uc3m | Universidad Carlos III de Madrid

Master in Big Data Analytics
2017/2018

Master's Thesis

Forecasting Spanish electricity prices

Aldo Ramón Franco Comas

Advisors:
Francisco Javier Nogales Martín
Carlos Ruiz Mora

February, 2017. Madrid

Abstract:
Forecast the electricity prices in short, medium and long term is very important for power portfolio managers, producers, utility companies or large industrial consumer. In the last years has been proposed a lot of models to try to forecast the electricity prices in different markets with the possible better accuracy. This research compares and combines different techniques in order to forecast the electricity price at four horizons (day, week, month and year). We consider support vector machine, seasonal ARIMA models, TBATS and dynamic factor models. We also developed a website that allows to visualize and download these predictions and their corresponding prediction intervals.

Key Words:
Electricity price forecast; TBATS; Linear models; Support vector machine; Dynamic factor models.

Contents

Index of figures

List of Tables

1 Introduction

Since the early 1990, the government-controlled and traditional monopolistic power sector have been changing due to the introduction of competitive markets and a process of deregulation. In many countries, the electricity price is traded under the rules of a spot market, where assets are sold for cash and delivered immediately. Spot markets provide both consumers and producers with greater flexibility in their trading decisions, since traders can adjust their trading programs until the day before the trade, on the day-ahead market. [45]

Electricity prices in Europe are set on a daily basis (every day of the year) at midday, for the twenty-four hours of the following day, in what we refer to as the Daily Market. The price and volume of energy over a specific hour are determined by the point at which the supply and demand curves meet, according to the marginal pricing model adopted by the European Union (EU) [1]. In essence the process of price formation follows the basic rule of microeconomics theory (Law of Supply and Demand) where the price of the underlying commodity in a competitive market should reflect the relative scarcity of the supply for a given demand level. In the case where the demand for a commodity is low, those suppliers with higher incremental costs must step out of competition (or make negative profits) and give way to suppliers with the lowest incremental costs.

OMI-Polo Español, S.A. (OMIE) manages the spot market on the Iberian Peninsula, in the same way that Nord Pool Spot does so in the Nordic countries, EPEXSPOT (the European Power Exchange) in Germany, France, the United Kingdom, the Netherlands, Belgium, Austria, Switzerland and Luxembourg, and GME in Italy. This market is operated in a transparent and non-discriminatory manner. In January, 1998 OMIE began their operations for the Spanish market, and in July, 2007 extended them to cover the Iberian Market.

OMIE manages transactions amounting over ten billion euros, accounting for more than 80% of the electricity supplied in Spain and Portugal. Iberian market operates 365 days at year, 24 hours at day, and is open to all those buying and selling agents that wish to trade on it. There are over 800 agents operating nowadays and are involved in a total of over 13 million transactions per year.

Buying and selling agents may trade on this market regardless of whether they are in Spain or in Portugal. Their purchase and sale bids are accepted according to their economic merit order, until the interconnection between Spain and Portugal is fully occupied. If at a certain time of the day the capacity of the interconnection is such that it permits the flow of the electricity traded by the agents, the price of electricity for that hour will be the same for Spain and Portugal. If, on the other hand, the interconnection is fully occupied at that time, the price-setting algorithm (EUPHEMIA)[28] is run separately so that there is a price difference between the two countries. In 2014, the price of electricity was the same in Spain and Portugal for 90% of the time, which confirms that the integration of the Iberian market is working properly.

The mechanism described for setting the price of electricity on the daily market in Spain and Portugal is referred to as market splitting, being the same mechanism as the one used across Europe. The results of the daily market, as determined by the free trade between buying and selling agents, are the most efficient solution from an economic perspective. Nonetheless, given the nature of electricity, this process also needs to be feasible in physical terms. Accordingly, once these results have been obtained, they are sent to Red Eléctrica de España, S.A. (System Operator), for their validation from the standpoint of technical viability. This process is known as management of the system technical limitations and ensures that the market results can be technically accommodated. This means that the daily market results may be altered slightly, affecting around 4-5% of the energy, in response to an analysis of the technical limitations conducted by the System Operator, giving rise to a viable daily program.

The Iberian Market is one of Europe more liquid ones, and their prices are comparable to those in the other markets. In fact, in most years, this market has recorded prices that are below the average for Europe's major markets. In addition, while Iberian prices fluctuate between 0 and 180 Euros by MWh, their European counterparts oscillate within a wider price bracket, from -500 to $+3000$ Euros by MWh.

However the electricity has its peculiarities, it is not storable and power system stability needs a constant balance between consumption and production. The electricity price depends a lot of factors like the weather (temperature, precipitation, wind speed, etc.), the price of products as carbon and crude oil, and the intensity of business and everyday activities. These specific and unique characteristics cause that is very difficult to forecast the electricity prices in differents markets.[37]

Forecast electricity prices from a few hours to a few months ahead is of the interest to the power portfolio managers. An utility company, producers or a large industrial consumer who can forecast the prices with a reasonable level of accuracy can adjust its own consumption or production schedule and its bidding strategy in order to maximize the profits or reduce the risk.[49]

A lot of methods and ideas have been developed in the last years for electricity price forecasting. Based on forecasting horizons and goals, we can categorized the Electricity Price Forecast (EPF) in three groups as follows [41][48]:

- *Short-term price forecasting:* Will be mainly used by the market players to maximize profits or minimize risk in the Spot Markets. Involves forecasts from a few minutes up to a few days ahead. In EPF this is the category where we found more papers and research than the rest of the categories.

  Variants of autoregressive and general ARMA process have been used to short-term forecast in the German EEX market[9] that is part of the EPEXSPOT market. On the other hand, in the California market also know

as CAISO from April to mid-June, 2000 was found that simple AR model structure when expanded to include a load forecast of the system operator, is a tough competitor among the ARX-GARCH,TARX and Markov regime-switching (MRS) models. ARX turns out to be the best in a relatively calm period in the California market (April to mid-June, 2000), and second best (after TARX) in a more volatile period (second half of 2000)[31]. For predicting hourly prices in Spain and California have been used transfer function (TF) and dynamic regression (DR) where the price at hour t is related to the values of past prices at hours $t-1, t-2, ...$ and to the values of demands at hours $t, t-1, t-2, ...$[33]

For the PJM market [1] have been used three time series specifications (ARIMA, TF and DR), a wavelet multivariate regression technique, and a multilayer perceptron (MLP) with one hidden layer. For a dataset comprising PJM prices from the year 2002 concluding that the ARIMA model is worse than the time series models with exogenous variables but better than the MLP[8]. In the same market an ARMAX model with the temperature, squared temperature and cubed temperature as explanatory variables was used for day-ahead EPF. It was found that all temperature variables to be highly statistically significant during the pre-crisis period from April 1st,1998 to April 30th,2000.[23]

Elman networks, that is a three-layer network, was used to obtain short-term price forecasts in the market of mainland Spain [4]. Other papers proposed for Spanish market in 2002 combine kernel PCA (for extracting features of the inputs) with a Bayesian local informative vector machine (for making the predictions) and suggested that the resulting technique is better than other methods, including ARIMA and artificial neural networks (ANN)[11]. Hybrid models have been developed for point and interval forecasting in Australia, Ontario, Spain and California markets. In this case have been used a FitzHugh–Nagumo (FHN) model, for mimicking the spiky price behavior, with an Elman network, for regulating the latter, and a feed-forward ANN, for modelling the residuals.[40]

- *Medium-term price forecasting:* Will allow the successful negotiations of bilateral contracts between suppliers and consumer. From a few days to a few months ahead are generally preferred for balance sheet calculations, risk management and derivatives pricing.

  For medium-term have been considered vector ARIMA (essentially VAR) and factor models in the Iberian market with a good accuracy [3]. Forward price models are the domain of mathematical finance and have been used in

[1] It is a regional transmission organization that coordinates the movement of wholesale electricity in all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia and the District of Columbia in the United States.

the Nordic countries; constructing smooth forward price curves in electricity markets can be difficult, however, the benefits of doing it are the readily available medium-term price forecasts for multiple horizons. These forecasts can be biased, though, and include the risk premium. [14]

Reduced-form models usually don't forecast hourly prices accurately, but is expected to recover the main characteristics of electricity spot prices, typically at the daily time scale. Such models provide a simplified, yet reasonably realistic picture of the price dynamics, and are commonly used for derivatives pricing and risk analysis.[6] Interestingly, when they have used to forecast volatility or price spike, the reduced-form models have been reported to perform reasonably well. However, mean-reverting jump-diffusions or Markov regime-switching have been criticized for forecasting in general.[30] On the other hand, at least for medium-term forecasts of average daily prices at the German EEX market some authors oppose to the conclusion that the accuracy is not good.[25]

- *Long-term price forecasting:* Will influence the decisions on transmission expansion and enhancement, generation and distribution plannings. The investment profitability analysis and plannings, such as determining the future sites or fuel sources for power plants. The usual lead times are months, quarters or even years.

  Agent based Computational Economics (ACE) has become a widely accepted approach to solving both theoretical and practical problems in energy economics. The basic tool of ACE is an Agent Based Model (ABM) wich is a class of computational structures and rules for simulating the actions and interactions of autonomous agents, with the ultimate objective being to assess their effects on the system as a whole.[24]

  The ABM approach is well positioned in the long-term electricity price forecasting. Perhaps with the development of more powerful processors and cloud computing, ABM will someday provide efficient tools for EPF. It has been proposed an algorithm which switches between the predictions of different models (neural networks, fuzzy regressions and a standard regression) based on some prespecified rules, and use them for long-term (annual time scale) EPF.[49]

  An integrated, multistep algorithm which combines three ANNs, seven fuzzy regressions and one standard regression model have been developed and proposed to provide a joint framework for long-term (annual time scale) EPF[5]. On the other hand some preliminary results have showed the usefulness of factor models for long-term predictions.[12] [3]

Additionally, electricity price forecasting methods can be classified in six categories[48]:

- *Multi-agent models:* Which simulate the operation of a system of heterogeneous agents (generating units, companies) interacting with each other, and build the price process by matching the demand and supply in the market.
- *Fundamental models:* Which describe the price dynamics by modeling the impacts of important physical and economic factors on the price of electricity.
- *Reduced-form models:* Which characterize the statistical properties of electricity prices over time, with the ultimate objective of derivatives evaluation and risk management.
- *Statistical models:* Which are direct applications of the statistical and econometric forecasting techniques.
- *Computational intelligence models:* Which combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic systems, and may be regarded as 'intelligent' in this sense.
- *Hybrid models:* Many of the modeling and price forecasting approaches considered in the literature are hybrid solutions, combining techniques from two or more of the groups listed above.

The main contribution of this thesis is to compare the performance of several models in the forecast of the electricity price at the Iberian market considering different forecast horizons. Four horizons will be considered 24 hours, 144 hours, 30 days and 12 months. Also, a web page using Shiny will be created and loaded in Amazon Web Services (AWS) where the forecasts can be obtained and downloaded.

The rest of this document is organized as follows. Section 2 explains how we can obtain our data using the ESIOSs API and the main R's packages that were used in this research. Section 3 presents the different algorithms and techniques that were used in our EPF exercise as well as the performance measures. In Section 4 we explain why we choose these models and not another one that were considered too. In Section 5 we explain how the web page was created and how we automatically compute the forecast values. Finally, Section 6 concludes with remarks, limitations and possible extensions.

2 Data Availability

2.1 Software tools

For the development of this thesis, it has been used the software R that is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. We have used the following R's packages that have been developed by the big community of R's users.

