decades, from simple methods to the application of complex AI algorithms and hybrid method that mix different techniques.
In the equator of the XX century, as more and more electric appliances, such as electric iron, radio and electric washer, were being invented and became popular, the forecasting problem gradually

turned to be non-trivial. In 1940's, it was observed that electricity demand was very importantly affected by weather, especially because of the penetration of air conditioners. Because there were no statistical software packages at that time, an engineering approach was developed to manually forecast the future load using charts and tables. Some of those elements, such as heating/cooling degree days, temperature-humidity index, and wind-chill factor, are inherited by today's load forecasting models. The similar day method, which derives a future load profile using the historical days with similar temperature profiles and day type (e.g., day of the week and holiday), is still used in many utilities' operations centers. Indeed, is used in this TFM to improve the performance of the forecast.

From the 1980s onwards, as computer applications started to appear and became popular in within the engineering community, a significant amount of research was dedicated to long term spatial load forecasting, which consist of predicting when, where and how much load growth will occur. The forecasting horizon ranges from several years to several decades. These forecasts have been widely used in transmission and distribution planning.

Later, short term load forecasting started to gain more and more importance. Researchers first tried to apply statistical techniques, such as regression analysis and time series analysis. Then Artificial Intelligence (AI) became one of the most discussed methods in the scientific community(Hong, 2013). The models based on AI techniques, such as ANN, fuzzy logic, and support vector machine, which are black-box models, are very used today and will be discussed on the following section.

3.3 Currently used forecasting techniques

There are multiple methods used in the field of electricity load forecasting. Some of the most used are the Autoregressive Integrated Moving Average (ARIMA), the Autoregressive Moving Average (ARMA), linear and multiple regression methods, Support Vector Machines (SVM) and Artificial Neural Networks (ANN). ANN will be explained deeper on Section 4.

- ARMA and ARIMA. These methods were introduced in 1970 (Box et al., 2015). The basic ARMA model is composed of an autoregressive model (AR) and a moving average model (MA). The autoregressive model is a linear regression of the current value based on one or more previous values. Just as an AR, the MA is a linear regression, at the difference that it regresses current values against the white noise or error of one or more past values (Kuster et al., 2017b). However, this model can only be accurate if the time series is stationary. If the process is dynamic, then ARIMA is used and transformation of the series to the stationary form is done first (Singh et al., 2012b).

- Regression analysis. This method is present in a lot of forecasting processes. The dependent variable or output can be defined by other independent variables. Linear regression links the output to the independent variable by the simple linear model:

$$y = \beta_0 + \beta_1 \cdot x + \epsilon \tag{1}$$

  Where β_0, β_1 and ϵ are the intercept and the slope of the line and the random "error", respectively (Kuster et al., 2017b). The difference between simple and multiple linear regression is the number of variables introduced as independent variables ranging from one variable in the simple model to several in the multiple.

- Support vector machines.This method was firstly introduced in 1995 (Cortes & Vapnik, 1995). The power of an SVM stems from its ability to learn data classification patterns with balanced accuracy and reproducibility. Although occasionally used to perform regression, SVM has become a widely used tool for classification, with high versatility that extends across multiple data science scenarios, including brain disorders research. An SVM decision function is more precisely an optimal "hyperplane" that serves to classify observations belonging to one class from another based on patterns of information about those observations called features. That hyperplane can then be used to determine the most probable label for unseen data (Pisner & Schnyer, 2020).

3.4 Error measures used in the forecasting methods

In order to evaluate the performance of a forecast method, different kind of error measurements are used. A wrong or incomplete selection of the error measure can lead to a inaccurate evaluation of the forecasting results. Usually, at least 2 error measures are chosen to perform a good performance analysis. In this section, different kind of error measures are explained (Shcherbakov et al., 2013). They can be divided in the following groups:

Absolute Forecasting Error: The first group is based on the absolute error calculation. They are all based on the calculation of the error e_h:

$$e_t = (y_t - \hat{y}_t) \tag{2}$$

Where:

- y_t is the measured value at time t.
- $\hat{y}_t$ is the forecasted value at time t.

This group of errors include, among others, the Mean Absolute Error (MAE), the Mean Square Error (MSE) and the Root Mean Square Error (RMSE). These errors are very used in various fields. However, they have the following limitations:

- Scale dependency (Hyndman & Koehler, 2006). If the forecast includes objects with different scales, absolute error can't be applied to obtain reliable results.
- High influence of anomalous data. If data contains errors like extremely high or low values, absolute error will offer a very altered value.

Measures based on percentage errors. These errors are all based on the calculation of p_t:

$$p_t = \frac{|y_t - \hat{y}_t|}{y_t} \tag{3}$$

This is the most used kind of error in the forecasting field. Among this group, the most extended one is the Mean Absolute Percentage Error (MAPE), which is one of the errors used in this TFM, as it was stated on 5. Its formula is presented on Equation 7. The limitations of the MAPE are the following ones:

- A division by zero can appear when y_t (the actual value of the magnitude) is zero.
- Non-symmetrical problem. The error values are different whether the forecasted value is bigger or smaller than the actual. In fact, predicted values ($\hat{y}_t$) below the actual one (y_t) can only offer a maximum MAPE of the 100%. However, values above the actual one are limitless (Goodwin & Lawton, 1999).
- Magnitude dependency. Deviations from the actual value of little importance in the forecasting process can offer a big MAPE that doesn't reflect reality. For example, if daily energy consumption wants to be predicted, it may happen that, for some energy uses, like lightning, that some hours in the day have consumption values near to 0, as it can be seen:

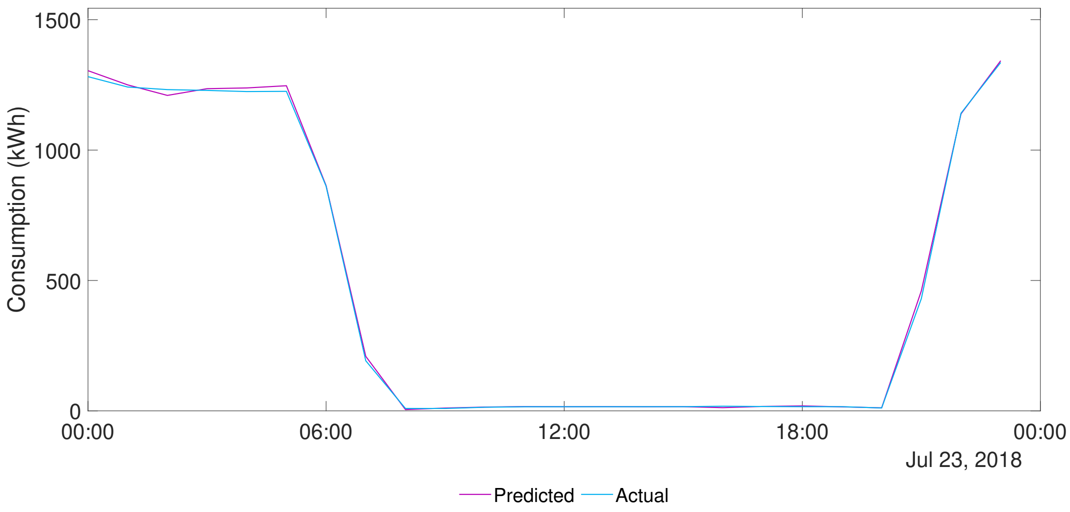


Figure 7: Typical lightning daily consumption profile.

As it can be seen in Figure 7, consumption during day hours is close to zero, while during the night, consumption reaches its maximum. The MAPE for the whole day, with the predicted value, is represented in the next figure:

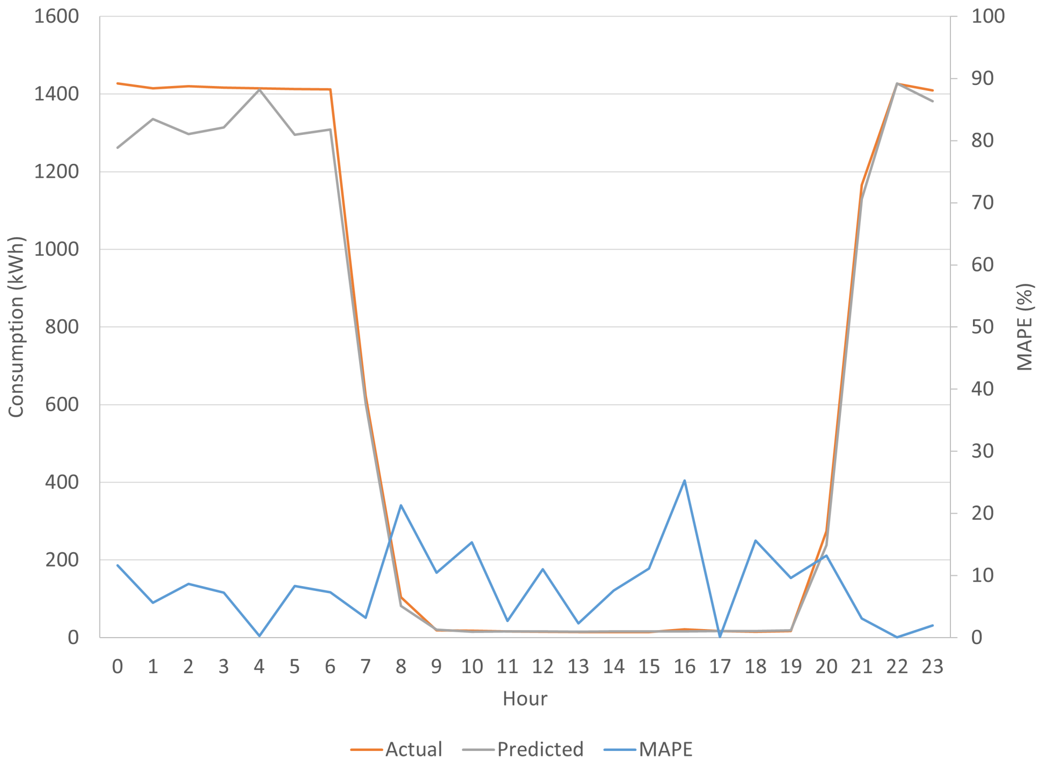


Figure 8: Behaviour of the MAPE for the evaluation of a daily forecast in PL.

Even though errors during the day hours are almost insignificant in terms of volume of energy, MAPE is higher during these hours than during the night. That could lead to think that the forecast is poor, which might be far from reality. Nevertheless, when calculating the MAPE for longer periods of time, like for example a whole week, this effect reduces its influence.

Scaled error: Scaled errors are based on the calculation of q_t:

$$q_t = \frac{|y_t - \hat{y}_t|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|} \tag{4}$$

Within this group, two main errors can be found; the Mean Absolute Scaled Error (MASE), which is calculated as the mean of q_t for the time horizon of the forecast evaluated and the Root Mean Scaled Error (RMSSE):

$$RMSSE = \sqrt{mean_{i=1,n}(q_i^2)} \tag{5}$$

Besides the MASE and de RMSSE, EME (Equation 8) could also be included in the category of scaled errors. This measure has been found in (Escrivá-Escrivá et al., 2011; Escrivá-Escrivá et al., 2014). EME is similar to q_t and is especially useful for the field of load forecast, since it calculated the forecast error taking into account the volume of the energy consumed during the time horizon of the forecast. That way, the issue that occurs on the calculation of the MAPE (Figure 8), doesn't take place with the EME. The main problem of the scaled errors is that, again, division by zero can occur. From all the error explained in this section, MAPE and EME have been chosen. MAPE is going to be used since most of researchers have measured the error of their forecast methodologies with it, so it would be easier to compare results of different methodologies with the one presented on this TFM. On the other side, EME has been chosen because it takes into account the volume of the energy consumed during the time horizon of the forecast, being a good support to the MAPE measure.

4 Artificial Neural Networks (ANN)

In this chapter, a basic explanation of how neural networks work is done. The explanation is dedicated to the most common type of ANN ensemble in the field of electric forecasting that is also the one used in this TFM; the multilayer perceptron (MLP) feed forward network.
The MLP consists of multiple layers of neurons that interact with each other using weighted connections. There is always an input and output layer, which are connected by a certain number of hidden layers, usually one or two. There are no interconnections within a layer. However, all neurons in a layer are fully connected to neurons in adjacent layers. Weights measure the degree of correlation between the input of a neurons and the output (Pal & Mitra, 1992). Figure 9 illustrates the shape of a MLP ANN:

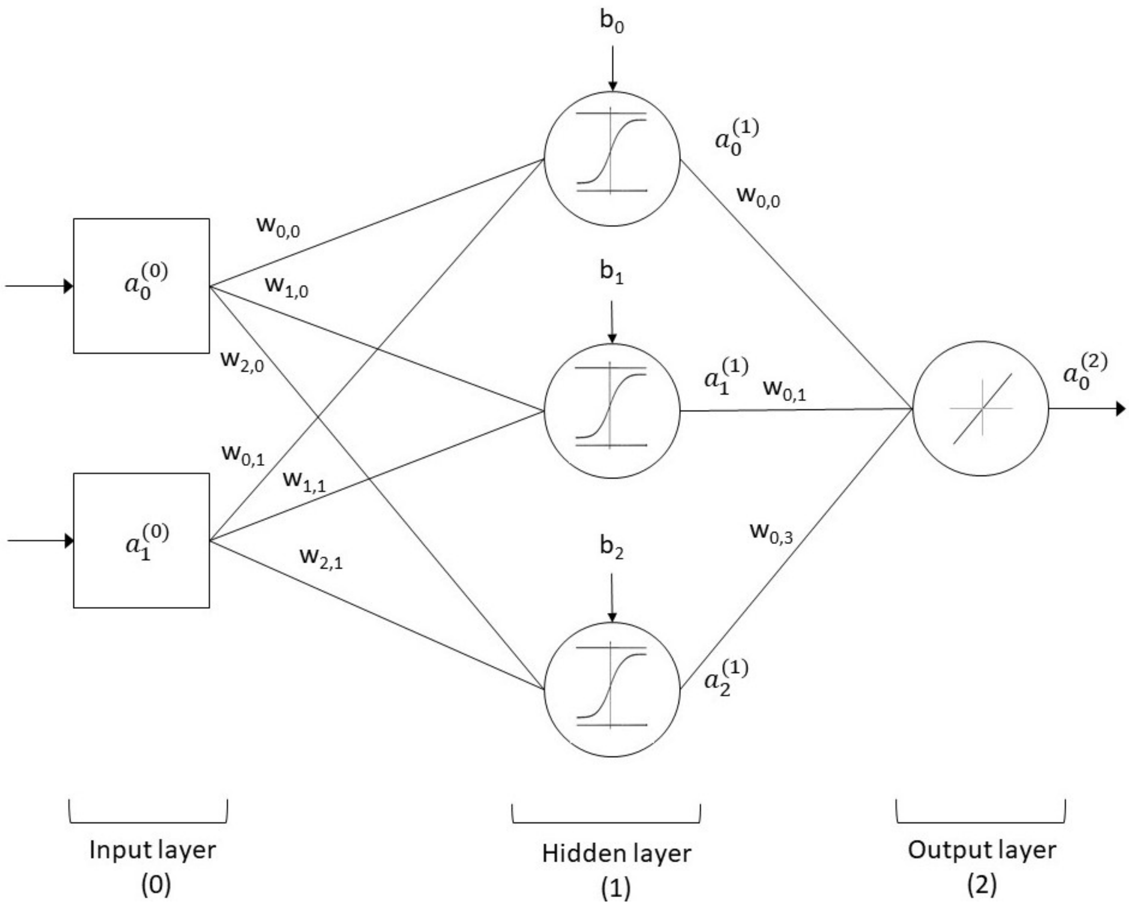


Figure 9: Scheme of a three-layer neural network with 3 neurons in the hidden layer. Own elaboration

The basic unit of an ANN ensemble is the artificial neuron, schematically represented in Figure 10:

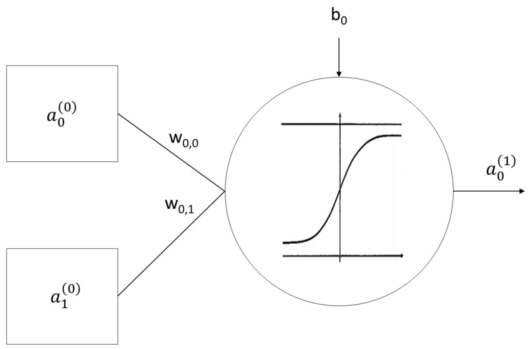


Figure 10: Example of a neuron within a ANN ensemble. Own elaboration

In the example, input layer and output layer are represented by a superscript ranging from (0) for the input layer to (1) for the output layer. The neuron receives input data from a certain number of input nodes, processes it internally and produces a response or output. The processing is usually done in two stages: first, the input values are linearly combined and then the result is used as the argument of a nonlinear function called the activation function. Every connection between input node and neuron has a weight ($w_{i,j}$) and every neuron has a certain bias associated (b_i). Hence, the output of a neuron is calculated as the weighted sum of the inputs passed through the activation function, as follows:

$$a_i^n = f\left(\sum_{j=0}^{N} w_{i,j} \cdot a_i^{n-1} + b_i\right) \tag{6}$$

where:

- a_i^n: Output of the ith neuron of the n layer
- $w_{i,j}$: Weight of the connection bewtween the ith and jth neurons
- b_i: Bias of the ith neuron of the n layer
- N: number of neurons in the n-1 layer

The activation function must be a nondecreasing and differentiable function; the most common choices are either the identity function (y = x) , or the sigmoid function:

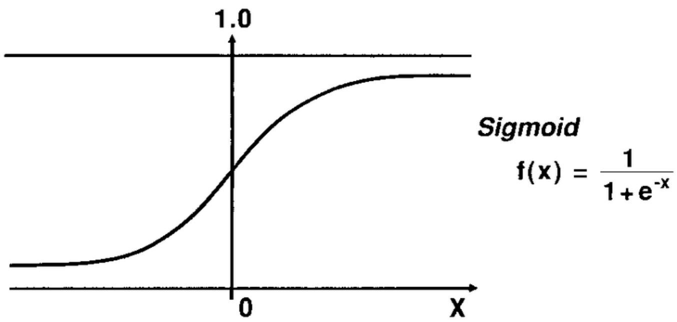


Figure 11: Sigmoid function.(Næs et al., 1993)

Since the range of the sigmoid function is [0,1], using it is a guarantee that the output of this unit will always be between 0 and 1. Also, as the sigmoid function is a non-linear function, the output of this unit would be a non-linear function of the weighted sum of inputs Saeed, 2021. Once the basic working principle of neurons is explained, it is important to know how are weights and biases calculated, since these are the parameters output of the network will depend on. This is done through a process called training. In this process, a set of experimental records in which inputs and outputs are known is used. This set is called the training set. Inputs are presented to the input layer of the network. This data is then multiplied by a weighting factor and fed forward to the nodes of the first hidden layer. The weighted outputs from the input layer are summed and transformed by the activation function, also called transfer function, as Equation 6 shows. The resultant output is also weighted and fed to the subsequent layer and so on. The output from the output layer represent the forecast of the dependent variables of the model. In this case, the electricity consumption. Then, outputs of the model are compared with the actual values, obtaining the prediction error from which a performance function (often the mean squared error) is derived. The performance function is used by a backpropagation training algorithm to adjust the weights and biases applied at the connections between nodes and to the neurons in each layer. By iteratively presenting the training set to the network and adjusting the weight and biases matrix, the performance function can be minimised. In this way, the network learns the relationships between ingredients and properties and develops a model capable of predicting the properties of any formulation/process problem lying within the model space (Plumb et al., 2005). The most common backpropagation training algorithm and the one used in this TFM is the Levenberg-Marquardt (Moré, 1978; Ranganathan, 2004).

