Modelling Week

Forecasting prices in electricity markets

Problem proposed by INDIZEN
V Modelling Week
Complutense University of Madrid

Students
Carlos González Salcedo
Diana Edith GüereLettich
Natalie Keating
Ignacio Martínez Calvo
CharmiDevendraPanchal
Víctor Puente García

Instructors
Juan Tejada (Complutense University of Madrid)
Hugo Marrao (INDIZEN)

Contents

1. Introduction to the problem.

Electricity is one of the most important goods for our society. Forecasting electricity prices at different time frames is very important for all industry stakeholders for cash flow analysis, capital budgeting and financial procurement as well as regulatory rule-making and integrated resource planning, among others. All factors determining the price can be classified as endogenous or exogenous to the market bring about uncertainty and volatility to the electricity prices.

The most important outcome of an electricity market is the formation of a price at which all power is traded, at least on a daily basis, by way of the so-called 'Spot' market. The daily (spot) electricity market serves as a marketplace of last resort for generators and demands to trade their remaining available not-contracted power. Almost all spot electricity markets currently under operation have implemented a mandatory day-ahead bidding framework, which may or may not be complemented with intra-day and real-time (balancing) markets. In terms of the amount of energy being traded, the day-ahead market is the most significant one among all spot (intraday or real-time) markets.

As a matter of fact, the economics of the whole electricity industry depends on a great deal on the electricity prices cleared at the market. In the short- run (from three to 24 hours) electricity price forecasting is especially important in electricity markets in which participants must optimize their positions (bidding price and quantity for the various markets, namely day-ahead and intraday) based on their perception of what the future hourly prices and incremental costs will be over the bidding period. Moreover, some agents, especially large consumers with self-power production, are able to decide which portion of their consumption is to be supplied by the market or by their own production and the corresponding timing. Nonetheless, the driving force behind the decision-making of all market participants is the maximization of profit.

Forecasting electricity prices is a challenging task not only because the prices are uncertain but, most importantly, because of the particularities of how these prices are brought into being. The process of price formation in electricity markets follows in essence the basic rule of microeconomic theory (Law of Supply and Demand) by which the price of the underlying commodity in a competitive market should reflect the relative scarcity of the supply for a given demand level. If 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. This process results in relatively low equilibrium prices. On the other hand, as the demand increases, those suppliers with the lowest incremental costs are the first ones to enter the market and use up their production capacity so more and more expensive suppliers have to come in to supply the increasingly scarce commodity, rising the equilibrium price.

This process is observed in electricity markets on a regular basis. The market clearing prices tend to follow closely the daily and seasonal swings in consumption. If consumption and price were determined by a one-to one deterministic relationship, anticipating the electricity prices would boil down to forecasting accurately the demand, which is one of the most investigated problems in powers systems operation and planning. The influence of the demand on the electricity prices is, however, far from being deterministic. There are a series of factors that bring about uncertainty to the price formation process even if the demand is known with certainty. One of the known factors that play a special role in electricity prices are the renewable energies. For example, in Spain the wind power is the most important factor to know electricity prices.

The aim of this work is the development of a prediction model of the hourly generation program and price of the Spanish power market. We also study Weighted Nearest Neighbor methods (WNN).

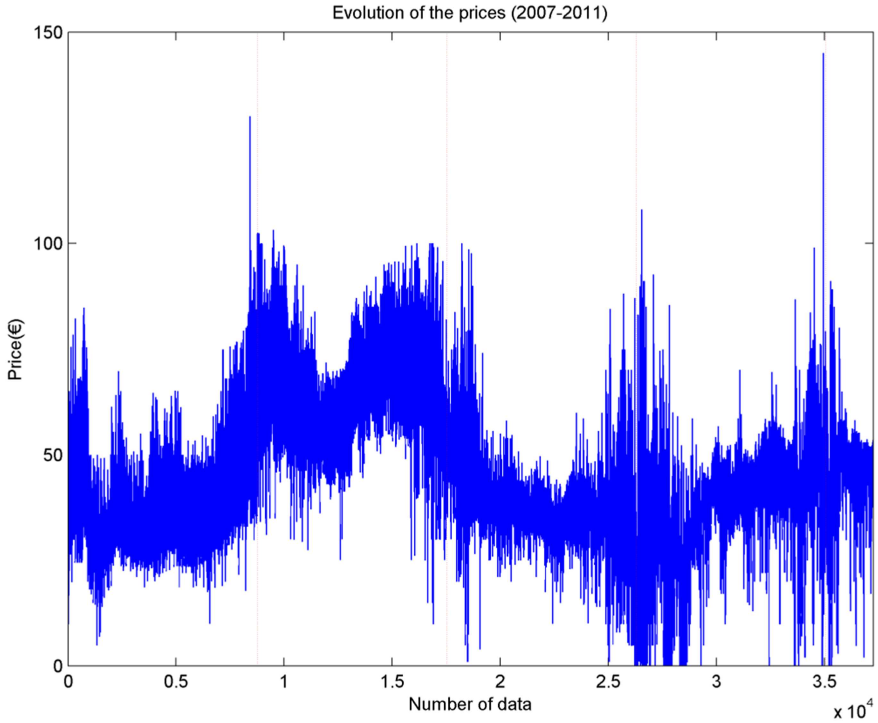
2. Exploratory analysis.