- *jsnolite:* A fast JSON parser and generator optimized for statistical data and the web. This package offers flexible, robust, high performance tools for working with JSON in R and it is particularly powerful for building pipelines and interacting with a web API. In addition this package can convert JSON data from/to R objects.
- *httr:* Useful tools for working with HTTP organized by HTTP verbs (GET(), POST(), etc). Configuration functions make it easy to control additional request components (authenticate(), add_headers() and so on).
- *lubridate:* It is an R package that makes it easier to work with dates and times.
- *MLmetrics:* A collection of evaluation metrics (MAE, MAPE, etc), including loss, score and utility functions, that measure regression, classification and ranking performance.
- *forecast:* Methods and tools for displaying and analyzing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modeling.
- *caret:* The caret package (short for Classification And REgression Training) is a set of functions that attempt to streamline the process for creating predictive models, it is frequently used for work with machine learning techniques.
- *shiny:* Makes it incredibly easy to build interactive web applications with R. It is used for R users who have zero experience with web development.
- *e1071:* Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier, among other statistical techniques.
- *dygraphs:* An R interface to the 'dygraphs' JavaScript charting library (a copy of which is included in the package). Provides rich facilities for charting time-series data in R, including highly configurable series- and axis-display and interactive features like zoom/pan and series/point highlighting.
- *markdown:* Convert R Markdown documents into a variety of formats.
- *cronR:* Create, edit, and remove 'cron' jobs on your unix-alike system. The package provides a set of easy-to-use wrappers to 'crontab'. It also provides an RStudio add-in to easily launch and schedule your scripts.

2.2 ESIOS

Red Eléctrica de España, S.A. (REE) has as its mission to ensure the global operation of the Spanish electricity system through two essential activities: the operation of the electrical system and the transmission of electricity in high voltage. As an operator of the electrical system, REE guarantees the continuity and security of the electricity supply and the permanent balance between the production and consumption of electricity.

REE has developed an information system known as "System Operator Information System (E-SIOS)", specially designed to run all the necessary processes to ensure an economic and reliable exploitation of the Spanish Power System in real time. This system is able to store, in its historical database all the information received or published as result of the different processes.

The System Operator (SO) must provide, to the market participants, the markets results following the confidentiality criteria and periods established in the current regulation. Also some information should be published for the general public, to ensure the transparency of the SO operations.

To address the above requirements the SO has developed two Websites:

- *https://sujetos.esios.ree.es:* Secure website that requires the use of electronic certificate, intended to interchange confidential information between the SO and the market participants.
- *https://www.esios.ree.es:* E-SIOS Public Website where the non confidential information, result of the SO market operations, or other information of public interest related to the electricity markets, is published by REE.

From the second website, we download the data for this thesis using the ESIOS-API (see section 2.3).

2.3 API

In computer programming, an application programming interface (API) is a set of subroutine definitions, protocols, and tools for building an application software. In general terms, it is a set of clearly defined methods of communication between various software components. Web APIs are defined interfaces through which interactions happen between an enterprise and applications that use its assets. An API approach is an architectural approach that provides programmable interfaces to a set of services for different applications serving different types of consumers. When used in the context of web development, an API is typically defined as a set of Hypertext Transfer Protocol (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language (XML) or JavaScript Object Notation (JSON) format.

The new E-SIOS public website makes available an API for data download, which is detailed at *https://api.esios.ree.es*. To use this API you must request a personal token sending an email to *consultasios@ree.es* with the following subject *Personal token request*, due to public current token changes often. The information downloaded from the API is clean, free of outliers and missing values.

Here is an example, in R, that shows how it can be downloaded the Daily Peninsular Demand Forecast from April 1st, 2014 at 0000 to June 30th, 2017 at 2300 in JSON format by hour and transform it into a dataframe.

```
rm(list=ls())
library(jsonlite)
library(httr)
start_day = "2014-04-01"
end_day = "2017-06-30"
address_demanda = paste0("https://api.esios.ree.es/
   indicators/460?start_date=",start_day,
                         "T00:00:00&end_date=",end_day,"T23
                            :00:00&time_agg=avg&time_trunc=
                            hour",
                         collapse = NULL)
resp_demanda=GET(url = address_demanda, add_headers('
   Authorization'= 'Token token="cecb6df8f7127018
   ..........................................488b"'))
demanda = content(resp_demanda, as ="text")
demanda = fromJSON(demanda)
demanda = demanda$indicator$values
```

In particular we obtained data for price, demand forecast, wind power generation forecast, solar PV generation forecast and solar thermal generation forecast.

Table 1: Variables downloaded from E-SIOS website.

Variable	Id	By Month	By Day	By Hour
Price	600	Yes	Yes	Yes
Demand Forecast	460	No	No	Yes
Wind Power Generation Forecast	541	No	No	Yes
Solar PV Generation Forecast	542	No	No	Yes
Solar Thermal Generation Forecast	543	No	No	Yes

3 Methodology

In this section we will present the different algorithms or modeling approaches used in this project. First we will introduce Support Vector Machines (SVM) with lineal kernel, next the linear time series models used in this research, the TBATS models, simple linear regression and the dynamic factor models. Finally, we present the model selection procedure and how the prediction intervals can be computed.

3.1 Support Vector Machines for Regression

Support vector machine (SVM) analysis is a popular machine learning tool for classification and regression, developed by Vladimir Vapnik and his colleagues in 1992 at the AT&T's laboratories [46]. This machine learning tool is considered a non-parametric technique because it relies on kernel functions.

Linear SVM Regression: Primal Formula
Given a training dataset of N observations, $\{(x_n, y_n) : 1 \leq n \leq N\}$, where x_n is a vector of explanatory variables and y_n is a real valued response, the linear SVM regression consists in finding a linear function $f(x) = x'\beta + b$, such that it has at most an ε-deviation from the response, y. The linear SVM regression can be formulated as the following convex optimization problem:

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\beta'\beta \\ &st: \begin{cases} \forall n : y_n - (x_n'\beta + b) \leq \varepsilon \\ \forall n : (x_n'\beta + b) - y_n \leq \varepsilon \end{cases} \end{aligned} \tag{1}$$

It is possible that no such function $f(x)$ exists to satisfy these constraints for all n, that is there is an n such that $y_n - (x_n'\beta + b) > \varepsilon$ or $(x_n'\beta + b) - y_n > \varepsilon$. To deal with those unfeasible constraints, slack variables ξ_n and ξ_n^* for each point are introduced into the objective function of problem (1).

The problem is reformulated to what is know as the primal formula [46]:

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\beta'\beta + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \\ &st: \begin{cases} \forall n : y_n - (x_n'\beta + b) \leq \varepsilon + \xi_n \\ \forall n : (x_n'\beta + b) - y_n \leq \varepsilon + \xi_n^* \\ \forall n : \xi_n^* \geq 0 \\ \forall n : \xi_n \geq 0 \end{cases} \end{aligned} \tag{2}$$

The constant C is a positive numeric value that controls the penalty imposed on observations that lie outside the ε margin and helps to prevent overfitting (regularization). This value determines the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than ε are tolerated.

Linear SVM Regression: Dual Formula
The optimization problem previously described is computationally simpler to solve in its Lagrange dual formulation. The solution to the dual problem provides a lower bound to the solution of the primal (minimization) problem. The optimal values of the primal and dual problems need not be equal, and the difference is called the duality gap. But when the problem is convex and satisfies a qualification constraint, the value of the optimal solution to the primal problem is given by the solution of the dual problem.

To obtain the dual formula, construct a Lagrangian function from the primal function by introducing nonnegative multipliers α_n and α_n^* for each observation x_n. This leads to the dual formula, where we minimize:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) \quad (3)$$

$$st: \begin{cases} \sum_{n=1}^{N}(\alpha_n - \alpha_n^*) = 0 \\ \forall n: 0 \leq \alpha_n \leq C \\ \forall n: 0 \leq \alpha_n^* \leq C \end{cases}, \quad (4)$$

where $K(x_i, x_j)$ is a kernel function.[32] In the present thesis a linear kernel ($K(x_i, x_j) = x_i' x_j$), polynomial kernel and radial kernel were considered.

The parameter β can be completely described as a linear combination of the training observations using the expression:

$$\beta = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)(x_n). \quad (5)$$

The Karush-Kuhn-Tucker (KKT) conditions of problem (4) are:

$$\begin{cases} \forall n: \alpha_n(\varepsilon + \xi_n - y_n + f(x_n)) = 0 \\ \forall n: \alpha_n(\varepsilon + \xi_n^* + y_n - f(x_n)) = 0 \\ \forall n: \xi_n(C - \alpha_n) = 0 \\ \forall n: \xi_n^*(C - \alpha_n^*) = 0 \end{cases}. \quad (6)$$

These conditions indicate that all observations strictly inside the ε-band have Lagrange multipliers $\alpha_n = 0$ and $\alpha_n^* = 0$. Observations with nonzero Lagrange multipliers are called support vectors. The function used to predict new values depends only on the support vectors since

$$f(x) = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)K(x_n, x) + b. \quad (7)$$

3.2 Time Series Linear Models

A time series is a set of observations, X_t, each one being recorded at a specific time t. Time series appear in signal processing, finance, economy, meteorology, astronomy, etc. Time series analysis is the set of techniques, procedures and models that

are used to extract valuable characteristics of the data. Time series forecasting is one of the fundamental tasks of this analysis and it consists in predicting future values based on the previously observed values. There are many kind of models that can be adjust to time series, in this section we are going to explain some of them. We will concentrate our attention on the family of linear models.

One of the most famous and simplest model is the AutoRegressive model, AR, that is frequently used in economics for stationary process. [2] This kind of models specified that X_t depends only on its own past values and on an independent term.

Definition 3.1. An AutoRegressive model of order p, AR(p), is defined as follows [7]:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + Z_t; \qquad t = 0, \pm 1, \pm 2, ... \tag{8}$$

where Z_t is a sequence of independent and identically distributed random variable with zero mean and constant variance, σ^2. Z_t is called white noise and it is denoted by $Z_t \sim WN(0, \sigma^2)$.

There are time series, X_t, that depends on the current and various past values of the noise. Those time series are usually called Moving Average MA model.

Definition 3.2. A Moving Average model of order q, MA(q), is defined as follows [7]:

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}; \qquad t = 0, \pm 1, \pm 2, ... \tag{9}$$

where $Z_t \sim WN(0, \sigma^2)$.

If we combine the AutoRegressive structure with the Moving Average ones, we obtain a general model called ARMA(p, q) where p is the order of the autoregressive component and q is the order of the moving average component. An ARMA($0, q$) model is equal to a MA(q) model and similarly for an ARMA($p, 0$) which is equal to an AR(p) model.

Definition 3.3. An AutoRegressive Moving Average or ARMA(p, q) model is given by [7]:

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}, \tag{10}$$

where $Z_t \sim WN(0, \sigma^2)$.

This equation can be written in the following form:

$$\boldsymbol{\phi}(B) X_t = \boldsymbol{\theta}(B) Z_t; \qquad t \in \mathbb{Z} \tag{11}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are polynomials of degrees p and q, respectively. That is,

$$\boldsymbol{\phi}(z) = 1 - \phi_1 z - ... - \phi_p z^p, \tag{12}$$

$$\boldsymbol{\theta}(z) = 1 + \theta_1 z + ... + \theta_q z^q \tag{13}$$

[2] A stationary process is a stochastic process whose joint probability distribution doesn't change when shifted in time. Consequently, parameters such as mean and variance do not change over time.

and B being the backshift operator also know as lag operator and it is defined by:

$$B^j X_t = X_{t-j}; \qquad j \in \mathbb{Z}. \tag{14}$$

The ϕ_i are the parameters of the AR polynomial, and the θ_j are the parameters of the MA polynomial.[3]

There are time series that are not stationary but with simple transformations they become stationary. Some of these simple transformation are the regular and seasonal difference.

Definition 3.4. A seasonal differencing with season $s > 1$ is a transformation applied to time serie in order to make it stationary. It is defined by[7]:

$$X_t^{`} = X_t - X_{t-s}. \tag{15}$$

Definition 3.5. A regular differencing is a transformation applied to time series in order to make it stationary. It is defined by [20]:

$$X_t^{`} = X_t - X_{t-1}. \tag{16}$$

Regular and seasonal differences are very useful because they remove trend and seasonality and stabilize the time series.

We have already discussed the class of ARMA models for representing stationary series. We can generalize the ARMA(p, q) model to an AutoRegressive Integrated Moving Average ARIMA(p, d, q), which are useful to handle series that shows evidence of non-stationarity.