5 Methodology

The methodology followed to forecast the electricity consumption of a municipality consists of three parts and it's summarized in the next figure:

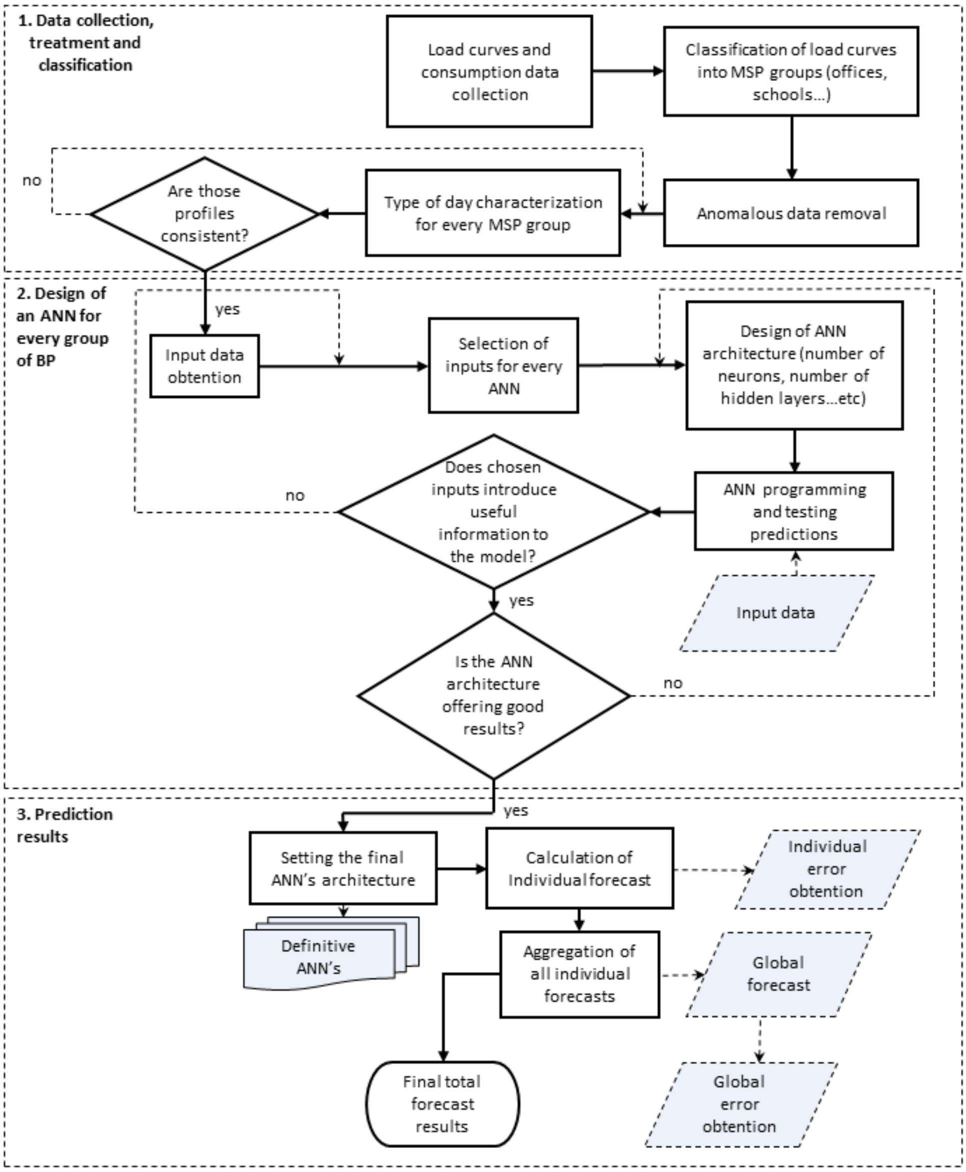


Figure 12: Methodology block diagram

1. Data collection, treatment and classification.

First, it is necessary to obtain as much consumption data as possible from at least one year, in order to appreciate the seasonal behaviour of demand so that ANN's can be trained properly. Granularity of data is also important. For the method proposed, at least a one-hour granularity is required. Also, consumption data should be as specific as possible. That means, having a repository of load curves from specific MSP instead of having consumption inventories that put together data from several consumption points. Furthermore, is necessary to classify every load curve available of every MSP by groups. As noted above, this groups should share a similar consumption scheme and variables influencing consumption. Also, the different types of day should be the same. In municipalities, some of this groups are usually public lightning, public offices, working buildings and schools, among others. Also, it is important to collect data from the potential inputs that might be introduced to the model. Some of these potential inputs are hourly temperatures, previous electricity consumption and sunrise and sunset time for the period of analysis. With this data, other inputs can be set, like daily average temperature.
Once data is collected, a removal of anomalous data should be done, with the aim of excluding from the analysis noise that could affect the performance of the forecast. Examples of anomalous data are sudden values drops to zero or extremely high values.
After collecting and filtering the available data, consumption data must be classified into different type of days for every group of MSP. A typical type of day classification consists of treating weekdays and weekends separately. However, depending on the MSP group, other types of day might be useful for the analysis. This classification should be done by observing and analysing the consumption patterns for every group. In figure 13, consumption behaviour for three different types of days is shown.

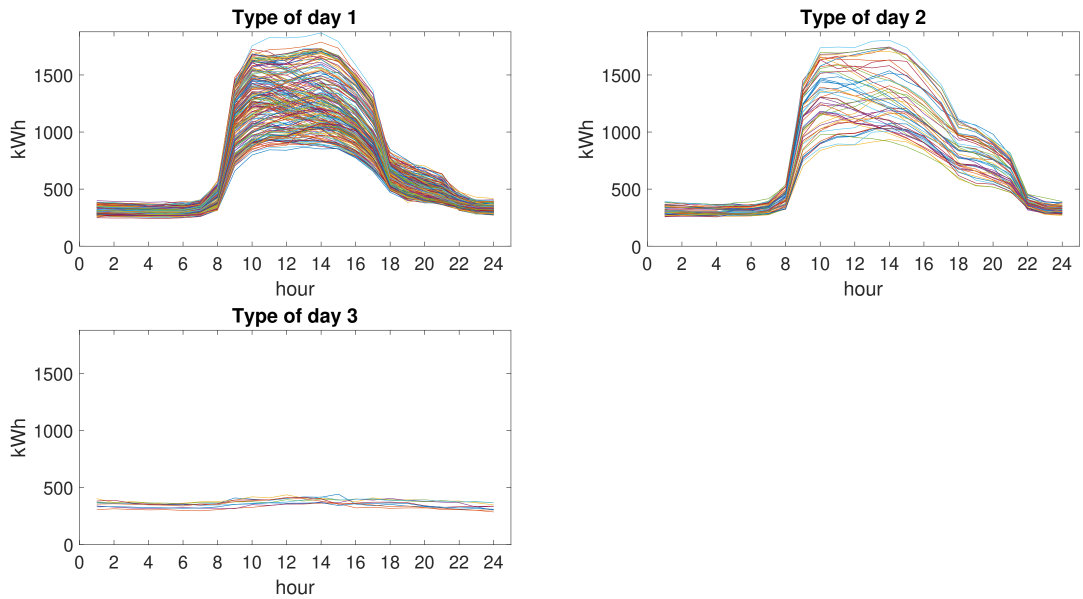


Figure 13: Type of day consumption daily profiles for an example MSP group

Also, it can be seen that daily load curves of the same type of day are quite different in magnitude. Since all the days of the year of the same type of day have been plotted, magnitude of consumption changes depending on the season. Climate equipment consume more in cooling mode (summer) than in heating mode. That's why magnitude of consumption is different. That way, even though the shape of the daily load curve is the same, the amount of energy consumed increases or decreases depending on the moment of the year.

2. Design of ANN's for every MSP group.

Step two consists of the design of ANN's for every MSP group. The first part of the ANN design is the preselection of inputs. Since almost all of the MSP groups represent different types of buildings, most inputs correspond to external temperatures and previous electricity consumption as both have a major correlation with future consumption (Roldán-Blay et al., 2013). Some authors also use outdoor solar radiation (Mena et al., 2014). In (Akarslan & Hocaoglu, 2018), the actual date, time and electricity consumption are employed as inputs. Nevertheless, to check whether the selected inputs for every MSP group would introduce useful information to the model or not, the relation between the possible inputs and electrical consumption has been analysed. For example, in figure 14, the influence of the electrical consumption of the previous hour to the hour of prediction (t) and the temperature at t over consumption is shown for an example MSP group. We can see that consumption at t and t-1 are linearly related. On the other side, we can also appreciate the influence of cooling and heating systems over electrical consumption. As temperature increases or decreases above or below 20ºC, heating and cooling systems start to work and consumption rises. Therefore, these two inputs will be introduced to the model for this particular MSP group. On the other side, some inputs are not related with consumption, as we can see in figure 15. In this case, the represented inputs are the sunrise and sunset time. For some MSP groups like public lightning, these inputs will have an important influence over consumption. However, for this example MSP group in the figure, the represented inputs and consumption are not related. Hence, these two inputs won't be introduced to the model.

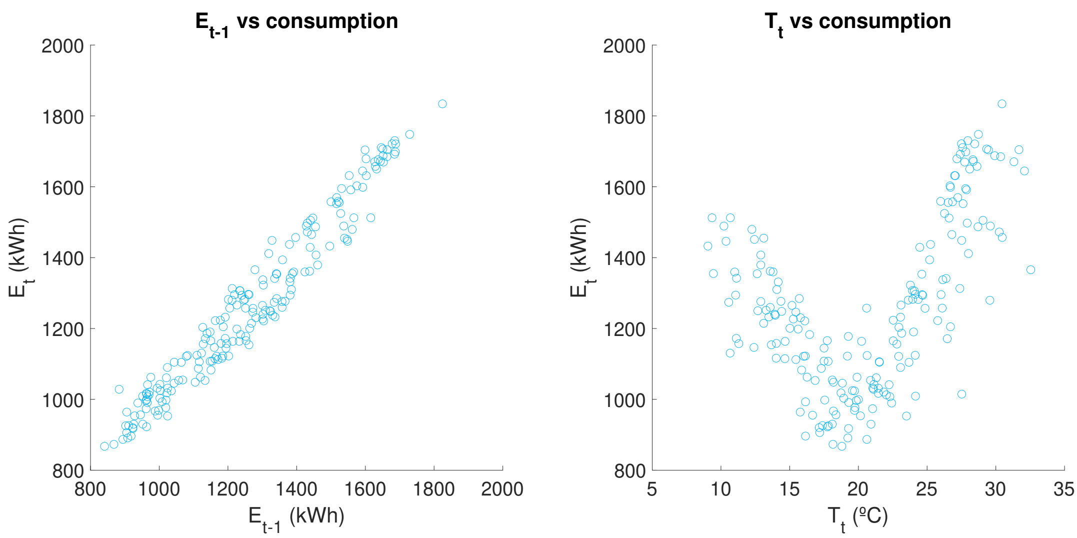


Figure 14: Influence of variables over electrical consumption.

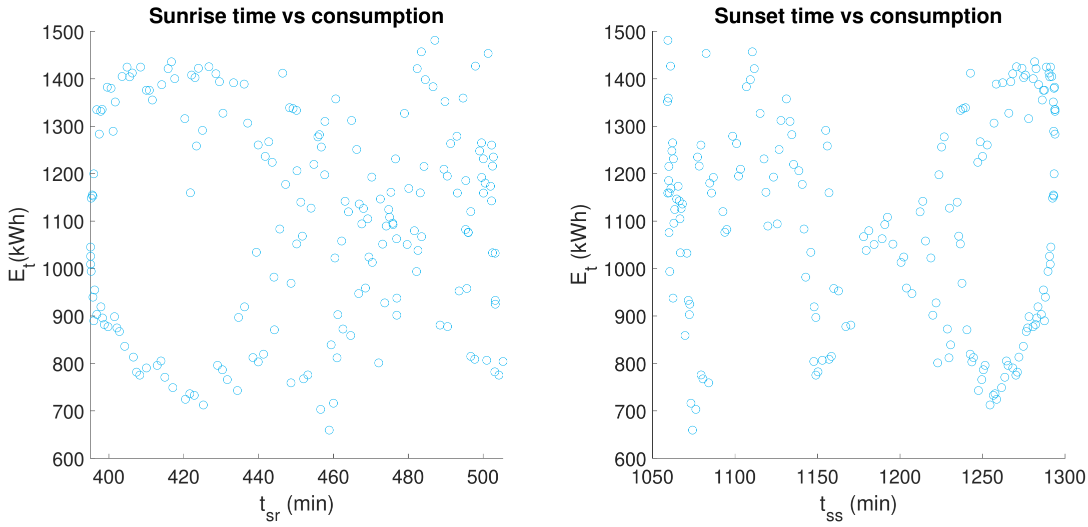


Figure 15: Influence of variables over electrical consumption (non related).

As inputs are defined, we start defining the ANN architecture itself. During this phase we make an initial choice of parameters such as number of hidden layers, number of neurons, training algorithm and activation functions (Mohandes et al., 2019). These parameters were explained on Section 3 and will be optimized after checking ANNs results and behaviour, since deciding ANNs architecture is usually an iterative process.

After several tests and bibliography revision, the ANN architecture is the one presented in Table 7.

Table 4: ANN architecture

Parameter	value
Type of network	Feed-forward
Number of hidden layers	1
Number of neurons in hidden layers	5
Training algorithm	Levenberg-Marquardt
Activation function	Sigmoid function

3. Prediction results

Once the training phase is completed, we feed the hourly networks with new input data, which corresponds with the data for the day of prediction (DOP). Then, we sum the hourly results to obtain a full day or a several days prediction. To check the network performance, we use the

Mean Absolute Percentage Error (MAPE) and the Energy Mean Error (EME).

$$MAPE = \frac{\sum_{t=1}^{N} \frac{\left|\hat{E}_t - E_t\right|}{|E_t|}}{N} \cdot 100 \quad (7)$$

$$EME = \sum_{t=1}^{N} \frac{\left|\hat{E}_t - E_t\right|}{\sum_{t=1}^{N} E_t} \cdot 100 \quad (8)$$

Where:
- N is the total number of hours within the period that wants to be predicted.
- $\hat{E}_t$ is the predicted consumption during the hour t.
- E_t is the actual consumption during the hour t.

At that point, we observe the results graphically and analytically. If prediction is not working properly, we check whether the architecture of the ANN or the input selection could be optimized. Finally, annual consumption for the different MSP groups is forecast. Then, all these individual yearly predictions are added to obtain the global prediction of the municipality. Yearly forecast performance is also evaluated through equations 8 and 7.

6 Case Study

In this section, the methodology explained on section 5 is applied to the case of study of Valencia's city municipality

6.1 Data collection, treatment and classification.

6.1.1 Data collection

First, data of actual electricity consumption of the Valencia municipality is collected. Data consists of a general register in Excel datasheet format which contains information about the 1.558 MSP owned by the Valencia municipality during the period 2017-2018 provided by the electricity supply contracting service of the City Council itself: CUPS [1], MSP group, location, contracted power, billing data for a specified period, etc. On the other hand, there are 888 Excel files which are the load curves with a 15 minutes granularity of some of the MSP that appear in this general register.

Since there are 1.558 MSP registered but 888 load curves, it is convenient to know to which MSP group these load curves belong. For this purpose, a script in matlab has been programmed, which reads all the data from the general register and at the same time reads the CUPS of the load curves datasheets. If the CUPS of the load curve examined coincides with a CUPS from the general file, it reads the MSP group that this load curve belongs to and add it to the rest of the load curves of the same MSP group.

In order to observe the effects of seasonality over consumption, a data period of at least one year is desired. However, not all load curve measures cover the same time period, so it is necessary to determine which time period is the one with more measures available. After analysing multiple load curves, it is observed that the most common period of measures among them is from 1/9/2017 to 1/9/2018. To check how many load curves have measurements in the described period, a script in Matlab has been programmed, which is given the start and end dates to be checked as input. The script stores the data from the general register, examines each of the load curves, checks if they have data in the period that has been specified and if so, stores its CUPS. If, on the other hand, it has no data on the period of analysis, it does exactly the same. After proceeding in this way with all the load curves, it obtains the number of meters that have and do not have data in the period described.

Finally, 525 load curves are available for the study from September 2017 to September 2018, which is the period of analysis (PA) established. Load curves of that time period represent a total consumption of 49 MWh.

6.1.2 MSP classification

One of the first tasks is regrouping the load curves by MSP groups. Initially, in the general register mentioned above, MSP were classified in 15 type of MSP groups and so were load curves. However, this number has been reduced to 8 by grouping some of the initial MSP groups according to its characteristics, schedules, uses of energy and most important variables affecting

[1] Universal Supply Point Code, a code to identify every supply point in Spain

consumption. Finally, the groups established are public lightning (PL), schools (SC), public parks and gardens (PG), public offices and working buildings (PO), cemeteries (CM), public markets (PM), libraries and museums (LM) and traffic lights and tunnel ventilation systems (TL). All the load curves available for these MSP groups have been added, obtaining a single load curve for every MSP group. The number of load curves available for each different MSP group is presented on table 5:

Table 5: MSP groups

n	MSP group		Available load curves
1	Public lightning	PL	144
2	Schools	SC	18
3	Parks and gardens	PG	56
4	Public offices and working buildings	PO	78
5	Cemeteries	CM	3
6	Public markets	PM	5
7	Libraries and museums	LM	3
8	Traffic lights and tunnel ventilation systems	TL	218
TOTAL			525

6.1.3 Anomalous data removal

Removing anomalous data is important in order to avoid ANN training with data that may introduce errors in the later prediction. Data considered anomalous is eliminated by two methods:

- Simple observation of load curves: First, consumption of every MSP group is calculated daily. When representing daily consumption for the whole PA for every MSP group, data that is obviously anomalous is identified and removed, like zero consumption days or extremely high and sudden consumption days.

- Average and thresholds approach. As in the previous approach, daily consumption is calculated. Then, a script is programmed in Matlab. This script analyzes each day individually and calculates the average consumption of the 7 previous of the same type of day. A threshold (TH) is established above and below the calculated average. If consumption of the day is greater or lower than the thresholds, consumption of that day is considered anomalous and is removed. Threshold is calculated as a percentage of the seven previous days' average, depending on the variance of the consumption of the MSP group analyzed.

$$TH = \pm x\% \cdot mean_{DOP-7}^{DOP}(\sum_{h=1}^{24} E_h) \qquad (9)$$

Where x is different for every MSP group, depending on the variance of the consumption, as said above.

6.1.4 Type of day classification

In order to make easier the ANN learning process, performing a type of day classification for every MSP group on Table 5 is very important (Cancelo et al., 2008; Chen et al., 2009; Senjyu et al., 2005), since consumption patterns can change a lot from weekdays to weekends, for example.
To do so, a cyclic process has been followed: First, total load curves of the MSP groups are observed individually to see the consumption profile over the week. Then, an initial type of day classification is done according to that initial observation of the profiles. For example, if different consumption behaviours are observed between Mondays and the rest of the weekdays, they are considered as different types of day. If after this classification non-coherent consumption profiles are found, the type of day separation is rethought, until the minimum types of days that fairly represent the consumption behaviour is reached. For example, for a certain MSP group, two of the types of day initially created are the ones on Figure 16. As it can be seen, the consumption pattern is exactly the same so no differentiation is needed between them.

