ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

MASTER’S DEGREE IN THE ELECTRIC POWER INDUSTRY (MEPI)

MEDIUM-TERM ELECTRICITY LOAD FORECASTING

Author: Pablo de Lera Figal
Director: Antonio Muñoz San Roque

Madrid
July 2016

Master's Thesis Presentation Authorization

THE STUDENT:

Pablo de Lera FIgal

…………………………………………………………

THE SUPERVISOR

Antonio Muñoz San Roque

Signed: ……………………… Date: 11…/ 07./ 2016

Authorization of the Master's Thesis Coordinator

Dr. Luis Olmos Camacho

Signed: ……………………… Date: ……/ ……/ ……

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

MASTER’S DEGREE IN THE ELECTRIC POWER INDUSTRY (MEPI)

MEDIUM-TERM ELECTRICITY LOAD FORECASTING

Author: Pablo de Lera Figal
Director: Antonio Muñoz San Roque

Madrid
July 2016

MEDIUM-TERM ELECTRICITY DEMAND FORECASTING

Author: de Lera Figal, Pablo
Director: Muñoz San Roque, Antonio
Collaborating institutions: ICAI Universidad Pontificia Comillas

ABSTRACT

In the context of competitive power markets and liberalization, electricity can be bought and sold at market prices like any other commodity. However, unlike most other commodities, electricity cannot be stored for future use in massive quantities. The process of power generation, transmission, distribution and consumption usually happens at the same time.

This unique characteristic makes electricity a complex commodity to handle. Additionally, seasonality has to be taken into account, at the daily, weekly and annual time scales. Moreover, there are many exogenous variables to be considered as key drivers of electricity demand patterns such as weather conditions, economic activity, regional market characteristics and working patterns.

Thus, traditional risk-averse electric utilities have to deal with a substantially increasing amount of risk. Managing a company in an efficient manner involves yet more and more statistical analysis, as well as careful forecasting both electricity demand and prices.

Deregulation has made forecasting a key necessity for all market agents to hedge their corresponding risk exposures. Traditionally, vertically-integrated utilities used short-term load forecasts to ensure security of supply, and long-term load forecasts for future capacity investments. However, since competition was introduced, this is no longer the case, and the minimization of volumetric risk has never been of such importance as it is today. For this reason, load forecasting is gradually becoming the most important stage for utilities, system operators, retailers and other market participants both at the planning and operation&maintenance levels.

Demand forecasting deals with hourly, daily, weekly and monthly values of the system load and peak system load. The forecasting of different horizons is important for different activities within a company. This distinction has typically leaded to a forecasting classification within time horizons. Although the thresholds that detach them differ within publications and authors, the common clusters are: short-term, medium-term and long-term. Other publications may also include very-short-term (real time) and very-long-term load forecasting as well.

This Master´s thesis addresses the problem in a comprehensive way. It first analyses the state of the art in energy forecasting, making a clear distinction between time horizons and main areas including price, load and renewable energy sources forecasting among others. Then, a main classification of different methods is done, in which both qualitative and quantitative approaches are reviewed. Within the latter group, both explanatory and time-series models are introduced. The section finishes by stating the main methods and methodologies used in electricity load forecasting.

Then, the models are presented following the same structure. Fist, the main variables are analysed, then the model is developed and the mathematical equation computed. Finally, results are obtained and the model is assess taking into account different statistical measures such as the MAPE (accounting for the model’s error) and the R squared (which gives insight on how much of the data is being explained by the model).

After comparing both models developed, a final section is devoted to the conclusions and future developments so improvement insights are given for further research on the topic.

Table of contents

List of figures

List of tables

Part I MEDIUM-TERM ELECTRICITY LOAD FORECASTING

Chapter 1 INTRODUCTION

In the context of competitive power markets and liberalization, electricity can be bought and sold at market prices like any other commodity. However, unlike most other commodities, electricity cannot be stored for future use in massive quantities. The process of power generation, transmission, distribution and consumption usually happens at the same time.

This unique characteristic makes electricity a complex commodity to handle. Additionally, seasonality has to be taken into account, at the daily, weekly and annual time scales. Moreover, there are many exogenous variables to be considered as key drivers of electricity demand patterns such as weather conditions, economic activity, regional market characteristics and working patterns.

Thus, traditional risk-averse electric utilities have to deal with a substantially increasing amount of risk. Managing a company in an efficient manner involves yet more and more statistical analysis, as well as careful forecasting both electricity demand and prices.

Deregulation has made forecasting a key necessity for all market agents to hedge their corresponding risk exposures. Traditionally, vertically-integrated utilities used short-term load forecasts to ensure security of supply, and long-term load forecasts for future capacity investments. However, since competition was introduced, this is no longer the case, and the minimization of volumetric risk has never been of such importance as it is today. For this reason, load forecasting is gradually becoming one of the most important stages for utilities, system operators, retailers and other market participants both at the planning and operation&maintenance levels.

Demand forecasting deals with hourly, daily, weekly and monthly values of the system load and peak system load. The forecasting of different horizons is important for different activities within a company. This distinction has typically leaded to a forecasting classification within time horizons. Although the thresholds that detach them differ within publications and authors, the common clusters are: short-term ranging from hours to a few weeks in the future; medium-term, from a month to a year; and long-term, covering between a year and ten in the future. Other publications may also include very-short-term (real time) and very-long-term load forecasting as well.

With supply and demand fluctuating, electricity prices enormous volatility and the widely concern about security of supply, energy forecasting has become increasingly important since the rise of competitive energy markets.

This Master´s thesis addresses the problem in a comprehensive way. It first recognizes and analyses the main drivers of electricity load patterns. Then it accomplishes data mining techniques to appropriately treat the information gathered. Next step is to build the model and obtain advantageous results. Finally the model is review, and so the results obtained, in order to extract key insights for future work.

The proposed methodology has been used to forecast the electricity load of Spain, using data with a sufficiently long time-frame. This is the case, as the daily, monthly and annual data reflects the Spanish demand for the period 2005-2015. The data was kindly proportioned by Red Eléctrica de España.

Concerns about energy security of supply, renewable sources management and overall system efficiency had traditionally led to a deep branch of electrical engineering devoted to load forecasting.

In this context, a variety of methods had been deployed to help system operators, generators and other important players, to perform their corresponding activities in a more advantageous and bold manner.

For the short term scope, we can find very interesting methods such as: semi-parametric additive models [1], multi-region regression models [2] or neural networks [3]. Moreover, the univariate and multivariate methods [4] developed by Box-Jenkins [5] can be applied both for the short and medium time-scales.

Research on the medium term load forecasting includes some really smart models such as: multilayer perceptron models (MLP) or support vector machines (SVM) [6] which will be discussed later in the state of the art section.

In order to address the risk management part of the business, utilities had also used other medium-term approaches as probabilistic forecasting methods. Framed in this context, the Quantile Regression Averaging [7] is particularly important [8] computing electricity spot price prediction intervals.

For the long term it's important to forecast the load in a probabilistic manner [9], as well as the expected peak demand [10].

To address the Spanish system idiosyncrasy, a number of papers and publications specially focused on the issue had been considered. In this context the review of the load forecasting from one day to one week ahead [11] and the modelling of non-linear response of Spanish electricity demand to temperature [12] were especially important for the development of the project.

The state of the art review cannot conclude without mentioning two major papers on energy forecasting that had guide the overall thesis and helped to tune some of its most relevant sections. The first one addresses a very similar problem in a deep mode [13]. The second could not be other than the widely famous Rafał Weron´s statistical approach to electricity loads and prices modelling and forecasting [14].

MOTIVATION OF THE PROJECT

One of the key reasons to develop comprehensive medium and long term load forecasting models, it is to guarantee the security of supply within these time horizons. Due to the complexity, cost and length that building new power plants and facilities involve, it is important to have a good insight on how the demand is expected to grow. In this sense, load forecasting is vital for our modern society and welfare, as it guarantees the good functioning of the supply in the long term.

Moreover, taking a financial point of view, investments carried out by utilities tend to have a deep impact on developed country's economy and PIB. Thus, long-term forecasting turns out to be relevant indicators to consider, not only within the electric sector, but to assess the overall evolution of a certain country's economy.

There are different agents among the electric power industry: generators, distributors, retailers, regulators, technical operators (including both system and market), ministries… For all of them, load forecasting is key, besides each one has its own objectives and typically tends to focus more in one specific time frame.

Regarding the long term, the regulator, the ministry and the system operator, should have a good load forecasting in order to send efficient economic signals to the generators. Only this way, generators will manage to perform their investments in an efficient manner, guarantying the future security of supply. In the medium term, generators should stock their future needs to operate and maintain their power plants optimally. Distributors should stablish, maintain and better their distribution grids, suiting them for the predicted demand. Finally, retailers should also know the future expected consumption, to hedge against the risk for their client portfolio.

Managing a portfolio of clients involves buying electricity in different time horizons. In the short term, retailers have to participate in the day-ahead and intraday markets, where different agents buy and sell electricity within the offer and demand rules. These markets have a predefined order, when the retailers look for their electricity needs within its price. The electricity price is obtained matching the offer and demand bidding curves.

The electricity price is the result of different agents' strategies, the portfolio of client behaviour and the system resources. Besides the uncertainty caused by the behaviour of the clients, it also exists uncertainty regarding the volatility of raw materials and commodity prices, so it is only natural that electricity prices suffer from a high volatility in those markets too. With the objective to minimize this risk on the prices, agents can also agree contracts to buy or sell energy in the medium or long term within different time limits.

There are mainly two types of long-term markets: physical (OTC) and financial. In order to attend this type of markets retailers should know their clients' electricity demand as good as possible. Only this way it would be possible for them to hedge the risk for the tight amount of energy they need. In this sense, they should develop forecasting models to foresee the medium and long term needs of their clients.

In general terms, a retailer's portfolio is composed of big industrial consumers, and small domestic clients who are generally connected to low voltage. The former ones tend to have much more stable profiles, while the domestic clients don't usually have hourly meters to monitor their consumption on a meaningful way.

Forecasting models have been evolving and adapting to the new market situations and regulatory framework. Moreover, socioeconomic activity and relevant events have to be taken into account. In this context, the situation of economic crisis has caused significate drops in the demand and its behaviour. This type of situation has to be taken into consideration within our model.

For this reason, it is necessary to re-design the previous models, taking into account not only notorious variables as working-day effect or temperature, but also these types of events. A good model has to look for all the variables that influence the demand patterns in order to make a prediction as close to reality as possible.

OBJECTIVES

The main objectives of the project are:

1. To carry a practical study of energy forecasting techniques. Identify the most suitable modelling trend within the diversity of approaches available (optimization, equilibrium and simulation models) for the medium term planning purposes.

2. To perform a suitable data mining taking daily and monthly data from the Spanish system operator. With this step we aim to develop an explanatory model for the medium-term electricity demand as a tool for the optimal planning of market agents.

3. To integrate the key drivers of electricity demand (i.e. meteorology, calendar effect, economic development and efficiency increase) within the model framework.

4. To adjust and validate the model within a reasonable time frame, and to perform a bold comparison with other models.

METHODOLOGY

The methodology to be applied consist of data-mining the daily time series available. The model will forecast the electricity load by decomposing its most relevant explanatory variables

First of all, we should correct the series affected intrinsically by labour patterns.

Moreover we should model the weather influence on electricity demand. In this section we should devote special attention to temperature and its non-linear influence and modelling possibilities.

Economic activity should also be introduced in the model. In this regard it could be done a distinction between sectors: industry, services, domestic consumers …

To address the demand decrease suffered in Spain during the last years, we need

to look for explanatory variables to introduce in our model. In this context, energy efficiency plays a key role and should be considered in addition to the aforementioned variables. Other variables can be considered as well to describe the load decline.

RESOURCES AND TOOLS APPLIED

The very first stage will include an in-depth research about the state of the art, models currently applied for load forecasting in the medium-term, as well as key factors affecting electricity demand. In this regard, it is know that economic activity, labour patterns and weather conditions (especially temperature) are important drivers of the demand time series.

Framed in this very beginning milestone, it should be found relevant information of the aforementioned factors on different time scopes.

The electricity demand was provided by Red Eléctrica de España (REE) on a daily basis. Daily labour patterns were based on national and regional holidays published in the Spanish Official State's Bulletin (BOE). Temperature was also analysed on a daily basis, while most of the economic activity indexes used were framed on a monthly on even annual time scope.

The following table sums up the data examined, scope and its sources:

Data	Scope	Sources
Electricity load	Daily	Red Eléctrica de España (REE)
Temperature	Daily	Agencia Estatal de Meteorología (AEMET)
Calendar effect	Daily	Boletín Oficial de Estado (BOE)
Economic activity (IPI, labour force, average disposable income…)	Monthly	Ministerio de Industria, Energía y Turismo (MINETUR), Instituto Nacional de Estadística (INE)
Energy efficiency (indexes, housing stock…)	Yearly	Instituto para diversificación y Ahorro de Energía (IDAE)

Table 1. Data, scope and sources analysed

Once all information has been gathered, the second milestone will comprise the calendar correction of the time series, the temperature explanation and the economic activity model. For this purpose, Matlab will be a useful tool, as well as Excel and VBA in particular.

The final milestone aims to collect the previous work on a single medium-tem explanatory model. The final results, future possible developments and conclusions have to be stated.

In this case the tool applied is the Intelligent Data Analysis Tool (IDAT) developed at IIT (Instituto de Investigación Tecnológica).

Chapter 2 STATE OF THE ART IN ELECTRICITY LOAD FORECASTING

INTRODUCTION: ENERGY FORECASTING

The term "energy forecasting" refers to both point and probabilistic/density forecast. The distinction is as follows:

·**Point forecasting:** forecast of the mean or median of the future energy data.

·**Probabilistic/density forecasting [9]:** provides estimates of the full probability distribution of possible future values of the energy data, which can be demand (in case of electricity load forecasting), price (energy price forecasting), RES (renewable energy sources forecasting) or other type of data.

MAIN AREAS

Energy forecasting comprises a wide variety of topics involving forecasting in the energy industry. The most notable areas (although not the only ones) of research include:

·**Electricity load forecasting [3]:** demand power (MW) or energy (MWh) forecast concerning hourly, daily monthly or yearly values of load and peak load.

·**Energy price forecasting [8]:** electricity and fossil fuels (gas, coal, oil…).

·**Renewable energy sources forecasting [18]:** mainly wind solar and hydro .

HORIZONS

Although there is no clear consensus in literature as what the threshold should actually be, it is common to distinguish between:

·**Short-term forecasting:** involves horizons from a few minutes up to few days ahead and is of particular importance in day-to-day market operations. In load forecasting, very-short term forecasting deals with times in minutes and is often considered separately.

·**Medium-term forecasting:** from a few days to a few months ahead, is generally used for risk management and derivatives pricing. This type of modelling has a long tradition in finance, where probability distributions on prices are used.

·**Long-term forecasting:** with time lead measured in months or years, it focuses on investment analysis and planning.

FORECASTING METHODS

Forecasting techniques can be classified in two main groups

·**Quantitative methods:** used when sufficient information about the past is available and the information can be set as a numerical time series. The main assumption is that historical data can be used to make predictions about future data

·**Qualitative methods:** used when the past does not provide direct information about the future behaviour of the time series. They have the advantage of being very flexible. However, they are subjective methods, which are biased by the opinion and experience of a group of experts, so it is difficult to quantify the accuracy of their forecasts.

QUANTITATIVE METHODS

Within quantitative methods we can find different types of models:

Explanatory models

Explanatory models assume that the variable to be forecasted exhibits an explanatory relationship with one or more other variables.

The general expression:

$$y = f(x_1, x_2, \dots x_n, noise)$$

In the case of electricity load forecasting we might model the demand as:

$$demand = f(temperature, economic\ activity, time\ of\ day, day\ of\ week, \dots, noise)$$

Models in this class include **regression models** [20], **decomposition methods** (most notably **additive** and **multiplicative models)**, and some kinds of **neural networks** [3].

Univariate time series models

They aim at predicting a variable based on its past time-sequence observations.

The general expression:

$$y(t) = f(y(t-1), y(t-2), \dots, noise)$$

In the case of electricity load forecasting:

$$demand(t) = f(demand(t-1), demand(t-2), \dots, noise)$$

The prediction of the future is based on past values of a variable and/or past errors, but not on explanatory variables which may affect the system. Time series models used for forecasting include **ARIMA** models [16] and **exponential smoothing** [19].