Definition 3.6. If d is a non-negative integer, X_t follows an Auto Regressive Integrated Moving Average ARIMA(p, d, q) if $Y_t := (1 - B)^d X_t$ is a stationary ARMA process.[7]

This definition means that X_t verifies a difference equation of the form:

$$\phi^*(B)X_t \equiv \phi(B)(1 - B)^d X_t = \theta(B)Z_t; Z_t \sim WN(0, \sigma^2). \tag{17}$$

It is easy to check that if $d = 0$ the model is an ARMA(p, q) process. It should be noticed that the roots of $\phi^*(B)$ polynomials are one or bigger than one on modulus.

Additionally, can extend the ARIMA models to seasonal time series, this is the special case of the general seasonal ARIMA(SARIMA) model defined as follows.

Definition 3.7. If d and D are nonnegative integers, then X_t follows a seasonal ARIMA$(p, d, q)(P, D, Q)_s$ process with season s if the differenced series $Y_t = (1 - B)^d(1 - B^s)^D X_t$ is a stationary ARMA process defined by [20]:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t; \quad Z_t \sim WN(0, \sigma^2) \tag{18}$$

where $\phi(z) = 1-\phi_1 z-...-\phi_p z^p$; $\Phi(z) = 1-\Phi_1 z-...-\Phi_P z^P$; $\theta(z) = 1+\theta_1 z+...+\theta_q z^q$ and $\Theta(z) = 1 + \Theta_1 z + ... + \Theta_Q z^Q$.

[3] An ARMA(p, q) model is stationary iff the modulus of roots of the autoregressive polynomial are bigger than one.

In order select the "best" model for a time series we can use the Akaike Information Criterion (AIC), the bias corrected version of Akaike information criterion (AICc) or the Bayesian Information Criterion (BIC) also known Schwarz criterion (SBC or SBIC).

Definition 3.8. The AIC value [2] of model M with k parameters to be estimated is defined by:

$$AIC = 2k - 2\ln(\widehat{L}), \tag{19}$$

where $\widehat{L}$ is the maximized value of the likelihood function of M.

Definition 3.9. AICc is AIC with a bias correction for finite sample sizes. The AICc value [17] of model M with k parameters to be estimated is defined by:

$$AICc = AIC + \frac{2k(k+1)}{n-k+1}, \tag{20}$$

where n is the sample size.

Definition 3.10. The BIC [29] value of model M with k parameters to be estimated is defined by:

$$BIC = \ln(n)k - 2\ln(\widehat{L}), \tag{21}$$

where $\widehat{L}$ is the maximized value of the likelihood function of M and n is the sample size.

Given a set of candidate models, the model with minimum AIC(or AICc or BIC) value will be selected.

3.3 TBATS Models

The time series models presented at the previous section are designed to accommodate simple seasonal patterns with an integer-valued seasonal period. However, there are time series that exhibit more complex seasonal patterns including multiperiods and non-integer seasonal period. In this section, we present a class of models that are able to handle these complex structure.

Single seasonal exponential smoothing methods, are the most widely used forecasting procedures in practice [42] and have been shown to be optimal for a class of state space models [34]. Among this class, one of the most commonly used seasonal model is the additive (or multiplicative) Holt-Winter procedure. The following method is an extension of the Holt-Winters method, that incorporates a second seasonal component [44]:

$$y_t = l_{t-1} + b_{t-1} + s_t^{(1)} + s_t^{(2)} + d_t \tag{22}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha d_t \tag{23}$$

$$b_t = b_{t-1} + \beta d_t \tag{24}$$

$$s_t^{(1)} = s_{t-m_1}^{(1)} + \gamma_1 d_t \tag{25}$$

$$s_t^{(2)} = s_{t-m_2}^{(2)} + \gamma_2 d_t, \tag{26}$$

where m_1 and m_2 are the seasonal periods, d_t is a white noise; the components l_t and b_t represent the level and trend components at time t, $s_t^{(i)}$ is the ith seasonal component at time t; $i = 1, 2$. The smoothing parameters are the coefficients α,β,γ_1 and γ_2 and the seeds or initial states variables are l_0,b_0,$s_{1-m_1}^{(1)}, ..., s_0^{(1)}$ and $s_{1-m_2}^{(2)}, ..., s_0^{(2)}$.

The above model can be extended by using a Box-Cox transformation [4], $ARMA$ errors and T seasonal patterns to handle a wider variety of seasonal patterns as follows[10]:

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{I} s_{t-m_i}^{(i)} + d_t \tag{28}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \tag{29}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t \tag{30}$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t \tag{31}$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{32}$$

where $y_t^{(\omega)}$ is the Box-Cox transformed series, $m_1, ..., m_I$ are the seasonal periods, l_t is the local level in period t, b is the long-run trend, b_t is the short-run trend in period t, $s_t^{(i)}$ represents the ith seasonal component at time t, d_t denotes an ARMA(p, q) and ε_t is a Gaussian white noise with zero mean and constant variance σ^2. The smoothing parameters are the coefficients α,β and γ_i for $i = \overline{1, I}$.

The model defined by (28)-(32) is denoted by BATS$(\omega, \phi, p, q, m_1, m_2, ..., m_I)$ where the acronym results from **B**ox Cox transformation, **A**RMA errors, **T**rend and **S**easonal components. It generalized the Holt-Winter procedure but still does not consider the non-integer seasonal period case.

In order to solve this last incovenience, the seasonalities are expressed by a trigonometric representation based on Fourier series as follows [50]:

[4]The Box-Cox transformation is a parametric function used to reduce anomalies such as non-additivity, non-normality and heteroscedasticity [38]. It is defined by the following expression:

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}; & \omega \neq 0 \\ \log y_t; & \omega = 0 \end{cases}. \tag{27}$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)} \tag{33}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \tag{34}$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t, \tag{35}$$

where $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ are the smoothing parameters and $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$. The terms $s_{j,t}^{(i)}$ and $s_{j,t}^{*(i)}$ are the stochastic level and the stochastic growth in the level of the ith seasonal component, respectively. The number of harmonics required for the ith seasonal component is $k_i = \frac{m_i}{2}$ for even values and $k_i = \frac{m_i - 1}{2}$ for odd values of m_i.

Now replacing the seasonal component $s_t^{(i)}$ in the BATS model and modifying the measurement equation $y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + d_t$ we get TBATS, where the T means Trigonometric. This model depends on the following arguments (ω,ϕ,p,q,$\{m_1, k_1\}$,...,$\{m_T, k_T\}$). The TBATS model handles non-integer seasonal periods as well as nested and non-nested seasonal components. [10]

3.4 Simple Linear Regression

In order to add some explanatory variables to our models we can use the simple linear regression to add one explicative variable. A linear regression model [43] is given by:

$$y_t = \alpha + \beta x_t + \varepsilon_t, \tag{36}$$

where ε_t is assumed to be a Gaussian white noise.

The target is to find the best values of α and β that give us the best fit, this can be done using least square estimation that minimize the sum of the squared errors[43] as follows:

$$\min_{\alpha,\beta} \sum_{t=1}^{n} (y_t - \alpha - \beta x_t)^2 = \sum_{t=1}^{n} \varepsilon_t^2. \tag{37}$$

3.5 Dynamic Factor Models

The models in the previous sections assume that the times series or the response variable are univariate. In this section, we present the dynamic factor models (DFM) a class of multivariate time series models that has been popularized in the last 20 years but was proposed in the eighties (see [13] and [35]). Those models have been used in EPF problem by [3] and [12], among others.

The DFM assumes that y_t, an m-dimensional time series, can be written as a linear combination of common factors plus an error term:

$$\underset{m\times 1}{y_t} = \underset{m\times r}{P}\ \underset{r\times 1}{f_t} + \underset{m\times 1}{\epsilon_t}\,, \tag{38}$$

where f_t is the r-dimensional vector of common factors, P is the weight matrix of the factors also known as loading matrix, and ϵ_t is the vector of specific factors or error term.
Additionally, it is assumed that the vector of common factors follows a seasonal VARIMA$(p,d,q)(P,D,Q)_s$ model defined by:

$$\underset{r\times r}{\phi(B)}\underset{r\times r}{\Phi(B)}\underset{r\times 1}{f(t)} = \underset{r\times r}{\theta(B)}\underset{r\times r}{\Theta(B)}\,\underset{r\times 1}{\upsilon_t}\,, \tag{39}$$

where B is the lag operator, $\phi(B) = I - \phi_1 B - ... - \phi_p B^p$ and $\theta(B) = I - \theta_1 B - ... - \theta_q B^q$ are $r \times r$ matrices of autoregressive polynomials and moving averages of the regular part respectively and $\Phi(B) = I - \Phi_1 B - ... - \Phi_P B^P$ and $\Theta(B) = I - \Theta_1 B - ... - \Theta_Q B^Q$ of the seasonal part. The innovations are assumed to be time independent, e.g., $E(\upsilon_t, \upsilon_t + h) = 0$ for $h \neq 0$ and independent of specific factors, e.g., $E(\upsilon_t, \epsilon_t + h) = 0$ for all h. For the specific factors, ϵ_t, we assume that each component follows an univariate seasonal ARMA model.

The factorial model defined by (38)-(39) is not identified since for any non-singular $r \times r$ matrix, Ω, it is possible to express the vector series y_t in terms of $(P\Omega)$ and $(\Omega^{-1} f_t)$ which are new sets of weights and factors, respectively. Several restrictions have been proposed to solve the problem of identification, e.g., $\sum_\epsilon$ or $P'P = I$ [36], and $P = [p_{i,j}]$ with $p_{i,j} = 0$ for $j > i$ [15]. In this thesis, we use the restriction $P'P = I$ and we assume that the factors are orthogonal, i.e, for, $f_{\cdot,i} \perp f_{\cdot,j}$ for $i \neq j$, as in [27] and [12].

Also, as in [12], we will consider a simpler version of the DFM where the common factor follows seasonal ARIMA models. A relevant step in DFM is the selection of the seasonal ARIMA (ARMA) models that approximate the common (specific) factors dynamical structure. In [12], the TRAMO-SEATS procedure [15] was used to select the order of those models but in this thesis, since our implementation is on R, we will use the auto.arima function which is included in the package forecast [21].

3.6 Forecasting Accuracy Measure

Many measures of forecast accuracy have been proposed in the literature. In EPF the most frequently used measures of accuracy are those based on absolute errors $|y_t - \widehat{y}_t|$ where y_t is the actual value at time t and $\widehat{y}_t$ is the predicted value. Since it is difficult to compare the accuracy of the models in different datasets using the absolute errors, many authors use measures based on absolute percentage errors as *MAPE* (Mean Absolut Percentage Error) defined by:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \widehat{y}_t}{y_t}\right|, \tag{40}$$

where y_t is the real price and $\widehat{y}_t$ is the corresponding forecast.

There are some inconveniences in using this accuracy measure even when is the most popular [49] in EPF. One of the disadvantage when we use MAPE is that it

put heavier penalty on positive errors than on negative errors. If y_t is close to zero probably $\widehat{y_t}$ is also close to zero. However the MAPE still involves a division by a number close to zero. Other disadvantage of measures based on percentage errors is that they assume a meaningful zero [22]. For example it doesn't have sense in measuring forecast error in the others European markets where there are negative prices. Due to this inconvenient we decided to use two more accuracy measure, *MAE* (Mean Absolute Error) and *MDAE* (Median Absolute Error) defined as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \widehat{y_t}| \tag{41}$$

and

$$MDAE = median(|y_t - \widehat{y_t}|). \tag{42}$$

We will show the accuracy of each model tested in MAPE, MAE and MDAE, selecting the best model according to MAE.

3.7 Model selection criteria based on prediction performance

When selecting models, it is common to use a portion of the available data for fitting (data train), and use the rest of the data for evaluating the model performance (data test). In order to choose the "best" model, we used a procedure known as rolling forecasting origin. It works as follows [20]:

1. Select the data train windows with a fixed size.
2. Train a model (in our case automatically without user intervention).
3. Forecast the period in which we are interested.
4. Obtain the prediction errors as the difference between the real values and the forecasted values.
5. Compute the prediction accuracy.
6. Update the train set with the real values of the period that was forecasted.
7. Remove in the train set the first values with length equal to the forecasted period.

The above seven step procedure is repeated until the end of the test set. Figure 1 illustrates the rolling forecasting origin procedure.