Data available comprise following variables:

- Electricity price.
- Temperature.
- Total demand of energy.
- Energy produced classified by technology.

These data are available from January 2007 to May 2011. There is a data per hour.

First, a graphic of the evolution of the prices was made. It shows a high volatility in its behavior as shown in the figure below.



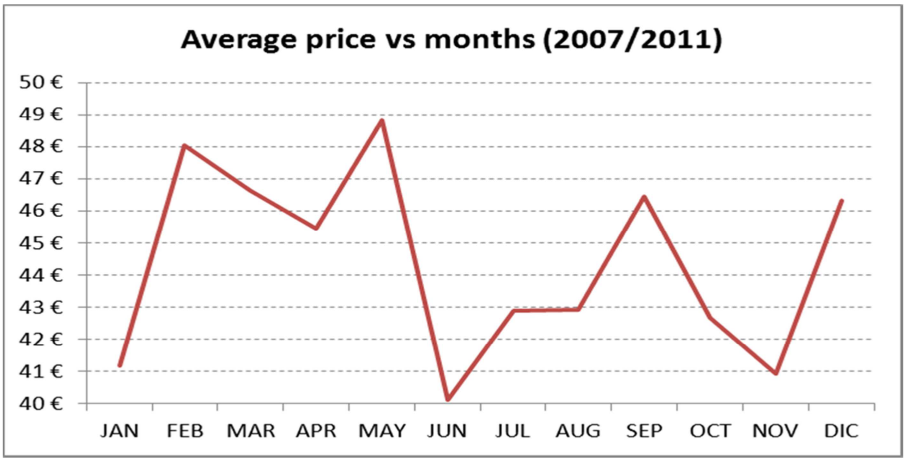
In order to understand the behavior of prices, a statistic analysis was carried out. The table below shows the summary of statistics.

Years	2007	2008	2009	2010	2011
Mean	39,35 €	64,43 €	36,96 €	37,01 €	46,01 €
Std. deviation	13,19 €	12,85 €	9,55 €	14,70 €	10,06 €
Maximum	130,00 €	103,15 €	100,00 €	145,00 €	91,01 €
Minimum	5,00 €	10,00 €	0,00 €	0,00 €	0,00 €

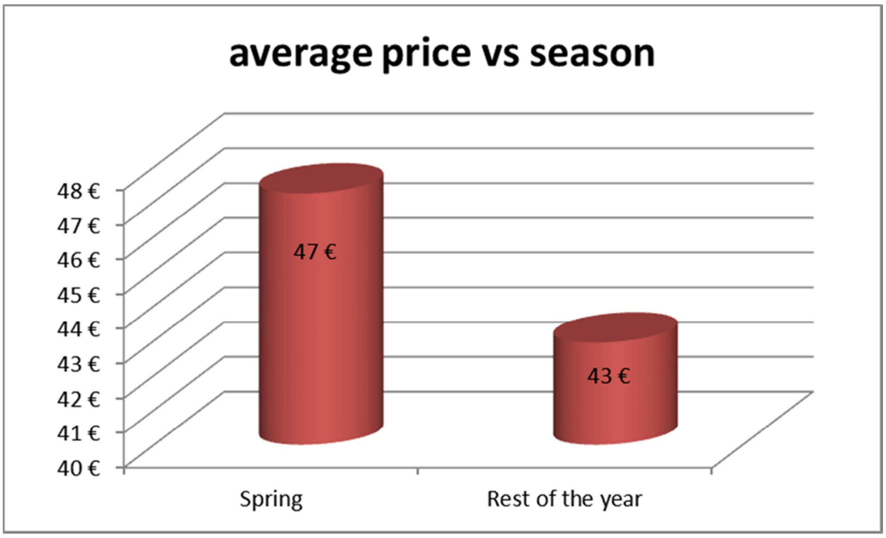
It is important to highlight that there are zero values in the price. This must be taken into account in case a transformation of the series is necessary to make it stationary.