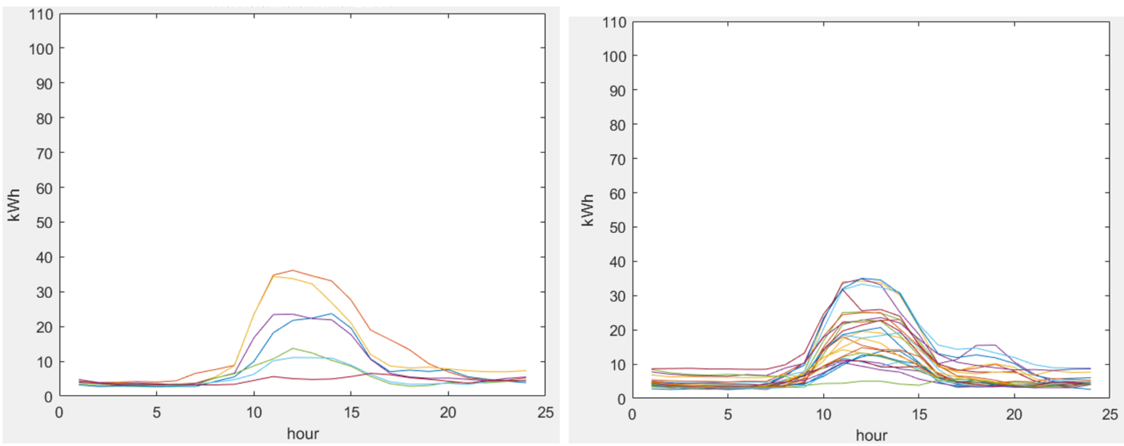


Figure 16: Similarities between consumption patterns for two different types of day of a generic MSP group.

In the following figures the different consumption profiles for some of the MSP groups that do have different behaviour depending on the type of day are shown.

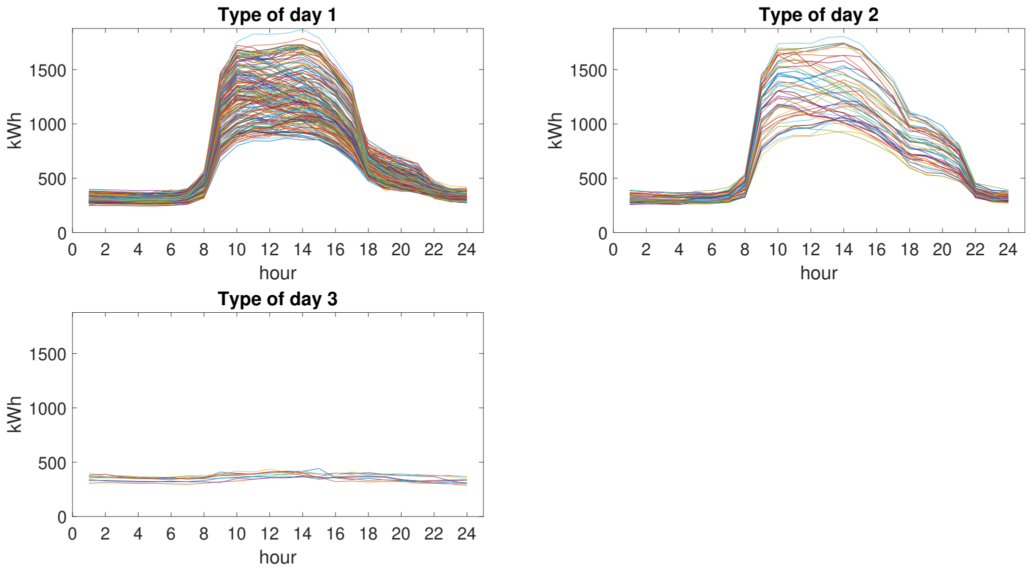


Figure 17: Daily consumption profiles for the PO group

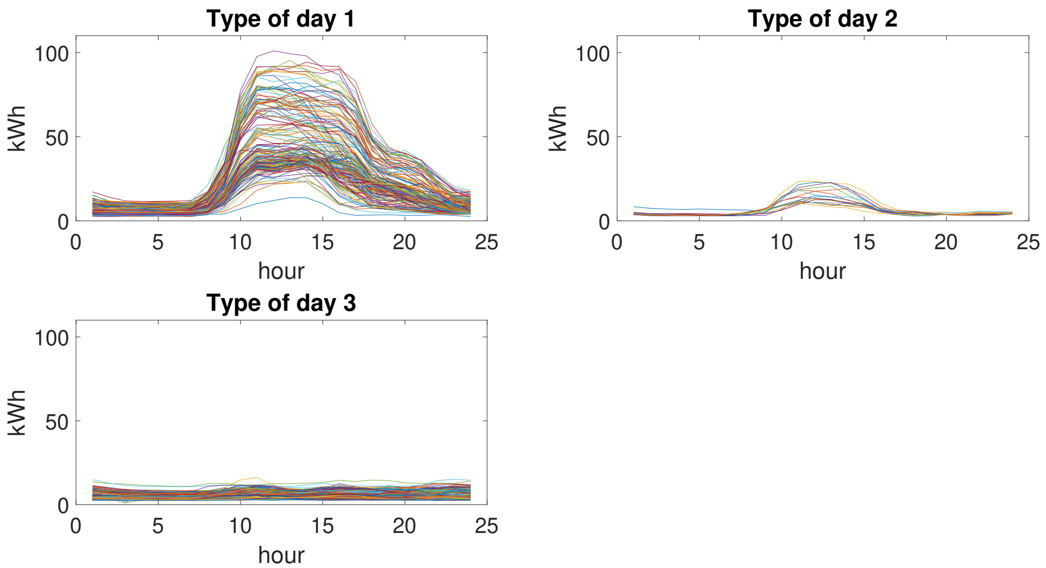


Figure 18: Daily consumption profiles for the SC group

The consumption of the rest of the MSP groups is quite constant along the whole year or its low consumption doesn't allow to create type of day profiles that are significantly different to others. In the case of the PM group, the type of day classification is not useful, since only 5 load curves are available and each one of them follow different consumption patterns, as it can be seen on Figure 19. Fortunately, the consumption of this group is low compared with the rest. Thus, the error introduced by this MSP group won't make a significant change in the precision of the total forecast.

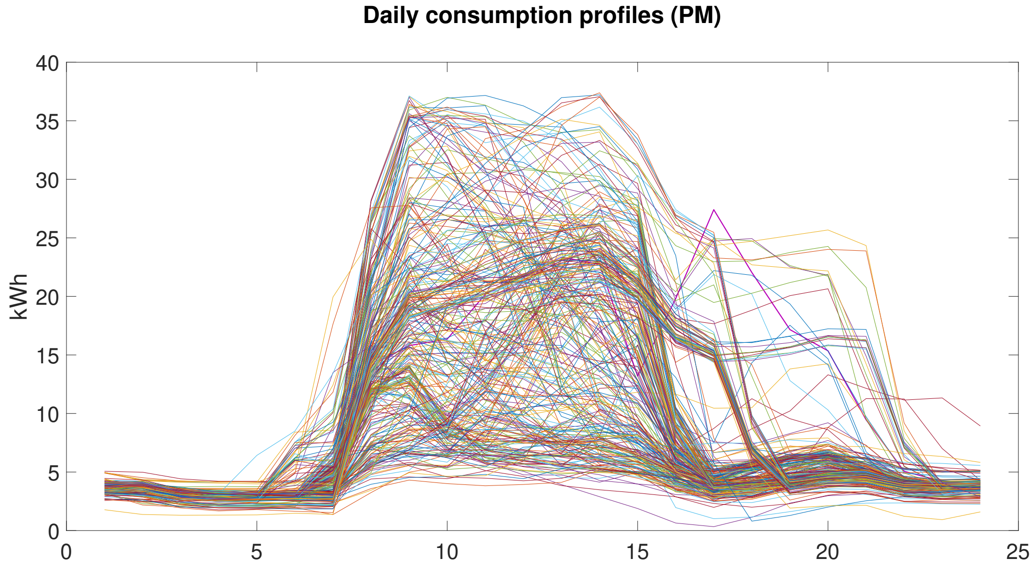


Figure 19: Daily consumption profiles for the PM group

6.1.5 Selection of ANN inputs

As stated on Section 5, most of the inputs correspond to external temperatures and previous electricity consumption. Also, for certain groups like PL, it is also necessary to know the moment at which the sunrise and sunset take place for the DOP, since it will indicate at what moment electricity consumption rises or drops. Temperature data is obtained from a measurement station at the Polytechnic University of Valencia (UPV). However, some data was unavailable, as it can be seen in Figure 21. Hence, the method presented in (Roldán-Blay et al., 2013) is used used to forecast the missing temperatures. The model used by authors enables the creation of an hourly temperature curve using the minimum and maximum temperatures of the DOP and the surrounding days. Implementing this method, it is possible to obtain the missing temperatures. The model uses a sinusoidal and a exponential segment for modelling the day, as shown:

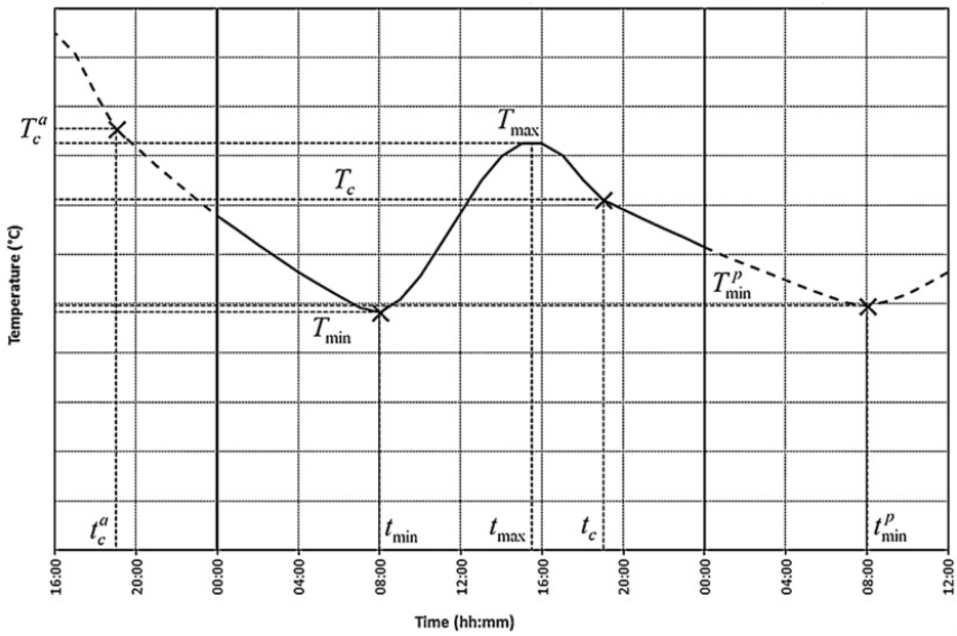


Figure 20: Temperature curve model. (Roldán-Blay et al., 2013)

The sinusoidal segment is used to simulate sun hours, since thermal inertia causes a gradual temperature rise. Exponential segment is used to simulate the progressive decrease of the temperature due to the lowering of radiation.

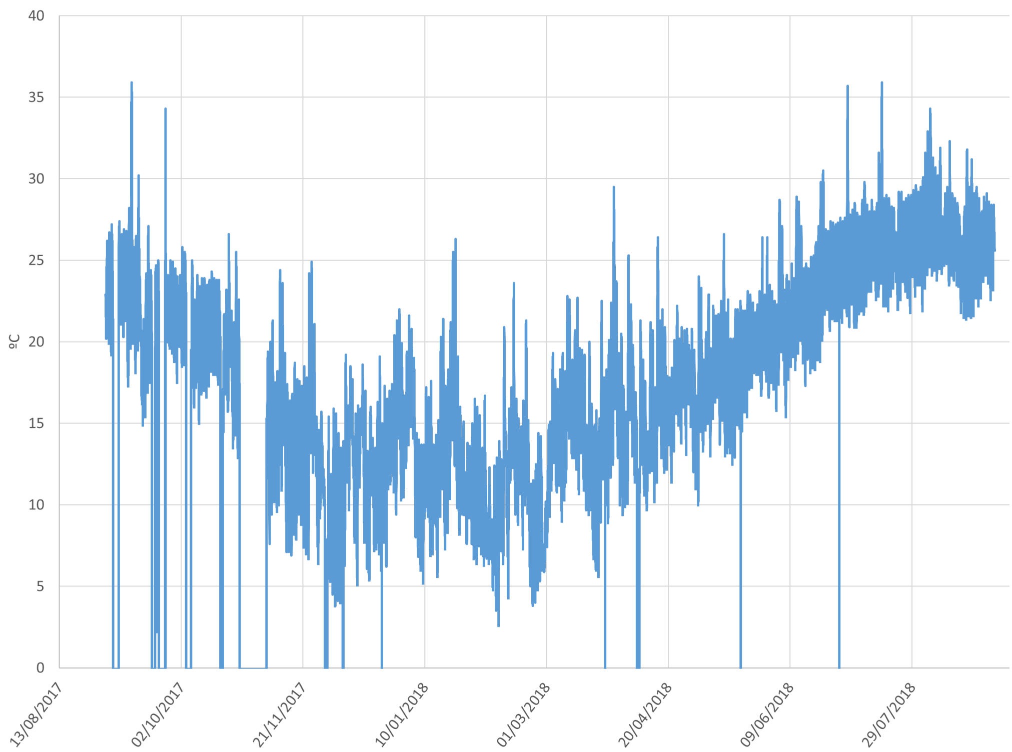


Figure 21: Initial hourly temperatures curve. Measurement station of the UPV

The sequence of equations implemented to calculate the temperature curve has been replicated from (Roldán-Blay et al., 2013) and is presented here:

$$\theta = \frac{2 \cdot \pi}{365} \cdot (dn - 1) \tag{10}$$

Where:

- θ is the daily angle in radians
- dn is the Julian day number of the year

$$et = (0.000075 + 0.001868 \cdot \cos(\theta) + 0.032077 \cdot \sin(\theta) - 0.014615 \cdot \cos(2 \cdot \theta) - 0.04089 \cdot \sin(2 \cdot \theta))229.18 \tag{11}$$

$$\delta = 23.45° \cdot \cos(2 \cdot \pi \cdot \frac{dn - 173}{365}) \cdot \frac{\pi}{180} \tag{12}$$

Where:

- et is the equation of time, which denotes the difference between true solar time and mean solar time
- δ is the solar declination angle between the equatorial plane and the line connecting the centres of the Sun and Earth in radians
- 173 corresponds to julian day number for june 22 (vernal equinox)

$$h = \cos^{-1}\left(\frac{sin(\frac{-0.833 \cdot \pi}{180} \cdot \sin(\delta))}{\cos(\varphi) \cdot \cos(\delta)}\right) \tag{13}$$

$$nd = \frac{h}{7.5} \cdot \frac{180}{\pi} \tag{14}$$

Where:

- h is the Solar angular hour
- φ is the latitude in radians
- n_d is the number of hours of daylight

$$t_{sr} = 12 - \frac{n_d}{2} - \frac{et}{60} \tag{15}$$

$$t_{ss} = 12 + \frac{n_d}{2} - \frac{et}{60} \tag{16}$$

Where:

- t_{sr} is the sunrise time in hours
- t_{sr} is the sunset time in hours

$$t^a_{min} = t^a_{sr} + h^a_d - \lambda \cdot \frac{24}{360} - \frac{et^a}{60} \tag{17}$$

$$t^a_{max} = \frac{t^a_{min} + t^a_{ss} + h^a_d - \lambda \cdot \frac{24}{360} - \frac{et^a}{60}}{2} + 1.5 \tag{18}$$

$$t_{min} = t_{sr} + h_d - \lambda \cdot \frac{24}{360} - \frac{et}{60} \tag{19}$$

$$t_{max} = \frac{t_{min} + t_{ss} + h_d - \lambda \cdot \frac{24}{360} - \frac{et}{60}}{2} + 1.5 \tag{20}$$

$$t^p_{min} = t^p_{sr} + h^p_d - \lambda \cdot \frac{24}{360} - \frac{et^p}{60} \tag{21}$$

Where:

- Superscript a refers to all variables of the day before the DOP
- Superscript p refers to all variables of the day after the DOP
- t_{min} is the time in the model at which minimum temperature of the day occurs in hours
- t_{min} is the time in the model at which maximum temperature of the day occurs in hours
- h_d is the hours of difference to the meridian for the location and date of the DOP. In Valencia, a value of +1h during winter and +2h during summer time is used
- λ is the longitude in radians
- t is the time of forecast (from t=0 to t=23)

$$t^a_c = \left(t^a_{ss} + h^a_d - \lambda \cdot \frac{24}{360} - \frac{et^a}{60} - 2\right) - 24 \tag{22}$$

$$t_c = \left(t_{ss} + h_d - \lambda \cdot \frac{24}{360} - \frac{et}{60} - 2\right) \tag{23}$$

where t_c is the time at which the sinusoidal temperature segment joins the exponential segment (Figure 20)

$$Tavg^a_{sin} = \frac{T^a_{max} + T^a_{min}}{2} \tag{24}$$

$$Tavg_{sin} = \frac{T_{max} + T_{min}}{2} \tag{25}$$

With this data, a value for the temperature can be calculated for every hour (from t=0 to t=23) following the next list of equations for the different time intervals presented:

For $t < t_{min}$

$$P_{sin}^a = 2 \cdot (t_{max}^a - t_{min}^a) \tag{26}$$

$$A_{sin}^a = T_{max}^a - T_{min}^a \tag{27}$$

$$T_c^a = Tavg_{sin}^a + \frac{A_{sin}^a}{2} \cdot \left((t_c^a - t_{max}^a) \cdot \frac{2 \cdot \pi}{P_{sin}^a} \right) \tag{28}$$

$$B = \frac{\ln(T_c^a) - \ln(T_{min})}{t_c^a - t_{min}} \tag{29}$$

$$A = \frac{T_c^a}{e^{B \cdot t_c^a}} \tag{30}$$

$$T(t) = A \cdot e^{B \cdot t} \tag{31}$$

For $\boldsymbol{t_{min} \leq t < t_c}$

$$P_{sin} = 2 \cdot (t_{max} - t_{min}) \tag{32}$$

$$A_{sin} = T_{max} - T_{min} \tag{33}$$

$$T(t) = Tavg_{sin} + \frac{A_{sin}}{2} \cdot \cos\left((t - t_{max}) \cdot \frac{2 \cdot \pi}{P_{sin}} \right) \tag{34}$$

For $\boldsymbol{t \geq t_c}$

$$P_{sin} = 2 \cdot (t_{max} - t_{min}) \tag{35}$$

$$A_{sin} = T_{max} - T_{min} \tag{36}$$

$$T_c = Tavg_{sin} + \frac{A_{sin}}{2} \cdot \cos\left((t_c - t_{max}) \cdot \frac{2 \cdot \pi}{P_{sin}} \right) \tag{37}$$

$$B = \frac{\ln(T_c) - \ln(T_{min}^p)}{t_c - t_{min}^p} \tag{38}$$

$$A = \frac{T_c}{e^{B \cdot t_c}} \tag{39}$$

$$T(t) = A \cdot e^{B \cdot t} \tag{40}$$

Where:

- P_{sin} is the period of time of the sinusoidal part of the temperature curve in hours
- A_{sin} is the amplitude of the sinusoidal part in kelvin
- A and B are parameters established to define the exponential segment

Finally, all this equations are applied through a script in Matlab. As only some temperature values were missing, the model described above in only used to obtain those values. Hence, the initial temperature curve in Figure 21 becomes the one presented on Figure 22 after adding the temperatures obtained to the temperatures initially available.