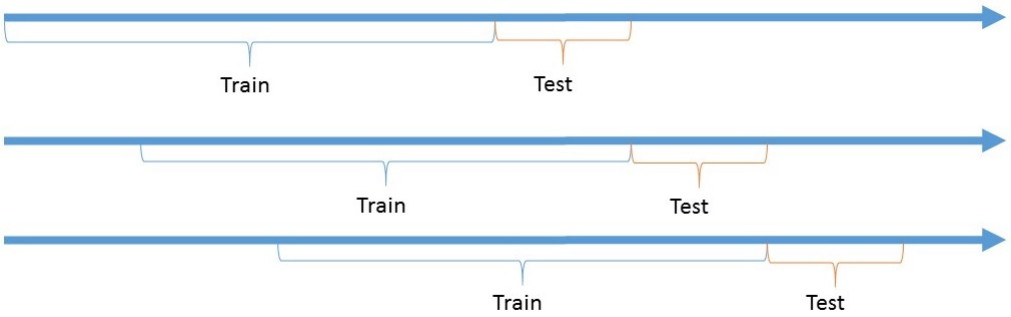

Figure 1: Illustration of the rolling forecasting origin procedure.

We are going to select the model whose MAE distribution have the lower third quartile (Q3). We prefer Q3 instead of the mean since the mean can be affected by outliers. The model associated to the better distribution taking in account Q3 will be selected as the best model and it will be used to forecast future values in the developed app (see section 5).

3.8 Prediction Intervals

In this section, we describe a simple procedure for obtaining prediction intervas. It makes use of the recorded prediction errors as well as bootstrap techniques. A prediction interval is an interval associated with a random variable yet to be observed, with a specified probability of the random variable lying within the interval.[19] For example, I can give $(1-\alpha)\%$ interval for the forecast of the electricity prices for tomorrow, so the prices for tomorrow should be on the interval with probability $1-\alpha$.[5]

Once we performed the rolling forecasting origin procedure, we have a set of prediction errors for each instant/origin t in the testing set and for the different horizons. For instance, in the case of prediction horizon of 24 hours, we have the following prediction errors set:

$$\begin{bmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,24} \\ e_{2,1} & e_{2,2} & \dots & e_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N,1} & e_{N,2} & \dots & e_{N,24} \end{bmatrix}, \tag{43}$$

where $e_{t,h}$ corresponds to the prediction error at origin t for hour h and N is the length of the testing set.

Then, a $(1-\alpha)\%$ prediction intervals for $y_{t,h}$ can be obtained by the following expression:

$$\hat{y}_{t,h} \pm z_{\alpha/2}\sigma(\varepsilon_{t,h}), \tag{44}$$

[5]Confidences and predictions intervals are not the same thing. Unfortunately both concepts are usually confused for example in econometrics is very common use "confidence intervals" as prediction interval.[47]

where $z_{\alpha/2}$ is the percentile of the standard Gaussian distribution and $\sigma(\varepsilon_{t,h})$ is the standard deviation of the prediction error, $\varepsilon_{t,h}$. In order to obtain a bias-corrected estimates of these standards deviations, a bootstrap procedure (with $B = 100000$ replicas) for each one of the horizon, h, was performed. That is, we generate B replicas of the prediction error set

$$\begin{bmatrix} e_{1,1}^{(*,b)} & e_{1,2}^{(*,b)} & \dots & e_{1,24}^{(*,b)} \\ e_{2,1}^{(*,b)} & e_{2,2}^{(*,b)} & \dots & e_{2,24}^{(*,b)} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N,1}^{(*,b)} & e_{N,2}^{(*,b)} & \dots & e_{N,24}^{(*,b)} \end{bmatrix}, \tag{45}$$

where $b = 1, 2, \dots, B$. At each replica, we calculate the standard deviation for each one of the horizon. The bias corrected estimate is obtained by

$$\sigma(\varepsilon_{t,h}) - \left(E^*\left[\sigma(\varepsilon_{t,h}^*)\right] - \sigma(\varepsilon_{t,h})\right) = 2\sigma(\varepsilon_{t,h}) - E^*\left[\sigma(\varepsilon_{t,h}^*)\right], \tag{46}$$

where E^* denotes the bootstrap mean.

4 Forecasting exercises

Our goal is to obtain automatic models that forecast the prices for different horizons size. We have decided to select the best model according to the data that we have and not select a fixed model that can't be effective over time. For example, given a time series X_t we can select a model M_1 that is slightly better than another model M_2, but after two months maybe the second model is better than the first one. So it is important highlight this aspect, we will not use a fixed model, that is, we will automatically select, using, R the best model according to a set of parameters and conditions.

In our case we are going to predict the electricity prices for the following horizons:

- 1-24 hours (One day ahead)
- 1-144 hours (Six days)
- 1-30 days (One month)
- 1-12 months (One year)

The algorithms and tools presented in the section of **Methodology** will be used independently or combined to reach our goal.

4.1 Forecasting 24 hours

Thanks to the website *https://www.esios.ree.es/es*, we can download the following explanatory variables using the API:

- Daily Peninsular Demand Forecast.
- Peninsular Wind Power Generation Forecast.
- Solar PV Generation Forecast.
- Solar Thermal Forecast.

The difference between Solar Termal and Solar PV is the following, solar thermal electric energy generation concentrates the light from the sun to create heat, and that heat is used to run a heat engine, which turns a generator to make electricity. On the other hand photovoltaic, or PV energy conversion, directly converts the sun's light into electricity. This means that solar panels are only effective during daylight hours because storing electricity is not a particularly efficient process. Heat storage is a far easier and efficient method, which is what makes solar thermal so attractive for large-scale energy production. Heat can be stored during the day and then converted into electricity at night. Solar thermal plants that have storage capacities can drastically improve both the economics and the dispatchability of solar electricity.[39]

For this exercise our test data begin on January 1st, 2016 until June 30th, 2017. We considered different machine learning algorithms like SVM with linear, polynomial and radial kernel, KNN and Gaussian process with radial and polynomial kernel. We will use different training windows sizes, DFM was considered too. The accuracy for each day forecasted was measured using MAE, MAPE and MDAE. In the machine learning case it was used the following relationship:

$$\begin{aligned} Price \sim & DailyPeninsularDemandForecast \\ & + PeninsularWindPowerGenerationForecast \\ & + SolarPVGenerationForecast \\ & + SolarThermalForecast. \end{aligned}$$

The SVM, Gaussian Process and KNN models were trained using the package *caret* (Classification And Regression Training) that contains tools for data splitting, pre-processing, feature selection, model tuning using resampling, variable importance estimation as well as other functionalities. It is known that are many different modeling functions in R. For model training and/or prediction some have different syntax. The package *caret* started off as a way to provide a uniform interface the functions themselves, as well as a way to standardize common tasks (such parameter tuning and variable importance) [26]. The tuning of hyper parameter was done using grid search.

In the next tables we can observe the accuracies of each method for different training windows size.

Table 2: Accuracies for SVM with linear kernel, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.57	2.13	2.94	3.39	3.96	18.26
	MAPE	1.26	4.64	6.61	12.46	10.65	272.45
	MDAE	0.38	1.77	2.61	3.12	3.83	17.22
14	MAE	0.72	2.17	3.04	3.64	4.46	14.33
	MAPE	1.59	4.64	7.08	13.88	11.43	325.62
	MDAE	0.45	1.79	2.72	3.39	4.23	16.52
21	MAE	0.74	2.24	3.15	4.01	4.70	17.00
	MAPE	1.72	4.77	7.40	15.06	12.87	347.72
	MDAE	0.52	1.89	2.78	3.76	4.50	17.30
42	MAE	0.62	2.57	3.63	4.96	6.25	20.77
	MAPE	1.55	5.46	9.13	17.24	17.08	339.39
	MDAE	0.51	2.23	3.37	4.78	6.21	21.56
84	MAE	0.63	2.90	4.81	6.15	8.78	22.16
	MAPE	1.46	6.38	12.17	21.44	20.53	405.32
	MDAE	0.54	2.48	4.24	5.95	8.43	21.81

Table 3: Accuracies for SVM with polynomial kernel, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.44	2.03	2.95	3.59	4.24	30.55
	MAPE	1.06	4.48	6.70	13.23	10.81	252.70
	MDAE	0.33	1.69	2.48	3.12	3.89	32.22
14	MAE	0.68	2.09	3.03	3.66	4.54	31.95
	MAPE	1.67	4.55	7.08	13.53	12.05	293.92
	MDAE	0.43	1.68	2.66	3.31	4.20	35.07
21	MAE	0.49	2.08	3.07	3.97	4.84	19.40
	MAPE	1.08	4.61	7.29	14.37	12.72	293.52
	MDAE	0.32	1.75	2.79	3.67	4.48	18.63
42	MAE	0.49	2.38	3.53	4.75	6.09	16.80
	MAPE	1.20	5.10	8.57	16.48	16.41	312.16
	MDAE	0.49	2.01	3.22	4.50	5.84	18.45
84	MAE	0.55	2.60	4.58	5.83	8.02	21.03
	MAPE	1.36	6.01	11.05	19.96	19.41	381.17
	MDAE	0.49	2.23	4.33	5.57	7.69	22.64

Table 4: Accuracies for SVM with radial kernel, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.71	2.38	3.36	4.19	5.03	27.49
	MAPE	1.40	5.22	7.68	16.75	13.81	399.84
	MDAE	0.55	1.97	2.95	3.88	4.74	34.02
14	MAE	0.79	2.28	3.36	4.13	5.07	25.95
	MAPE	1.88	5.12	7.80	15.93	13.09	349.40
	MDAE	0.52	1.83	2.96	3.81	4.94	33.45
21	MAE	0.76	2.35	3.35	4.37	5.50	27.97
	MAPE	1.82	5.12	7.74	16.66	14.56	344.37
	MDAE	0.43	1.92	2.96	4.03	5.15	37.05
42	MAE	0.83	2.50	3.82	5.00	6.44	24.11
	MAPE	1.58	5.50	9.12	17.86	17.72	311.99
	MDAE	0.54	2.09	3.44	4.70	6.15	31.86
84	MAE	0.93	2.87	4.73	5.94	8.12	26.34
	MAPE	2.27	6.45	11.08	20.84	20.58	364.50
	MDAE	0.62	2.39	3.98	5.64	7.82	25.13

Table 5: Accuracies for Gaussian Process with polynomial kernel, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.46	1.98	2.77	3.45	4.12	34.22
	MAPE	1.10	4.33	6.49	12.70	10.44	247.05
	MDAE	0.39	1.66	2.51	3.06	3.81	31.44
14	MAE	0.69	2.05	2.96	3.57	4.44	19.92
	MAPE	1.49	4.54	6.76	13.29	11.58	281.98
	MDAE	0.42	1.70	2.60	3.25	4.04	20.49
21	MAE	0.54	2.08	3.00	3.89	4.66	18.25
	MAPE	1.19	4.59	7.06	14.07	12.33	283.66
	MDAE	0.40	1.73	2.75	3.60	4.43	17.39
42	MAE	0.56	2.37	3.53	4.68	5.98	16.35
	MAPE	1.30	5.24	8.58	16.12	15.91	297.25
	MDAE	0.52	2.02	3.28	4.46	5.98	17.85
84	MAE	0.57	2.76	4.79	5.81	8.03	20.01
	MAPE	1.40	6.21	11.74	19.71	19.45	368.53
	MDAE	0.45	2.44	4.45	5.57	7.74	21.61

Table 6: Accuracies for Gaussian Process with radial kernel, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.54	2.26	3.25	4.02	5.01	27.29
	MAPE	1.25	4.89	7.54	15.92	13.04	393.21
	MDAE	0.36	1.86	2.83	3.71	4.61	29.99
14	MAE	0.72	2.17	3.12	3.95	4.88	22.66
	MAPE	1.74	4.77	7.36	15.14	13.31	302.48
	MDAE	0.41	1.80	2.83	3.68	4.68	28.57
21	MAE	0.78	2.19	3.13	4.17	5.11	23.72
	MAPE	1.60	4.82	7.38	15.71	13.88	314.52
	MDAE	0.46	1.88	2.87	3.89	4.93	30.75
42	MAE	0.63	2.45	3.57	4.82	6.14	23.09
	MAPE	1.55	5.41	8.81	17.02	16.67	288.01
	MDAE	0.39	2.08	3.27	4.59	5.91	29.91
84	MAE	0.69	2.77	4.52	5.82	8.18	25.00
	MAPE	1.70	6.25	11.04	20.23	19.75	341.77
	MDAE	0.56	2.46	4.09	5.57	7.72	21.54