We make different plots to represent the distribution of the variables during the last four years:

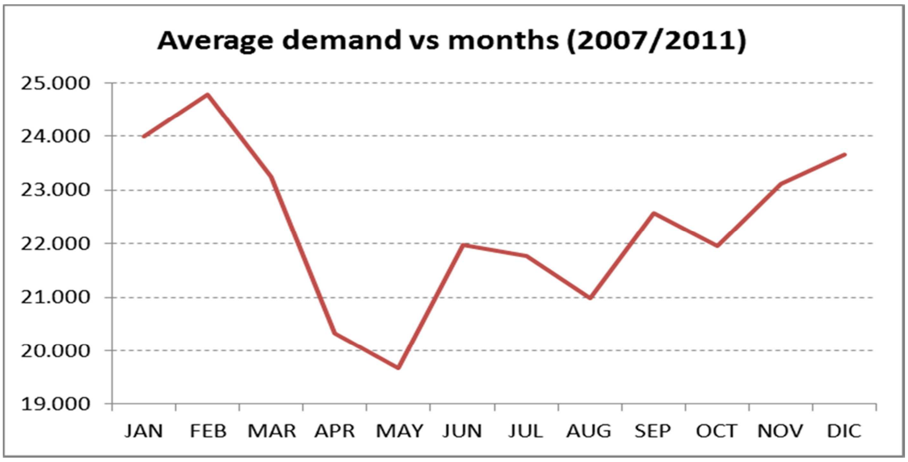
Here we have a graphic that represents the distribution of the average of the prices for the different months during the last four years. The goal is to recognize some differences in the electricity price between months or seasons.



As we can see there is a big difference in the price between February, March, April and May(high prices) and the rest of the year. For this reason we consider to create a new variable call "Season" that classifies each observation in "Spring" and "Rest of the year".

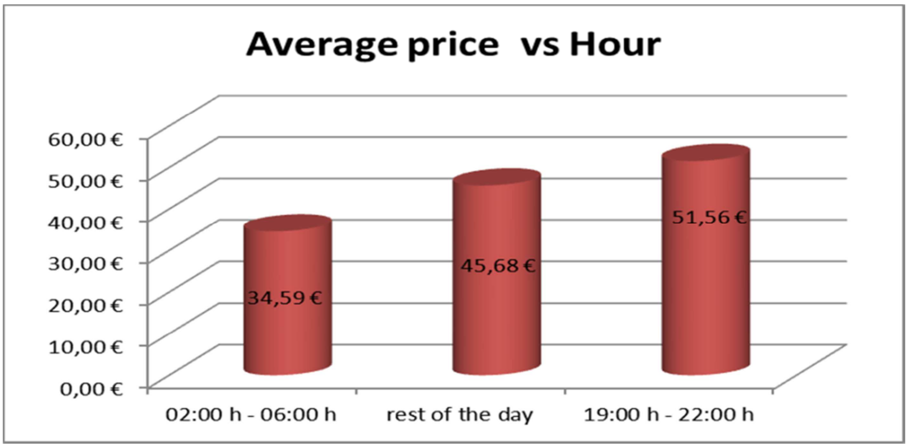


In the same way we create a graphic with the average of the demand per month during the last four years.



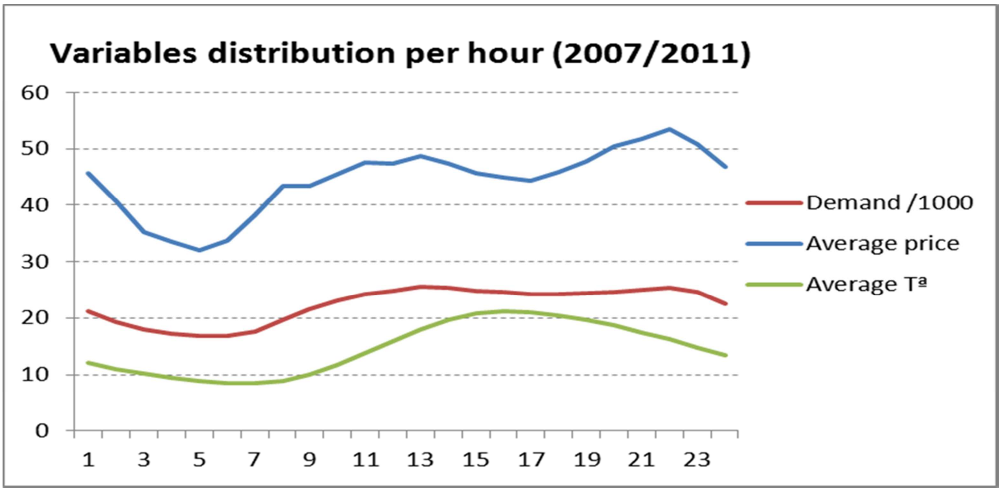
As we expected the demand is higher in Winter than in the rest of the year because the consume of electricity is higher.