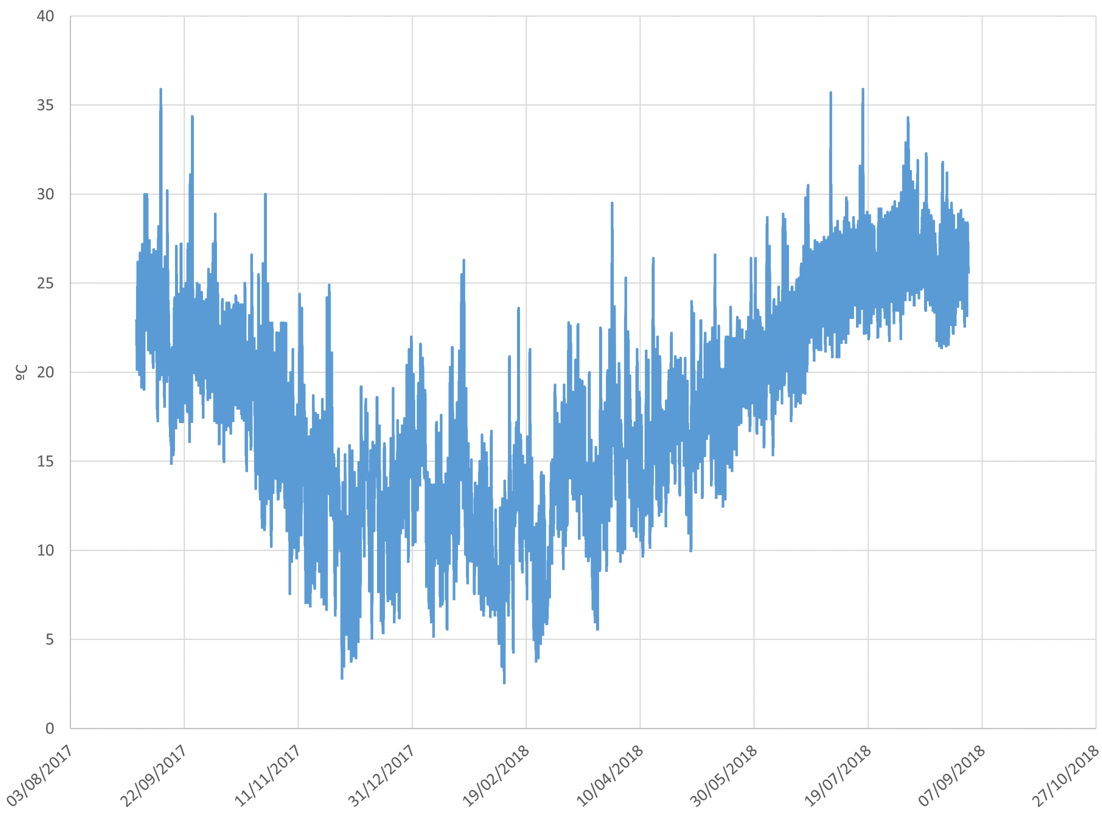


Figure 22: Hourly temperature curve after adding the temperatures obtained with the method presented on (Roldán-Blay et al., 2013) to the initial curve.

Another inputs used in the model, like sunset and sunrise time data were obtained from (Salida y puesta del sol, 2017). Electrical consumption was obtained from the load curves given by the council's electricity supply managers after treating all the available material, as explained on 6.1. Finally, the inputs used for SC, PO, CM, PM and LM are the energy consumed at t-1 (E_{t-1}), the average temperature of the three previous hours ($T_{Avg,t-3}$) and the temperature at t (T_t). PG and TL groups only receive as input the previous consumption, since their daily consumption profile remains constant, having a little variation depending on the type of day. Finally, PL receive the previous consumption and also the minute of the day at which the sunrise and sunset takes place for the DOP (tsr and tss). A summary of the inputs together with the type of day classification for every ANN is presented on Table 6.

Table 6: ANN parameters for every MSP group

MSP group	n	Type of day	input 1 (kWh)	input 2 (ºC)	input 3 (ºC)
LM	1	Working days	E_{t-1}	T_t	$T_{Avg,t-3}$
	2	Weekends, holidays and August			
SC	1	School days			
	2	Working days that are not school days	E_{t-1}	T_t	$T_{Avg,t-3}$
	3	Weekends, general holidays and August			
PO	1	Working days except Tuesdays			
	2	Tuesday working days	E_{t-1}	T_t	$T_{Avg,t-3}$
	3	Weekends and holidays			
PM	1	Working days (including saturdays)	E_{t-1}	T_t	$T_{Avg,t-3}$
	2	Sundays and holidays			
PG	1	Working days	E_{t-1}	-	-
	2	Sundays and holidays			
PL	1	All	E_{t-1}	tsr	tss
TL	1	All	E_{t-1}	-	-
CM	1	All	E_{t-1}	T_t	$T_{Avg,t-3}$

In the SC MSP group, working days that are not school days are referred to those days in which students are in holidays, but teachers and staff still go to work.

6.2 Design of ANN's for every MSP group

After explaining how ANN's work and its more important parameters, the architecture decided for the ANN's implemented in this TFM is the following [2]:

Table 7: ANN architecture

Parameter	value
Type of network	Feed-forward
Number of hidden layers	1
Number of neurons in hidden layers	5
Training algorithm	Levenberg-Marquardt
Activation function	Sigmoid function

[2] all the parameters involved were explained on Section 4

In order to apply the ANN's, matlab's Deep Learning Toolbox is used. A general script for every MSP group is used. Doing different scripts for every single MSP group is necessary, since the type of days and inputs used vary from one to another (Table 6).
Within every script, a single ANN is used for each one of the hours of the day, as explained below:

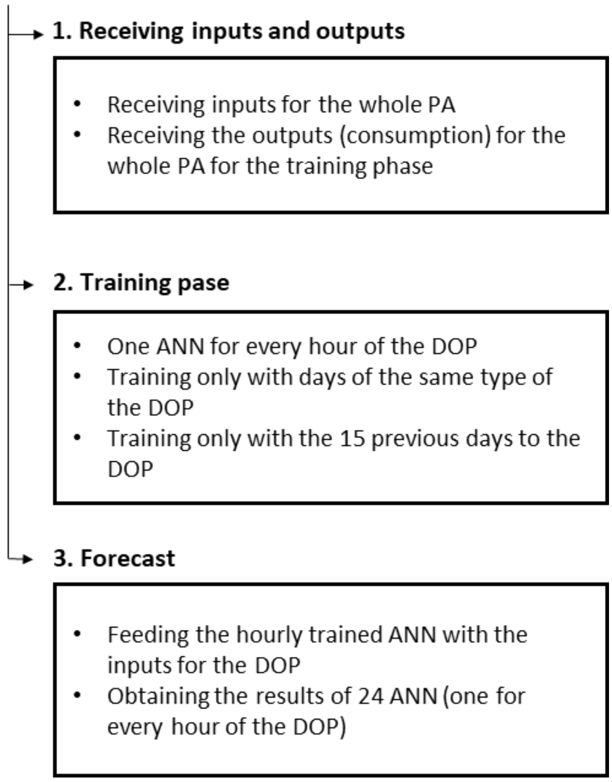


Figure 23: Working principle of the ANN's used for every single MSP group

First, the script receives the corresponding ANN inputs for all the period of analysis (PA). The granularity of the prediction is one hour so the ANN receives 8760 values for each input. Script is also fed with actual hourly consumption for the PA, which will be ANN's output during the training phase.
Secondly, only the days that are of the same type of the DOP are selected .
Then, a single ANN for every hour of the DOP is trained. Thus, the prediction of a whole day will be the group of the results of 24 ANN. Every hourly network is trained only with data from the moment t of the 15 previous days , adopting so a sliding time window strategy (Yang et al., 2005).That means that if the ANN of the hour 13 is being trained, ANN will only train with data from the hour 13 of the 15 previous days to the DOP. If many measures are used in the

training phase, then random and uncontrolled training data may be introduced (Escrivá-Escrivá et al., 2011). In order to train the networks, data has to be split into 3 sets; training, validation and testing. The division is made randomly between the 15 selected days, establishing only the percentages of data dedicated to each set. 70% for training, 15% for validation and 15% for testing has been used, which means that 12 days are used for training, 2 for validating and 1 for testing. The training set is used in parallel with the validation set in a way that during the training phase, outputs obtained with the ANN are compared with the actual outputs given and weights and biases are adjusted to match inputs with outputs. Validation set is used to check the performance of the ANN continuously by calculating the error of the prediction. When that error reaches a minimum, the training phase is finished and an error obtained through the testing set is obtained. The testing set is not strictly necessary, since the prediction itself is going to be a test, but Matlab doesn't allow to set the testing set percentage to 0, so a minimum percentage of 5% has been considered.
Once the training phase is completed, the hourly networks are fed with new input data, which corresponds with the input data for the DOP

6.3 Prediction results

To obtain a full prediction, it is necessary to forecast separately the consumption of every single MSP group for the same time period, as explained above. If the time horizon of the forecast is one single day, 24 ANN's are calculated for every MSP group, since a single ANN is trained for every hour of the day.
Once the consumption of every MSP is predicted separately, results are added, obtaining the total forecast for the whole electricity consumption of the Valencia municipality.
To evaluate how good the forecast is, MAPE and EME are calculated for the time horizon of the forecast. As results will be presented for some example weeks on Section 7, MAPE and EME equations for that time period are specified:

$$MAPE_{week} = \frac{\sum_{t=1}^{24\,hours\cdot 7\,days} \frac{\left|\hat{E}_t - E_t\right|}{|E_t|}}{24\,hours \cdot 7\,days} \cdot 100 = \frac{\sum_{t=1}^{168} \frac{\left|\hat{E}_t - E_t\right|}{|E_t|}}{168} \cdot 100 \tag{41}$$

$$EME_{week} = \sum_{t=1}^{24\,hours\cdot 7\,days} \frac{\left|\hat{E}_t - E_t\right|}{\sum_{t=1}^{24\,hours\cdot 7\,days} E_t} \cdot 100 = \sum_{t=1}^{168} \frac{\left|\hat{E}_t - E_t\right|}{\sum_{t=1}^{168} E_t} \cdot 100 \tag{42}$$

7 Results

In this section the different forecast results obtained through the MSP classification and the ANN application are shown. First, forecast results of every individual MSP group are calculated. Then, these individual forecast are aggregated, obtaining a complete forecast for all the Valencia municipality consumption. All the results are shown for the same example week selected randomly.

7.1 MSP groups forecast results

On the following figures, results of the forecast methodology applied to all the MSP groups on Table 5 are shown. Results are displayed in order of importance according to its volume of annual electricity consumption.

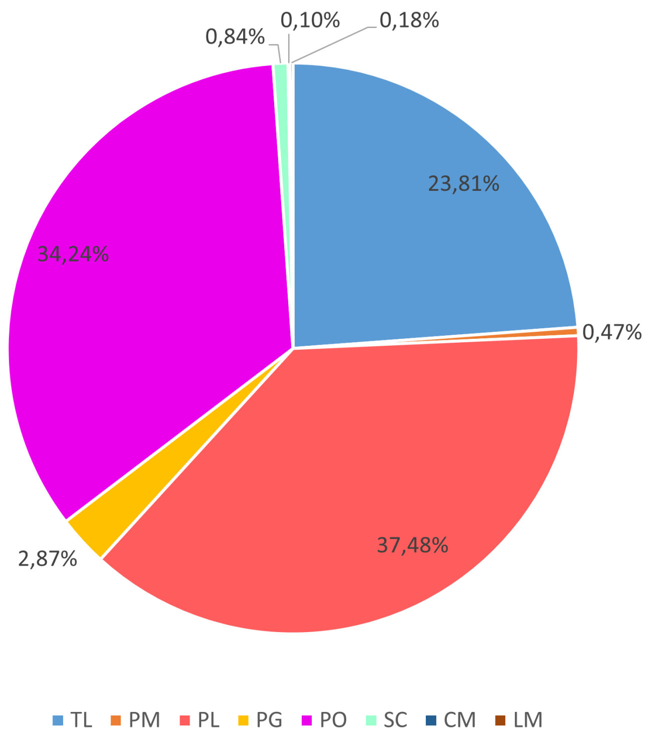


Figure 24: Annual energy consumption distribution of every MSP group

As it can be seen on Figure 24, the MSP groups that represent a highest annual electricity consumption are PL (37,48%), PO (34,24%) and TL (23,81%). Hence, achieving a good forecast on these groups is vital to obtain a good total forecast of the Valencia municipality electricity consumption.

7.1.1 Public lightning (PL)

The firs forecast results presented are those of the PL MSP group:

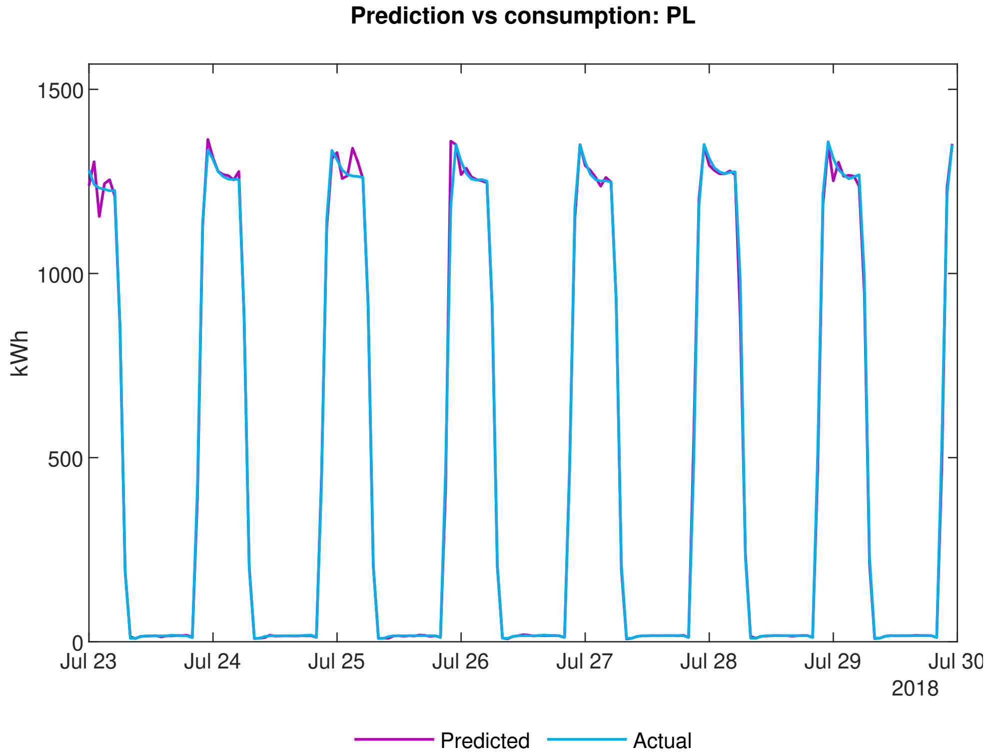


Figure 25: Prediction vs consumption for the PL MSP group

Table 8: MAPE and EME (%) of the PL forecast for the week shown on Figure 25.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	6,60	3,91	5,41	3,97	2,83	4,63	2,91
EME	2,96	1,12	3,47	1,14	1,40	2,11	2,34

Figure 25 shows that the forecast of the PL group is almost perfect. As it can be seen, there's a moment in which consumption is residual. This hours correspond to the day hours, in which no lightning is needed. However, during the night hours, electricity consumption reaches its maximum. Lightning is a type of consumption that is easy to predict, since installed equipment has a constant consumption during the hours of use. One of the biggest variation that this type of consumption can have is the one caused by the minute of the day in which sun rises or sets, which varies along the year. Because of that, time of sunrise and sunset is given as an input to the ANN's of this MSP group.
Even though forecast is almost perfect, the MAPE is quite high. This is due to the effect of not considering the volume of energy of the period of calculation to calculate the hourly MAPE, as it was commented on Section 3.4. High MAPE values on day hours cause the daily MAPE to be higher than expected.

7.1.2 Public offices (PO)

Now, PO results are presented. The PO MSP group is the second group in terms of annual electricity consumption, as seen before.

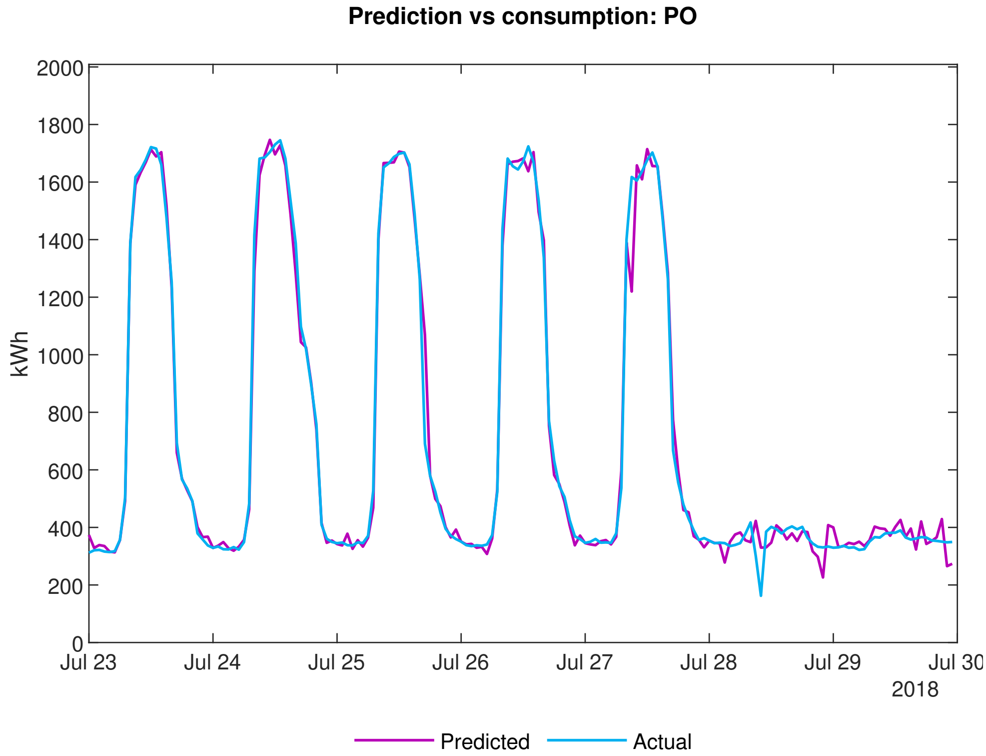


Figure 26: Prediction vs consumption for the PO MSP group

Table 9: MAPE and EME (%) of the PO forecast for the week shown on Figure 26.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	3,06	2,88	5,08	3,12	4,72	14,75	8,31
EME	2,04	2,94	3,54	2,78	4,80	12,34	8,29

Forecast is really good during weekdays, as it can be seen on Figure 26 and Table 9. Consumption on these days starts to rise at 7:00, as people begin entering the workplaces and starts to decrease at 17:00 as workers leave. The consumption of the rest of the day is the residual consumption (some lightning, equipment like some PC's...etc):