Mixed models

They use both time series and explanatory variables.

The general expression:

$$y(t) = f(y(t-1), x_1, x_2, \dots x_n, \dots, noise)$$

An example expression:

$$demand(t) = f(demand(t-1), temperature, economic\ activity, time\ of\ day, \dots, noise)$$

These types of models have been given various names in different disciplines. They are known as **dynamic regression models** (where two main methods apply: traditional Box-Jenkins [5] based on cross-correlations; and Linear Transfer Function proposed by Liu and Hanssens: **LTF** [21]), **panel data models**, **longitudinal models**, **transfer function models**, and **linear system models** (assuming f is linear).

In all cases the observation integrates two components:

·**Pattern:** accounts for the predictable part that repeats with a regular manner.

·**Noise:** accounts for the error due to random variation and the effects of relevant variables not included in the model.

The objective of the modelling process is to separate both components, in order to use the pattern for forecasting and the noise for characterizing the prediction errors

ELECTRICITY LOAD FORECASTING

As it was stated before, electricity load forecasting is one of the main topics of energy forecasting.

The classification of electricity load forecasting can be also done following the same criteria as above: point vs probabilistic, quantitative vs qualitative methods, explanatory vs time series models …

Regarding time scopes; it is typical to include very-short term load forecasting, so the classification would be as follows:

·**Very-short term load forecasting (VSTLF):** only requires past loads.

·**Short term load forecasting (STLF):** usually required past loads and weather information.

·**Medium term load forecasting (MTLF):** requires weather and economic information.

·**Long term load forecasting (LTLF):** needs weather, economic, demographic and further information.

TECHNIQUES, MODELS AND METHODS

Before going any further it might be useful to clarify the following terms**:**

·**Technique:** group of **models/methods** that fall in the same family, such as Multiple Linear Regression (MLR) models, Artificial Neural Networks (ANN).

·**Methodology:** represents the general solution framework that can be implemented with **multiple techniques**. For example, a variable selection methodology may be applicable to both MLR models and ANNs.

Some of the most notable models are now analysed.

Multiple Linear Regression (MLR)

Regression analysis is an explanatory, mathematical process that states the relationship among dependent and independent (explanatory) variables.

The general expression:

$$y(t) = a_0 + a_1 * x_1(t) + \cdots + a_n * x_n(t) + \varepsilon(t)$$

In case of just one explanatory variable the model is called simple linear regression, opposed to multiple linear regression (MLR) where there are more than one explanatory variable.

The regression coefficients $(a_0, a_1, \ldots a_n)$ are often estimated by the least squares method.

Load is often the dependent variable while whether and calendar effect (working day effect-WD) are the independent variables.

As an example:

$$demand(t) = a_0 + a_1 * WD(t) + a_2 * Temp(t) + \varepsilon(t)$$

Multiple linear regression models have been used in long and short term load forecasting. Hong 2010 [20] proposed applying MLR to STLF computing trend and calendar effect. Linear regression models have also been applied for LTLF in [22], where various scenarios where taken into account to generate a long term probabilistic load forecast. The authors showed that the models based on hourly data had less ex post forecasting errors than the ones based on monthly or daily data.

Artificial Neural Networks (ANN)

ANNs are also explanatory models framed in the artificial intelligence (AI) area together with fuzzy logic and Support Vector Machines [6].

They are black-boxes that learn from empirical examples. They do not require the forecaster to model the underlying physical system. A mapping between input variables and electricity demand is adopted for the prediction.

A number of architectures have been used, including back propagation, Hopfield, Boltzmann machine…Literature on ANN is extensive, and includes Multiple-Layer Perceptron [27] (MLP) and Radial Basis Function Network [15] (RBFN).

The steps of ANN are as follows:

1. Select the training data and accuracy.

2. Select the network architecture. Decisions to be taken: number of layers that make up the network, number of neurons that make up each layer, and transfer functions corresponding to each one of the neurons in the hidden layer and output layer.

3. Network training. To do so we minimize the difference between the output obtained and the actual, in order to obtain the values of the parameters contained in the network. This process is repeated until the difference is small enough. In case this objective is not reached, the network's architecture is redefined.

4. Forecast from the values of the parameters that make up the network, obtained in the previous step.

5. Validation of the results. To do so, we compute the error between the value obtained by the network and the actual value. If the error is small, the network is considered adequate enough. If the error is too big, the process is repeated.

Artificial Neural Networks have been used for load forecasting mainly in the short term. Markou, Kyriakides and Polycarpou [27] proposed applying MLP to STLF identifying those factors that affect load demand the most. These drivers depend greatly upon the distinct characteristics of the power utility including its size and geographic location. The results revealed differences between mild and extreme weather conditions with MAPEs varying between 2.16% and 5.62% respectively.

Autoregressive (AR) model

Random series describing a time-varying process. The output variable depends linearly on its own previous values and on a stochastic term (white noise); thus the model is in the form of a stochastic difference equation:

$$y(t) = \Phi_1 * y(t-1) + \Phi_2 * y(t-2) + \cdots + \Phi_p * y(t-p) + \varepsilon(t) \rightarrow \Phi(L)y(t) = \varepsilon(t)$$

Where:

$$\Phi(L) = 1 - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$

And the lag operator (L) is defined as:

$$Ly(t) = y(t-1)$$

Moving-average (MA) model

Together with the AR model, the MA model is a key component of the more general ARMA and ARIMA models. The model equation is as follows:

$$y(t) = \varepsilon(t) - \varepsilon(t-1) * \theta_1 - \varepsilon(t-2) * \theta_2 - \cdots - \varepsilon(t-q) * \theta_q \rightarrow y(t) = \theta(L)\varepsilon(t)$$

Where:

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^q$$

ARMA model

An ARMA model combines AR and MA models to describe a stochastic process. The general expression:

$$y(t) - \Phi_1 * y(t-1) - \cdots - \Phi_n * y(t-n) = \varepsilon(t) - \varepsilon(t-1) * \theta_1 - \cdots - \varepsilon(t-q) * \theta_q$$
$$\rightarrow \Phi(L)y(t) = \theta(L)\varepsilon(t)$$

ARIMA models

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These time-series models are applied in some cases where data show evidence of non-stationarity.

As an example, the equation of an ARIMA(3,2,1) model:

$$(1 - \Phi_1 * L - \Phi_2 * L^2 - \Phi_3 * L^3)(1 - L)^2 y(t) = (1 - \theta_1 L)\varepsilon(t)$$

The ARIMA model was developed by Box-Jenkins [16], and is one of the most used methods mainly in STLF. A.E. Clements, A.S. Hurn and Z. Li [27] developed a forecasting method using ARIMA to predict day-ahead electricity load using multiple equation time series.

T steps to follow in order to build an ARIMA model are:

1. Represent the time series and look for possible outliers.

2. Stabilize the variance carrying out a logarithmic or Box-Cox transformation.

3. Analysis of the stationarity of the transformed series. If the data maintains a constant level, and both autocorrelation functions of simple and composed cancel quickly, it can be considered stationary. If the series is not stationary it must be differentiated.

4. Analysis of the seasonality of the transformed series. If the time series is non-seasonal, we apply regulate seasonal differentiation. If the series is seasonal, we differentiate first seasonally and then regularly.

5. Identification of the seasonal pattern analysing the coefficients of the autocorrelation functions both simple and composed.

6. Identify the regular component of the model analysing the coefficients of the autocorrelation functions simple and composed.

7. Check the level of significance of the coefficients.

8. Analysis of the residuals. Outlier detection, Ljung-Box test.

9. Compare different models using AIC or SBC.

Exponential Smoothing

As pointed by Weron [14], exponential smoothing is a pragmatic approach to forecasting, whereby the prediction is constructed from an exponentially weighed average of past observations.

This time-series model assigns exponentially decreasing weights to the past observations over time. It does not rely on variables other than lagged loads, meaning less data requirements than other widely used techniques such as MLR and ANN. Since the electricity demand is highly driven by weather, changes in weather patterns can greatly affect the load profiles. When the weather condition is quite volatile, the techniques that do not leverage meteorological forecasts are often in a disadvantage situation.

The general expression for the simplest exponential smoothing model would be:

$$\hat{y}(t+1) = \hat{y}(t) + \alpha * (y(t) - \hat{y}(t))$$

Exponential smoothing has been used mainly in STLF [19], where forecasts derived in this way are compared with more conventional approaches.

Semi-parametric additive models

Semi-parametric additive analysis is an explanatory model in the regression framework but designed to account for non-linear relationships.

An example expression [1] could be as follows:

$$\log\big(demand(t)\big) = WD(t) + Temp\left(Temp_{Madrid}(t), Temp_{Barcelona}(t)\right) + demand(t-1) + \varepsilon(t)$$

Goude, Nedellec, & Kong (2014) [23] applied generalized additive models to model electricity demand over more than several substations of the French distribution network, at both short (STLF) and middle (MTLF) term horizons. These generalized additive models estimated the relationship between load and the explanatory variables including temperatures, calendar variables, and so forth.

METHODOLOGIES

Similar Day Methodology

Used to find a day in history similar to forecasted day. Similarity is often based on weather patterns, season of the year, day of the week… As mentioned by Weron [14] and Hong [24], this methodology was one of the earliest of load forecasting, and still continues to be one of the most popular ones. Currently it can be also applied following clustering techniques.

Variable selection

This methodology is useful to determine which explanatory variables to use. Hong 2010 [25] proposed a variable selection methodology and applied it to compare different techniques as linear regression, neural networks and fuzzy regression.

Hierarchical forecast

Although the literature is limited, there are some papers related to hierarchical forecasting such as Hong 2008 [26], where he implemented a hierarchical trending method for spatial load forecasting at medium size utilities, which involved fitting S curves for their aggregated levels through a constrained multi-objective optimization formulation.

TIME HORIZONS IN ELECTRICITY LOAD FORECASTING

Models cannot be strictly divided into time horizons by themselves, as they might be used for different time scopes. This is the case of MLR (applied both for STLF and LTLF) or semi-parametric additive models (used for STLF and MTLF).

Transforming a model forecasting a certain time horizon into a different time frame can be done by adding the variables needed in the new time scope and focusing on the desired time period. As an example, if we are to transform STLF to MTLF, we will add econometrics variables to the STLF model and extrapolate the model to a longer time horizon.

Once this clarification has been done, it might be useful to sort the different time scopes:

Very-Short term

Deals with time in minutes and requires past loads.

VSTLF information is used in real-time market for short term unit commitment and economic dispatch of generating units.

Among the methods explained earlier neural networks are widely use within this time scope.

Short term

Horizons ranges from a few minutes up to few days ahead and data required include past loads and weather information.

STLF looks for the optimization of tactical operations, which might influence the decisions within the daily markets.

MLR, ANNs, ARIMA and exponential smoothing models are usually classified under the STLF label according to Weron [14].

Medium term

Horizons considered vary within authors from: few-days-to-months, to a month-to-year ahead; and data required include past loads, weather and economic information.

The main objective of MTLF is to stablish an optimal schedule for generating plants as well as to improve the efficiency of fuel supply in power stations.

The vast majority of linear models are usually classified under the MTLF label, although others (semi-parametric additive models, SVMs) have also been used under this time scope.

José Ramón Cancelo, Antoni Espasa and Rosmarie Grafe [11] developed a model to forecast the electricity load from one day to one week ahead for the Spanish TSO. This method, ranging between STLF and MTLF, was of particular importance for the model done within this Master's thesis paper. The basic strategy was to formulate an additive logarithmic model including trend, seasonality, weather and disturbance effects.

Long term

Time lead measured in months or years (from one to ten), and data required include past loads, weather, economic and further information.

LTLF is of vital importance when planning future expansion of both the transmission system, and the desirable energy mix and its location. It must also be related to the energy policy of the nation and be consistent with the regulation and planning of electricity tariffs.

Probabilistic/density models are usually classified under the LTLF label [9, 10].

WEATHER VARIABLES

Electricity load is highly affected by meteorological conditions, especially in those areas with high air-conditioning system penetration.

Although temperature stands out as the most important meteorological factor, the use of data on humidity, wind speed, cloudiness, rainfall and solar radiation to come up with an explanatory climate variable has also been done [12].

Temperature

Temperature has a non-liner relationship with electricity load as both increases and decreases of temperature passing the corresponding threshold result in an increase of electricity consumption.

The response is caused by the difference between the ambient/outdoor temperature and the comfort/indoor temperature.

Literature dealing with the non-linearity of the temperature effect tends to segment temperature variations in terms of heating degree days functions and cooling degree days functions.

The cooling and heating degree day function is defined as follows:

$$CDD = \sum_{j=1}^{nd} max(0; t_j - T^{HOT*})$$

$$HDD = \sum_{j=1}^{nd} max(0; T^{COLD*} - t_j)$$

Where nd the number of days is considered, T^{HOT*} and T^{COLD*} are the thresholds temperatures of cold and heat, and t_j is the observed temperature a day j.

Further insights on temperature and demand’s response to temperature will e analized in the next Chapter.

Chapter 3 EXPLANATORY ANALYSIS

The objective of this chapter is to present the different data that has been gathered, and at the same time assess the correct way to process it so it can be used for further developing the forecasting model.

ELECTRICITY DEMAND

Electric energy consumption is the actual energy demand made on existing electricity supply.

For the purpose of this project we have to take a sufficiently long time period, in order to tackle all the possible future scenarios that may arise. We have decided to choose a horizon covering from 2005 to 2015, in which we will get not only the effect of growing demand of the buoyant years of the economic bubble, but also the consumption recession caused by the financial crisis.

The following picture shows the monthly and yearly data of the 2005-2015 period. It is clear that the season and month of the year affect directly to the electricity consumption. Moreover, we can slightly intuit the effect of the economic crisis, with an increasing demand from 2005 until 2008. Starting in this year, the demand drops enormously. Then, in 2009 it seems like a recovery has taken place. But not for too long, since the demand will fall down again, and this time even deeper, reaching levels as low as the ones of 2005.

Although the seasonality is pretty clear in Figure 1, perhaps the overall trend can be better perceives in Figure 2.