To see if there are differences between the different periods of a day we create this graphic:



Taking in account this plot we create a new variable called “period of the day” that classifies the day in three periods. As we can see in the plot the period that goes from 19:00 to 22:00 are the ones with higher prices because is the moment of the day with a higher consume.

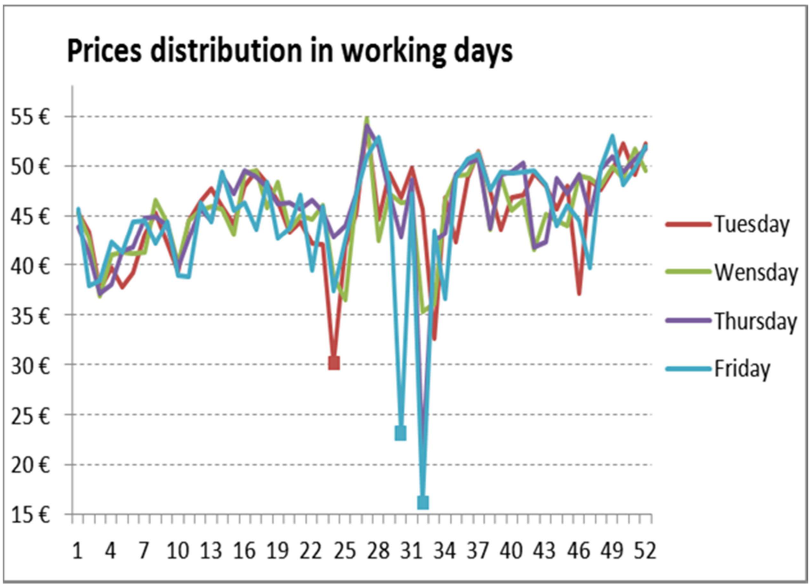
Other part of the analysis consist in create the distribution of the main variables for the 24 hours.



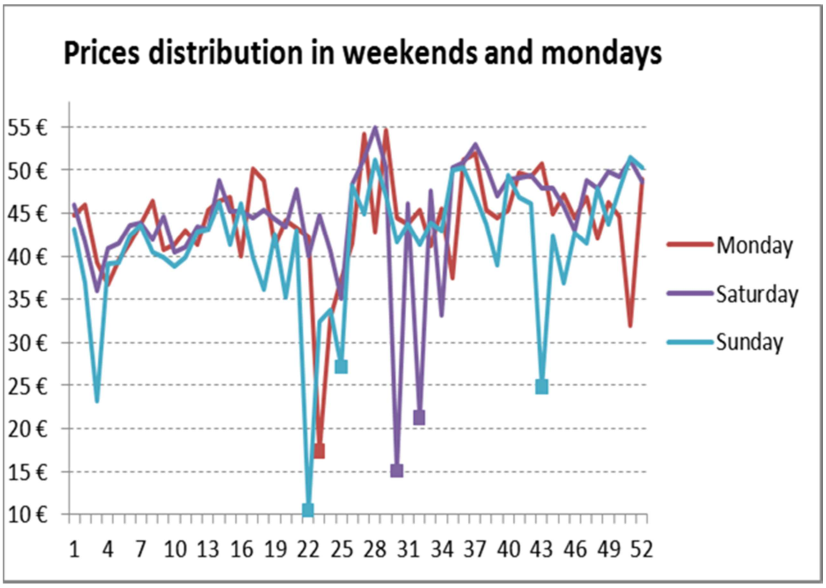
It´s clear that the highest prices are taken from 19:00 h to 23:00 h and that there is a high correlation between temperature and demand of electricity.

In the next plot we identify the outliers of the prices distribution. The horizontal axis represents all the weeks form 1st January to the 31of May

First table we represent the average of the prices of the working days during the last four years



In the same manner we represent the average of the prices of weekends and Mondays during the last four years.