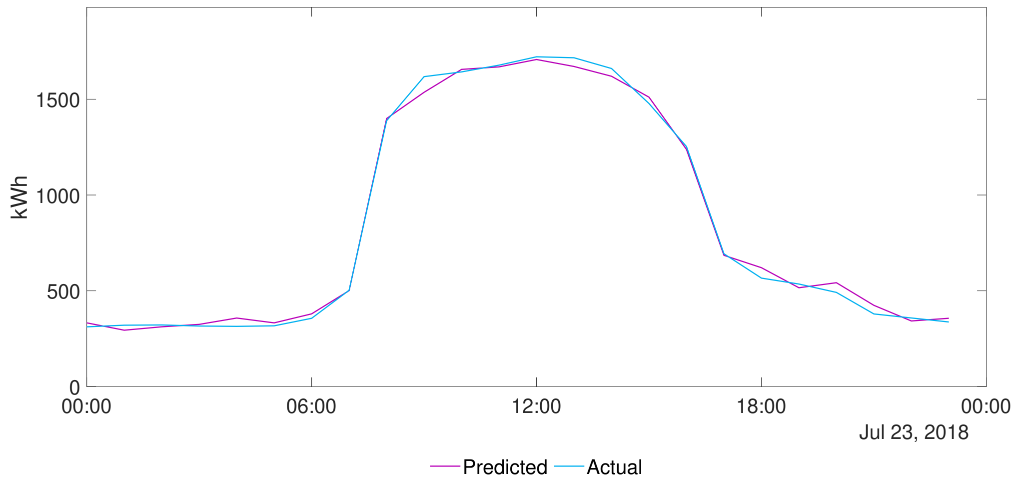
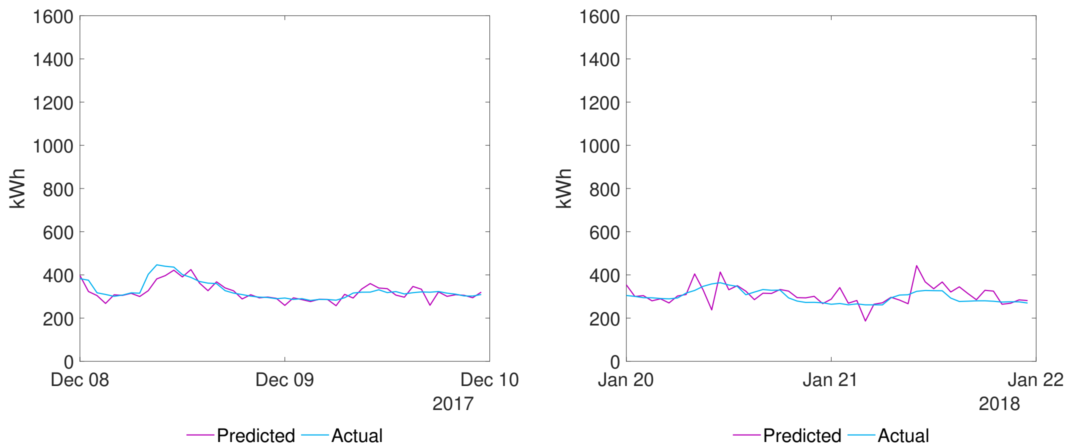


Figure 27: Prediction vs consumption for a single day for the PO MSP group

On the weekends, however, forecast is relatively poor, this is because during these type of days, consumption is very changeable. Furthermore, ANN's train with the 15 previous days of the same type. That means that ANN's of the DOP of a weekend is trained with days from some weeks ago, in which conditions can be different from that ones of the DOP. However, prediction on the weekends in winter is way better than in the week shown on Figure 26, as it can be seen:



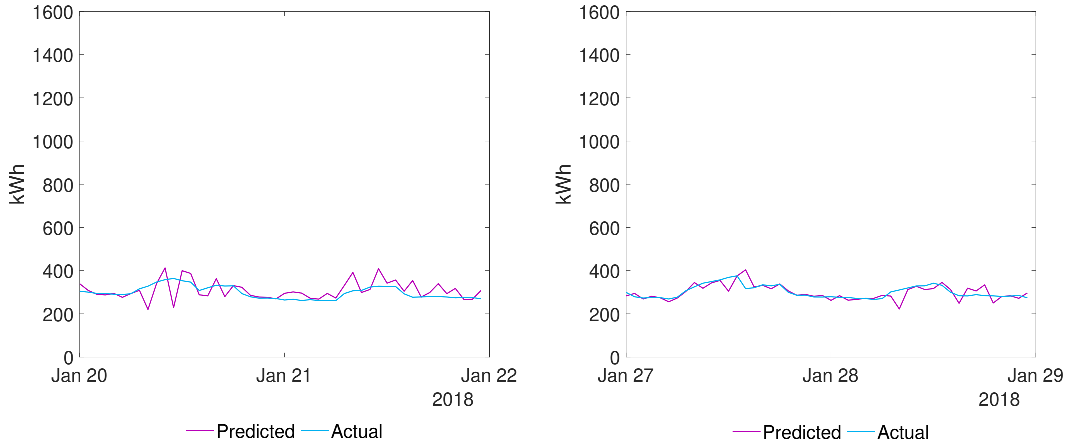


Figure 28: Consumption vs prediction for winter weekends (PO MSP).

7.1.3 Traffic lights (TL)

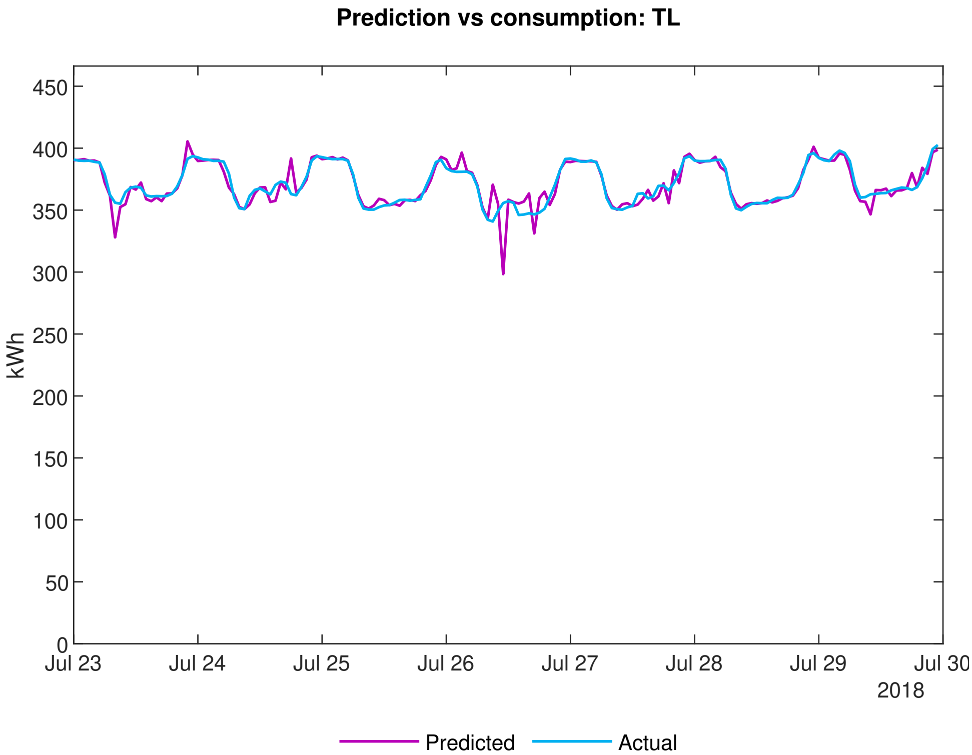


Figure 29: Prediction vs consumption for the TL MSP group

Table 10: MAPE and EME (%) of the TL forecast for for the week shown on Figure 29.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	1,016	1,179	0,549	2,572	0,962	0,499	1,079
EME	1	1,17	0,54	2,52	0,95	0,50	1,07

Prediction for the TL MSP group works fine, with the exception of some unexpected peaks happening on days 1 and 4 of the displayed week. This might be caused by peaks happening on the previous week, which ANN's of the displayed week train with. Despite that, MAPE and EME errors on Table 10 are very low, since peaks happen only in two hours of the whole week. The following MSP groups represent low levels of consumption compared with the previous ones, as it was seen on Figure 24. Since the number of load curves is much lower that in the previous MSP groups and the consumption is less, errors are bigger in general. However, because of its importance in relation to the previous MSP groups, bigger errors in the following groups won't introduce a significant error in the general forecast.

7.1.4 Public gardens (PG)

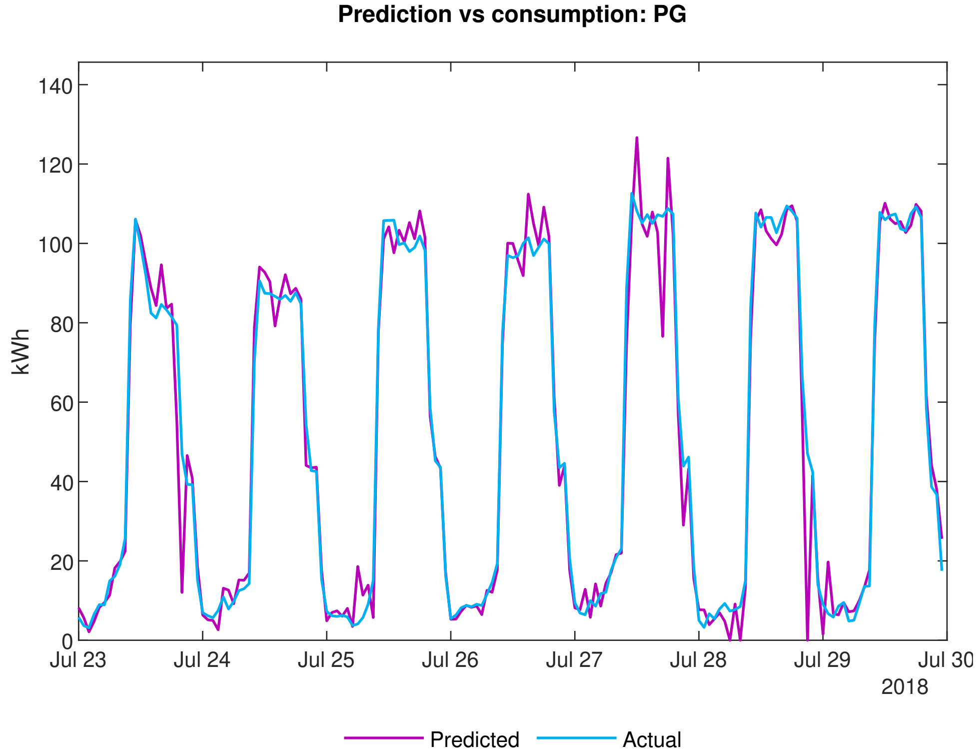


Figure 30: Prediction vs consumption for the PG MSP group

Table 11: MAPE and EME (%) of the PG forecast for for the week shown on Figure 30.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	17,346	13,761	30,693	7,413	18,784	28,540	22,645
EME	11,067	6,789	6,910	5,652	11,200	9,876	5,723

As it was expected, forecast in the PG MSP group is worse than in the previous groups. However, bigger errors take place on the moment of maximum or minimum consumption, while the rising or decrease moments of electricity consumption are well foretold. Even though daily errors are a little high, in terms of energy, the maximum error that happens, which takes place on july 9, is less than 50 kWh. And again, MAPE error is much higher than EME because of the same effect commented for the PL MSP group. In day 3 that effect can be appreciated properly. EME in that day is less 7%. However, MAPE exceeds a 30%. The moment that is causing that difference is the one that happens in the early morning of july 26, when consumption is low. Since MAPE calculates the error without taking into account the moment of the day, a small error in terms of energy (less than 10 kWh) is causing a big MAPE value.

7.1.5 Schools (SC)

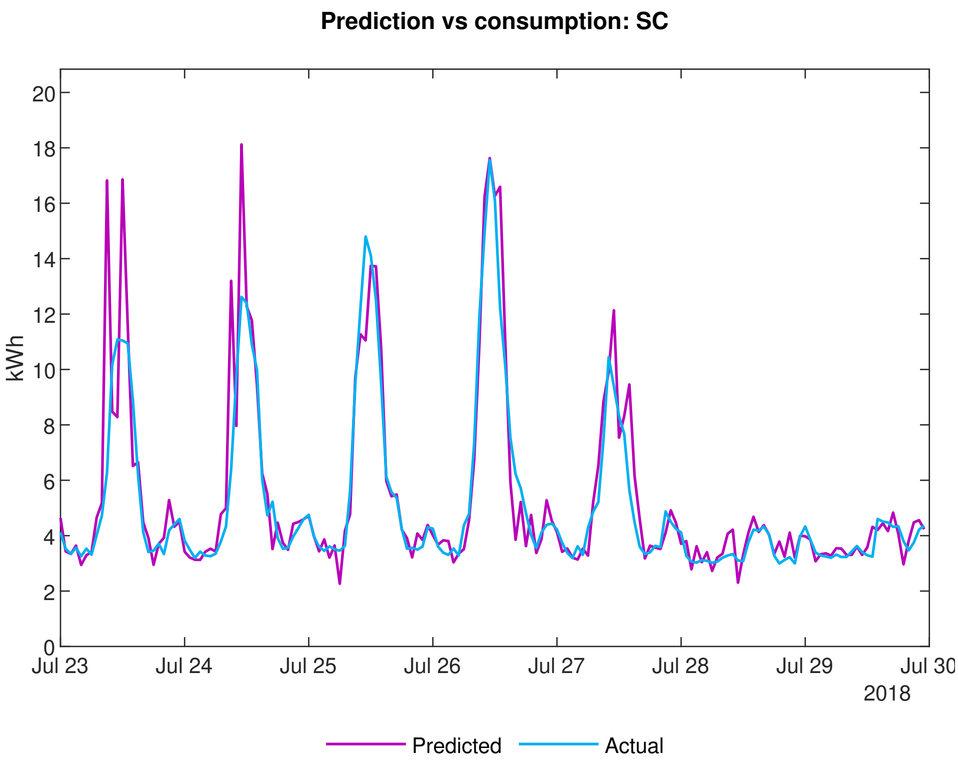


Figure 31: Prediction vs consumption for the SC MSP group

Table 12: MAPE and EME (%) of the SC forecast for for the week shown on Figure 31.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	18,783	14,995	8,862	12,159	14,197	10,808	7,908
EME	23,203	17,203	9,205	11,526	15,320	10,349	7,843

Since only 18 load curves are available for the analysis of the SC MSP group, error are quite high. Despite of that, the shape of the forecast is very similar to the actual consumption, with the exception of some unexpected peaks happening specially in the moments of higher consumption. As only 18 load curves are included in the analysis, a change in the consumption pattern at some particular moment in some particular school can introduce a peak in the consumption of the MSP group. ANN's are trained with all data of the previous 15 days, including these little peaks of consumption. Hence, ANN's might tend to replicate that peak, causing the error observed specially on July 23 and 24. Nevertheless, the error is very low in terms of energy, reaching a maximum of about 10 kWh.

7.1.6 Cemeteries (CM)

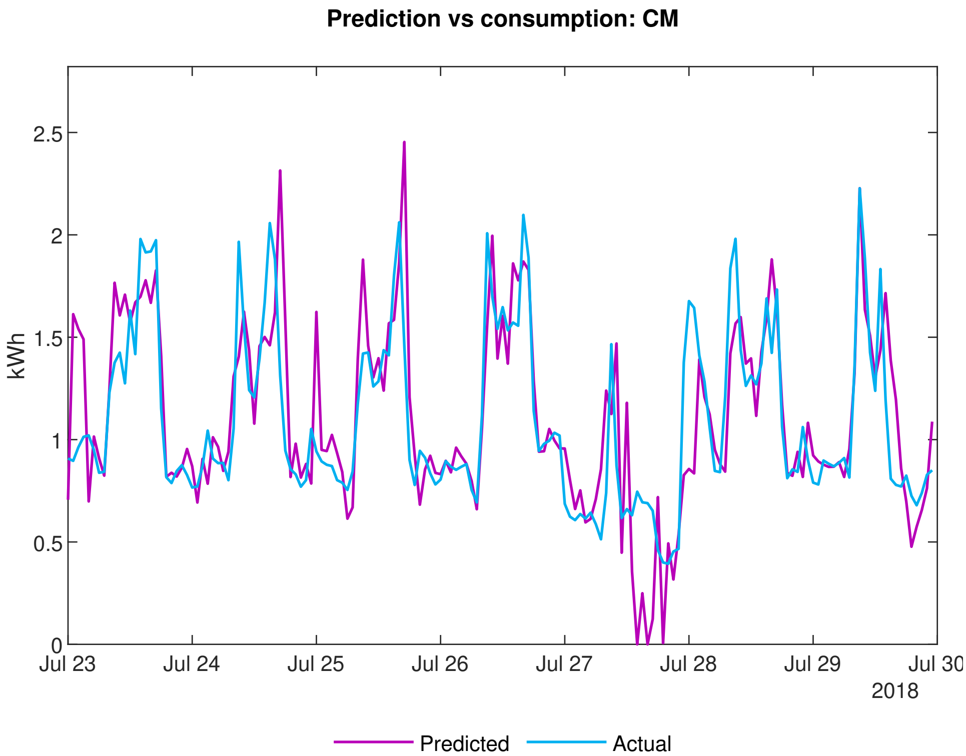


Figure 32: Prediction vs consumption for the CM MSP group

Table 13: MAPE and EME (%) of the CM forecast for for the week shown on Figure 32.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	18,635	18,391	18,784	7,655	46,231	15,002	16,841
EME	17,584	19,386	18,765	8,852	45,910	16,326	15,290

Cemeteries forecast is less accurate than the first MSP groups. Only 3 load curves are included in this analysis and as it can be seen, a maximum of 2,5 kWh of daily electricity consumption is reached. Hence, the error introduced by this MSP group is negligible. Regardless of that, the model is pretty capable to imitate the consumption pattern of the group.

7.1.7 Libraries and museums (LM)

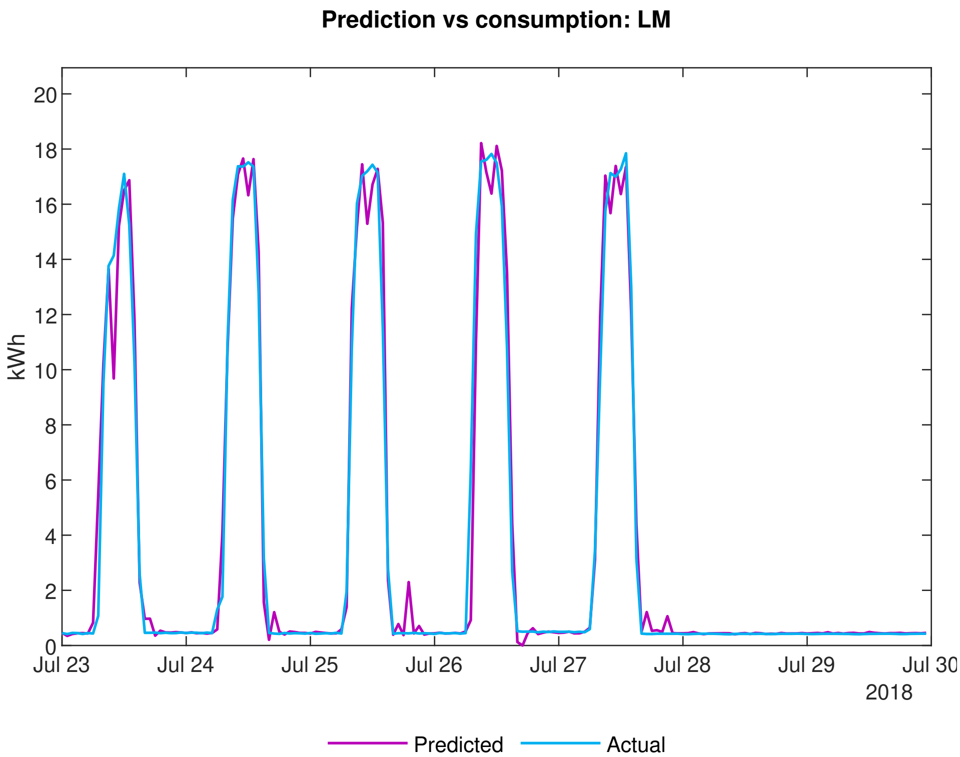


Figure 33: Prediction vs consumption for the LM MSP group

Table 14: MAPE and EME (%) of the LM forecast for for the week shown on Figure 33.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	37,530	24,251	32,735	20,480	24,376	4,747	6,554
EME	15,224	8,587	11,179	15,261	9,207	4,727	6,541

Even though only 3 libraries are included in the analysis, the model predicts quite fine the consumption. This is because libraries consumption is very stable. Errors introduced are mainly due to libraries having different schedules and working days. This factors introduce noise in the model that then ANN's can replicate in some way. Maximum error in the displayed weeks is about 8 kWh.