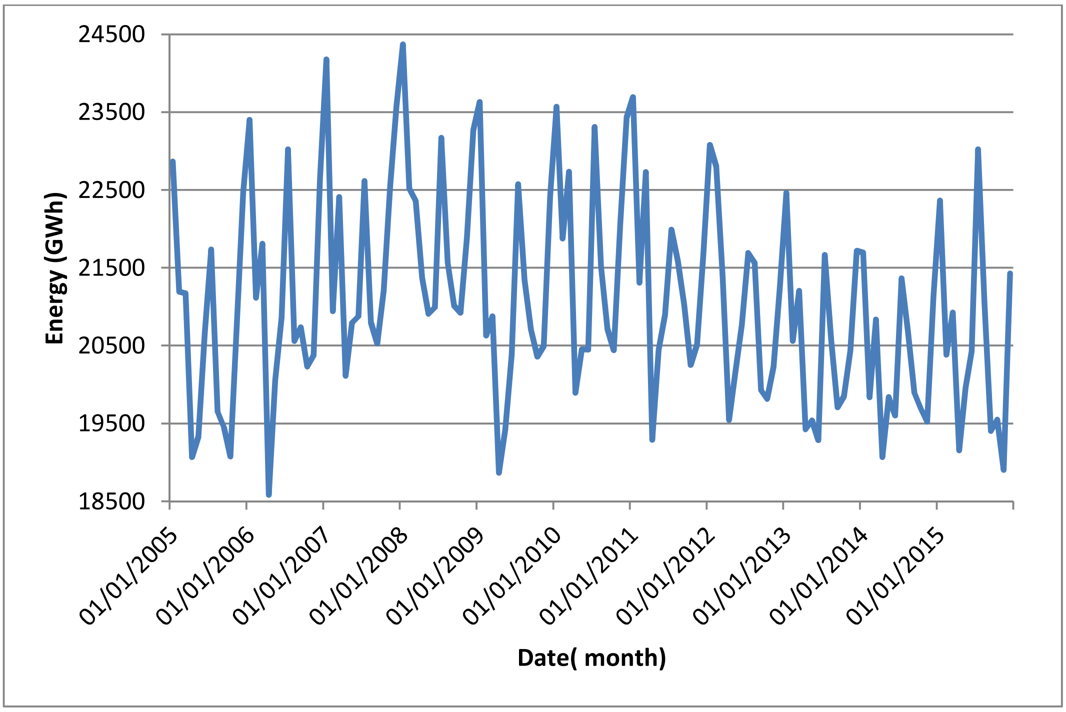

Figure 1. Monthly evolution of the Spanish electricity demand (GWh)

The following figure shows the electricity demand time series on a yearly basis, so the impact of the economic crises is better captured:

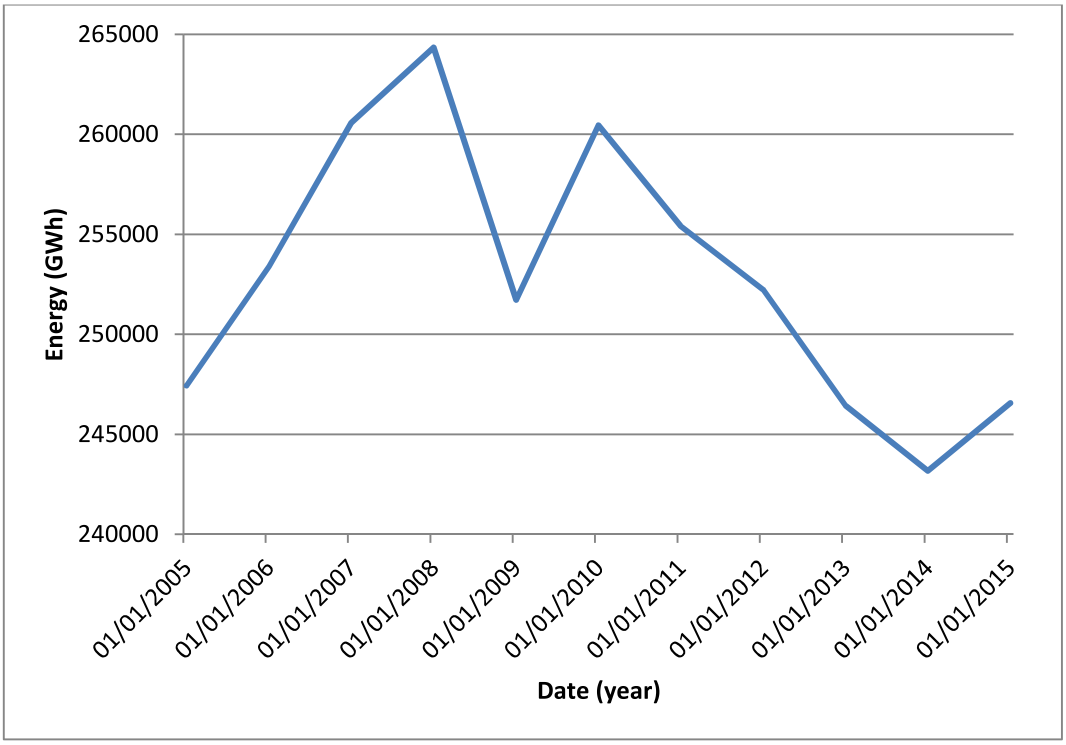

Figure 2.Yearly evolution of the Spanish electricity demand (GWh)

Finally, we can also represent the data on a daily basis. This way what we obtain is a much fuzzier picture, which can be slightly confusing at the very first glance.

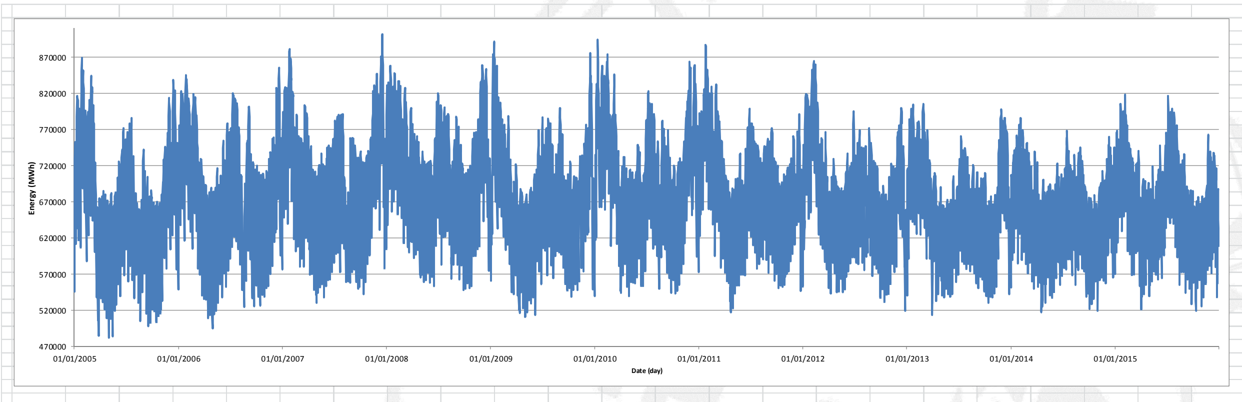

Figure 3. Daily evolution of the Spanish electricity demand (MWh)

Although the seasonality and overall downwarding or upwarding trends can be somehow appreciated, they are much clearer in the previously shown figures.

TREND, CALENDAR AND TEMPERATURE EFFECT MODEL

The objective of this section is to develop a linear regression model of the Spanish electricity demand. To do so, we will first analyse the main variables: calendar effect, trend and temperature; and then evaluate the model and its results.

VARIABLE ANALYSIS

Calendar effect

As observed before, the variable analysed is the daily Spanish demand (MWh), and the period studies is that between 2005 and 2015. Electricity demand shows a persistent trend and seasonality; also, given the daily frequency of the data, the predominance of the working day effect is notorious.

The calendar effect is collected by the variable “working day effect”, which represents the effect of the calendar in a particular daily demand as a percentage of electricity demand on a representative day. This effect is independent of the rest of the effect such as temperature, season, trend and economic activity. [12]

In this project, we have chosen Wednesdays as the reference day of the week, and therefore we have assigned them a calendar effect equal to one. The method used to obtain the calendar effect factor consists of the following steps:

1. We get the weekly calendar variation index (WCVI), with the formula:

$$WCVI = D_{d,w}/W_w$$

Where $D_{d,w}$ is the electricity demand of day d of week w, and W_w is the electricity demand of the Wednesday of the same week. Since the effects of trend, seasonality, temperature and economic activity conditions are smooth and both $D_{d,w}$ *and* W_w are near, it is clear that all effects cancel out except from the calendar effect. In case the Wednesday happens to be holiday we will pick the data from the previos non-holiday Wednesday.

2. Determination of homogeneous types of days. To do so, we will carry a regression of the WCVI variable on a combination of dual variables reflecting the day of the week and the type of holyday. For further details it would be useful to note Table 2, where it can be seen the different categories analysed: non-holidays, Easter, Christmas, national holidays…

As a result of this process we obtain the so called "working day effect" variable, which captures the different demand fluctuation caused by different levels of activity on holidays and on working days in domestic, industrial and commercial sectors. The values are reflected in Table 2, and represent the coefficient of equivalence of each type of day. If we multiply a random day by this factor, we will automatically “transform” this day into a typical day. With this variable we reach the objective we were looking for, we eliminate from the demand series the effect of the calendar or working day effect.

The results obtained in Table 2 are reasonable and congruent with those obtained by Julián Moral Carcedo and José Vicens Otero [12], whose methodology was followed despite the differences in input data.

The study reflects that typically, days of the week that tend to be less energy demanding are Sundays, Saturdays and Mondays. Days with high demand are Tuesdays, Wednesdays and Thursdays, with similar electricity demand levels and above Fridays.

Moreover, holidays are associated with deep declines in electricity consumption. This fact can be seen in Table 2.

Finally, it is important to note that the effect of each type of holiday might vary according with the type of Labour Day they are replacing. Thereby, holidays falling on Mondays and Fridays will have the effect of a lengthen weekend, while holidays falling on Wednesday won't have this impact.

WORKING DAY EFFECT								
		MON	TUE	WED	THU	FRI	SAT	SUN
Non-holidays		0.98	1.00	1.00	1.00	0.99	0.88	0.85
Easter	Week before Easter day	1.00	1.02	1.00	0.87	0.80	0.83	0.79
	Easter Monday and Tuesday	0.86	0.98					
Christmas	21-dec, 22-dec, 3-jan,4-jan	1.07	1.03	1.00	1.00	1.00	0.94	0.90
	23-dec, 27-dec, 28-dec, 29-dec	1.12	0.83	0.83	0.80	0.86	0.97	0.87
	30-dec	1.24	1.08	1.00	0.99	0.96	0.89	0.96
	25-dec	0.79	0.84	1.00	0.85	0.72	0.74	0.71
	1-jan	0.80	0.78		0.86	0.83	0.75	0.76
	6-jan		0.78	1.00	0.83	0.82	0.92	0.87
	26-dec, 2-jan,7-jan	0.90	0.93	1.00	1.30	1.01	0.87	0.82
	24-dec	0.95	1.18	1.00	0.84	0.84	0.81	0.76
	31-dec	0.89	1.17	1.00	0.92	0.90	0.84	0.82
	5-jan	0.86	1.16	1.00	0.98	1.04	0.95	
National holidays	01-may	0.74	0.82	1.00	0.79	0.79	0.79	0.75
	15-aug	0.82	0.90	1.00	0.89	0.82	0.87	0.73
	12-oct	0.79	0.83	1.00	0.81	0.80	0.86	0.79
	01-nov	0.79	0.81	1.00	0.83	0.85	0.86	0.78
	6-dec	1.04	0.90	1.00	0.88	0.86	0.80	0.80
Day after national holidays	16-aug	0.91	0.98	1.00	1.13	0.96	0.83	0.84
	13-oct	0.91	0.97	1.00	1.19	0.86	0.80	0.82
	02-nov	0.92	0.97	1.00	1.19	0.90	0.86	0.81
	7-dec	0.87	1.09	1.00	1.07	0.92	0.86	0.75
Monday before national holiday		0.83						
Regional and local holidays	25-jul, 24-jun	0.90	0.94	1.00	0.98	0.90	0.84	0.79
	11-sep	0.98	0.96	1.00	0.96	0.94	0.89	0.77
	28-feb, 23-apr	0.94	0.98	1.00	0.98	0.97	0.86	0.77
	9-oct, 15-may	0.93	0.98	1.00	0.97	0.96	0.87	0.76
	09-nov	0.94	0.98	1.00	0.97	0.98	0.88	0.78
	24-sep	0.99	0.98	1.00	0.99	0.95	0.86	0.73
	19-mar, 2-may	0.86	0.96	1.00	1.01	0.87	0.82	0.78
	8-dec	0.80	0.83	1.00	0.88	0.97	0.83	0.83
Day after regional and local holidays	26-jul, 25-jun	0.95	1.00	1.00	1.03	0.99	0.85	0.76
	12-sep	0.95	1.02	1.00	1.04	0.92	0.84	0.81
	29-feb, 1-mar	0.91	0.99	1.00	1.00	0.98	0.87	0.82
	24-apr, 10-oct	1.10	0.75	1.00	1.01	0.96	0.85	0.76
	16-may, 10-nov	0.94	0.99	1.00	1.02	0.98	0.88	0.80
	25-sep	0.94	1.02	1.00	1.00	1.00	0.83	0.76
	20-mar, 3-may	0.93	0.98	1.00	1.02	1.05	0.84	0.76
	9-dec	0.94	0.98	1.00	1.16	0.94	1.01	0.82

Table 2. Working day effect

Trend correction

The basic steps that were implemented are based on the regression models explained.

We assume the consumption can be modelled as follows:

$$demand(t) = a_0 + a_1 * t + a_2 * t^2 + a_3 * t^3 + a_4 * WD + \varepsilon(t)$$

Where t represents the time sequence, $a_0 \dots a_4$ are the coefficients of the regression model, WD is the working day variable obtained previously and ε_t is the noise which also accounts for the temperature effect.

Using Matlab we can build the matrix expression corresponding to the aforementioned regression formula for the whole period of time studied.

.

If we neglect the effect of the noise (ε_t) and knowing the consumption data ($demand(t)$), the given time sequence value (t) and the computed calendar effect (WD), the coefficients ($a_0, a_1, a_2, a_{3,}, a_4$) can be obtained. Having obtained the coefficients we can compute the estimated consumption.

We developed a program in Matlab to implement the reasoning mentioned beforehand. The actual code can be found at the end of this document.

Using IDAT we can distinguish between the trend and the calendar effect, where the original data is plotted in blue (demanddaily), as well as the calendar effect in green (demanddaily_WD) and the tendency effect in red (demanddaily_trend).

In fact we can even add the two effects so we get what is called estimated demand that can be compared with the original data.

The following figure shows two graphs illustrating the demand, trend, calendar effect and estimated demand that accounts for the two aforementioned effects.

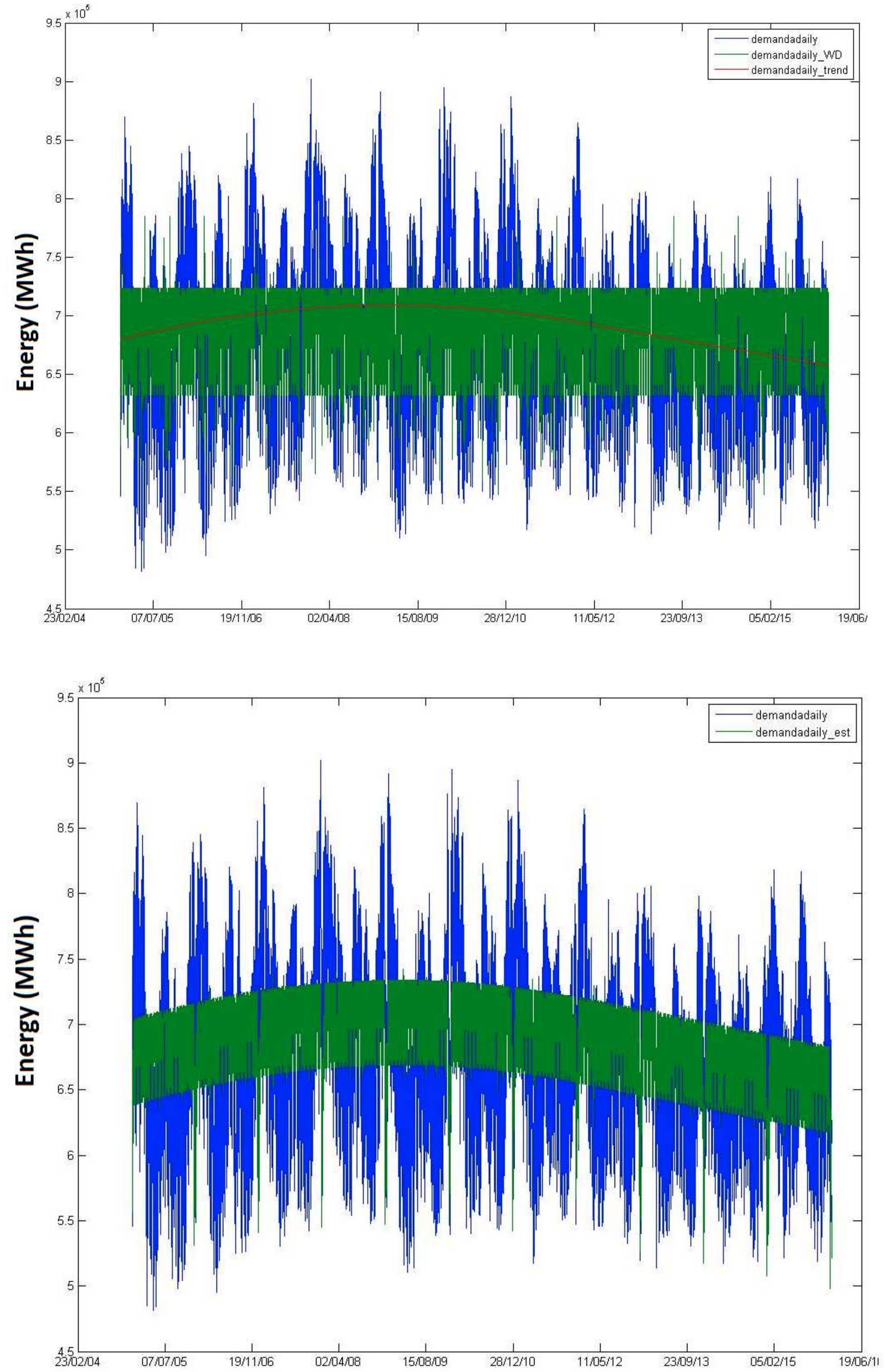

Figure 4. Electricity consumption, calendar effect, trend and estimated demand

The following picture shows a detail for year 2006 so as to compare the demand data (green) with the estimated demand (blue). We can see clearly how the model has captured both trend and calendar effect, but still there are key drivers missing. Especially noticeable is the effect of temperature that will be analysed in the next section.