Table 7: Accuracies for KNN, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
7	MAE	0.73	2.37	3.21	3.89	4.66	21.43
	MAPE	1.59	4.95	7.60	15.00	13.06	327.68
	MDAE	0.43	1.89	2.81	3.45	4.14	22.67
14	MAE	0.75	2.32	3.21	3.96	4.98	19.10
	MAPE	1.42	5.03	7.58	15.05	12.71	307.05
	MDAE	0.47	1.82	2.80	3.55	4.49	21.63
21	MAE	0.79	2.33	3.38	4.25	5.30	20.04
	MAPE	1.66	5.09	8.06	15.85	14.23	303.76
	MDAE	0.52	1.93	3.01	3.88	4.96	22.23
42	MAE	0.75	2.55	3.93	4.94	6.20	20.16
	MAPE	1.56	5.64	9.13	17.23	17.00	303.38
	MDAE	0.58	2.18	3.27	4.57	5.87	22.69
84	MAE	0.87	2.96	4.69	5.98	8.30	23.86
	MAPE	1.72	6.45	11.17	20.85	20.39	384.46
	MDAE	0.65	2.57	4.25	5.62	8.03	25.68

Table 8: Accuracies for DFM, 1-24 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	0.75	2.91	4.34	5.27	6.56	28.75
	MAPE	1.52	6.58	10.18	18.05	17.27	360.30
	MDAE	0.53	2.51	3.98	5.05	6.45	32.47
200	MAE	0.99	2.89	4.40	5.16	6.40	27.18
	MAPE	2.21	6.26	10.20	18.00	16.65	356.61
	MDAE	0.63	2.49	3.96	4.97	6.32	30.27
300	MAE	0.88	2.76	4.14	5.01	6.30	27.87
	MAPE	2.12	6.15	9.90	17.49	16.32	364.77
	MDAE	0.56	2.40	3.76	4.83	6.17	30.90
400	MAE	0.57	2.72	4.20	4.94	6.31	28.13
	MAPE	1.36	6.24	9.85	17.35	15.83	365.14
	MDAE	0.46	2.34	3.71	4.73	6.12	31.57
500	MAE	0.56	2.71	4.16	4.96	6.28	28.12
	MAPE	1.33	6.18	9.64	17.75	16.13	368.16
	MDAE	0.35	2.32	3.72	4.77	6.25	31.66

The best machine learning algorithm according to Q3 in the MAE's distribution is SVM with linear kernel using the last 7 days. Figure 2 shows the hourly predictions in a day with a MAE close to the median, this means that approximately the half of the cases will have a MAE smaller or equal to this graph. On the other hand in Figure 3 we can observe a poor prediction with a high MAE, we can observe how the MAPE is high because the real values are close to zero.

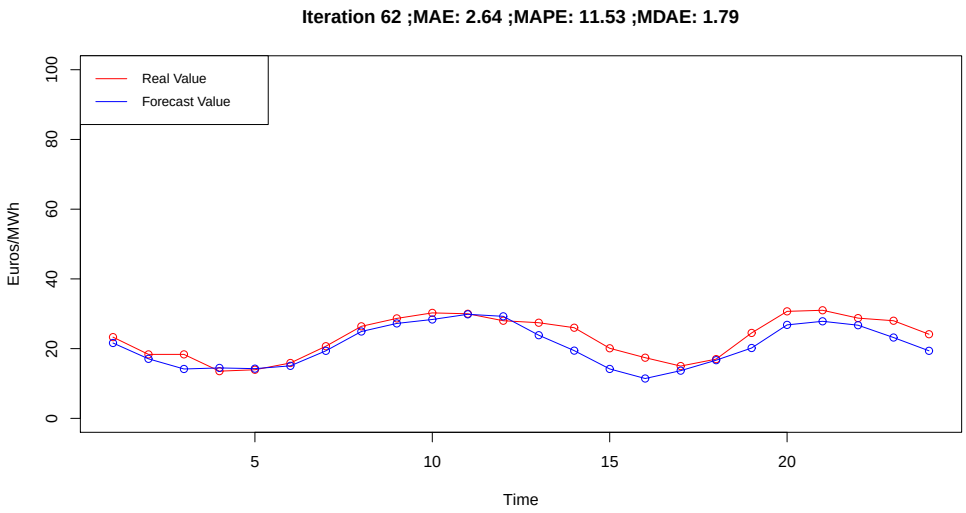

Figure 2: Real and predicted prices when a low MAE is obtained using SVM with linear kernel for the next 24 hours.

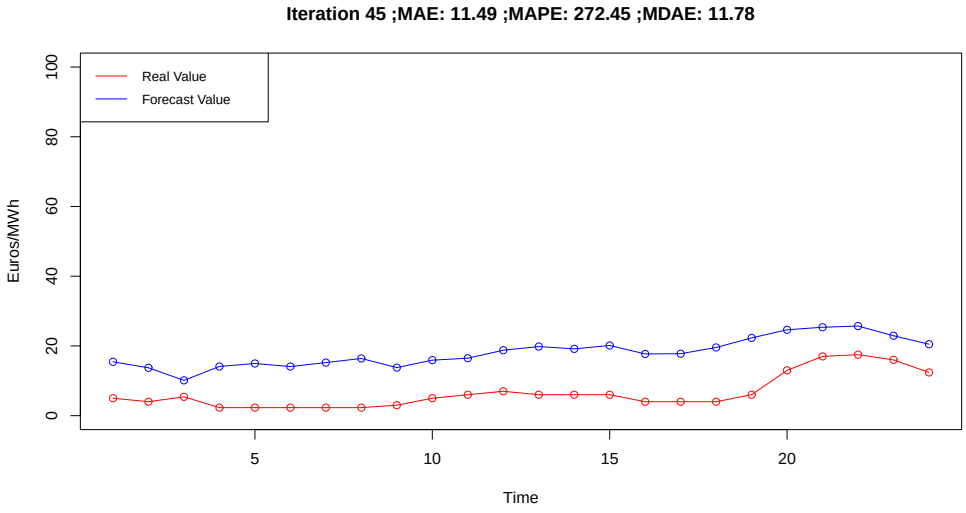

Figure 3: Real and predicted prices when a high MAE is obtained using SVM with linear kernel for the next 24 hours.

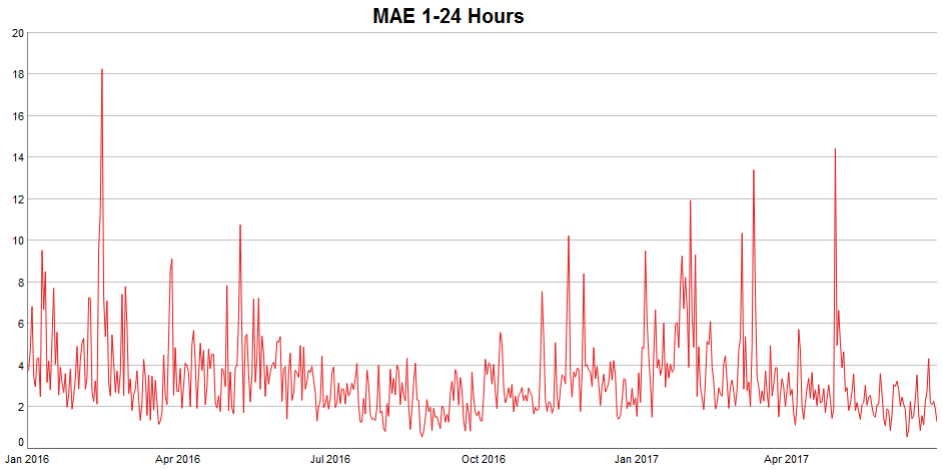

Figure 4: Daily MAE for 1-24 Hours from January 1st, 2016 to June 30th, 2017.

4.2 Forecasting 144 hours

For this exercise our data test begin January 1st, 2016 to June 30th, 2017. On the other hand, we face with the inconvenient that E-SIOS only provides the prediction for the demand of the next 168 hours, so the only variable that we can use is the Demand, obtaining the following relationship **Price** $\sim$ **Demand**. In this context, we decide to use the dynamic factor model approach that has been proven to be effective in the weak-ahead horizon. Additionally, we will combine the DFM with a simple linear regression between price and demand. We first obtain the residuals of model (36) where y_t is the price at time t and x_t is the corresponding demand. Then, we fix a DFM to the obtained residuals.

Table 9: Accuracies for DFM with r=2, 1-144 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	1.75	4.28	5.82	6.85	8.08	28.53
	MAPE	4.25	9.34	13.43	23.96	27.30	195.63
	MDAE	1.36	3.58	5.04	6.15	7.24	30.10
200	MAE	1.58	4.38	5.73	6.65	7.58	28.75
	MAPE	3.84	9.33	13.56	23.93	26.71	218.38
	MDAE	1.20	3.67	4.94	5.95	6.84	29.68
300	MAE	1.61	4.13	5.45	6.41	7.22	25.96
	MAPE	3.94	8.68	12.59	22.88	27.30	210.04
	MDAE	1.33	3.46	4.78	5.74	6.37	27.90
400	MAE	1.99	3.78	5.09	5.78	6.95	20.50
	MAPE	4.31	7.89	11.46	21.76	24.08	167.82
	MDAE	1.32	3.22	4.41	5.17	6.05	21.39
500	MAE	1.57	4.05	5.49	6.19	7.12	24.46
	MAPE	3.80	8.65	12.48	23.31	25.86	183.01
	MDAE	1.28	3.43	4.85	5.57	6.41	24.48

Table 10: Accuracies for DFM with r=3, 1-144 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	1.84	4.15	5.69	6.77	7.95	28.33
	MAPE	4.63	8.95	13.28	23.72	27.05	195.06
	MDAE	1.60	3.52	4.94	6.08	7.11	29.91
200	MAE	1.63	4.24	5.63	6.60	7.58	28.76
	MAPE	3.98	9.06	13.37	23.69	26.50	218.29
	MDAE	1.33	3.58	4.80	5.86	6.80	29.64
300	MAE	1.58	4.01	5.38	6.40	7.14	26.30
	MAPE	3.88	8.45	12.28	22.69	26.73	211.18
	MDAE	1.29	3.39	4.71	5.68	6.37	27.94
400	MAE	1.58	4.06	5.43	6.07	7.00	21.76
	MAPE	3.85	8.41	12.31	21.84	25.19	156.97
	MDAE	1.33	3.47	4.72	5.44	6.26	23.55
500	MAE	1.58	3.97	5.43	6.13	7.06	24.69
	MAPE	3.85	8.52	12.40	23.07	25.42	180.99
	MDAE	1.37	3.42	4.83	5.52	6.34	25.54

Table 11: Accuracies for LR + DFM with r=2, 1-144 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	1.95	4.08	5.47	6.33	7.62	21.80
	MAPE	3.58	8.57	12.83	21.85	26.60	177.91
	MDAE	1.40	3.43	4.82	5.78	6.92	22.09
200	MAE	2.39	4.19	5.31	6.39	7.59	22.58
	MAPE	5.58	8.82	12.32	23.50	26.36	183.20
	MDAE	1.76	3.50	4.60	5.75	6.84	23.13
300	MAE	2.11	3.75	5.02	5.80	6.91	21.24
	MAPE	4.19	7.98	11.50	21.71	24.09	170.84
	MDAE	1.61	3.20	4.32	5.19	6.12	21.55
400	MAE	1.99	3.78	5.09	5.78	6.95	20.50
	MAPE	4.31	7.89	11.50	21.76	24.08	167.82
	MDAE	1.32	3.22	4.41	5.17	6.05	21.39
500	MAE	1.97	3.65	4.96	5.73	6.68	20.61
	MAPE	4.55	7.60	11.57	21.73	24.04	165.55
	MDAE	1.49	3.09	4.30	5.13	6.09	21.53

Table 12: Accuracies for LR + DFM with r=3, 1-144 Hours.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	1.94	4.05	5.46	6.32	7.60	21.76
	MAPE	3.58	8.53	12.83	21.84	26.52	177.76
	MDAE	1.38	3.40	4.77	5.76	6.86	22.13
200	MAE	2.41	4.17	5.27	6.37	7.58	22.54
	MAPE	5.35	8.82	12.33	23.41	26.36	183.11
	MDAE	1.80	3.52	4.52	5.70	6.88	23.05
300	MAE	2.05	3.80	4.97	5.79	6.90	21.28
	MAPE	3.87	7.98	11.45	21.71	23.82	170.82
	MDAE	1.51	3.13	4.33	5.18	6.07	21.33
400	MAE	1.99	3.80	5.04	5.77	6.91	20.55
	MAPE	4.16	7.92	11.47	21.72	23.98	167.36
	MDAE	1.34	3.24	4.34	5.17	6.11	21.52
500	MAE	1.98	3.63	4.92	5.72	6.65	20.59
	MAPE	4.50	7.60	11.52	21.67	24.13	165.36
	MDAE	1.50	3.10	4.31	5.12	6.06	21.55

In this particular exercise the best model was LR + DFM with $r = 3$ using the last 500 days. Figures 5 and 6 illustrate the performance of this model in a period with low and high MAE, respectively.