7.1.8 Public markets (PM)

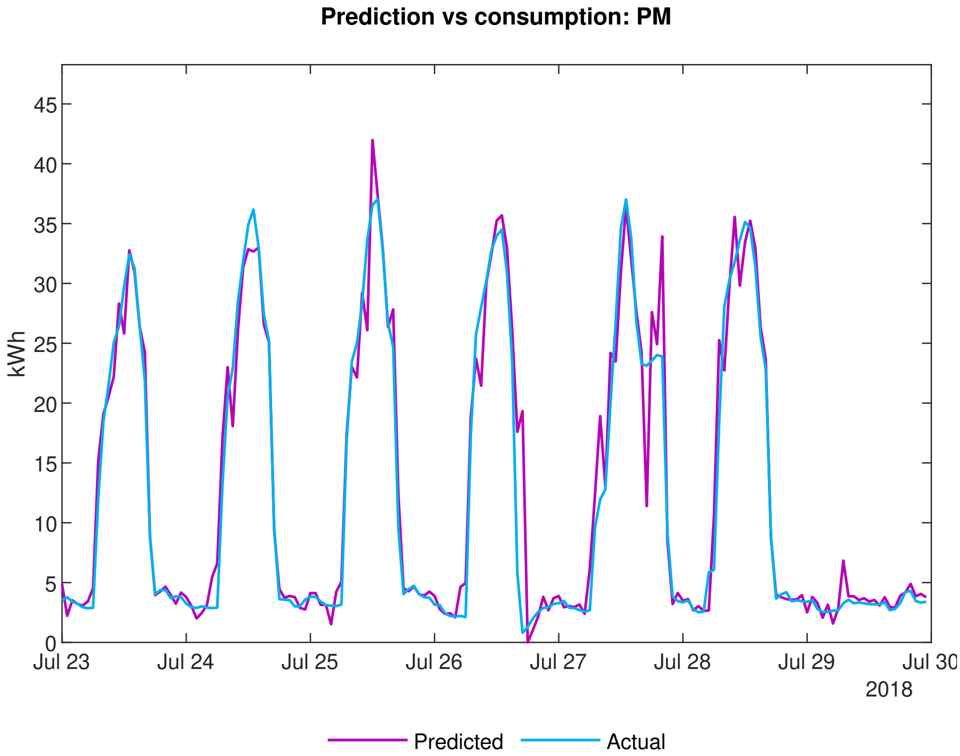


Figure 34: Prediction vs consumption for the PM MSP group

Table 15: MAPE and EME (%) of the PM forecast for for the week shown on Figure 34.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	12,049	20,031	12,994	128,472	20,862	13,087	17,545
EME	7,911	10,083	9,207	20,015	16,183	10,429	17,213

Only 5 load curves are available for the analysis of the PM MSP group. Furthermore, every market consumption is very different. They have different daily schedules and different holidays, so the training process of the ANN's is difficult. In the following figure, the electricity consumption of the 365 days of the year is shown:

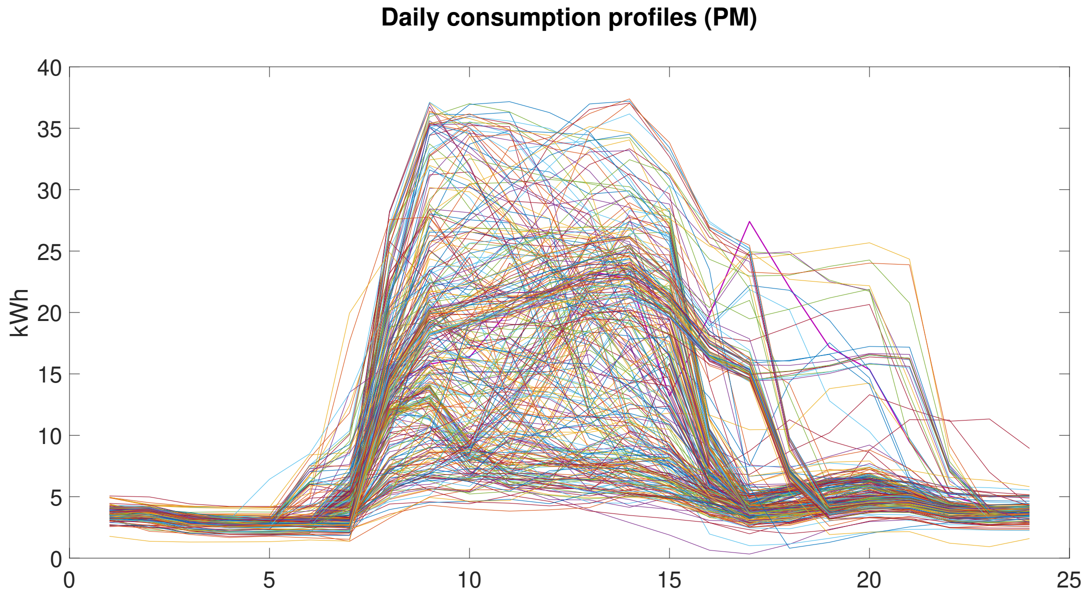


Figure 35: Daily profile of the PM electricity consumption

As it can be seen, almost every day's consumption is different. That way, ANN's are difficulty able to calculate reproducible patterns. Even though forecast in Figure 34 is quite fine taking into account what's been explained, forecast in some weeks might me worse, depending on how good the training process is.

7.2 Aggregated forecast

Once the forecast of every individual MSP group is done, they are added to achieve the forecast of the total consumption of the Valencia municipality. To appreciate the effect of seasonality over consumption profile and consumption, results for one week of winter and another of summer are shown.

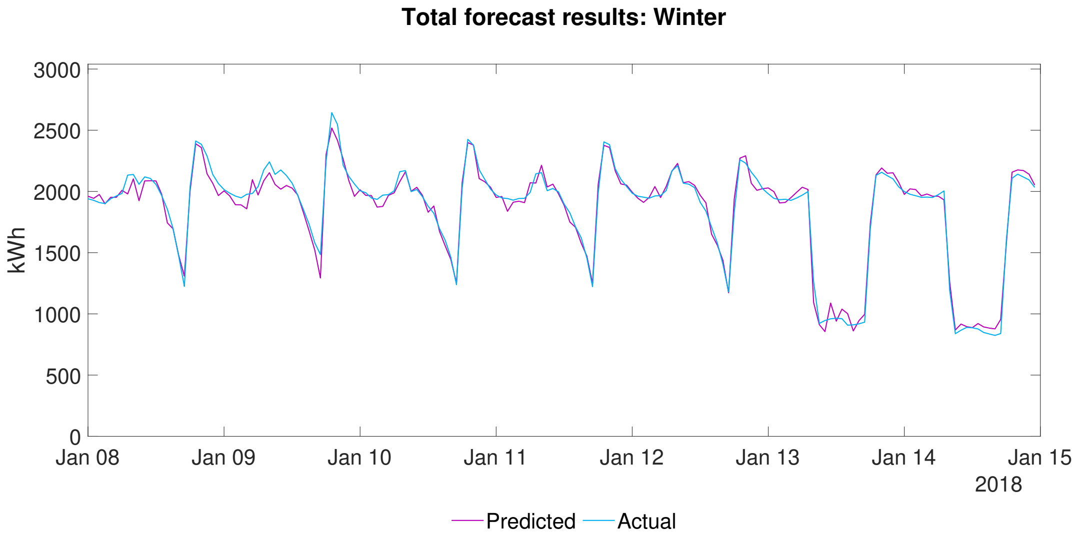


Figure 36: Aggregated prediction for a winter week

Table 16: MAPE and EME (%) of the total forecast for the week shown on Figure 36.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	2,474	3,770	1,587	1,871	1,876	4,113	3,179
EME	2,455	3,709	1,572	1,860	1,863	3,273	2,566

As it can be seen, forecast is very accurate and error measures don't surpass a 4%. Most of the days, both MAPE and EME errors are below the 3%.

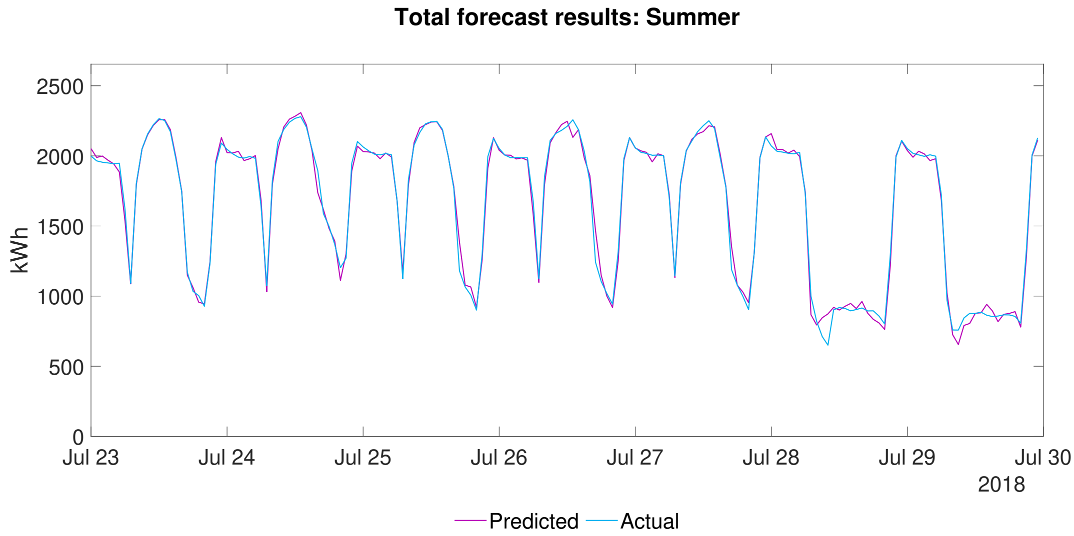


Figure 37: Aggregated prediction for a summer week

Table 17: MAPE and EME (%) of the total forecast for the week shown on Figure 37.

error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
MAPE	1,407	2,045	2,013	2,545	1,521	5,074	3,360
EME	1,270	1,867	1,570	2,181	1,216	3,551	2,603

In the summer week, forecast results look even better, with the exception of some moments on the 28th and 29th of July.
As it can be seen on Figures 36 and 37, consumption patters for winter and summer are quite different. This can be appreciated better by plotting one weekday of winter and summer:

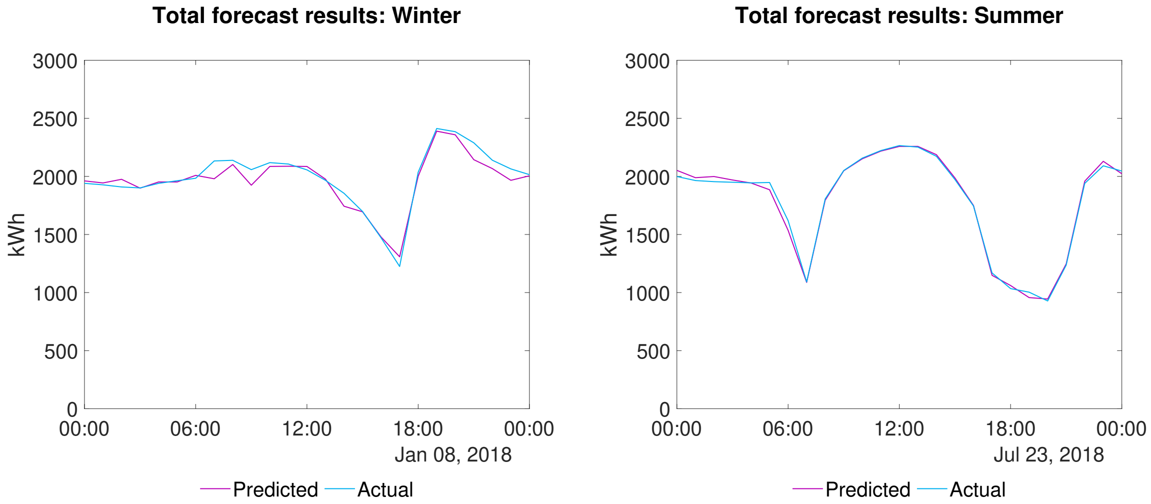


Figure 38: Comparison between one weekday in winter and summer.

The differences between the two profiles are due to the following factors:

- In winter, public lightning needs to be on for longer during the night. That way, in the summer profile can be appreciated a decrease on the consumption at 6:00 am that can't be seen on the winter profile. This is because in summer lights are turned off earlier.
- Again, since in summer there are more sun hours, a delay in the increase of the electricity consumption can be seen in the afternoon. Winter consumption increase starts at 18:00 approximately while summer increase starts at 20:00.
- Climate equipment consume more in cooling mode (summer) than in heating mode. That's why a higher consumption can be seen during summer in the working hours. However, difference it's not very noticeable.

As results show, forecast of the electricity consumption of the Valencia municipality is quite accurate. As it was commented before, by adding the individual MSP groups forecast results, errors are considerably reduced. That way, by plotting the MAPE and EME errors for a sample week of every MSP, it is possible to appreciate the improvement:

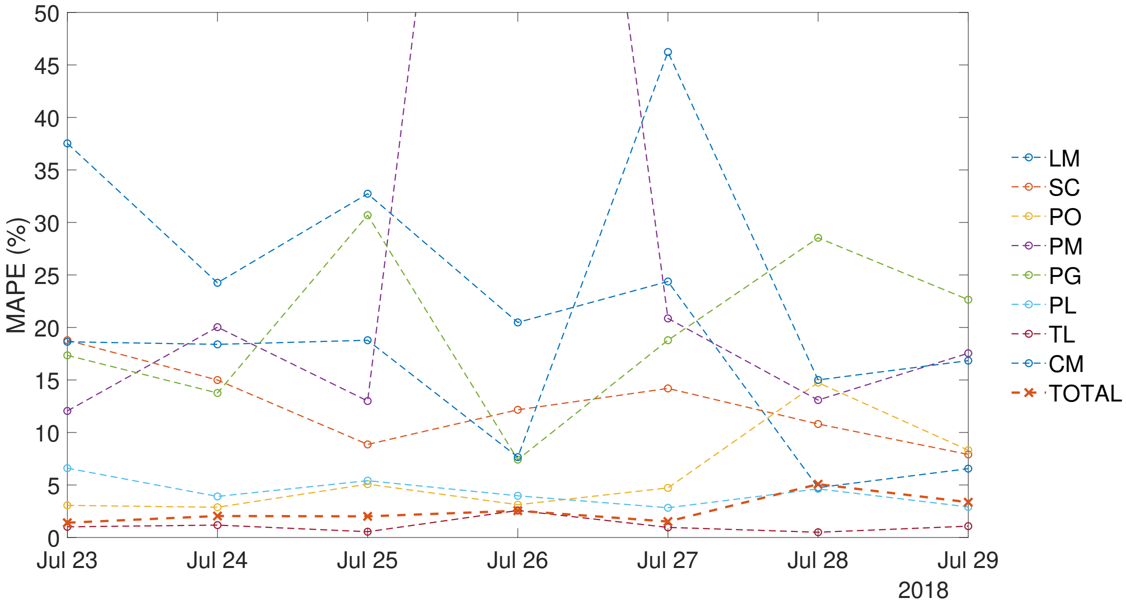


Figure 39: MAPE error for the different MSP groups.

One of the values is out of the graph because it was so high that including it would have done difficult to see the others. As Figure 39 shows, when performing the total forecast by adding the results of the different MSP groups, errors decrease significantly. The only MSP group that has a lower error by itself than the total forecast is the TL MSP group, although difference is negligible.

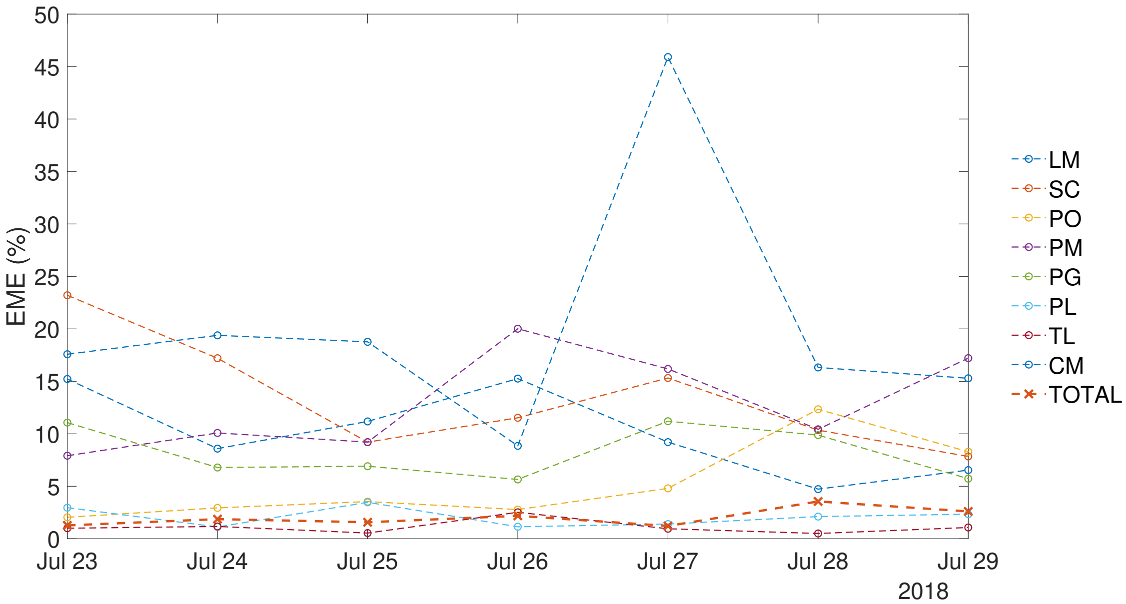


Figure 40: EME error for the different MSP groups.

EME error also decreases when performing the total forecast of the Valencia municipality electricity consumption. Both Figures 39 and 40 show that even though the forecast of some of the MSP groups is poor, it doesn't affect the total forecast. This is because, as it was explained, these MSP groups don't represent a significant level of electricity consumption in comparison with MSP groups like PL, PO or TL.

7.3 Justification of the methodology

To justify the MSP classification in groups that share a similar consumption scheme and variables influencing consumption, a comparison between the results with and without classifying the MSP into groups is done. Again, to appreciate the effects of seasonality, one week of winter and summer has been analysed. This has been done by programming a script in Matlab. The script takes as inputs all the variables on Table 6 and receives also the total consumption of the Valencia Municipality, without classifying MSP into groups. For the E_{t-1} variable, it takes the total consumption of the previous hour. The same goes to $T_{Avg,t-3}$. Type of day classification is also done. To do so, days have been classified in all the different days specified on Table 6. The types of day finally used are the following ones:

Table 18: Types of day used on the non-classified ANN's

n	Type of day
Type 1	Working days except Tuesdays
Type 2	Tuesday working days
Type 3	Saturdays
Type 4	Sundays, August and general holidays
Type 5	Working days that are not school days

The ANN's architecture and training methodology is the same explained on Section 6.2 but without previously classifying MSP into groups and using as inputs all the variables on Table 6 and types of day adapted.