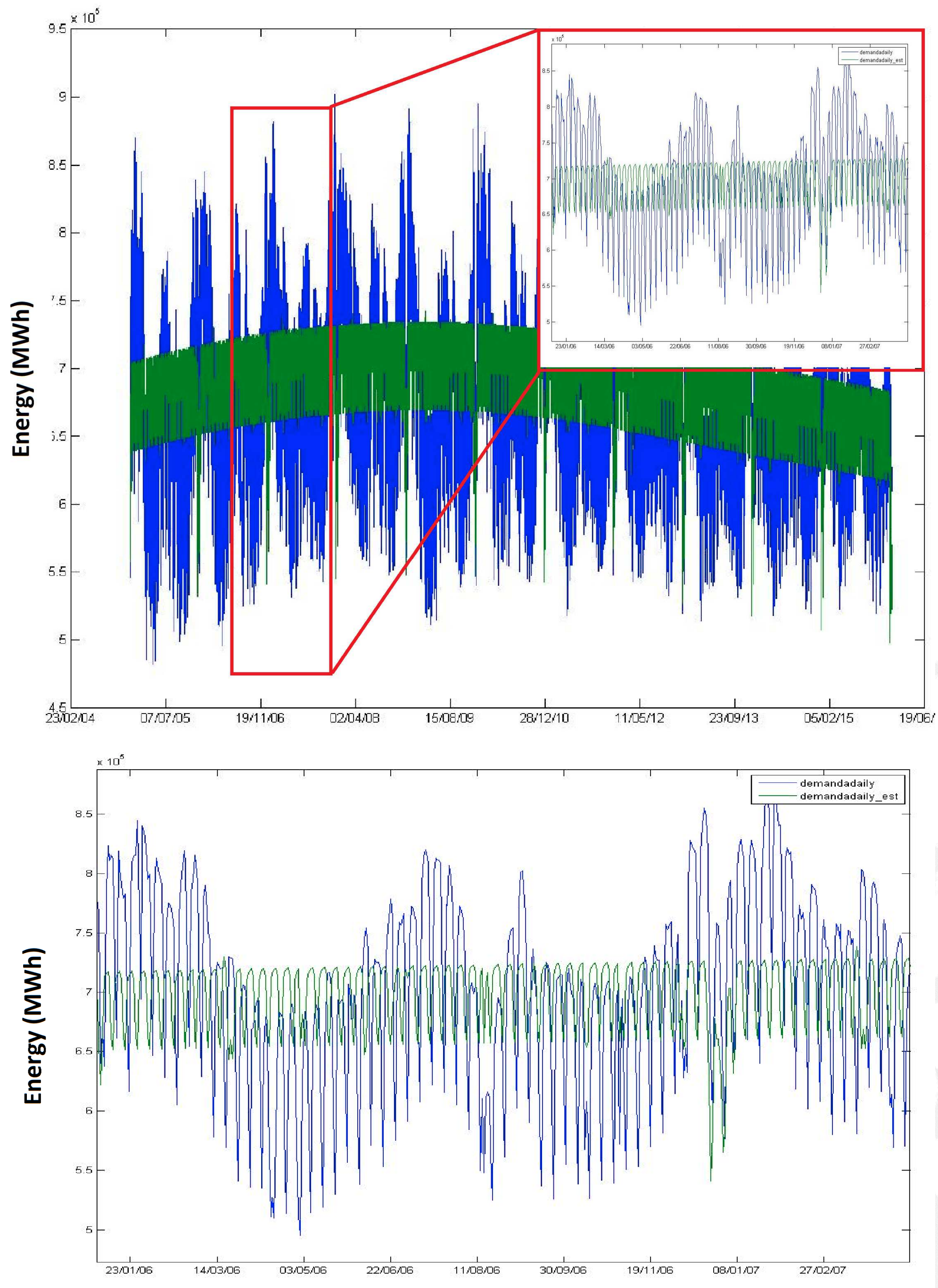

Figure 5. Detail of the daily and estimated demand 2006

Temperature

The electricity consumption of a certain country is deeply dependant on the climatological conditions of the former, which explain to a large degree the inter-annual variation that the energy consumption experiments within a year. Among the meteorological factors that affect the demand, the temperature plays an important role, given that it is one of the most significate variables when explaining the volatility of the demand of electricity. However, there have been a number of publications that consider also other interesting variables such as humidity, wind velocity, luminosity, precipitations... Notwithstanding, considering only the temperature has been the most commonly used practise, as it captures huge information of the other variables as well.

As a consequence that the equipment (not only the one in households, but also in the industrial and commercial sectors) shows an increasing pattern, not only on its availability, penetration and energy requirements, but in the technology incorporate thereto, it can be stablished that the demand response to the weather changes has to evolve naturally with the years to come, in such a way that considering such a wide time period as the one framed in this study, the curve response might be affected substantially.

To prove our hypothesis it is only necessary to compare the demand response to temperature within years sufficiently separated in time. This individual treatment of each one of the years, permits to observe the changed experimented by the demand curve, independently of any other factor, specially hose derived from the economic activity, which affect directly to the level of demand, and not to the shape of the curve, and how it respond to the temperature variations, which can be associated solely to the changes produced on the equipment level sensible to temperature variations.

The following curves show the aforementioned considerations, where we have plotted the corrected demand (demand data minus estimated demand) and its relationship with temperature. It is important to note that the corrected demand is being used instead of the raw demand data as the calendar and trend effect have to be removed for a better analysis.

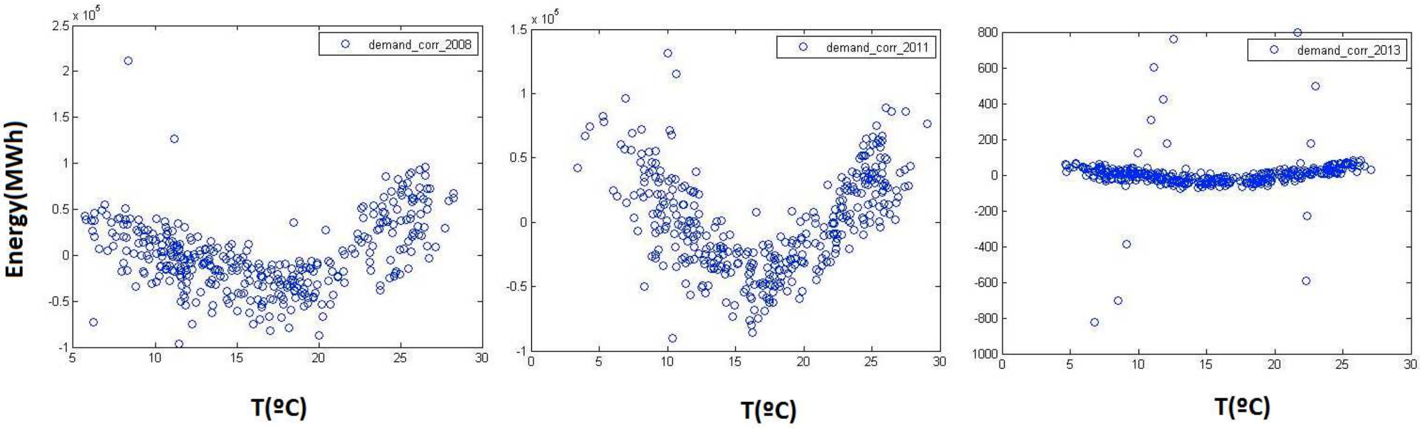

Figure 6. Demand response to temperature 2008, 2011 and 2013

The evolution of demand's response to temperature within time is now clear, being 2013's behaviour particularly steady compared to 2008 and especially to 2011. Certain otlier can be observed within the years, being particularly noticeable in 2013.

Within the study of Spanish electricity demand response to temperature, it is unavoidable to have a temperature variable that reflects the aggregate variations of the temperature fluctuations that have been registered within the peninsular territory.

Unfortunately, there is not a so-called "national temperature", and the climatological characteristics of each region of the Spanish territory are different and unique in themselves.

The process adopted to capture these variations is as follows:

1. Determination of homogeneous climatological zones. Depending on the mean temperature values registered within a determined year, we will have to identify those observatories that have similar climatological conditions, and therefore, delimit climatological zones. We have to assume that the values obtained at the observatory are representative of determined geographical area.

2. Selection of the meteorological observatories representing each climatological zone. Once the climatological zones have been determined, it is time to select an observatory, which represents the whole area. For the purpose of this project, we will identify seven climatological zones and one representative observatory within each zone.

3. Obtaining the variable: "national temperature". Once the observatories have been selected, we obtain a mean of the temperatures in an aggregated form, as a weighted mean of the temperatures resisted in each one of the representative observatories.

Based on the results obtained by the Ward technique, we come up with seven different groups which include the whole set of observatories taken into account for the period 2005-2015. Moreover, it has been taken into consideration both the geographical proximity as well as the continuity as classification criteria.

The aforementioned climatological classification can be seen in Table 3 and the following picture:

GROUP 1	ALBACETE, CIUDAD REAL, LÉRIDA, TOLEDO, GRANADA, ZARAGOZA Y MADRID.
GROUP 2	ALICANTE, CASTELLÓN, VALENCIA, MURCIA YALMERÍA
GROUP 3	ÁVILA, BURGOS, LEÓN, PALENCIA, SORIA, ÁLAVA, CUENCA, GUADALAJARA, HUESCA, LOGROÑO,NAVARRA, SEGOVIA, SALAMANCA, TERUEL, VALLADOLID Y ZAMORA.
GROUP 4	BADAJOZ, JAÉN, CÁCERES, SEVILLA Y CÓRDOBA
GROUP 5	LUGO, CORUÑA, ORENSE, PONTEVEDRA, SANTANDER, GUIPÚZCOA, OVIEDO Y VIZCAYA
GROUP 6	GERONA, TARRAGONA Y BARCELONA
GROUP 7	CÁDIZ, HUELVA Y MÁLAGA

Table 3. Spanish climatological classification

According to this definition of climatological areas and the selection of observatories previously detailed, we proceeded to the computation of the weighted average national temperature, as a mean of each temperature observed in each observatory.

In order to weight the temperature in a meaningful manner, we can consider different parameters: population, energy consumption…

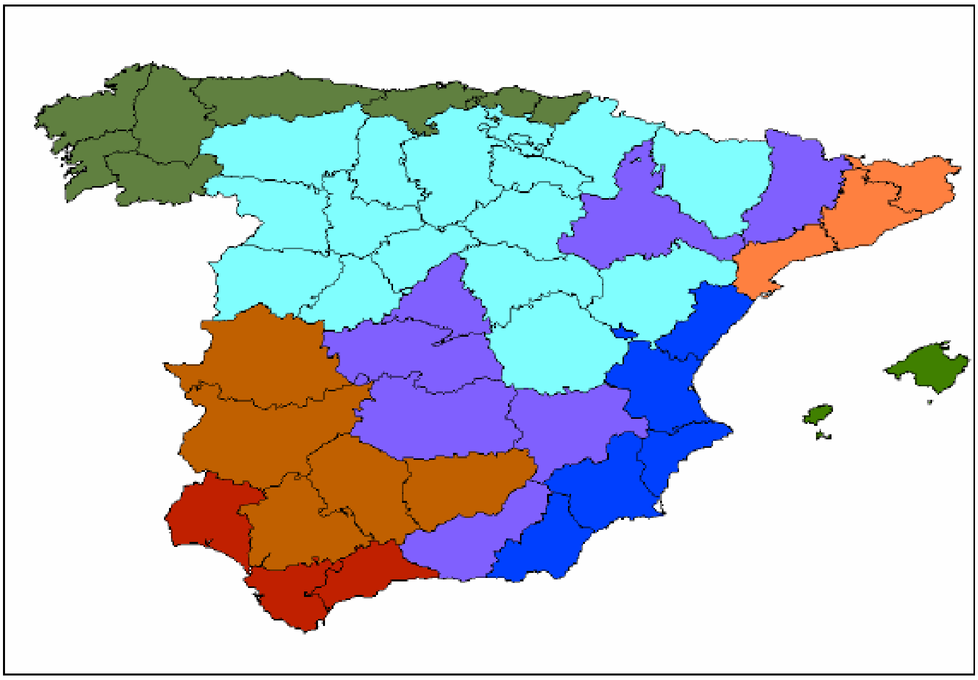

Figure 7. Spanish climatological areas

The following table reflects the number of Spanish clients by geographical zone and the energy consumption within regions:

	ENERGY [MWH]	ENERGY [pu]	CLIENTS	CLIENTS[pu]
Madrid	27.347.461	0,30	3.247.631	0,30
Valencia	11.317.532	0,12	1.578.763	0,15
Valladolid	2.553.510	0,03	333.074	0,03
Sevilla	8.098.976	0,09	962.179	0,09
Bilbao	7.437.045	0,08	656.903	0,06
Barcelona	28.795.154	0,31	2.868.495	0,27
Malaga	6.137.297	0,07	1.031.587	0,10
	91.686.975		10.678.632	

Table 4. Spanish energy and number of clients

Both criteria seems reasonable, however we have chosen the number of client standard as the most appropriate one. Therefore, the weighing of each group will be as follows: 0.32

for Madrid, 0.15 for Valencia, 0.09 for Sevilla, 0.06 for Bilbao, 0.28 for Barcelona and 0.1 for Málaga.

The formula is the following:

$$TEMP^{dd/mm/aa} = \sum_{i=1}^{7} w_i T_i^{dd/mm/aa}$$

Where we compute the weighted national temperature of each day of the month and year considered. The temperatures of each particular zone are weighted according to the number of clients' criteria.

To conclude this section devoted to the clarification of the different climatic zones, it might be useful to look at the following pictures, which show the average temperature registered in the different observatories for the 2005-2015 period.