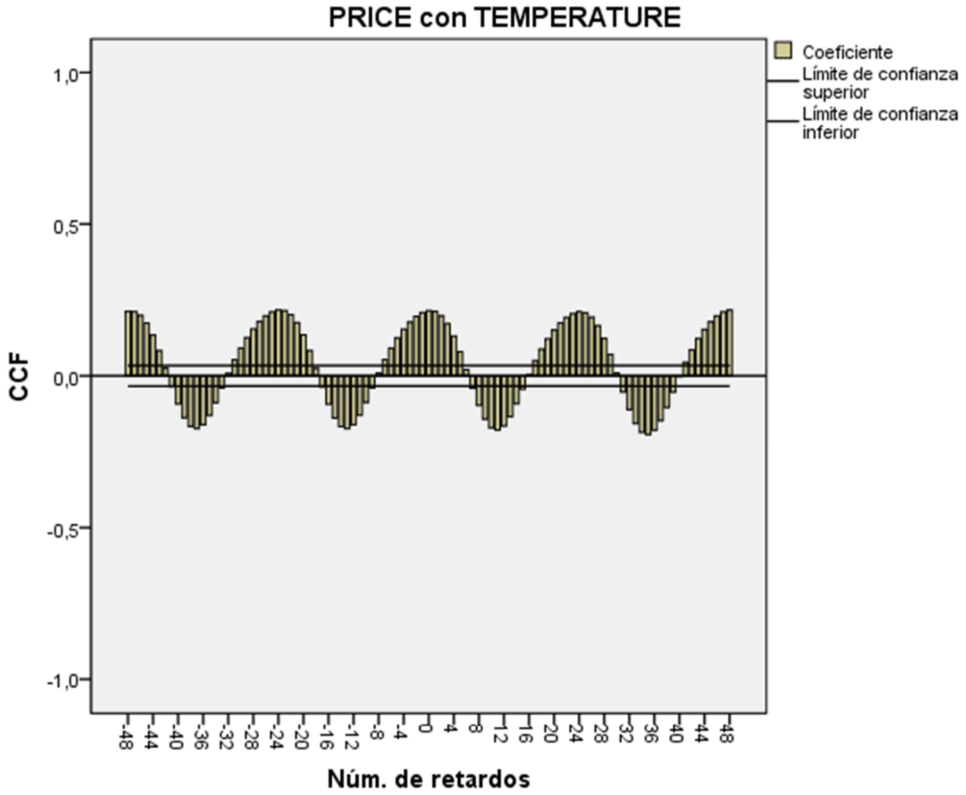


The outliers of both graphics correspond to several Holidays in Spain (Christmas day, Easter,etc…)For this reason during these days there is a lower consume of electricity and for instance lower prices

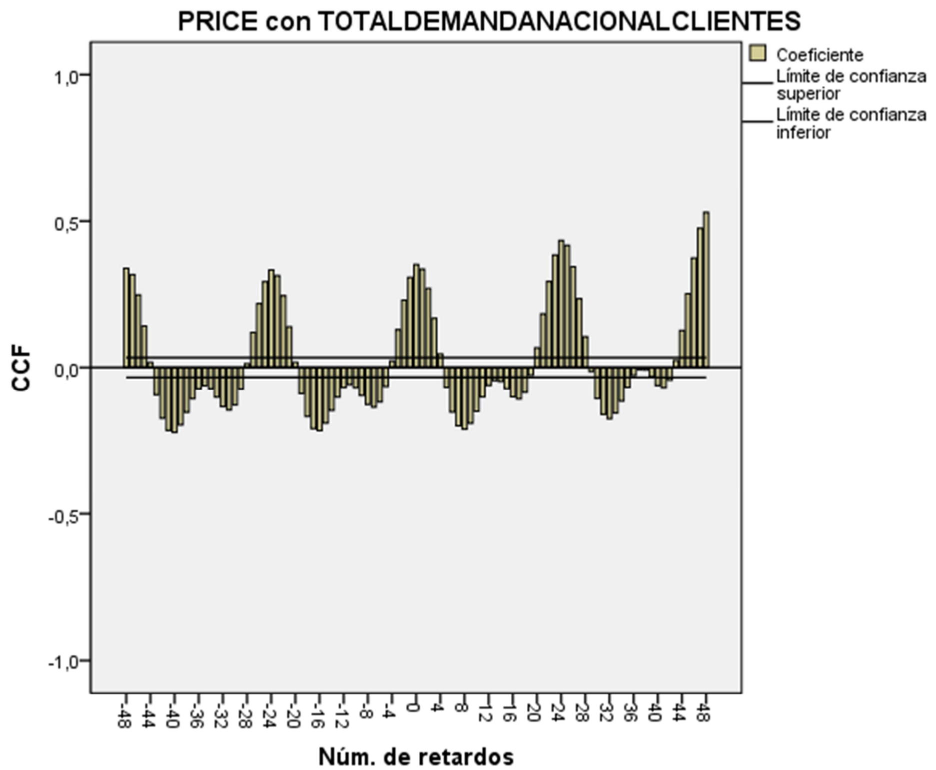
Because of the huge amount of data, we focused on one year data. Training dataset, used to build the model, comprise data from 31st of January 2010 to 29th of May 2011. Validation dataset comprise two weeks from 30th of May to 12th of June.