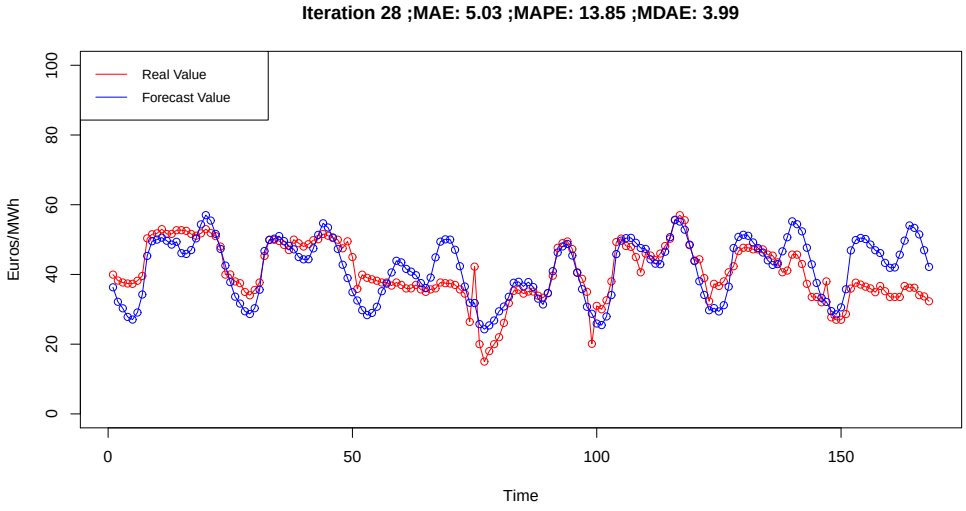

Figure 5: Real and predicted prices when a low MAE is obtained using LR+DFM with r=3 for the next 144 hours.

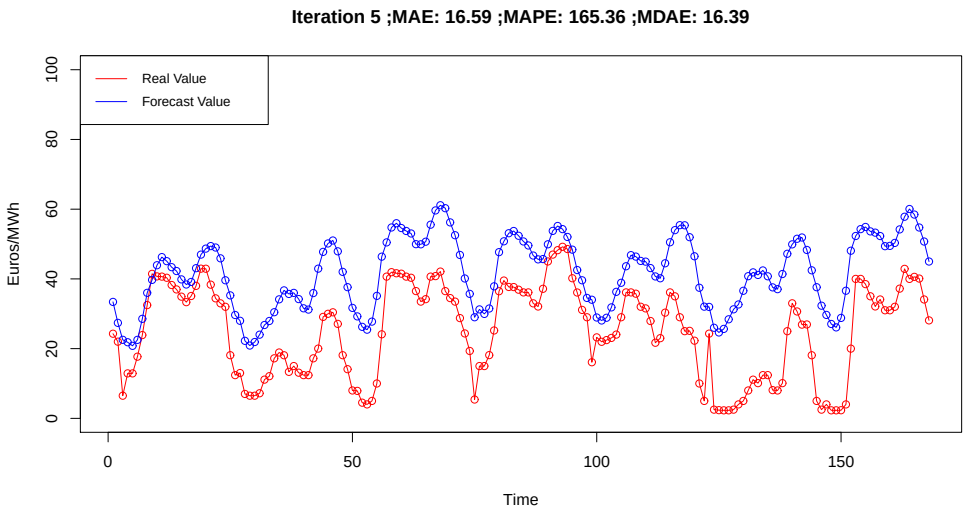

Figure 6: Real and predicted prices when a high MAE is obtained using LR+DFM with r=3 for the next 144 hours.

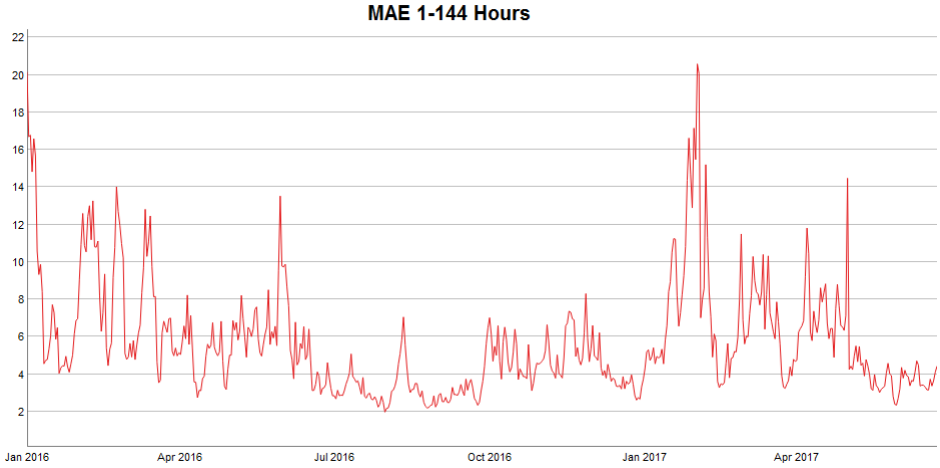

Figure 7: Daily MAE for 1-144 Hours from January 1st, 2016 to June 30th, 2017.

4.3 Forecasting 30 days

For this particular case we are going to considered the price as a time series without regressor variables. We will use the period from January 1st, 2016 until June 30th, 2017 as test set. The models studied in this case were tested using the package *forecast* without user intervention, TBATS and ARIMA models were tested in order to select the best.

Table 13: Accuracies for $ARIMA(p,1,q)(P,1,Q)_7$, 1-30 Days.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	1.45	4.56	6.73	9.53	11.51	70.03
	MAPE	3.66	9.39	18.21	28.13	37.67	234.08
	MDAE	0.87	3.71	6.13	8.99	10.78	61.02
200	MAE	1.72	4.58	6.81	9.76	11.77	109.05
	MAPE	4.25	10.33	19.02	28.21	34.42	508.52
	MDAE	1.34	3.95	6.17	9.19	10.81	86.77
300	MAE	2.25	4.56	6.00	9.37	10.86	152.05
	MAPE	4.36	9.67	16.27	27.32	29.79	703.59
	MDAE	1.55	3.80	5.58	8.85	9.66	119.58
400	MAE	2.32	4.24	5.52	8.17	8.93	177.72
	MAPE	4.59	8.65	14.06	24.12	29.87	822.49
	MDAE	1.67	3.49	5.00	7.61	8.11	138.69
500	MAE	2.37	4.40	5.67	9.55	9.56	216.26
	MAPE	5.15	9.17	14.10	29.19	30.35	1004.57
	MDAE	1.84	3.57	5.20	8.81	8.61	165.78

Table 14: Accuracies for $TBATS_7$, 1-30 Days.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
100	MAE	2.09	4.10	5.44	7.16	9.28	47.96
	MAPE	4.15	8.38	14.72	21.24	27.45	105.19
	MDAE	1.43	3.41	5.11	6.67	8.65	51.29
200	MAE	2.02	3.94	5.32	6.86	8.45	31.23
	MAPE	4.62	8.53	13.91	20.67	26.66	121.92
	MDAE	1.35	3.35	5.00	6.39	7.65	35.83
300	MAE	2.45	4.07	5.36	6.88	8.68	30.21
	MAPE	4.47	8.59	13.76	20.72	26.93	124.27
	MDAE	1.74	3.40	4.97	6.41	8.13	34.50
400	MAE	2.30	3.88	5.28	6.80	8.72	25.70
	MAPE	4.50	8.29	13.96	21.16	26.80	132.87
	MDAE	1.37	3.31	4.87	6.29	8.05	27.47
500	MAE	2.18	4.14	5.39	6.75	8.41	25.33
	MAPE	4.34	8.57	13.82	20.79	26.44	109.32
	MDAE	1.23	3.47	4.87	6.20	7.85	27.65

Although we only consider one seasonality value to the TBATS model, we believe that this approach is very interesting, in fact we didn't find any kind of information about EPF using TBATS models and it is a model that outperforms traditional approach. The best model is $TBATS_7$ using the 500 days as training window size. Figures 8 and 9 illustrate the performance of this model in a period with low and high MAE, respectively.

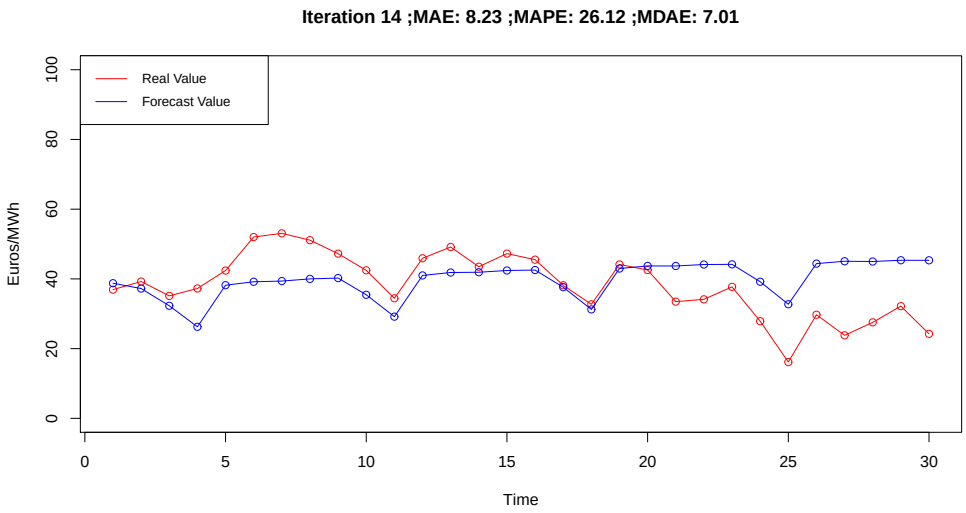

Figure 8: Real and predicted prices when a low MAE is obtained using $TBATS_7$ for the next 30 days.

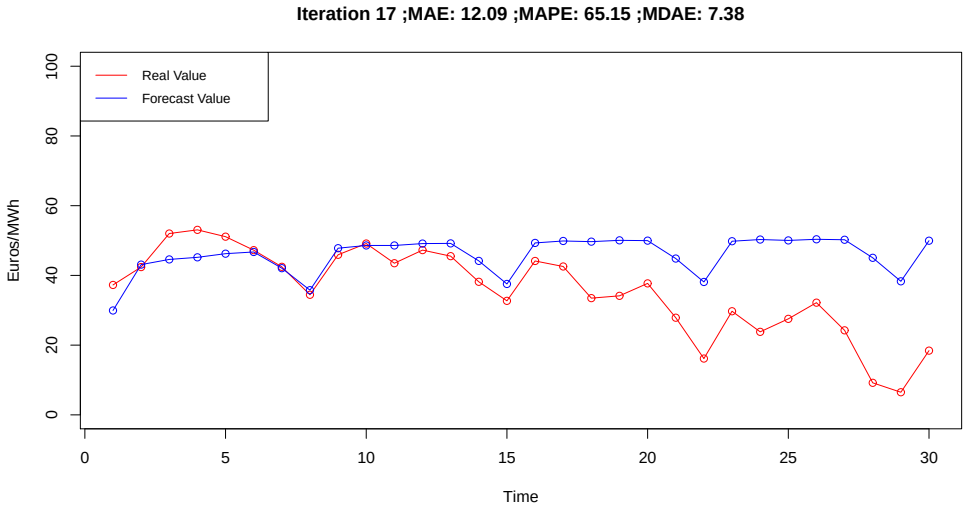

Figure 9: Real and predicted prices when a high MAE is obtained using $TBATS_7$ for the next 30 days.