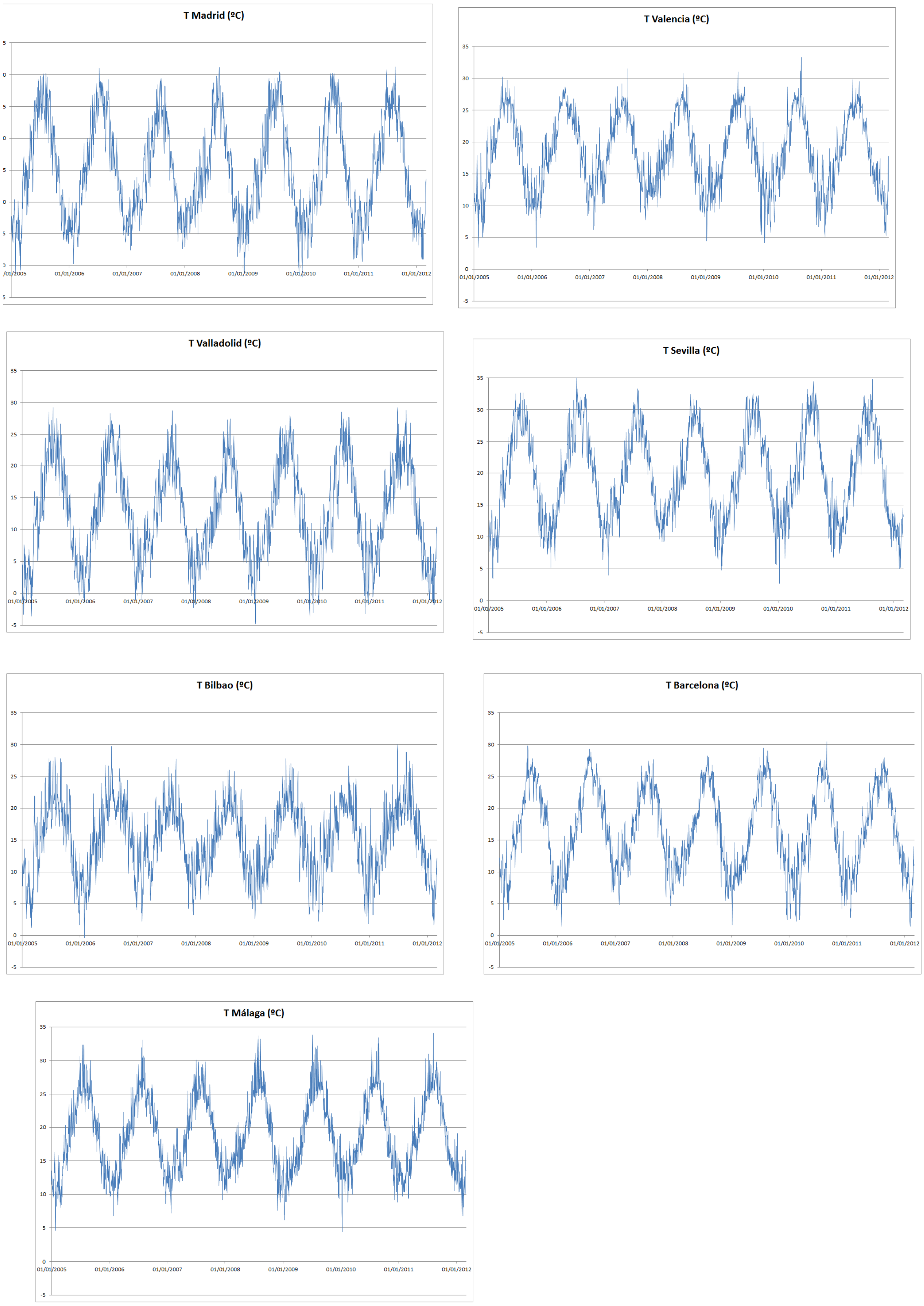

Figure 8. Average temperatures of each zone

Demand's response to temperature

As observed before, the variable analysed is the daily Spanish demand (MWh),

Demand's response to temperature changes presents a clear non-linear pattern. This non-linearity is due to the fact that both the temperature drops and rises have a positive effect in electricity consumption. Demand increases in both situations as a result of the increasing use of heating appliances during cold periods and air conditioner when temperatures are high.

The response is shown in the following picture:

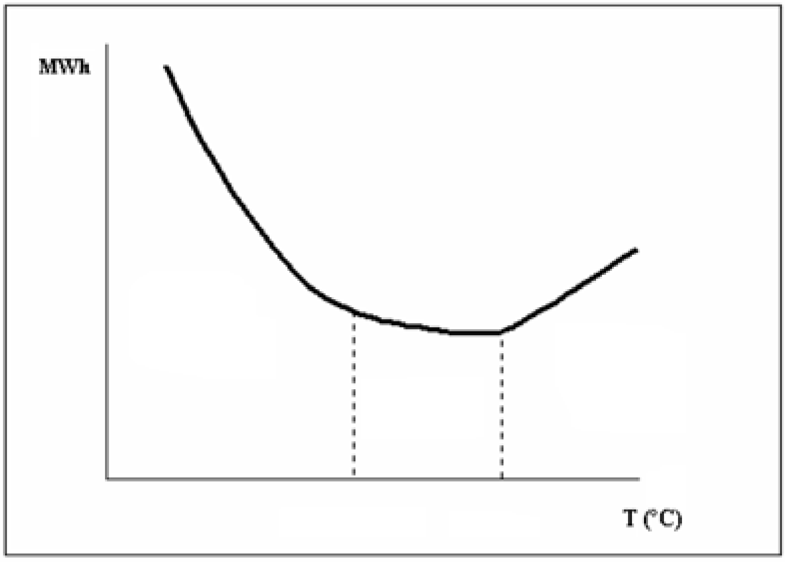

Figure 9. Demand's response to temperature [13]

Despite the merely physic reasoning, the response to the temperature variations might also be affected by other variables, that can affect to the convexity of the curve and its overall shape. When the climate is mild and has typically no abrupt variations, the response should be less pronounced. Moreover, technical specifications of heating and cooling systems affect the degree of response to temperature variations, as they affect the household rent and energy prices, as well as the range of the comfort-zone temperatures.

To sum up, we can conclude that the shape of the curve might slightly vary, and therefore we can simplify it in a meaningful manner as follows:

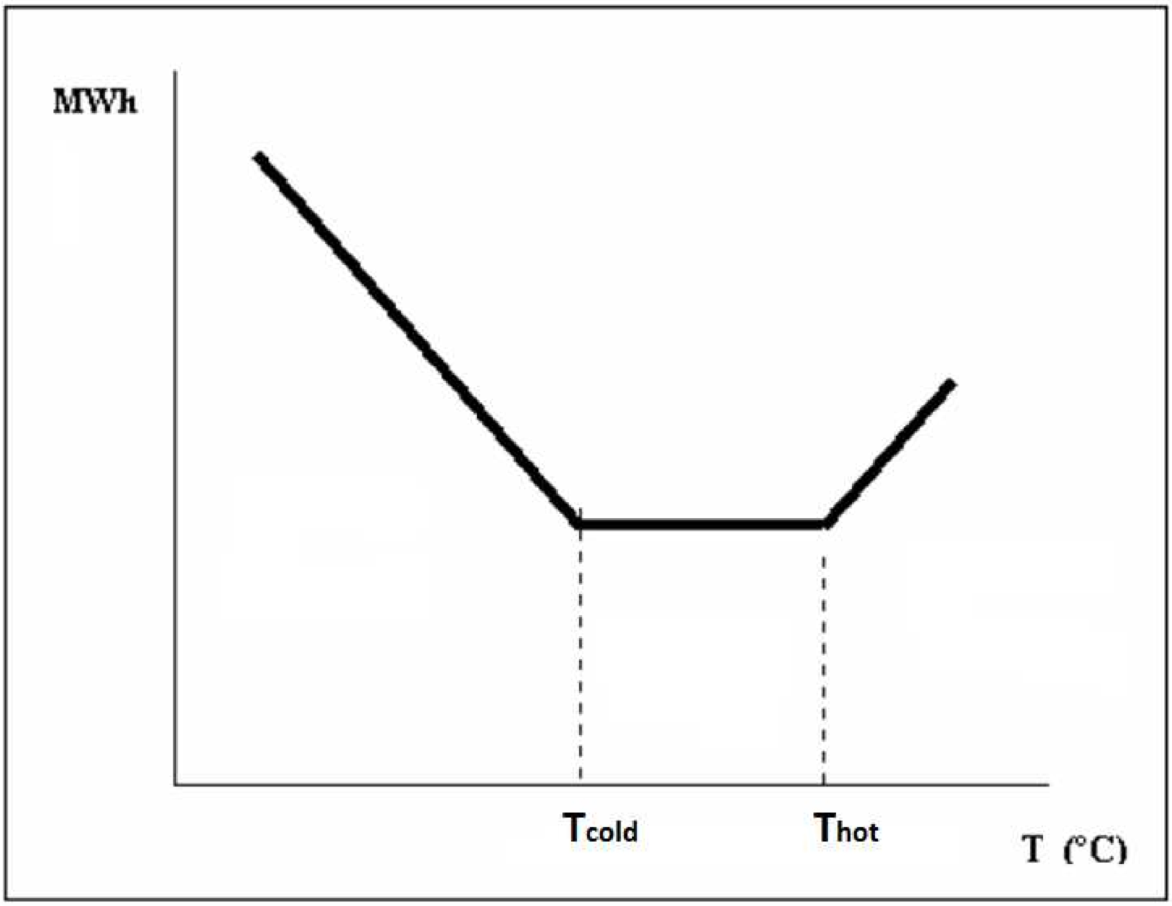

Figure 10. Demand's response to temperature simplification [13]

The previous assumption means that the difference between the observed temperature and the threshold is the most relevant aspect of the curve. As it assumes that the demand response is similar in a situation of 1ºC difference or other in which the difference is 10ºC, that is to say, independent of the level of temperature.

The previous reasoning leads us to the definition of the so-called "degree-days" as the difference between the average daily temperature and the cold or hot thresholds:

$$CDD = \sum_{j=1}^{nd} max(0; t_j - T^{HOT*})$$

$$HDD = \sum_{j=1}^{nd} max(0; T^{COLD*} - t_j)$$

Where *nd* is the number of days of the month or any number of days desired in order to represent the inertia (this means that the first cold/hot day the consumption might not react as if there have been a couple of cold/hot days and therefore there is time enough for the customers to turn in the heating/cooling air conditioner system), T^* is the threshold temperature either cold or hot, and t_j is the observed temperature of a certain day j.

These functions collect the number of days in which the temperature is below or above the temperature thresholds and the magnitude measured in degrees of the deflection.

We will have to define the threshold before going any further in our study of HDD and CDD functions.

Temperature thresholds have to represent a value in which the variable analysed (in this case the electricity demand), change considerably. In this case, the threshold would be the temperature for which the relationship between electricity consumptions and temperature will change its sign for the sensitivity parameter.

Threshold determination could be done just by looking at the different graphs representing demand versus temperature. By doing so it is clear that demand has evolved within its response to temperature, and it seems reasonable to establish 15ºC as the lower temperature and 18ºC as the higher, although a possible future improvement might include evolving temperatures that show demand evolution within time.

Needless to say, that there have been plenty of research on this field, and typically these threshold determination problems are addressed using what is known as the Linear Hinges Model (LHM) [17].

It might also be useful to evaluate the heating and cooling degree days to the power of two and three and check whether or not they explain the model.

We used the IDAT tool in order to compute the linear regression model. We start by defining the output variable that we want to estimate, which is the daily demand. The input variables for the model would be the working day, the average temperature, daily HDD and CDD, daily HDD and CDD to the power of two and three, the HDD and CDD each two and five days and the HDD and CDD each two and five years to the power of two and three respectively.

Moreover, the variables "Frio" (Cold) and "Calor" (Hot) were introduced to represent temperatures colder and hotter than 5 and 25ºC respectively. The following figure shows those variables during 2010:

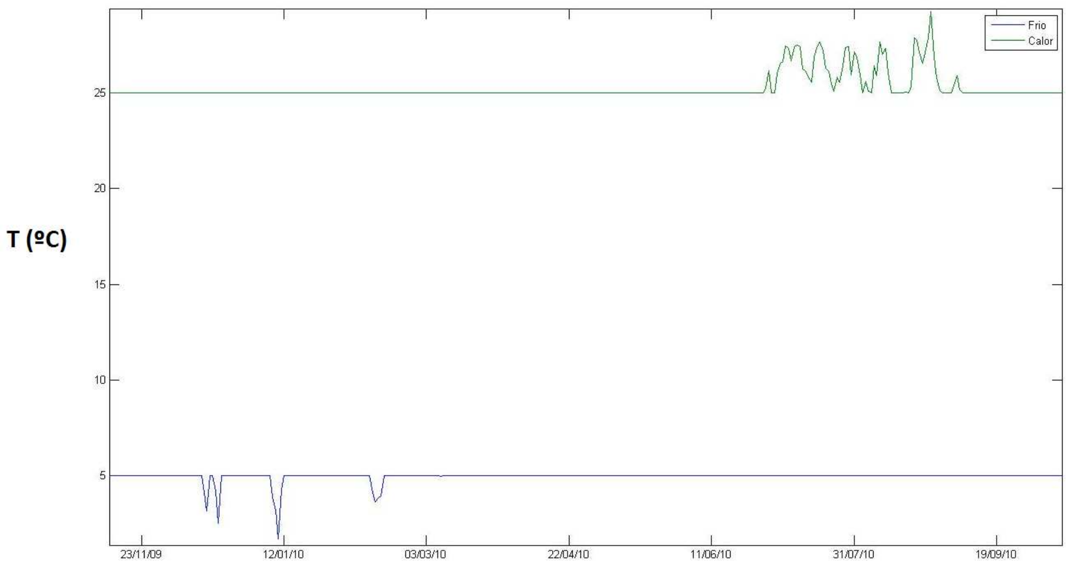

Figure 11. Hot and cold variables

MODEL DEVELOPMENT

As an example we will now analyse the series between 2005 and 2015, although it will be analysed later including variables representing economic activity and efficiency that will explain the trend.

The following picture shows how much of the series is explained, in this case 69.82%.

Estimación mediante mínimos cuadrados. Variable: 'demandadaily'

	Valor
R-squared	0.69828
Rbar-squared	0.69738
Sigma^2	1772412599.0922
Durbin-Watson	1.473
N°observaciones	4017
N° variables no podadas	13
N° variables podadas	15
Error absoluto medio	29548.948
Error absoluto porcentual medio (%)	4.4306

Table 5. Linear regression model evaluation 2005-2015

Besides the R squared, we will compute the mean absolute percentage error (MAPE), which is a typical statistical measure that expresses the prediction accuracy of a prediction method.

The following picture shows which variables are representative for the model and which variables are not:

Coeficientes de regresión y estadísticos asociados

	Coeficiente	Estadistico-t	Probabilidad-t
WD	320985.1567	26.0437	0
CDD5days	2236.0824	6.0977	1.178e-009
2CDD	890.5677	4.4381	9.3186e-006
2CDD5days	-64.1275	-3.1833	0.0014674
3CDD	-56.1484	-3.1539	0.0016232
3CDD2days	3.833	6.4005	1.7276e-010
3CDD5days	0.94937	3.2699	0.0010849
HDD2days	2208.3929	14.1914	0
t	44778.4133	5.3673	8.4448e-008
t2	-24440.7093	-4.1073	4.0829e-005
t3	3619.9882	3.0395	0.0023851
Frio	NaN	NaN	0.99903
2HDD5days	NaN	NaN	0.79543
3HDD5days	NaN	NaN	0.93839
CDD2days	NaN	NaN	0.79709
2HDD	NaN	NaN	0.74617
HDD5days	NaN	NaN	0.46664
3HDD2days	NaN	NaN	0.32149
HDD	NaN	NaN	0.42747
CDD	NaN	NaN	0.42723
2HDD2days	NaN	NaN	0.30781
TempMed	NaN	NaN	0.44401
2CDD2days	NaN	NaN	0.049739
constante	NaN	NaN	0.03518
Calor	NaN	NaN	0.14336
3HDD	NaN	NaN	0.059625

Table 6. Input variables evaluation

We will only include in our model the variables with probabilities less than 0,05.

Following that rule the equation based on the above results would be as follows:

$$c_{2005-2015} = 320985.15 * WD + 2236.08 * CDD_{5days} + 890.56 * CDD^2 - 64.12 * CDD_{5days}^2 - 56.14 * CDD^3 + 3.83 * CDD_{2days}^3 + 0.94 * CDD_{5days}^3 + 2208.39 * HDD_{2days} + 44778.41 * t - 24440.70 * t^2 + 3619.98 * t^3$$

RESULTS AND DIAGNOSIS

We can see now how good our model is, just by plotting both the original series (demandadaily) and the estimated one (demandadailynueva).