Once the data were chosen, we searched for correlations between the price and other variables. Cross-correlation graphics were plotted and correlation coefficients were computed.

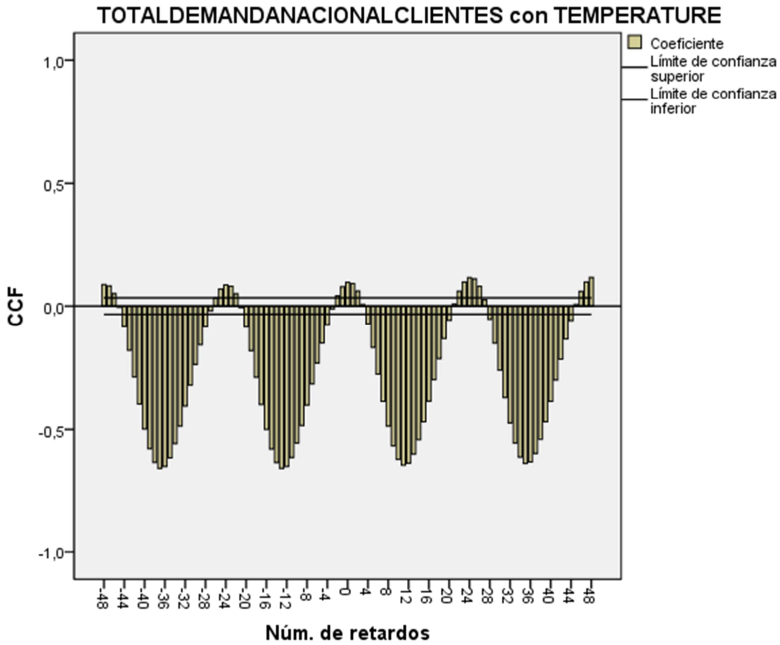
Regarding correlation between price and temperature, there is a significant correlation as shown in the picture below.



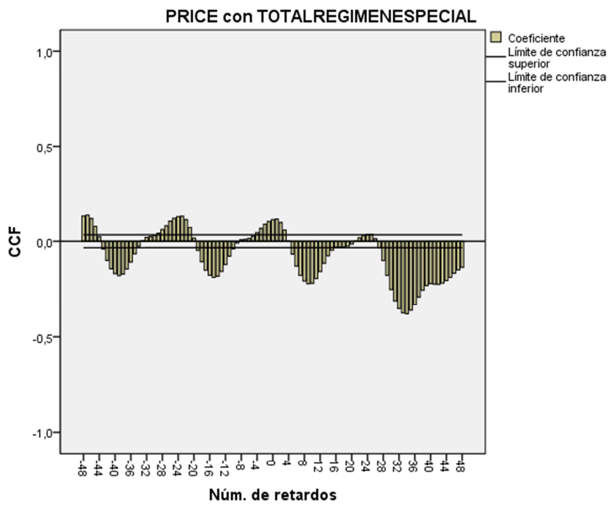
Something similar happened to total demand, but in this case the correlation is higher than in the previous case.