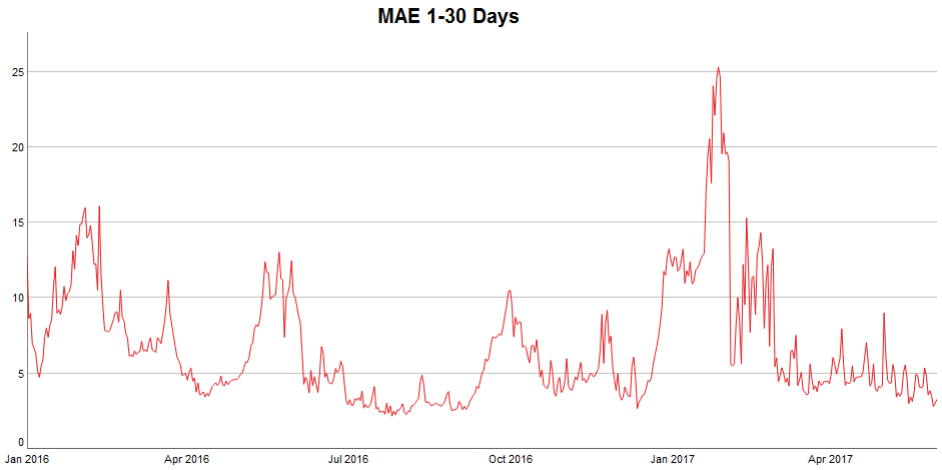

Figure 10: Daily MAE for 1-30 Days from January 1st, 2016 to June 30th, 2017.

4.4 Forecasting 12 months

In order to forecast the following 12 months by months was selected as test period the dates from January, 2010 to June, 2017. We used different training windows sizes (Table 15 and 16).

Table 15: Accuracies for $ARIMA(p,1,q)(P,1,Q)_{12}$, 1-12 Months.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
60	MAE	2.47	7.49	10.18	10.75	13.21	30.00
	MAPE	4.79	18.01	26.14	27.15	35.30	65.30
	MDAE	1.88	5.39	9.88	10.14	13.67	31.63
80	MAE	2.75	6.31	8.75	9.86	12.27	29.64
	MAPE	5.38	15.45	23.44	24.96	30.82	68.36
	MDAE	2.24	5.12	7.49	9.22	12.97	30.86
100	MAE	2.93	6.39	8.87	9.69	12.12	29.57
	MAPE	5.91	14.62	23.08	24.89	32.69	67.21
	MDAE	1.71	4.64	7.56	8.91	11.41	30.72
120	MAE	2.68	5.42	8.65	9.29	11.74	29.47
	MAPE	5.23	11.97	22.25	23.81	31.50	69.02
	MDAE	2.41	4.11	6.40	8.53	11.84	30.54
140	MAE	2.30	5.55	8.42	9.30	11.36	29.49
	MAPE	4.51	13.37	21.95	24.12	32.51	69.10
	MDAE	1.81	4.52	7.15	8.61	11.54	30.72

Table 16: Accuracies for $ARIMA(p,1,q)(P,1,Q)_{12}(Auto\_BoxCox)$, 1-12 Months.

Days	Measure	Min	Q1	Median	Mean	Q3	Max
60	MAE	3.73	7.18	10.06	11.21	14.12	31.35
	MAPE	7.25	19.15	26.40	28.28	35.20	70.26
	MDAE	3.51	5.35	9.23	10.62	14.78	32.28
80	MAE	2.97	6.04	9.35	9.94	12.29	30.62
	MAPE	5.73	14.08	25.66	25.54	34.44	69.24
	MDAE	2.14	4.99	7.24	9.32	12.68	31.25
100	MAE	2.38	5.61	8.54	9.22	11.33	31.46
	MAPE	4.64	13.05	22.75	23.96	32.44	70.25
	MDAE	2.11	4.30	6.47	8.40	11.11	32.71
120	MAE	2.69	5.38	8.12	9.05	10.92	30.26
	MAPE	5.24	11.68	21.85	23.50	32.66	69.71
	MDAE	2.07	3.98	6.74	8.25	10.44	31.98
140	MAE	2.53	5.36	8.23	9.22	10.90	30.00
	MAPE	4.85	12.10	20.41	24.05	33.69	70.06
	MDAE	1.76	3.86	7.08	8.41	11.02	31.68

The best model is $ARIMA$ with a BoxCox transformation using Guerrero's method [16]. This model used the last 140 months as training windows size. Using this model we can observe two cases one with good accuracy (Figure: 11) and another with bad accuracy (Figure: 12).

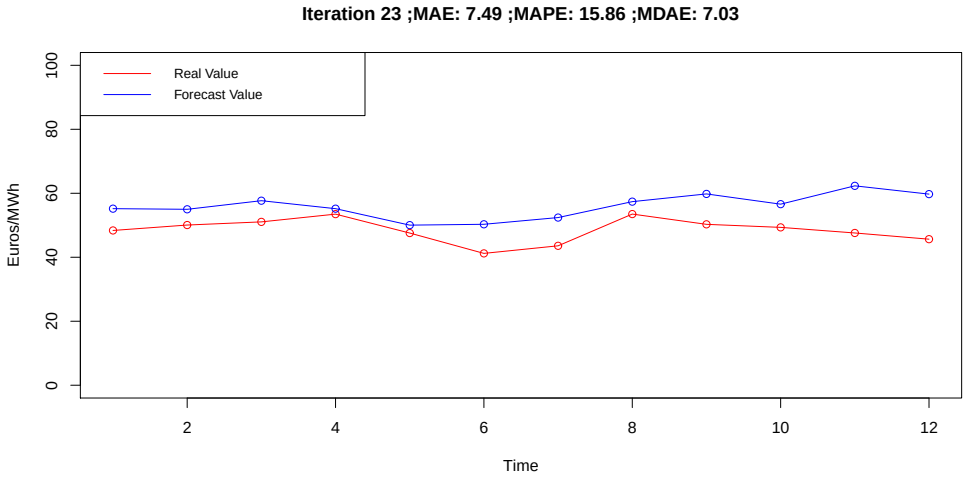

Figure 11: Real and predicted prices when a low MAE is obtained using $ARIMA$ for the next 12 months.

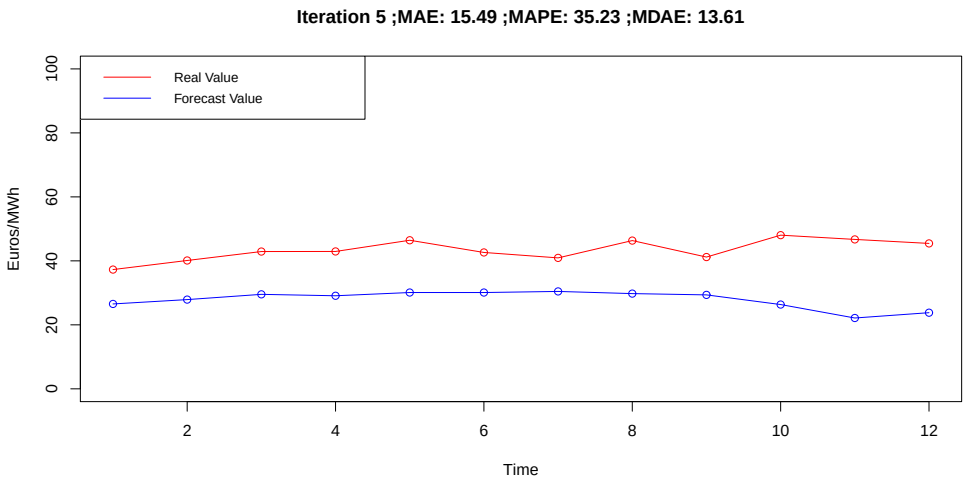

Figure 12: Real and predicted prices when a high MAE is obtained using $ARIMA$ for the next 12 months.

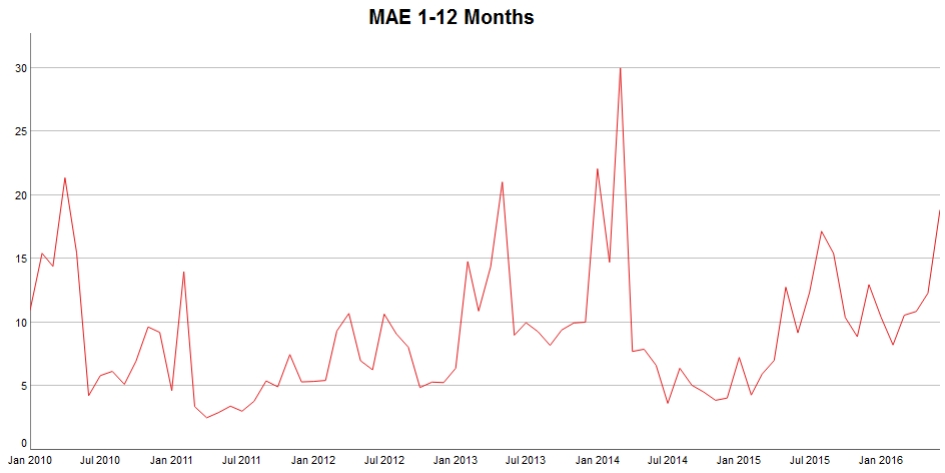

Figure 13: Monthly MAE for 1-12 Months from January, 2010 to June, 2017.

5 Automatization and Visualization

In order to implement a public access to the developed forecasting procedures, we will need a server that allows frequent running of our R'scripts. We will use a server in the cloud provided by Amazon Web Services (AWS).

Thanks to funding from Universidad Carlos III de Madrid we were able to create an account at AWS and to select an instance with RStudio + Julia server (experimental) for statistical computation, that can be accessed by public DNS in a web browser (standard port 80). Using this instance, we created a web page in order to visualize and download the forecasted prices with their prediction intervals. This web page was done using *shiny*. The instance used in particular is free for 12 months.

We can access to the web page using the link `http://35.157.17.182/shiny/rstudio/sample-apps/Forecast/`. If there is any problem you can send an e-mail to the following address alfranco@est-econ.uc3m.es with the following subject *Can't access to Forecast/AWS*, because the IP address can change if the instance that was created has a crash. Using the package *cronR* we run our scripts automatically without user intervention, this package was developed for Ubuntu but for Windows there exists a version called *schedule*.

In the app we can observe and download in a *csv* file the electricity price forecast and the prediction interval for the following 24 hours, 144 hours, 30 days and 12 months. The app is very easy and intuitive of use as can be observed in Figure 14, where in the drop-down list denoted by 1 it can be selected the period of interest; in the button denoted by 2 we can download the forecast values and in menu option denoted by 3 the contact information of the App's developers is available. Also, Figure 14 illustrates the forecast and the prediction intervals for date January 31th, 2017.