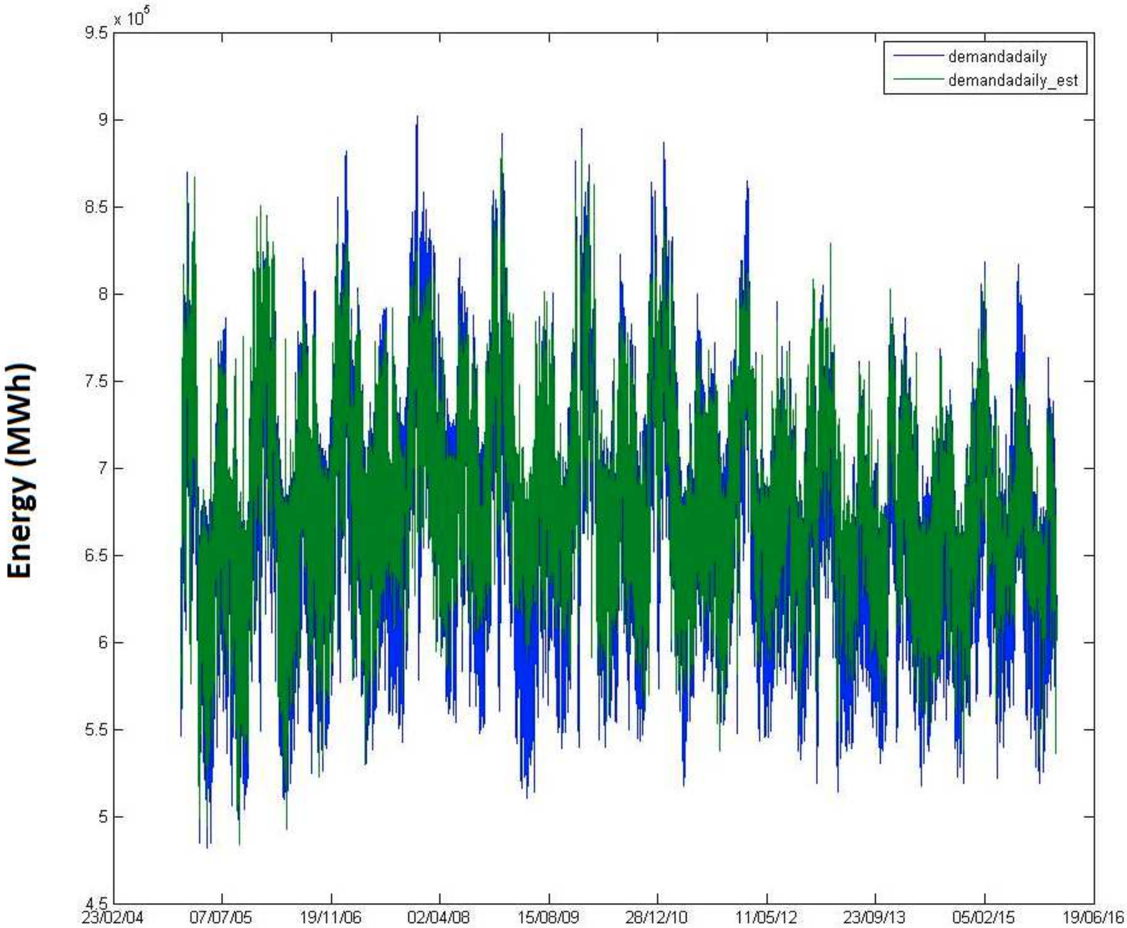

Figure 12. Original and estimated daily demand 2005-2015

If we make a zoom we can see that our model tends to fail on the peaking point, both in the low and the high range:

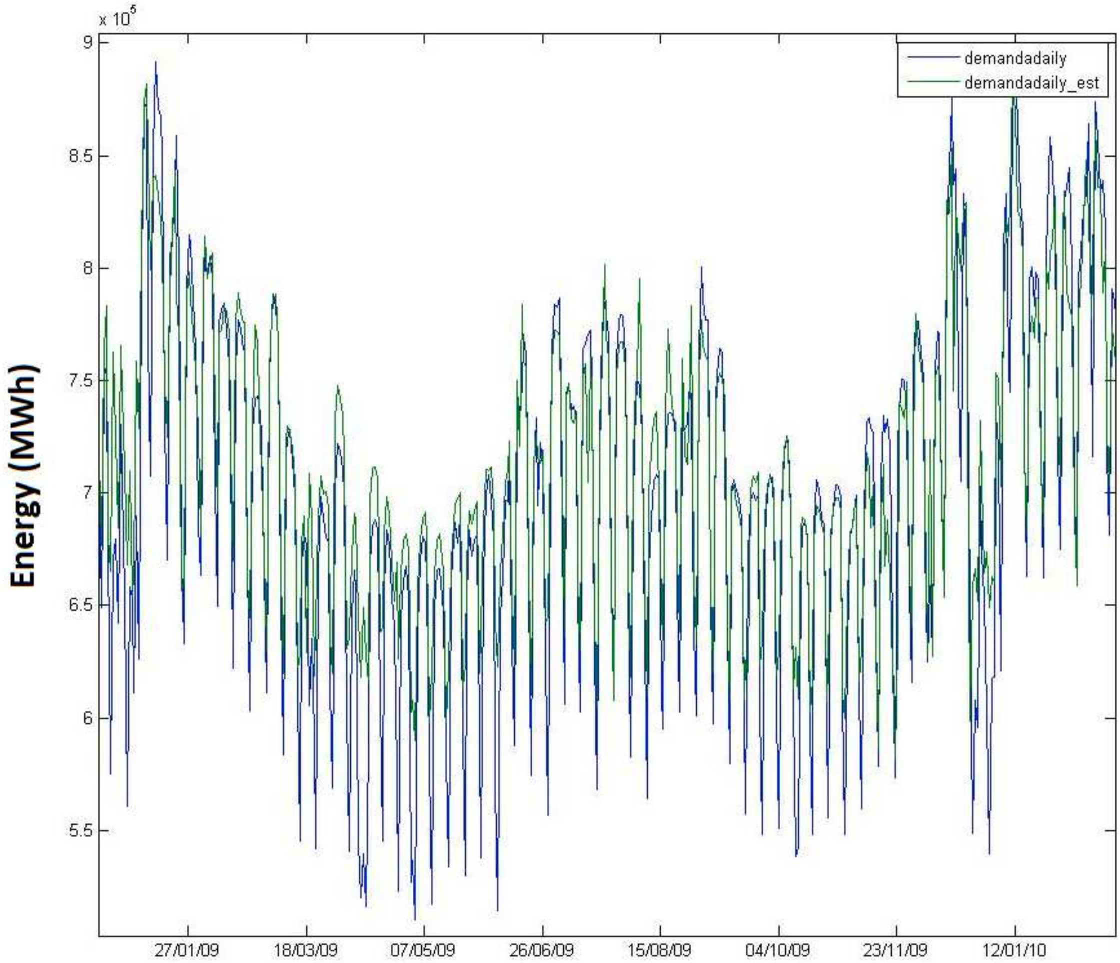

Figure 13. Detail of daily and estimated demand

To explain this misbehaviour it might be useful to carry out the study year by year, as the demand's response to temperature is somehow changing.

The results for 2005 are shown in the following table. We can see that now the model is explained almost 75% and the most important variables are the working day, CDD each five days, HDD to the power of two and CDD each two and five days to the power of two.

Estimación mediante mínimos cuadrados. Variable: 'demandadaily'

	Valor
R-squared	0.73419
Rbar-squared	0.73124
Sigma^2	1897146189.1101
Durbin-Watson	1.3043
Nºobservaciones	365
Nº variables no podadas	5
Nº variables podadas	21
Error absoluto medio	30699.492
Error absoluto porcentual medio (%)	4.6688

Coeficientes de regresión y estadísticos asociados

	Coeficiente	Estadistico-t	Probabilidad-t
WD	667736.6589	165.2816	1
CDD5days	3421.1286	5.9838	5.2684e-009
2HDD	745.4764	7.4152	8.7774e-013
2CDD2days	239.1464	4.2121	3.2012e-005
2CDD5days	-45.4933	-2.9229	0.0036871
CDD2days	NaN	NaN	0.99753
3CDD	NaN	NaN	0.99579
Calor	NaN	NaN	0.96505
HDD2days	NaN	NaN	0.84184
t	NaN	NaN	0.81049
TempMed	NaN	NaN	0.77607
HDD	NaN	NaN	0.74616
t3	NaN	NaN	0.73664
CDD	NaN	NaN	0.56326
2CDD	NaN	NaN	0.51744
constante	NaN	NaN	0.35478
Frio	NaN	NaN	0.31618
3HDD	NaN	NaN	0.26923
3HDD2days	NaN	NaN	0.24016
2HDD2days	NaN	NaN	0.25549
3CDD2days	NaN	NaN	0.049474
3CDD5days	NaN	NaN	0.11767
3HDD5days	NaN	NaN	0.029671
2HDD5days	NaN	NaN	0.44894
HDD5days	NaN	NaN	0.41343
t2	NaN	NaN	0.010926

Table 7. Linear regression model evaluation 2005

It is also clear that the estimated variable resembles much more the original daily demand.

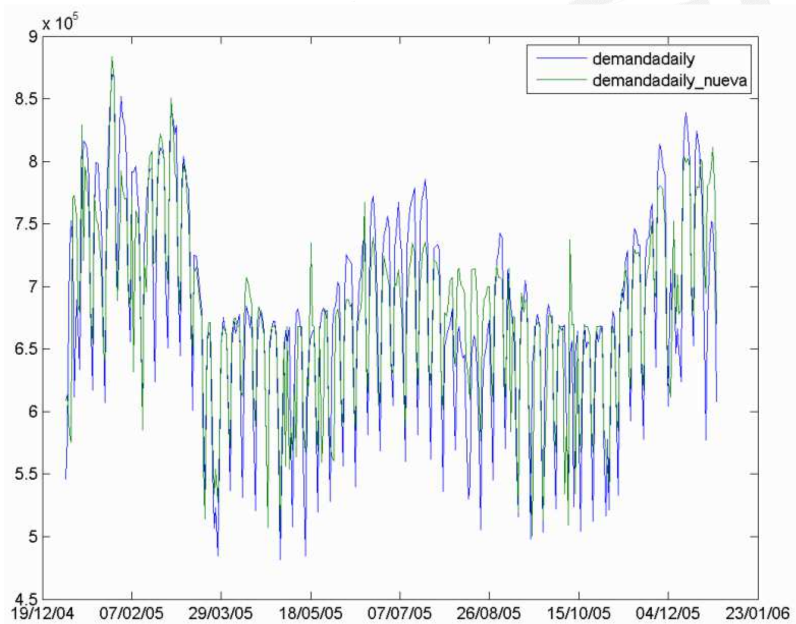

Figure 14. Detail of daily and estimated demand 2005

We repeat the process year by year in order to see the different demand response to temperature along the time.

The following table sums up the results during the period 2005-2015 taking into account the calendar, trend and temperature effects:

year	R squared	MAPE
2005	0.73419	3.524
2006	0.72541	3.238
2007	0.71076	3.640
2008	0.63144	3.494
2009	0.75159	3.877
2010	0.76458	2.988
2011	0.64215	2.981
2012	0.87940	4.646
2013	0.64255	4.296
2014	0.71577	7.210
2015	0.70148	8.845

Table 8. Yearly results 2005-2015

It is now clear that certain years are better explained than others. This is the case of 2012 with an “R-squared” of 87.9% and 2010 with 76.4% On the other hand; years that are not properly estimated include 2008 and 2011 with an “R-squared” of 63.1% and 64.2% respectively. Taking into account the MAPE we can see how the error ranges from a reasonable value up to an alarming 7.21 or even 8.84 the last two years.

Finally the next picture shows the MAPE evolution as computed using IDAT:

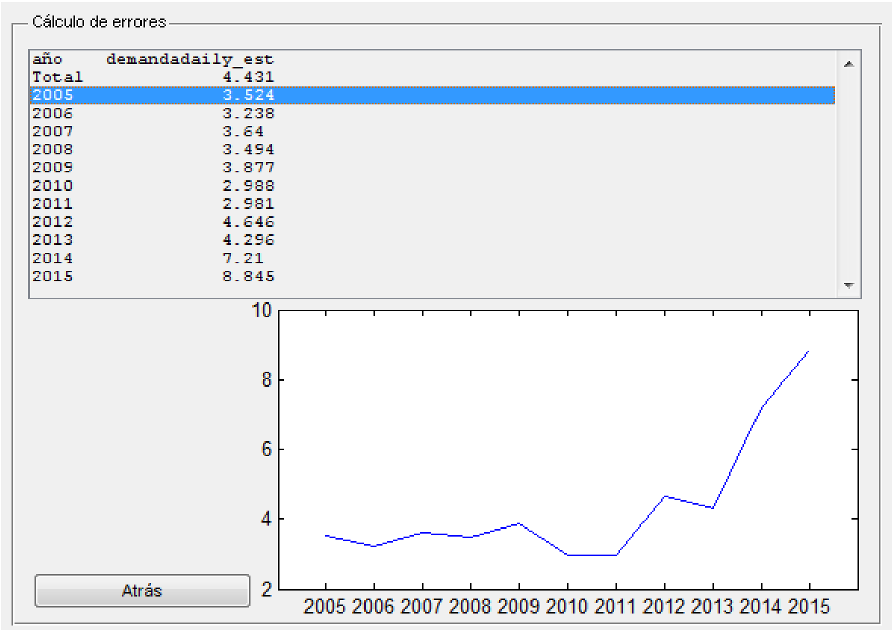

Figure 15. MAPE evolution 2005-2015

ECONOMIC ACTIVITY AND EFFICIENCY EFFECT MODEL

The objective of this section is to develop a linear regression model of the Spanish electricity demand. However, as opposed to the first model, this one will focus its attention on trying to account for the tendency effect based on economic and efficiency explanatory variables. After this new variables are analysed, we will develop our model and evaluate the results.

VARIABLE ANALYSIS

Economic activity

The special nature of electricity demand and the different ways in which it is affected by economic activity are widely known and can be detected linking the economic development of a country and its energy consumption.

Electricity demand, as any other energy product, can be incorporated to the economic system within a double framework. On the one hand, as a product encompassed by the final household's demand, and on the other, as an input for the process of industrial and commercial companies.

GDP is the most relevant economic variable, as it reflects the value of the final production generated by and economy during a specific period of time. It synthetises the level of activity of activity of all sectors. Moreover, given the correlation between GDP and rent, it allows approximating the variations of the disposable income, and it seems a reasonable indicator for the households' spending capability. However, we should note that there is not such a direct correlation between GDP and electricity demand. As an example, electricity demand registered higher growth as GDP during 1980-2000 period.

Moreover, not all sectors have identical energetic requirements, and therefore, electricity consumption does not present a direct correlation with the weight each sector has on the GDP. Therefore it is necessary to develop an explanatory indicator that reflects properly the electricity correspondence. With this purpose, we developed a synthetic indicator reflecting the economic impact of each sector on the electricity demand.

In order to obtain this indicator, the procedure was as follows:

1. Selection of a set of indicators for each one of the sectors chosen
2. Analysis of each indicator and selection of the key driver of each sector.

3. Selection of the weight of each partial indicator for the global index
4. Obtainment and validation of the synthetic indicator for the economic activity.

Before beginning with the aforementioned procedure we should determine the sectors to be taken into account. Given the data provided by the Spanish Ministry of Industry, Energy and Tourism (MINETUR) and the information by sectors provided by the National Institute of Statistics (INE), we have divided the electricity consumption in ten sectors, as shown in the following table.

1. Extractives	Carbon extraction
	Mining
2. Energy by-products	Oil and gas production
	Water
	Electricity generation
	Gas production
	Oil and nuclear waste treatment
3. Agriculture, cattle raising, fishing	Agriculture, cattle raising and fishing
4. Minerals	Steel and casting
	Non-ferrous metallurgy
5. Construction	Other construction materials
	Construction and public works
	Non-metallic minerals industry
6. Chemistry	Chemistry and petro-chemistry
7. Industry: metals	Machinery and metal-processing
	Ship-building and repair
	Automobile
	Other conveyance construction
8. Industry: manufacturing	Food, beverage and snuff
	Clothing, leather and footwear
	Wood and cork industry
	Pulps and paper
	Graphic arts and publishing
	Plastic and rubber industry
9. Services	Retail services
	Rail transport
	Other transport companies
	Hostelry

	Commerce and services
	Services not intended for sale
	Public administration
	Public lighting
10. Domestic use	Domestic use
--- Non specified	Non specified

Table 9. Classification by sectors

For the initial selection of indicators we have to build up a sufficiently wide range of partial parameters, selected based on the following characteristics: economic meaning, source reliability, rapid attainment, represent a given sector of economic activity.

The economic activity analysis carried out was based on the period 2005-2011, mainly because statistical homogeneity criteria with the working day effect analysis.

The following table shows the partial indicators chosen for each one of the sectors based on the aforementioned criteria.

Sector	**Description**	**Unit**
1. Extractives	Workers in the mining sector	Number of employees
2. Energy by-products	Energy IPI	Index
3. Agriculture, cattle raising, fishing	Workers in the agriculture/cattle raising/fishing sector	Number of workers
4. Minerals	Metallurgy IPI	Index
5. Construction	IPI construction materials	Index
6. Chemistry	IPI chemistry industry	Index
7. Industry: metals	IPI machinery construction	Index
8. Industry: manufacturing	Retail trade index	Index
9. Services	Retail trade index	Index
10. Domestic use	Real wage income indicator	Index

Table 10. Partial indicator selected for each activity

Additionally, instead of considering all sectors, an alternative and simpler method was proposed, in which only four sectors were considered. The aim of this approach was to simplify the process as much as possible. The next table shows the indicator for the most relevant sectors.