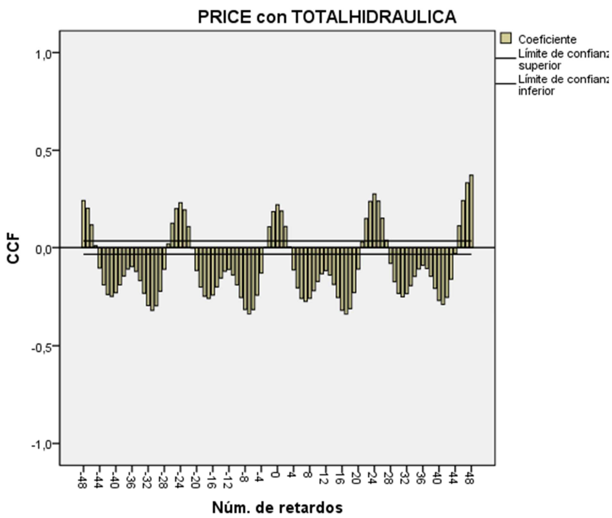


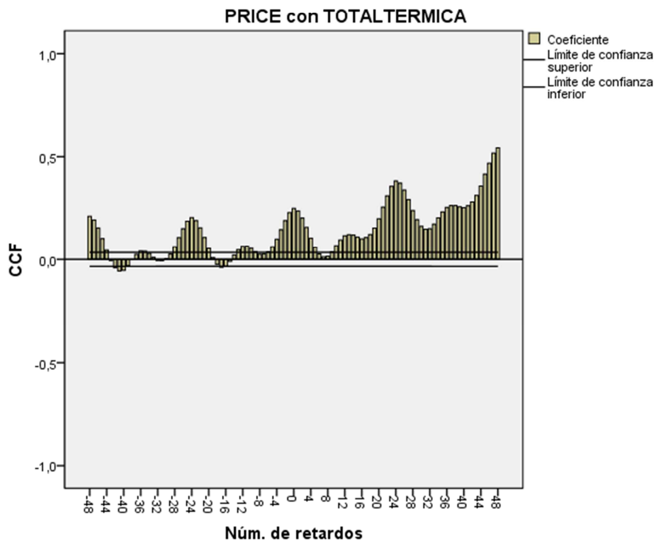
As expected, temperature and demand are highly correlated.



Concerning correlation between prices and sources of energy, cross correlation graphics are shown below. Prices are supposed to be correlated to wind production, which is the first image, but in this case it does not show a clear relation.







To conclude this part of the task, it may be convenient to use total demand as external variable in building prediction model, because it is a well-predictable variable by Red Electrica of Spain.

3. Prediction model building.

In order to build the prediction model, we have used two different methods:

- The well-known ARIMA technique.
- The K-Nearest Neighbors model which will be developed later.

We used as training data prices from one year, from May 30th, 2010, to May, 29th, 2011. For assess the prediction, we used the two first weeks of June, 2011.

To assess the prediction performance of the models, different statistical measures can be utilized. In this work, we used the following measures (Conejo et al. 2005; Weron 2006):

- Mean Daily Error:

The absolute error can be normalized by the average price attained duringthe day.

$$MDE = \frac{1}{24}\sum_{h=1}^{24} \frac{\left|P_h - \widehat{P_h}\right|}{\overline{P_h}}$$

Where P_h is the actual price at time h, $\widehat{P_h}$ is the forecast price, and $\overline{P_h}$ is the average of the actual prices.

- Mean Weekly Error:

It corresponds to the MDE when the value 24 is replaced by 168 (1 week).

$$MWE = \frac{1}{168}\sum_{h=1}^{168} \frac{\left|P_h - \widehat{P_h}\right|}{\overline{P_h}}$$

Both errors are usually reported in percent.

In order to compare our results to previous work we used a paper about forecasting electricity prices in California (Weron, 2008). In this work, MWE around 10-15% was achieved.

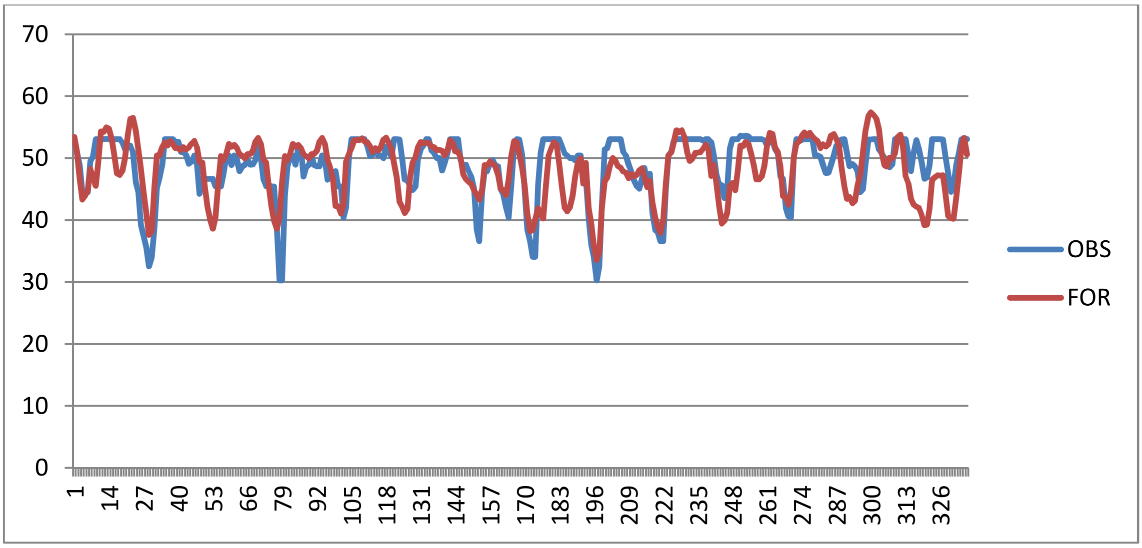
3.1. ARIMA Model.