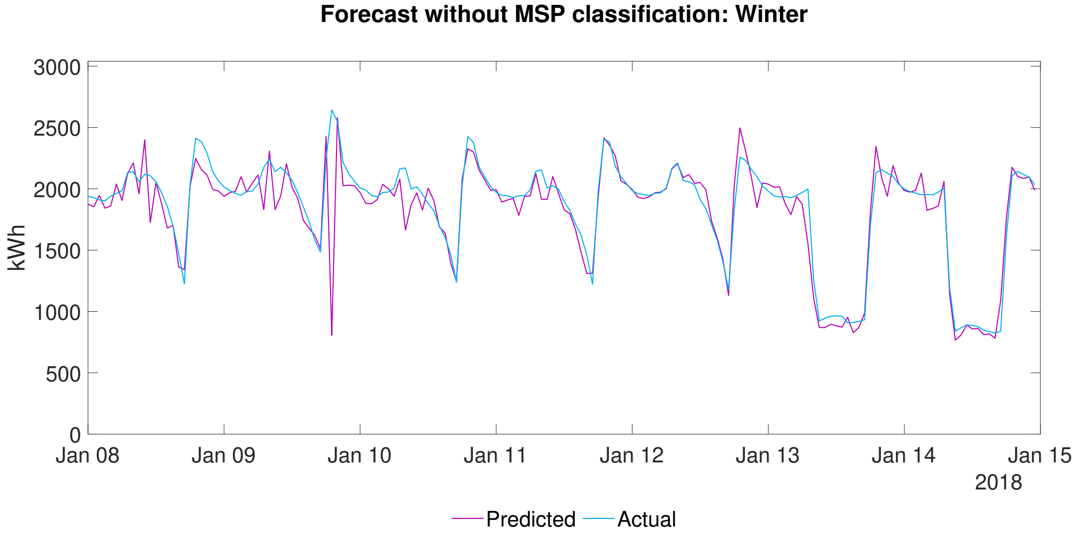


Figure 41: Total forecast results for a winter week without classifying consumption into MSP groups

Forecast without classifying in the winter week is inaccurate, specially during the moments of maximum consumption, where a lot of peaks appear. The peak that happens on the 9th of January is very high and introduces an error in terms of energy of more than 1500 kWh, which can have severe consequences if the forecast methodology is used to buy electricity in the wholesale market, for example.

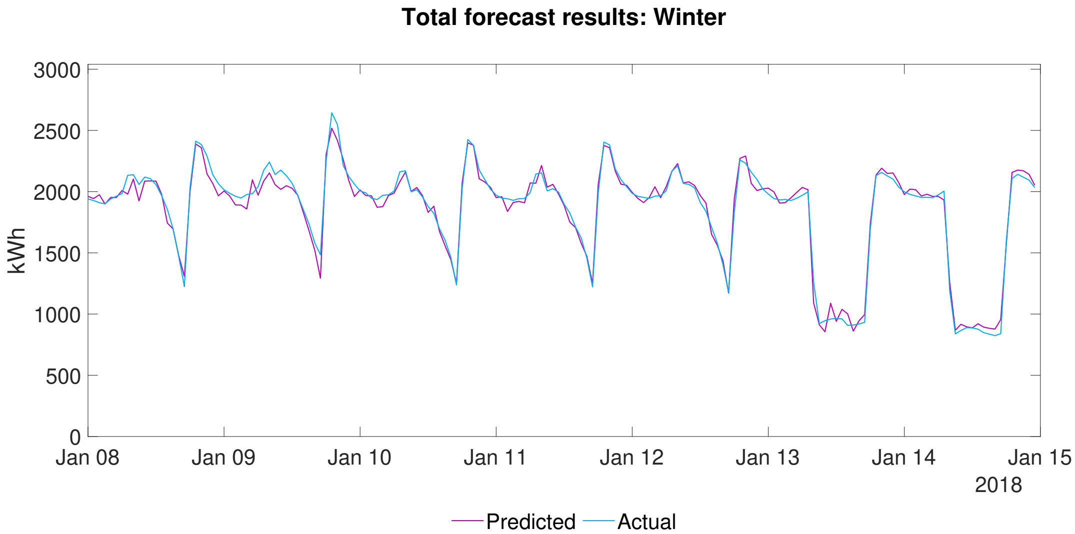


Figure 42: Total forecast results for a winter week classifying consumption into MSP groups

The improvement when classifying consumption into groups is obvious. On Figure, the errors are much more moderate 42. Later, the differences will be quantified.

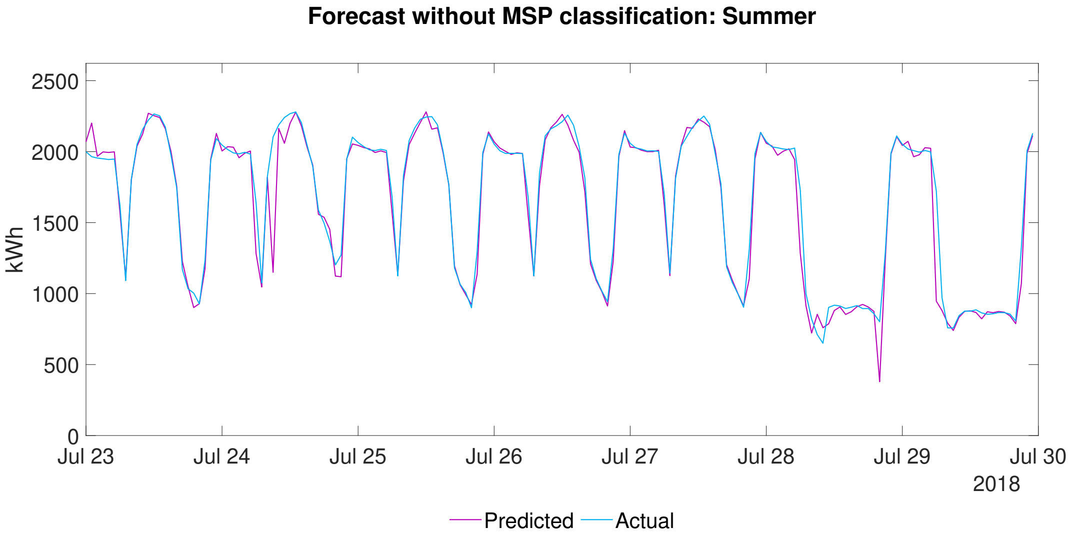


Figure 43: Total forecast results for a summer week without classifying consumption into MSP groups

Forecast without classifying is much better in the summer week than in the winter one. However, some peaks still appear. Even though they happen just on some particular hours, they could have severe consequences depending on the use given to the forecast methodology.

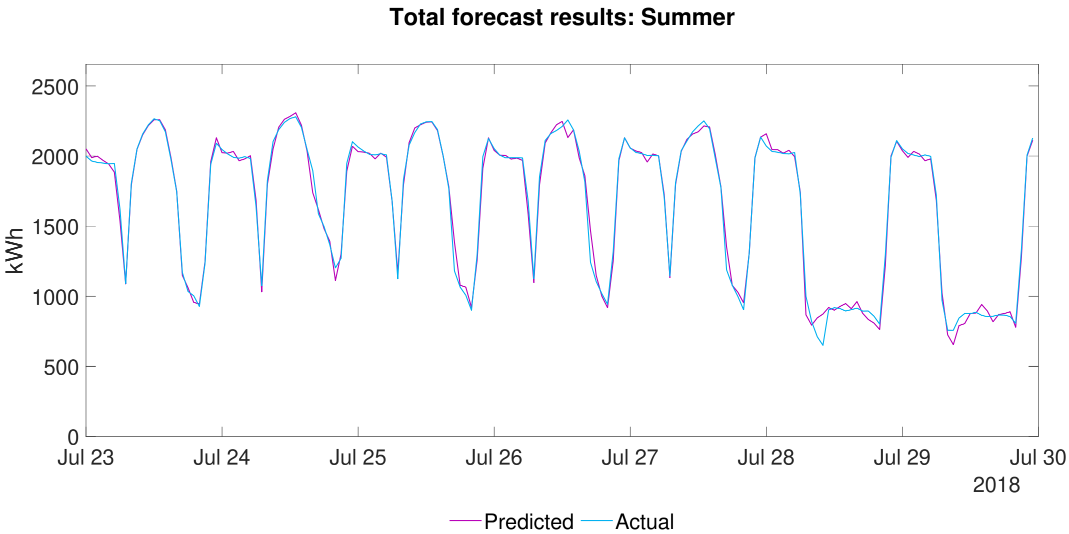


Figure 44: Total forecast results for a summer week classifying consumption into MSP groups

Again, the improvement is obvious. What's observed mainly is that the forecast is smoother when the classification is done, instead of having a more saw-shape like the one observed on Figures 41 and 43.
The improvement on the forecast when MSP are classified by groups can be observed visually and easily by comparing Figures 41, 42, 43 and 44. When not performing the previous MSP classification, the model takes as inputs some variables that for most of the types of consumption of the Valencia Municipality can act as noise in the training process. That would be the case of the t_{sr} and t_{ss} variables that would only be necessary in the case of public lightning. The same goes the other way around: variables related with temperature are totally unnecessary to study the consumption of public lightning, causing the ANN training process to be more erratic and as a consequence, obtaining worse results:

Table 19: Comparison between the results performing or not a MSP classification

Season	MSP classification	Error	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Summer	NO	MAPE	2,540	5,179	1,823	2,177	1,748	7,573	4,479
		EME	2,342	5,057	1,672	2,111	1,555	5,844	4,878
	YES	MAPE	1,407	2,045	2,013	2,545	1,521	5,074	3,360
		EME	1,270	1,867	1,570	2,181	1,216	3,551	2,603
Winter	NO	MAPE	5,457	7,485	4,010	3,466	3,123	6,553	4,890
		EME	5,489	8,379	4,132	3,265	3,144	6,338	4,212
	YES	MAPE	2,474	3,770	1,587	1,871	1,876	4,113	3,179
		EME	2,455	3,709	1,572	1,860	1,863	3,273	2,566

In the following figures, the increase of error measures when not classifying into MSP groups can be appreciated better. The difference is specially noticeable in the displayed winter week, when maximum EME error is reduced from 8% to 4% for January 9.

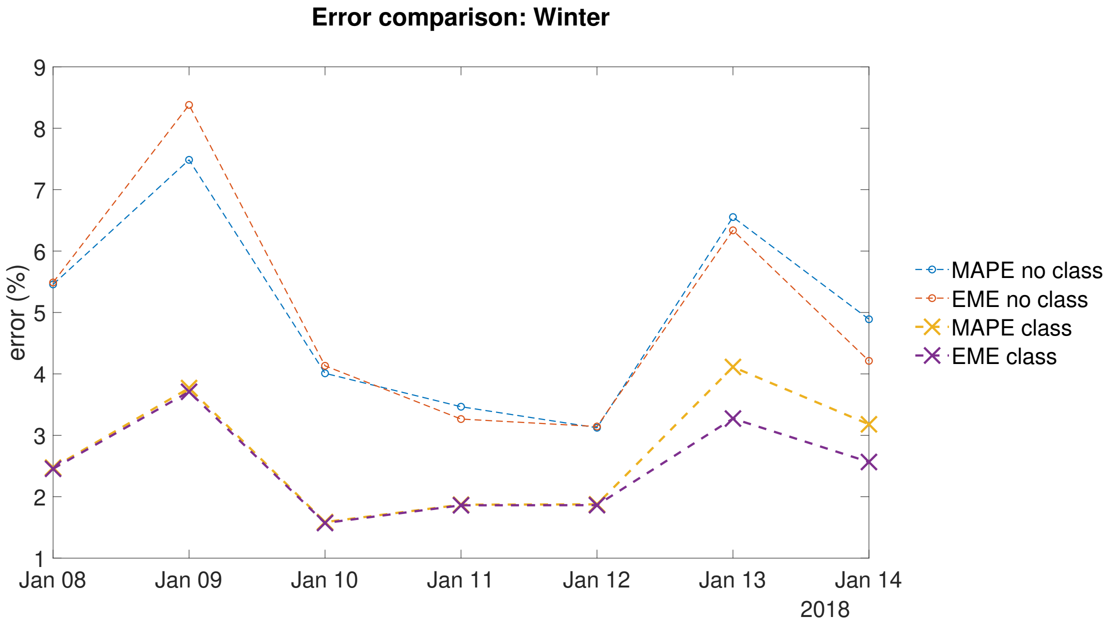


Figure 45: Comparison between the results performing or not a MSP classification in winter

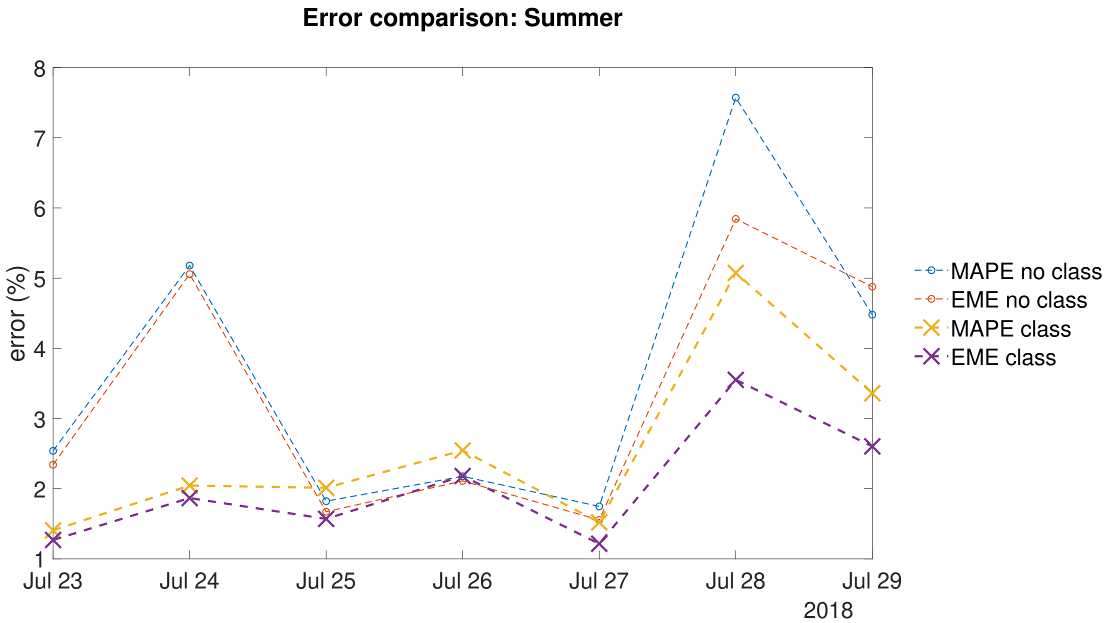


Figure 46: Comparison between the results performing or not a MSP classification in summer

Definitively, it has been proven that performing a MSP classification improves the results significantly. Depending on the use given to the methodology, it could be discussed if accuracy could be sacrificed in order to decrease the amount of efforts put into the classification of the MSP.

7.4 Discussion of results

Once seen the results of the methodology, in this section, the possibilities of the forecast procedure will be explained together with the risks that its implementation might have.
The methodology has been though mainly to provide municipalities with a tool to forecast its own consumption a day-ahead, to buy their own electricity in the daily wholesale market, whether they do it through starting a municipal retail company or by becoming wholesale consumers. In the daily market, as it was explained on Section 2.2.2, the electricity used in the day D must be purchased on day D-1. Hence, a procedure to forecast the consumption a day-head accurately is helpful and necessary. In (Mateo Barcos et al., 2020), the risks of not counting on a good forecast system to purchase electricity in the wholesale markets were studied and remarked on Section 3.1. It was seen that forecast error could lead to economic losses overcoming 500.000 € and even 1M € in more pessimistic scenarios (Figure 6).
Even though the methodology is applied and designed for the Valencia municipality, one of the advantages is that it can be replicated relatively easily to other municipalities. Results will vary depending on a lot of factors:

- The number of MSP that the municipality has. As bigger the municipality is, more consumption points can be added to the analysis, reducing considerably the error, as it was seen on Section 7.1, where the MSP groups where less load curves were available had way worse results. Hence, methodology will work better for big cities than for little ones. A good possibility for little cities or towns would be to aggregate their demand and purchasing electricity together in the wholesale markets. That way, data available would be higher and the results of their forecasts would improve, allowing smaller consumers to eliminate the cost of the retailing services.

- Distribution of the percentages of electricity consumption dedicated to the different MSP groups. For example, if a municipality count on less public lightning MSP and more offices, results could be a little bit worse, since PL is more stable along the year and is easier to predict. However, approximately the 60% of the electricity consumption of a municipality is dedicated to lightning. Percentages can vary but usually, the predominant consumption is PL.

- Results can also be different from one municipality to another depending on the data acquisition system that they have to measure temperature or other variables being introduced in the model.

It is also important to take into account that, even though results are good, they could be improved. Since this work is a master thesis, all has been done with the available data but in a context of a real implementation of the methodology, data availability could be much higher. As it was said on Section 6.1, from the 1.558 MSP points that the municipality had for the period of analysis, only 525 load curves were collected and valid for the analysis. By increasing data availability, more load curves could be included for every MSP group, improving its results. As was seen on Section 7.1, forecast of MSP groups like SC, CM, LM and PM have a considerable error that was assumed because of its low importance in relation to the total annual energy consumption. However, with higher data availability, this MSP groups could represent a more important share of the total consumption and a much lower error could be achieved, improving that way the model.

A forecast methodology like the one presented could also be used as a security tool to check that any fault is happening at the municipality dependencies. If the forecast methodology predicts a certain consumption on a MSP group in the day D-1 and then this consumption is highly exceeded on day D, that could indicate that something is not working properly. Several things could be happening like for example some problem with the installed equipment: malfunctioning of cooling or heating installations, fault of any type of machinery...etc. But also, it could indicate that the electricity meter is having some kind of issue and it is not giving a correct measure of the consumption, what should be immediately corrected not only to avoid problems when paying the bills but to avoid also training the ANN's with incorrect values that could affect future forecasts. However, the implementation of the presented forecast methodology could have some risk associated:

- Changes in the analysed building or dependencies can affect considerably the forecast. If new equipment is installed, ANN's will have a certain delay in being able to forecast correctly the new consumption. For example, if new offices are incorporated to the municipality dependencies, that would introduce a higher consumption to the PO MSP group. If no supervision or preparation of the forecast model is performed, prediction will work poorly for some days until ANN's are trained with enough days in the new situation to achieve good forecast results.

- Introduction of efficiency measures. As in the previous point, efficiency measures can also affect the performance of the forecast model. If for example all the public lightning is changed to LED, consumption will be drastically affected, affecting also the relation between inputs to the model and electrical consumption for some days. To avoid that, model should be prepared with data corresponding to the new situations. Definitively, every sudden change in the municipality dependencies can affect the performance of the forecast. If no supervision of the model is done, this changes could lead to important losses if electricity is being purchased in the wholesale markets, since imbalances could be high during the first days after any important change. However, this changes are usually gradual, so the model should be capable to adapt to new situations by itself.

- Need of reliable temperature data. Bad measures of temperature could lead to important errors in the model, since almost all of the MSP groups use as input temperature at hour t and temperature of the previous hours. If temperature measurement system has a failure and is not able to measure any temperature for a certain moment and gives a 0 value, forecast consumption will be much lower for almost all the MSP group, causing the total forecast to have an important error. This could be corrected easily with human supervision or by programming a security model to ensure that, if a situation like the one explained happen, temperature value doesn't drop to 0 but instead is calculated as the average of the previous hours, for example.