Sector	**Description**	**Unit**
1. Agriculture	Workers in the agriculture/cattle raising/fishing sector	Percentage of workers in relation with total workforce
2. Industry	General IPI	Index
3. Services	Retail trade index	Index
4. Domestic use	Real wage income indicator	Index

Table 11. Indicators for four sectors

Once the key drivers have been selected, the next step is to develop an indicator of the economic activity that explains the electricity demand.

The following expression shows the synthetic index formula:

$$IAEj = \sum_{i=1}^{10} \omega_{ij} I_i$$

$$\sum_{i=1}^{10} w_i = 1$$

Where IAE is the indicator of economic activity w_{ij} is the weighing coefficient for each one of the partial indicators; and I_{ij} is the partial indicator corresponding to each sector.

It is clear that the number of possible synthetic indicators is non-limited, depending on both the weighs and the number of sectors being considered (and therefore partial indicators to be considered). This study will be limited to the simple approach of four indicators instead of ten as explained above.

Weights corresponding to each sector can be set based on different options. It seem reasonable to look after the electricity consumption of each sector during the period analysed. The following table shows the participation of each economic sector related to the final electricity consumption based on the information provided by MINETUR.

Sectors	Mean(%) 2005-2011
1. Extractives	1.5
2. Energy by-products	1.9
3. Agriculture, cattle raising, fishing	2.7
4. Minerals	13.2
5. Construction	5.4
6. Chemistry	6.5
7. Industry: metals	5.1
8. Industry: manufacturing	13.2
9. Services	25.6
10. Domestic use	24.9

Table 12. Electricity consumption by sectors 2005-2011

Another alternative is to use the value of the final purchases that each sector made to the energy generators. We can see major differences in certain sectors such as the domestic one that represents a major weigh according to INE's data, while others such as the industrial sector represent a lower share.

Sectors	Mean(%) 2005-2011
1. Extractives	3.1
2. Energy by-products	1.9
3. Agriculture, cattle raising, fishing	2.1
4. Minerals	4.6
5. Construction	4.5
6. Chemistry	3.8
7. Industry: metals	7.9

8. Industry: manufacturing	11.2
9. Services	29.8
10. Domestic use	31.1

Table 13. Electricity production destiny (input-output)

Once the weighs were selected according to the aforementioned input-output criteria, the synthetic indicator was successfully added to the model as another input parameter able to explain the demand.

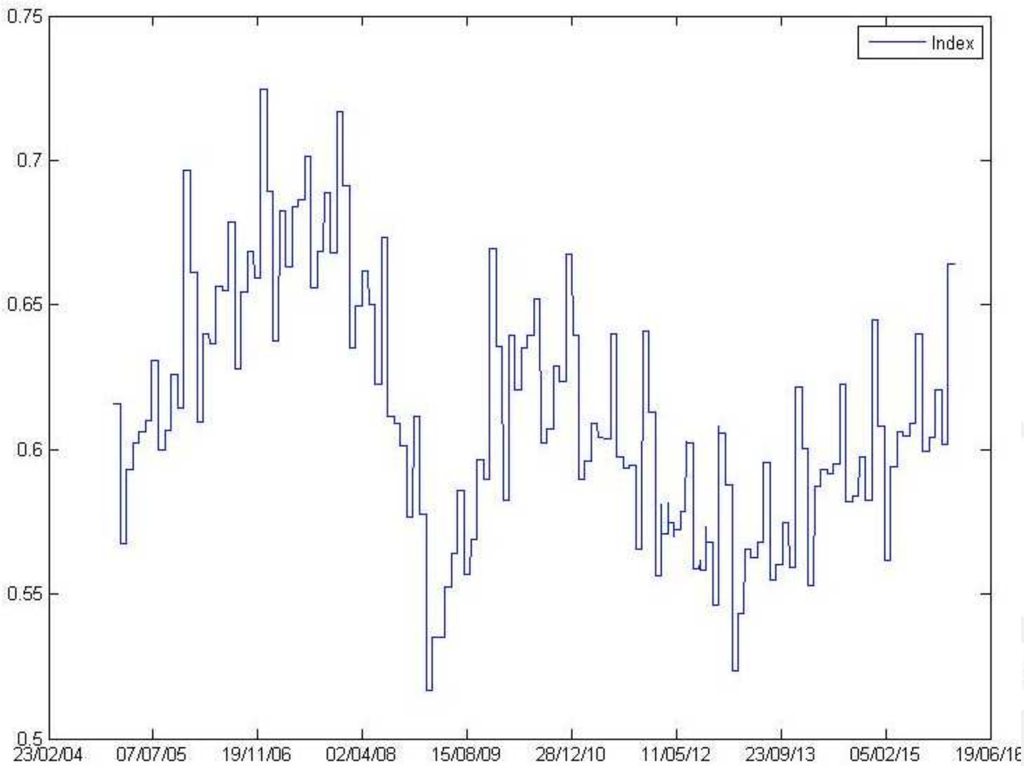

Figure 16. Economic activity index

The objective is to substitute the trend modelled previously with the time sequence variables (t, t^2, t^3), by the economic activity index (as well as the energy efficiency drivers analysed next). However, making this change could lead to potential errors as the IAE index is not intrinsically upwarding as it was the time sequence.

Energy efficiency and stock

As it was said previously on this paper, if we are to eliminate from the electricity demand the effects associated with the calendar patterns and temperature, the only remaining effect will be the economic one. Besides, it is known that GDP synthesizes the level of activity of all production sectors, and due to the relationship between GDP and income, we can approximate part of the economic drivers of electricity demand. However, we must note that approximating the demand just with GDP will bias the forecast, and therefore it is necessary to add variables that capture the evolution that demand might suffer inter-annually.

The upwarding tendency of electricity demand can have multiple causes, but it is mainly driven by the evolution of long term variables such as the stock that needs continuous electricity supply. It is clear that almost all households have a refrigerator and a relevant proportion have also freezer. The vast majority of commercial premises own cooling systems, which are also needed in restaurant and catering sector.

Due to the lack of information regarding the evolution of cooling systems or similar equipment, the housing stock (i.e. the total number of houses within the Spanish market) was chosen as an approximation. This will capture the long term variables. Moreover, the demographic upwarding trend of the population and the increasing number of clients provided with electricity will also be captured by the housing stock variable.

Moreover energy efficiency patterns have been considered, as they are one of the key drivers and objectives set at the European level. To do so, we took annual data from IDAE regarding the primary energy intensity on a global basis and for the industry and domestic sectors in particular. It should be clarified that the primary energy intensity is ratio between the gross inland consumption of energy (or total energy consumption) and Gross Domestic Product (GDP).

Drivers	Unit
1. Housing stock	Number of houses
2. Primary Energy Intensity (General)	Kep/€
3. Primary Energy Intensity (Industry)	Kep/€
4. Primary Energy Intensity (Domestic)	Kep/€

Table 14. Energy efficiency drivers

Needless to say, all those drivers are typically recorded on an annual basis. This is why we had to interpolate the annual data to obtain monthly data. This was done using Excel.

The following picture shows the procedure:

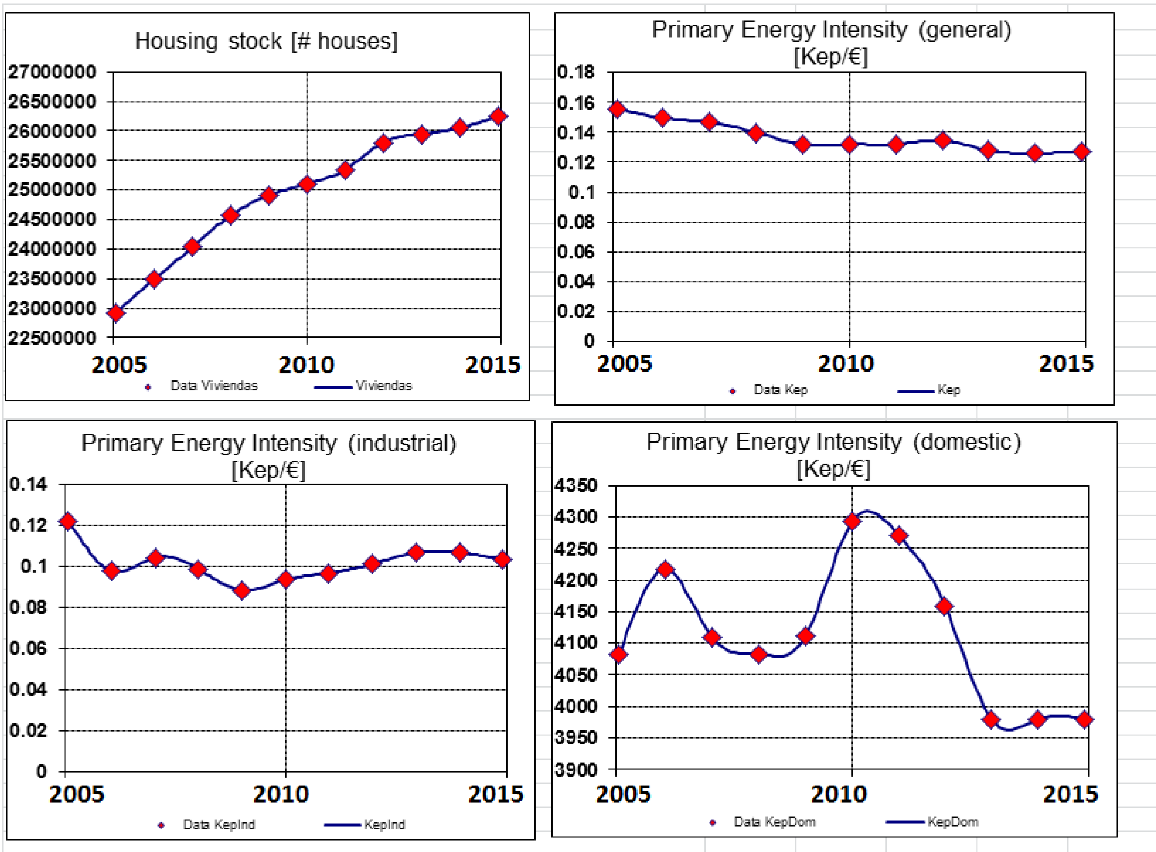

Figure 17. Annual to monthly transformation

If we are to substitute the trend modelled previously with the time sequence variables (t, t^2, t^3), we would have to consider that only the housing stock variable ("viviendas") has an upwarding effect.

In fact, all primary energy indexes present a downwarding pattern, which is particularly noticeable in the domestic index. This issue could lead to certain forecasting errors that will be evaluated in the next section.

EXPLANATORY VARIABLES RECAP

It will be clear now that the variables used in both models include temperature and calendar effect, while the trend is treated as a time sequence in the first model and explained through economic and efficiency variables in this second model.

The steps to develop the second model are similar as the ones done in the first one, but this time the trend will be explained using the economic activity index and the energy efficiency drivers.

The following table sums up the explanatory variables that have been used corresponding to each model.

Model	Factor	Variable	
1	Time sequence (trend)	t	
		t^2	
		t^3	
1&2	Calendar Effect	WD	
1&2	Temperature	CDD	HDD
		CDD_{2days}	HDD_{2days}
		CDD_{5days}	HDD_{5days}
		CDD_{2days}^2	HDD_{2days}^2
		CDD_{5days}^2	HDD_{5days}^2
		CDD_{2days}^3	HDD_{2days}^3
		CDD_{5days}^3	HDD_{5days}^3
		Calor	Frio
2	Economic activity (trend)	$IAE = f(IPI, retail\ trade\ index, real\ wage\ income\ indicator, workers)$	
2	Energy efficiency & stock (trend)	Housing stock	
		Primary Energy Intensity (General)	
		Primary Energy Intensity (Industry)	
		Primary Energy Intensity (Domestic)	

Table 15. Explanatory variables

MODEL DEVELOPMENT

The model is built as shown previously in the temperature section, but now we include all parameters and carry out the study for the whole period 2005-2015. Instead of explaining the trend with the time sequence variables (t, t^2, t^3), we will now include the economic and efficiency drivers to account for this upwarding effect of the time in electricity's demand.

The following picture shows how much of the series is explained. In this case 66.67%, which is less than the result we obtained using the trend time sequence variables (t, t^2, t^3), that was 69.82%.

Estimación mediante mínimos cuadrados. Variable: 'demandadaily'

	Valor
R-squared	0.66679
Rbar-squared	0.66604
Sigma^2	1956378973.083
Durbin-Watson	1.2853
Nºobservaciones	4016
Nº variables no podadas	10
Nº variables podadas	18
Error absoluto medio	32402.9969
Error absoluto porcentual medio (%)	4.8609

Table 16. Linear regression model evaluation 2005-2015 with economic and efficiency drivers

It is clear that the economic and efficiency drivers explain worse the time series than the time sequence variables.

We can also compute the equation based on the coefficients of the explanatory variables with probabilities lower than 0.05.

Coeficientes de regresión y estadísticos asociados

	Coeficiente	Estadístico-t	Probabilidad-t
WD	9845104.6308	66.4588	0
TempMed	-2375.9465	-4.4927	7.2313e-006
CDD5days	3169.1847	12.1493	0
HDD2days	5199.9653	14.5557	0
2CDD5days	-18.5449	-3.8073	0.00014262
3CDD2days	3.7136	6.2812	3.7187e-010
Index	281762.5795	14.5297	0
Kep	853060.9259	8.9063	0
KepInd	-905401.4065	-7.9228	2.8866e-015
constante	-465132.1977	-18.7008	0
2CDD2days	NaN	NaN	0.88932
Viviendas	NaN	NaN	0.79395
3HDD2days	NaN	NaN	0.72836
2HDD2days	NaN	NaN	0.94744
3CDD	NaN	NaN	0.65355
2CDD	NaN	NaN	0.7144
3CDD5days	NaN	NaN	0.54251
2HDD	NaN	NaN	0.42588
CDD	NaN	NaN	0.43158
HDD	NaN	NaN	0.58903
HDD5days	NaN	NaN	0.41817
2HDD5days	NaN	NaN	0.72951
3HDD5days	NaN	NaN	0.55188
Calor	NaN	NaN	0.22756
Frio	NaN	NaN	0.19865
CDD2days	NaN	NaN	0.20451
KepDom	NaN	NaN	0.062051
3HDD	NaN	NaN	0.020537

Table 17. Input variables evaluation

The formula will be as follows:

$$c_{2005-2015} = 9845114{,}53 * WD - 2375{,}94 * Temp + 3169{,}18 * CDD_{5days} + 5199{,}96 * HDD_{5days} - 18{,}54 * CDD_{5days}^{2} + 3{,}21 * CDD_{2days}^{3} + 281762{,}57 Index + 853060{,}92 * Kep - 905401{,}40 * Kep_{Ind} - 465132{,}19$$

RESULTS AND DIAGNOSIS

The results of the model can be easily seen if we compare both the original data (blue) and the predicted one based on the model (green).