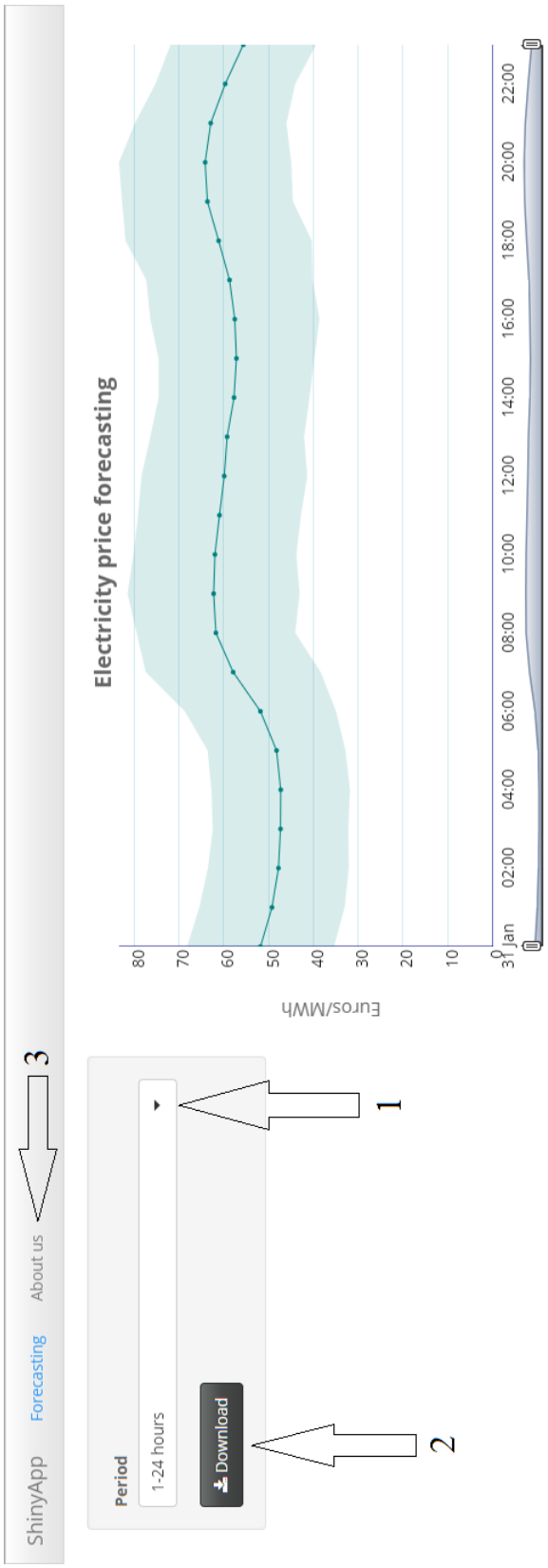

Figure 14: Capture of the app developed.

6 Conclusions and future research lines

The conclusions from this thesis are the following:

- After running all the models for predicting from one to 144 hours ahead and compared them using different kind of accuracies measure; we conclude that adding explicative variables will improve the accuracy of the models. At the end, the price depends on a lot of variables, the limitation of using this idea is that when we add variables as regressors we should know the forecast values of them in order to forecast the electricity price.
- For forecasting from one to 24 hours ahead, the SVM with linear kernel using the last 7 days is the best model. The hyper parameter tuning was done by a grid-search and for estimating the performance of the model cross-validation (with $K = 10$) was used.
- It is very important to remark that using few days was better than using longer training sets. Something that goes against what is believed of machine learning that the more data you have is better.
- For forecasting from one to 144 hours ahead, the combination of DFM and linear regression was a little better than only using DFM.
- For forecasting from one to 30 days ahead, the TBATS model show better results than using a seasonal ARIMA model, even when was considered only one seasonality in TBATS.
- For forecasting up to 12 months, an ARIMA with Box-Cox transformation obtain the better results.

Some future lines of research that we plan to pursue are the following:

- For forecasting up to 24 hours we can add more explanatory variables like hydraulic variables (amount of rain and capability of the reservoirs). We guess that adding this kind of variables will improve the accuracy of the models.
- Try to apply more models in order to forecast up to 30 days and up to 12 months and compare the results.
- It could be interesting to do cluster (unsupervised learning) in order to group similar periods and, after that, to train and test different models for each cluster.

7 Acknowledgments

I want to thank my advisors Francisco Javier Nogales Martín and Carlos Ruiz Mora for his invaluable guidance and support in the making of this project. Also to Andrés M. Alonso for his helpful suggestions to earlier versions of the thesis. Finally, I would like to thank my family and friends for their continuous support.

8 Appendix: Thesis Work Schedule

In the next table will show the time invested in order to complete this thesis.

Table 17: Thesis Work Schedule

Task	Duration	Start	End
Obtaining the token	4 days	14-4-2017	17-4-2017
Understanding how does it work the API	8 days	20-4-2017	27-4-2017
Studying interest articles	61 days	1-5-2017	30-6-2017
Creating the instance	10 days	1-9-2017	10-9-2017
Creating the Shiny App	11 days	11-9-2017	21-9-2017
Writing and revision of the thesis	57 days	22-9-2017	17-11-2017
Testing the models	123 days	1-7-2017	31-10-2017

References

[1] About our market. http://www.omie.es/en/home/markets-and-products/about-our-market. Accessed: 2017-08-19.

[2] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 6 (December 1974), 716–723.

[3] Alonso, A. M., Rodríguez, J., García-Martos, C., and Sánchez, M. J. Seasonal dynamic factor analysis and bootstrap inference: Application to electricity market forecasting. *Technometrics 53*, 2 (2011), 137–151.

[4] Anbazhagan, S., and Kumarappan, N. Day-ahead deregulated electricity market price forecasting using recurrent neural network. *Engineering Applications of Artificial Intelligence 26*, 7 (2013), 866–872.

[5] Azadeh, A., M., M., Mahdi, M., and Seyedmahmoudi, S. H. Optimum long-term electricity price forecasting in noisy and complex environments. *Energy Sources, Part B: Economics, Planning, and Policy 8*, 3 (2013), 235–244.

[6] Benth, F., Benth, J., and Koekebakker, S. *Stochastic Modelling of Electricity and Related Markets*. Advanced series on statistical science and applied probability. World Scientific, 2008.

[7] Brockwell, P., and Davis, R. *Introduction to Time Series and Forecasting*. Springer, 2002.

[8] Conejo, A. J., Contreras, J., Espínola, R., and Plazas, M. A. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting 21*, 3 (2005), 435–462.

[9] Crespo Cuaresma, J., Hlouskova, J., Kossmeier, S., and Obersteiner, M. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy 77*, 1 (2004), 87–106.

[10] De Livera, A. M., Hyndman, R. J., , and Snyder, R. D. Forecasting time series with complex seasonal patterns using exponential smoothing, October 2010. https://robjhyndman.com/papers/ComplexSeasonality.pdf.

[11] Elattar, E. E. Day-ahead price forecasting of electricity markets based on local informative vector machine. *IET Generation, Transmission and Distribution 7* (October 2013), 1063–1071.

[12] García-Martos, C., Rodríguez, J., and Sánchez, M. Forecasting electricity prices by extracting dynamic common factors: application to the Iberian market. *IET Generation, Transmission and Distribution 6*, 1 (January 2012), 11–20.

[13] Geweke, J. *The Dynamic Factor Analysis of Economic Time Series Models*. Social Systems Research Institute, University of Wisconsin-Madison, 1978.

[14] Gjolberg, O., and Brattested, T.-L. The biased short-term futures price at nord pool: can it really be a risk premium? *Journal of Energy Markets 4*, 1 (2011), 3–19.

[15] Gómez, V., and Maravall, A. Programs tramo and seats, instruction for user (beta version: september 1996). Working papers, Banco de España, 1996.

[16] Guerrero, V. M. Time-series analysis supported by power transformations. *Journal of Forecasting 12*, 1 (1993), 37–48.

[17] Hurvich, C. M., and Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika 76*, 2 (1989), 297–307.

[18] Hyndman, R., and Athanasopoulos, G. *Forecasting: principles and practice:*. OTexts, 2014.

[19] Hyndman, R. J. The difference between prediction intervals and confidence intervals. `https://robjhyndman.com/hyndsight/intervals/`, 2013. Accessed: 2017-08-20.

[20] Hyndman, R. J., and Athanasopoulos, G. *Forecasting: principles and practice*. OTexts, 2014.

[21] Hyndman, R. J., and Khandakar, Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software 27*, 3 (July 2008).

[22] Hyndman, R. J., and Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting 22*, 4 (2006), 679–688.

[23] Knittel, C., and Roberts, M. An empirical examination of restructured electricity prices. *Energy Economics 27*, 5 (2005), 791–817.

[24] Koritarov, V. S. Real-world market representation with agents. *IEEE Power and Energy Magazine 2*, 4 (2004), 39–46.

[25] Kosater, P., and Mosler, K. Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. *Applied Energy 83*, 9 (2006), 943–958.

[26] Kuhn, M. Building predictive models in R using the caret package. *Journal of Statistical Software 28*, 5 (2008), 1–26.

[27] Lee, R. D., and Carter, L. R. Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association 87*, 419 (1992), 659–671.

[28] León-Salas Bujalance, C. El mercado eléctrico español compra de energía eléctrica por parte de un consumidor directo, 2014.

[29] Liddle, A. R. Information criteria for astrophysical model selection. *Monthly Notices of the Royal Astronomical Society: Letters 377*, 1 (2007), L74–L78.

[30] Marie, B., and Othman, B. What causes the forecasting failure of Markov-switching models? A Monte Carlo study. *Studies in Nonlinear Dynamics and Econometrics 9*, 2 (June 2005), 1–24.

[31] Misiorek, A., and Weron, R. Interval forecasting of spot electricity prices. HSC research reports, Hugo Steinhaus Center, Wroclaw University of Technology, 2006.

[32] Murphy, K. P. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[33] Nogales, F. J., Contreras, J., Conejo, A. J., and Espínola, R. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems 17*, 2 (May 2002).

[34] Ord, J. K., Koehler, A. B., and Snyder, R. D. Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association 92*, 440 (1997), 1621–1629.

[35] Pena, D., and Box, G. E. Identifying a simplifying structure in time series. *Journal of the American Statistical Association 82*, 399 (1987), 836–843.

[36] Peña, D., and Poncela, P. Forecasting with nonstationary dynamic factor models. *Journal of Econometrics 119*, 2 (2004), 291–321.

[37] Rose, K. The impact of fuel costs on electric power prices. *American Public Power Association* (2007).

[38] Sakia, R. M. The Box-Cox transformation technique: A review. *Journal of the Royal Statistical Society. Series D (The Statistician) 41*, 2 (1992), 169–178.

[39] Schmalensee, R., Bulovic, V., Armstrong, R., Batlle, C., Brown, P., Deutch, J., Jacoby, H., Jaffe, R., Jean, J., Miller, R., O'sullivan, F., Parsons, J., Pérez-Arriaga, J. I., Seifkar, N., Stoner, R., and Vergara, C. The future of solar energy an interdisciplinary MIT study. Tech. rep., Energy Initiative Massachusetts Institute of Technology, 2015. `https://energy.mit.edu/wp-content/uploads/2015/05/MITEI-The-Future-of-Solar-Energy.pdf`.

[40] Sharma, V., and Srinivasan, D. A hybrid intelligent model based on recurrent neural networks and excitable dynamics for price prediction in deregulated electricity market. *Engineering Applications of Artificial Intelligence 26*, 5-6 (2013), 1562–1574.

[41] Singh, N., and Mohanty, S. A review of price forecasting problem and techniques in deregulated electricity markets. *Journal of Power and Energy Engineering* (2015), 1–19.

[42] Snyder, R. D. *Estimation of a dynamic linear model : another approach / R.D. Snyder*. Monash University, Department of Econometrics and Operations Research Clayton, Vic, 1985.

[43] Stigler, S. M. Gauss and the invention of least squares. *The Annals of Statistics 9* (1981), 465–474.

[44] Taylor, J. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society 54*, 8 (August 2003), 799–805.

[45] Termini, V., and Cavallo, L. Spot, bilateral and futures trading in electricity markets. implications for stability. Tech. rep., Nota di Lavoro, Fondazione Eni Enrico Mattei, 2007.

[46] Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.

[47] W. J. Granger, C. *Forecasting in Business and Economics*. Economic Theory, Econometrics and Mathematical Economics Series. Academic Press, 1989.

[48] Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Wiley, Chichester, 2006.

[49] Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting 30* (2014), 1030–1081.

[50] West, M., and Harrison, J. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag New York, Inc., 1997.