- Need of other inputs. If the model wants to be used in other municipality, some considerations should be taken into account. Other inputs could be needed. For example, some municipality's public lightning doesn't work with a fixed schedule depending on the time of the year, but instead has a system that measures solar radiation and if the measure is lower than a certain value, lights are tuned on. If that was the case for a municipality that wanted to use the forecast methodology, they should introduce to the model the solar radiation, instead of the variables used in PL for the case of study of Valencia, which are t_{sr} and t_{ss}.

- Need for technical experts. To check that everything works fine in the forecast, at least one expert in electricity systems and artificial intelligence is needed to avoid problems with the model and its possible consequences.

8 Conclusions

8.1 Summary of the results

The results presented on Section 7 show that the forecast methodology works well for the case of study of the Valencia municipality. The most accurate forecasts are those of the MSP groups that have more data availability, since they are trained with more data and their consumption is more aggregated, limiting the variance in the consumption patterns and facilitating the prediction of future behaviours.
The results of the total forecast are the ones presented on the following table for two randomly selected weeks of summer and winter:

Table 20: Results of the total forecast of the Valencia municipality consumption

Season	Error (%)	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Summer	MAPE	1,407	2,045	2,013	2,545	1,521	5,074	3,360
	EME	1,270	1,867	1,570	2,181	1,216	3,551	2,603
Winter	MAPE	2,474	3,770	1,587	1,871	1,876	4,113	3,179
	EME	2,455	3,709	1,572	1,860	1,863	3,273	2,566

As it can be seen, results are really good, with a maximum daily error of a 5% and a minimum error of a 1,27%. Taking into account that the methodology has been designed mainly to forecast the electricity consumption to purchase energy in the daily wholesale market, these results offer an error pretty low.

8.2 Limitations of the approach

Even though the methodology works fine, if it is used to purchase electricity in the daily wholesale market, it could have some limitations:

- The methodology is not ready yet for a real implementation of the prediction model. Even though the conditions of a real situation have been simulated, all the code involved has been done working with past data and it should be adapted if the methodology wanted to be actually used.

- Before a real implementation, the methodology should be tested with all the MSP data of a municipality. As seen on Table 5, for most of the MSP groups, few load curves are available. A simulation with all the data from a municipality should be done to know how different results are compared to the one presented on this work. However, error results shouldn't be much different, since most of the consumption for a municipality is PL, whose forecast has been proven to be very accurate. Nevertheless, for a real implementation of the methodology, this should be ensured.

- The methodology is designed for short-term forecast. In particular, all has been designed for a day-ahead prediction. If the methodology wanted to be used for another time horizon like week-ahead or another time scope, it should be adapted. Also, results should be checked since probably they would be worse. Probably, another forecast technique should be used instead of ANN.

8.3 Further work

The presented methodology has a lot of potential and work that could be done in the future. The following ideas are left to future development of the methodology and its possible applications:

- If the methodology has been mainly designed to give municipalities a tool to purchase electricity in the daily wholesale market, a calculation of the penalties due to imbalances should be done. To know exactly the penalties due to imbalances in the case of study, a further analysis should be done. Hourly imbalances obtained with the forecast methodology should be put in common with the hourly prices of the imbalances for the period of analysis. Furthermore, it should also be calculated the status of the power system at every hour as explained on Section 2.4.2, since depending on it, imbalances will result in losses or not. That way, an accurate calculation of the cost overruns due to imbalances for municipalities would be done, instead of an estimation based on a percentage of hourly imbalances (Mateo Barcos et al., 2020).

- A more complex software could be developed to create a forecast system that was continuously training with new data. That way, ANN would always train with the most recent data, adapting future forecast to changes in temperature, consumption patters...etc. Furthermore, security mechanisms could also be incorporated to the model if it's wanted to be continuously training. For example, some mechanism should be added to avoid that failures in the temperature data acquisition system had a major impact in future forecast and as a consequence on the imbalances.

- A field test would be very important to validate if the forecast methodology developed could have an implementation in the real world. That is, performing a forecast of some municipality in real-time or with current data instead of doing it with past data.

9 References

Ajuntament de Barcelona. (2020). Memòria explicativa pressupostos. https://ajuntament.barcelona.cat/pressupostos2021/docs/2021/nous/Mem%5C%C3%5C%B2ria%5C%20explicatva%5C%20pressupost%5C%202021%5C%20VF.pdf

Ajuntament de València. (2021). Resum pressupost municipal. https://www.valencia.es/documents/20142/9892427/PTO+2021+TOMO+RESUMEN.PDF/37aa3ade-141c-a5e7-dfe3-6e57ffdd8763?t=1610628477835

Akarslan, E., & Hocaoglu, F. O. (2018). Electricity demand forecasting of a micro grid using ann. *2018 9th International Renewable Energy Congress (IREC)*, 1–5. https://doi.org/10.1109/IREC.2018.8362471

Alcázar Ortega, M., Cañasas Peñuelas, C. S., Escrivà Escrivà, G., Fuster Roig, V. L., & Roger Folch, J. (2019). *Generación, transporte y distribución de energía eléctrica*. Editorial Universitat Politécnica de València.

Ayuntamiento de Madrid. (2020). Cuenta anual. https://www.madrid.es/UnidadesDescentralizadas/Presupuestos/EjecucionPresupuestaria/Cuentas_aprobadas/FicherosCuentas/CUENTA_AYTO_2020_WEB.pdf

Barcelona Energía. (2021). *Operadora eléctrica metropolitana*. https://www.barcelonaenergia.cat/

Bartelj, L., Gubina, A., Paravan, D., & Golob, R. (2010). Risk management in the retail electricity market: The retailer's perspective. *IEEE PES General Meeting, PES 2010*. https://doi.org/10.1109/PES.2010.5589578

Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in italy using linear regression models. *Energy*, *34*(9), 1413–1421. https://doi.org/https://doi.org/10.1016/j.energy.2009.06.034

Boroumand, R. H., & Zachmann, G. (2012). Retailers' risk management and vertical arrangements in electricity markets [Strategic Choices for Renewable Energy Investment]. *Energy Policy*, *40*, 465–472. https://doi.org/https://doi.org/10.1016/j.enpol.2011.10.041

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Bristol Energy. (2021). *Electricity and gas for social good*. https://www.bristol-energy.co.uk/

C40 cities. (2021). *Why cities?* https://www.c40.org/why_cities

Cabello García, S. (2020). Análisis de los desvíos del sistema eléctrico español en el 2019.

Cambranos, T. (2019). Avilés amplía la compra directa de energía a 53 edificios municipales y colegios. *La Nueva España*. https://www.lne.es/aviles/2019/02/27/aviles-amplia-compra-directa-energia-18464426.html

Cancelo, J. R., Espasa, A., & Grafe, R. (2008). Forecasting the electricity load from one day to one week ahead for the spanish system operator [Energy Forecasting]. *International Journal of Forecasting*, *24*(4), 588–602. https://doi.org/https://doi.org/10.1016/j.ijforecast.2008.07.005

Carbajo, A. (2007). Los mercados eléctricos y los servicios de ajuste del sistema. *Red Eléctrica de España*.

Chen, Y., Luh, P. B., Guan, C., Zhao, Y., Michel, L. D., Coolbeth, M. A., Friedland, P. B., & Rourke, S. J. (2009). Short-term load forecasting: Similar day-based wavelet neural networks. *IEEE Transactions on Power Systems*, *25*(1), 322–330.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Diario de Rivas. (2018). Rivas se ‘independiza’ de las eléctricas. https://www.diarioderivas.es/rivas-compra-electricidad-ayuntamiento/

Eléctrica de Cádiz. (2021). *Empresa pública de electricidad*. https://comercializadora.electricadecadiz.es/

Entso-e. (2021). *Single intraday coupling (sidc)*. https://www.entsoe.eu/network_codes/cacm/implementation/sidc/

Escrivá-Escrivá, G., Álvarez-Bel, C., Roldán-Blay, C., & Alcázar-Ortega, M. (2011). New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. *Energy and Buildings*, *43*(11), 3112–3119.

Escrivá-Escrivá, G., Roldán-Blay, C., & Álvarez-Bel, C. (2014). Electrical consumption forecast using actual data of building end-use decomposition. *Energy and Buildings*, *82*, 73–81. https://doi.org/https://doi.org/10.1016/j.enbuild.2014.07.024

Farinaccio, L., & Zmeureanu, R. (1999). Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings*, *30*(3), 245–259. https://doi.org/https://doi.org/10.1016/S0378-7788(99)00007-9

Generous, N., Fairchild, G., Deshpande, A., Del Valle, S. Y., & Priedhorsky, R. (2014). Global disease monitoring and forecasting with wikipedia. *PLoS computational biology*, *10*(11), e1003892.

Gestore mercati energetici. (2021). *Price coupling of regions*. https://www.mercatoelettrico.org/en/mercati/mercatoelettrico/PCR.aspx

Ghisi, E., Gosch, S., & Lamberts, R. (2007). Electricity end-uses in the residential sector of brazil. *Energy Policy*, *35*(8), 4107–4120. https://doi.org/https://doi.org/10.1016/j.enpol.2007.02.020

Goodwin, P., & Lawton, R. (1999). On the asymmetry of the symmetric mape. *International Journal of Forecasting*, *15*(4), 405–408. https://doi.org/https://doi.org/10.1016/S0169-2070(99)00007-2

Haida, T., & Muto, S. (1994). Regression based peak load forecasting using a transformation technique. *IEEE Transactions on Power Systems*, *9*(4), 1788–1794.

Hamburg energie. (2021). *Hamburg energie*. https://www.hamburgenergie.de/

Hong, T. (2013). Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*, 43–48.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679–688.

IBM Cloud Education. (2020). *Neural networks*. https://www.ibm.com/cloud/learn/neural-networks

Intergovernmental Panel on Climate Change. (2014). *Climate change 2013 – the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

Kumar, R., Aggarwal, R., & Sharma, J. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, *65*, 352–358. https://doi.org/https://doi.org/10.1016/j.enbuild.2013.06.007

Kuster, C., Rezgui, Y., & Mourshed, M. (2017a). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, *35*, 257–270.

Kuster, C., Rezgui, Y., & Mourshed, M. (2017b). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, *35*, 257–270.

Li, K., Hu, C., Liu, G., & Xue, W. (2015). Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy and Buildings*, *108*, 106–113.

Magnus commodities. (2022). *Continuous intraday market europe electricity*. https://www.magnuscmd.com/continuous-intraday-market-europe-electricity/

Mateo Barcos, S., Alcázar Ortega, M., & Ribó Pérez, D. (2020). Estudio de viabilidad de la constitución de una comercializadora pública de energıa en la ciudad de valencia.

Mena, R., Rodrıguez, F., Castilla, M., & Arahal, M. R. (2014). A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy and Buildings*, *82*, 142–155.

MIBEL. (2021). *El mercado ibérico de la electricidad – mibel*. https://www.mibel.com/es/home_es/

Moghaddam, M. P., Abdollahi, A., & Rashidinejad, M. (2011). Flexible demand response programs modeling in competitive electricity markets. *Applied Energy*, *88*(9), 3257–3269. https://doi.org/https://doi.org/10.1016/j.apenergy.2011.02.039

Mohandes, S. R., Zhang, X., & Mahdiyar, A. (2019). A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing*, *340*, 55–75.

Moré, J. J. (1978). The levenberg-marquardt algorithm: Implementation and theory. In G. A. Watson (Ed.), *Numerical analysis* (pp. 105–116). Springer Berlin Heidelberg.

Murthy, K. N., Sumithra, G. D., & Reddy, A. K. (2001). End-uses of electricity in households of karnataka state, india. *Energy for Sustainable Development*, *5*(3), 81–94.

Næs, T., Kvaal, K., Isaksson, T., & Miller, C. (1993). Artificial neural networks in multivariate calibration. *Journal of Near Infrared Spectroscopy - J NEAR INFRARED SPECTROSC*, *1*. https://doi.org/10.1255/jnirs.1

Ojanen, O. (2002). Comparative analysis of risk management strategies for electricity retailers. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.2966

OMIE. (2020). *Información general pcr*. http://m.omie.es/inicio/mercados-y-productos/informacion-general-pcr?m=yes

OMIE. (2021). *Electricity market*. https://www.omie.es/es/mercado-de-electricidad

Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, classifiaction.

Pina, A., Silva, C., & Ferrão, P. (2012). The impact of demand side management strategies in the penetration of renewable electricity [23rd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, ECOS 2010]. *Energy*, *41*(1), 128–137. https://doi.org/https://doi.org/10.1016/j.energy.2011.06.013

Pisner, D. A., & Schnyer, D. M. (2020). Chapter 6 - support vector machine. In A. Mechelli & S. Vieira (Eds.), *Machine learning* (pp. 101–121). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Plumb, A. P., Rowe, R. C., York, P., & Brown, M. (2005). Optimisation of the predictive ability of artificial neural network (ann) models: A comparison of three ann programs and four classes of training algorithm. *European Journal of Pharmaceutical Sciences*, *25*(4-5), 395–405.

Radiotelevisión del Principáu d'Asturies. (2014). Avilés se ahorrará cerca de 130.000 euros en energía. https://www.rtpa.es/noticias-asturias:Aviles-se-ahorrara-cerca-de-130.000-euros-en-energia_111399895261.html

Ranganathan, A. (2004). The levenberg-marquardt algorithm. *Tutoral on LM algorithm*, *11*(1), 101–110.

Roldán-Blay, C., Escrivá-Escrivá, G., Álvarez-Bel, C., Roldán-Porta, C., & Rodrıguez-Garcıa, J. (2013). Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model. *Energy and Buildings*, *60*, 38–46.

Saeed, M. (2021). *A gentle introduction to sigmoid function*. https://machinelearningmastery.com/a-gentle-introduction-to-sigmoid-function/

Salida y puesta del sol. (2017). *Hora de salida y puesta del sol en valencia*. https://salidaypuestadelsol.com/sun/valencia_(spain)

Senjyu, T., Mandal, P., Uezato, K., & Funabashi, T. (2005). Next day load curve forecasting using hybrid correction method. *IEEE Transactions on power systems*, *20*(1), 102–109.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., Kamaev, V. A., et al. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, *24*(24), 171–176.

Singh, A. K., Khatoon, S., Muazzam, M., Chaturvedi, D., et al. (2012a). Load forecasting techniques and methodologies: A review. *2012 2nd International Conference on Power, Control and Embedded Systems*, 1–10.

Singh, A. K., Khatoon, S., Muazzam, M., Chaturvedi, D., et al. (2012b). Load forecasting techniques and methodologies: A review. *2012 2nd International Conference on Power, Control and Embedded Systems*, 1–10.

Som energia. (2016). *Análisis viabilidad comercialización eléctrica y posibilidades*. https://www.pamplona.es

Strbac, G. (2008). Demand side management: Benefits and challenges. *Energy policy*, *36*(12), 4419–4426.

Urban Climate Change Research Network (UCCRN). (2018). The future we don't want. how climate change could impact the world's greatest cities.

Wei, T., Wu, J., & Chen, S. (2021). Keeping track of greenhouse gas emission reduction progress and targets in 167 cities worldwide. *Frontiers in Sustainable Cities*, *3*. https://doi.org/10.3389/frsc.2021.696381

Yang, J., Rivard, H., & Zmeureanu, R. (2005). Building energy prediction with adaptive artificial neural networks. *Ninth International IBPSA Conference Montréal*, 15–18.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, *14*(1), 35–62.

Part II
PROJECT BUDGET

10 Previous remarks

In this section the cost of performing the design and analysis of artificial neural networks for the prediction of the electricity demand of public administrations based on type of consumption decomposition is quantified. The study was performed over a period of approximately 1 year, taking into account the literature review on the field of electricity load forecast, the data acquisition, methodology design and development, study of the Spanish electricity market, methodology application to the case study, writing of the report and supervision.

11 Budget

11.1 Human resources

The human resources involved in this TFM were a junior engineer (JE), a doctoral student in electrical engineering (DS) and a specialist professor in electricity systems and markets (SP). The tasks developed by the junior engineer include all the literature review, the data acquisition, the methodology design, development and application to a case of study and writing. Tasks carried by the doctoral student include the conceptualization of the work and continuous supervision on all the aspects of the TFM. Lastly, the specialist professor has provided feedback at different stages of the TFM and has suggested important changes in the methodology, results delivery and structure of the final document. The following table summarizes the total human resources costs:

Table 21: Human resources costs

Task	Dedicated time	Position	Unit cost	Cost
Literature review	100 h	Junior engineer	30 €/h	3.240
Data acquisition	100 h	Junior engineer	30 €/h	3.240
Methodology design	280 h	Junior engineer	30 €/h	9.750
Simulation (application)	350 h	Junior engineer	30 €/h	12.960
Writing	54 h	Junior engineer	30 €/h	1.620
Revision and feedback	65 h	Doctoral student	55 €/h	3.575
Revision and feedback	25 h	Professor	70 €/h	4.550
TOTAL	974 h	–	–	31.845

11.2 Software and hardware amortization costs

The hardware equipment used in this TFM consists of a computer and a laptop belonging to the junior engineer and two more computers belonging to the doctoral student and the specialist professor. Engineer's equipment has been amortised in 12 months, while the remaining equipment has been amortised for the hours of use remarked on Table 21, which have been approximated to one month.

The software used in the project consists of a Microsoft 365 Personal package for the three people working on the TFM and a Matlab license for the junior engineer, together with the Matlab's Deep Learning Toolbox, used to program the ANN's. Also, the document has been written in Latex trough the editor Overleaf. However, this editor is free nowadays so no cost is associated to it.

Table 22: Software and hardware costs

Equipment	Price	Amortization period	Time of use	Cost
PC de sobremesa LG con procesador Intel Core 2 Duo (JE)	600 €	6 years	12 months	100 €
Lenovo ideapad 3i5 (JE)	950 €	6 years	12 months	158 €
Macbook Pro13' (DS)	1.000 €	6 years	1 months	14 €
PC Pentium IV2 GHz (SP)	1.500 €	6 years	1 months	21 €
Microsoft 365 Personal (JE)	62 €/year	–	12 months	62 €
Microsoft 365 Personal (DS)	62 €/year	–	1 month	5,17 €
Microsoft 365 Personal (SP)	62 €/year	–	1 month	5,17 €
MATLAB and Simulink Student Suite (JE)	69 €	2 years	12 months	34,5 €
MATLAB's Deep Learning Toolbox (JE)	7 €	2 years	12 months	3,5€
TOTAL				403,34 €

11.3 Other costs

Here, the following costs have been considered:

- The cost of transport to the workplace. The unit cost of the transport has been calculated taking into account that the junior engineer travels to its workplace by car, the average diesel price in 2022 and the distance from the engineer's home to the workplace. Also, the transport cost of the Doctoral student and the Specialist professor has been calculated as a 10% of the student's transport cost.
- The cost of the workplace itself (office material, cost of the rent...etc). This will be calculated as a 1% of previous costs.
- Indirect costs like internet bills, heating and cooling in the offices...etc. This will be calculated as a 0,5% of the previous costs.

Table 23: Transport cost

Component	Unit cost	Cost
Transport to workplace (JE)	0,091 €/km	874 €
Transport to workplace (DE and SP)		87,4 €
TOTAL		961,4 €

11.4 Budget summary

Now all the costs explained in the previous sections are added to obtain an aggregated gross budget. A profit margin is also calculated, considering a realistic commercial situation. Also, taxes are included in the budget. The final budget is eighty thousand seven hundred and fifty euros with fifty cents.

Table 24: Final budget

Component	Cost
Human resources	31.845 €
Software and hardware	403,34 €
Transport	961,4 €
Workplace costs	332 €
Indirect costs	166 €
Industrial margin (6%)	2.022,5
Taxable budget	35.730,4
IVA (21%)	7.503,4
TOTAL BUDGET	80.750,5