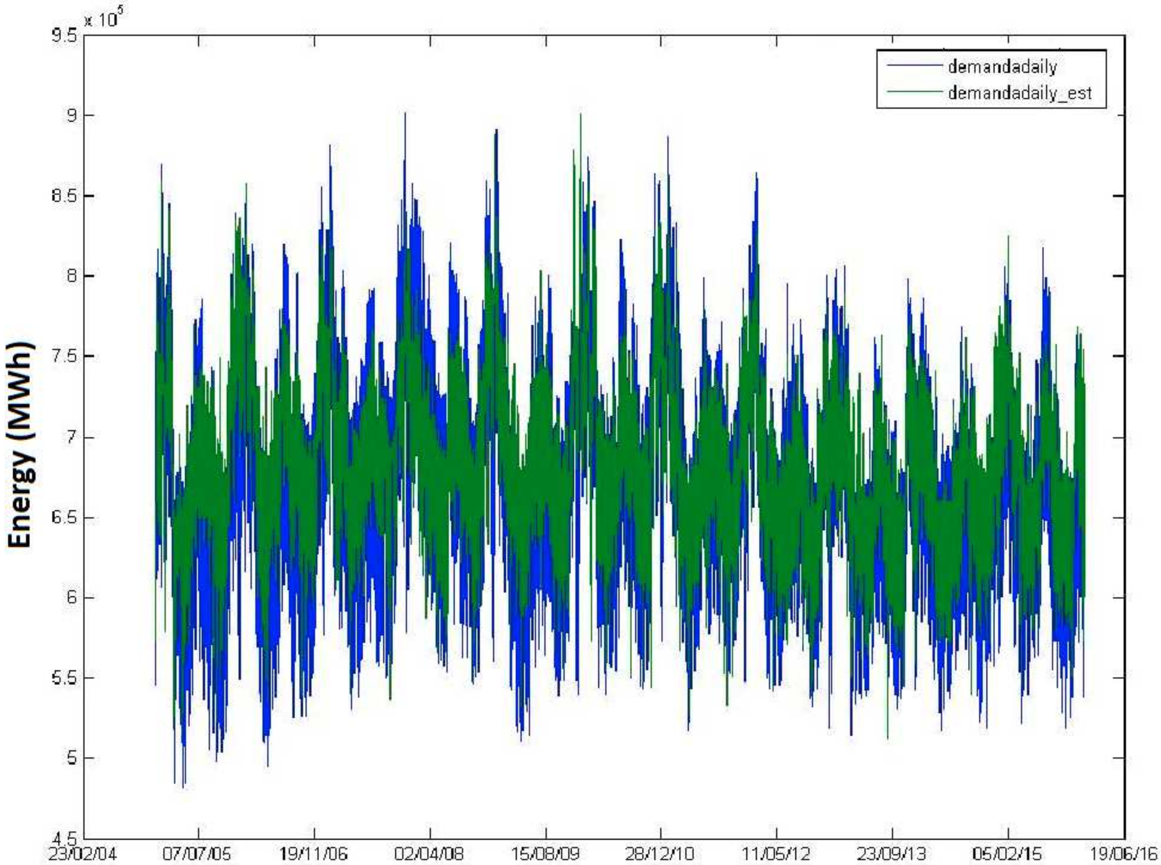

Figure 18. Original and estimated demand 2005-2015

To diagnose the model it is necessary to pay attention not only to the R squared, but also to the MAPE and its evolution within the years.

The following figure shows this statistical measure as it is presented by IDAT on a yearly basis:

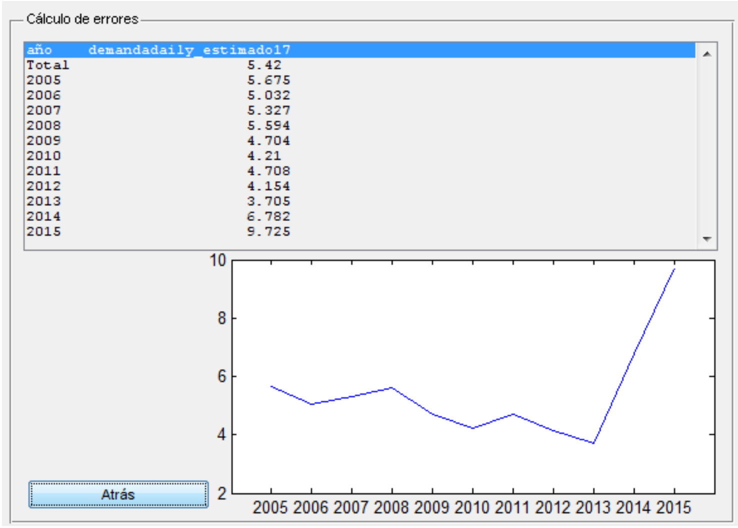

Figure 19. MAPE yearly evolution 2005-2015

It is obvious that the two last years (2014 and 2015) present a clear misbehaviour, and it could be useful to see the MAPE's evolution on a monthly basis too.

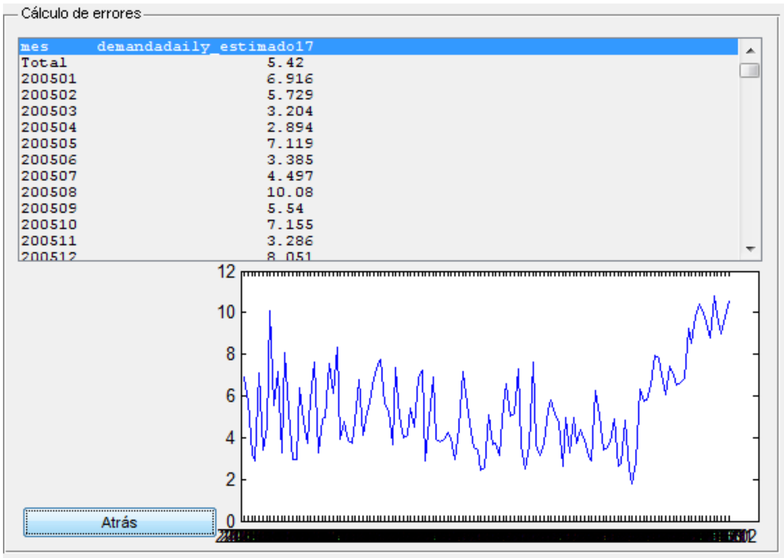

Figure 20. MAPE monthly evolution 2005-2015

Either way it is verified that the model's mean error is huge compared with the error of the model in which we used the trend time sequence variables instead of the economic and efficiency drivers to explain the upwarding trend.

This is particularly alarming the last year

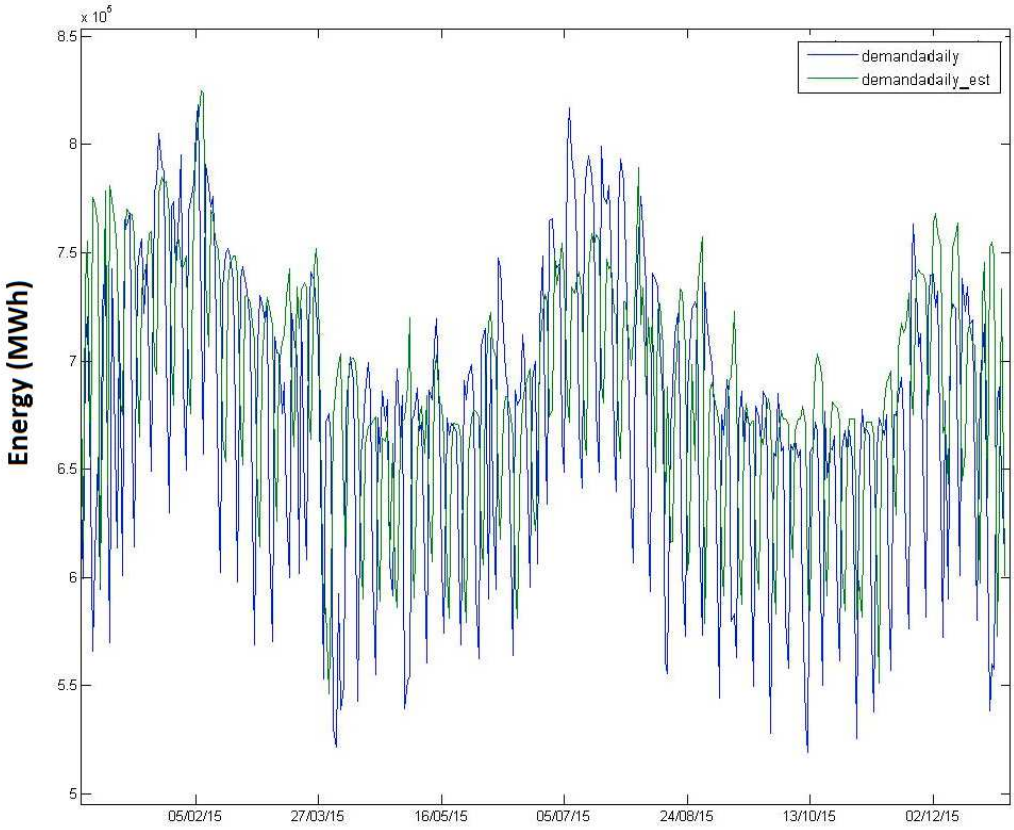

Figure 21. Detail of daily and estimated demand 2015

With this model diagnosis it is clear that, although the R squared is not significantly different between the two models, the MAPE is much worse using the economic and efficiency explanatory drivers instead of the time sequence trend proposed earlier.

Chapter 4 CONCLUSIONS

Just before discussing directly the conclusions it might be useful to recall the procedure that has been followed.

First of all, a review of the state of the art was done in order to determine the most suitable methodology to do the study. Then, the explanatory analysis was carried out. To do so, the electricity demand was explained using two possible models.

The first one includes calendar effect, temperature and time sequence variables as input data.

The second model includes also calendar effect and temperature variables, but this time the trend is explained through economic and efficiency indicators.

The structure of both models follows the same pattern, where an initial analysis of the explanatory variables is done, followed by the model development in which the equation is obtained, and the final results and diagnosis based on different statistical measures.such as R squared and MAPE.

Results and measures were estimated and compared within both models. It was clear that the second model presented a clear misbehaviour specially the last year. These results could be explained due to the fact that almost all drivers trying to explain the upwarding time effect of the trend were not intrinsically upwarding themselves.

By following this procedure, the most important conclusion we have noted, is the importance of well capturing the trend.

Besides, it is clear that certain parameters should always be part of these types of linear regression models, as it is the case of the working day, and temperature variables. Other drivers such as the housing stock or domestic primary energy intensity have resulted negligible.

It should be also noted that the first model has a reasonable error within most of its time framework, but it tends to fail during the two last years of the horizon analysed.

On the other other hand, the second model's error is clearly unacceptable, particularly the last year. Trend has not been well captured by the economic and efficiency variables, which lead to outrageous results.

To sum up, it is important to note the lack of demand's response to temperature evolution. Neither of the models proposed have analysed this important fact which should certainly improve the forecast.

Chapter 5 FUTURE DEVELOPMENTS

Although the project was completed satisfactorily, it is necessary to consider further developments and possible future improvements.

First of all, as it has been said in the previous section, future developments should include temperature variables that evolve within time in order to capture the demand response to temperature and how it changes within a time period.

Moreover, dealing with the temperature analysis, the thresholds that were set were stated qualitatively; this is, just by looking at the graphs showing demand's response to temperature. A possible improvement in this regard would be to carry out a quantitative method to state those thresholds, and a suitable possibility would be to use the Linear Hinges Model (LHM).

Other possible review would go through other methodology besides the simple linear regression model. The state of the art could be useful to find the technique that suits best for this purpose, whether it is a neural network (MLP, RBFN…), an ARIMA model or something else.

Finally, as a possible practical application of the project, we could use the forecasted demand for planning purposes and therefore as a tool for both utilities and system operator to know what should be expected in the following years.

BIBLIOGRAPHY

[1] Shu Fan and Rob J. Hyndman (2012) Short-term load forecasting based on a semi-parametric additive model. IEEE Transactions on Power Systems.

[2] Shu Fan, K. Methaprayoon, and Wei-Jen Lee (2009) Multiregion load forecasting for system with large geographical area. IEEE Transactions on Industry Applications.

[3] H. S. Hippert, C. E. Pedriera and R. C. Souza (2002) Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems.

[4] William W. S. Wei (2005) Time Series Analysis: Univariate and Multivariate Methods. Addison Wesley.

[5] George Box, Gwilym M. Jenkins, and Gregory Reinsel (1994) Time Series Analysis: Forecasting and Control. Prentice Hall.

[6] Bo-Juen Chen, Ming-Wei Chang and Chih-Jen Lin (2004) Load forecasting using support vector Machines: a study on EUNITE competition 2001. IEEE Transactions on Power Systems.

[7] Nowotarski, Jakub; Weron, Rafał (2015) Computing electricity spot price prediction intervals using quantile regression and forecast averaging. Computational Statistics .

[8] Weron, Rafał (2014) Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting.

[9] Tao Hong, Jason Wilson and Jingrui Xie (2014) Long term probabilistic load forecasting with hourly information. IEEE Transactions on Smart Grid.

[10] Rob J. Hyndman and Shu Fan (2010) Density forecasting for long-term peak electricity demand. IEEE Transactions on Power Systems.

[11] José Ramón Cancelo, Antoni Espasa, Rosmarie Grafe (2008) Forecasting the electricity load from one day to one week ahead for the Spanish system operator. UC3M Working papers. Statistics and Econometrics.

[12] Julián Moral Carcedo, José Vicens Otero (2005) Modelling the non-linear response of Spanish electricity demand to temperature variations. Energy Economics, Elsevier.

[13] Julián Moral Carcedo, José Vicens Otero (2003) Un modelo de previsión de demanda de energía eléctrica: THOR II.

[14] Weron, Rafał (2006) Modeling and forecasting electricity loads and prices. John Wiley & Sons, Ltd.

[15] Haykin, Simon (1998) Neural Networks A Comprehensive Foundation. Prentice Hall.

[16] Box, G. Jenkins (1970) Time Series Analysis, Forecasting and Control.

[17] Sánchez-Úbeda, Eugenio (1998) The Hinges model: A one-dimensional continuous piecewise polynomial model. Information Processing and Management of Uncertainty in Knowledge-based Systems

[18] Wen-Yeau, Chang Haykin, Simon (2014) A Literature Review of Wind Forecasting Methods.

[19] Infield , David; Hill, D.C. (1998) Optimal smoothing for trend removal in short term electricity demand forecasting. IEEE Transactions on Power Systems

[20] Tao Hong; Min Gui; Mesut E. Baran; H. Lee Willis (2010) Modelling and forecasting hourly electric load by multiple linear regression with interactions. PES General Meeting

[21] Liu and Hanssens (1982) Identification of multiple-input transfer function models. Communications in Statistics

[22] Hong, Wilson and Xie (2014) Long term probabilistic load forecasting and normalization with hourly information. IEEE Transactions on Smart Grid

[23] Goude, Nedellec and Kong (2014) Local Short and Middle Term Electricity Load Forecasting With Semi-Parametric Additive Models. IEEE transaction on smart grids

[24] Tao Hong (2014) Energy forecasting: past, present and future. Foresight: The International Journal of Applied Forecasting

[25] Tao Hong (2010) Short Term Electric Load Forecasting. PhD dissertation, North Carolina State University

[26] Tao Hong (2008) Long-Term Spatial Load Forecasting Using Human Machine Co-construct Intelligence Framework. Master thesis, North Carolina State University

[27] Markou, Kyriakides and Polycarpou (2008) 24-Hour ahead short term load forecasting using MLP. DEMSEE, Cyprus

[28] A.E. Clements, A.S. Hurn and Z. Li (2015) Forecasting day-ahead electricity load using a multiple equation time series approach School of Economics and Finance. Queensland University of Technology

Part II MATLAB CODE

Chapter 1 CODE

Here is the brief program that was used to compute the estimated demand (c_est) from the original data (c), based on the calendar effect (WD) and tendency (time sequence), and taking into account the regression model explained.

The difference between the original data and the estimated one was called corrected (c_corr) and represents the share of the electricity consumption that can no longer be explained just by the trend and calendar effect. This is somehow the noise, and would be rather cleared up using the rest of the drivers proposed (i.e. temperature variables).

```
clear all
clc
close all
%%%%%%%%%
%global D
%D=3287
%%%%%%%%%

% Datos diarios desde 01/01/2005(1) hasta 31/12/2013(3287)
load('d.mat')
load('c.mat')
load('Demand.mat')
load('t.mat')
load('wd.mat')
load('Matrix.mat')

a=Matrix\c
save a;
%a=1.0e+05[7.0247;0.3031;-0.0403;-0.0218;0.1507]=[a0;a1;a2;a3;a4]
%D=a0*WD+a1*t+a2*t^2+a3*t^3+a4
%D=1.0e+05[7.0247*WD+0.3031*t-0.0403*t^2-0.0218*t^3+0.1507]
%D'=D-...


%consumo estimado
c_est = Matrix * a;
save c_est;

figure(1)
plot([c c_est])
xlabel('time[days]')
ylabel('energy[MWh]')


%consumo corredigo
c_corr = c-c_est;
save c_corr;

figure(2)
plot([c c_corr])
xlabel('time[days]')
ylabel('energy[MWh]')
```