Other ARIMA models were built using data divided by days. For forecasting prices on Mondays, only Mondays were taken into account, and so on. Therefore seven series were achieved. Moreover, we used data of total demand as external variable.

The table below shows the results:

Results Day to Day	MDE	MDE
Monday-Monday	0,05	0,09
Tuesday-Tuesday	0,07	0,07
Wednesday-Wednesday	0,06	0,03
Thursday-Thursday	0,07	0,07
Friday-Friday	0,03	0,04
Saturday-Saturday	0,05	0,05
Sunday-Sunday	0,04	0,05
Mean Weekly Error	**0,05**	**0,06**

This picture shows the forecast for two weeks, and the actual prices.



The problem of this model is that it does not take into account the previous days. For instance, if we only use Monday data for predicting Mondays we do not have any information about what happened on Sunday. Maybe prices could be increasing but we do not know it.

3.2. Weighted neighbors.

Weighted nearest neighbors (WNN) algorithms are techniques for pattern classification that are based on the similarity of the individuals of a population. The members of a population are surrounded by individuals that have similar properties.

According to the WNN methodology, the 24 hourly prices of day are predicted by linearly combining the prices of the k days succeeding those in the neighbor set, that is

1. Calculate the distances between the load of day *d*, D_d and that of preceding points { D_{d-1}, D_{d-2}, D_{d-3} ,……}. Let v_1,….., v_k be the k nearest days to the day d, sorted by ascending distance.

2. The prediction is :

$$\widehat{D} = \frac{1}{\alpha_1+\alpha_2+\cdots+\alpha_k} \sum_{l=1}^{l=k} \alpha_l * D_{v_{l+1}}$$

Where,

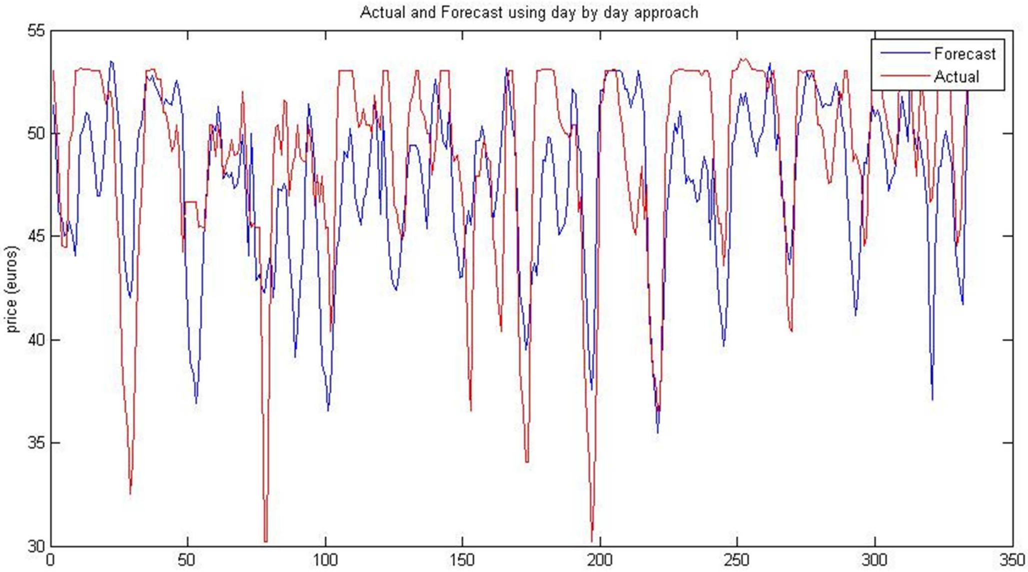
$$\alpha_j = \frac{d_w\left(D_d , D_{v_k}\right) - d_w\left(D_d , D_{v_j}\right)}{d_w\left(D_d , D_{v_k}\right) - d_w\left(D_d , D_{v_1}\right)}$$

Also note that Notice that $0 \le \alpha_j \le 1$ i.e., the weight is null for the most distant day and is equal to one for the nearest day.

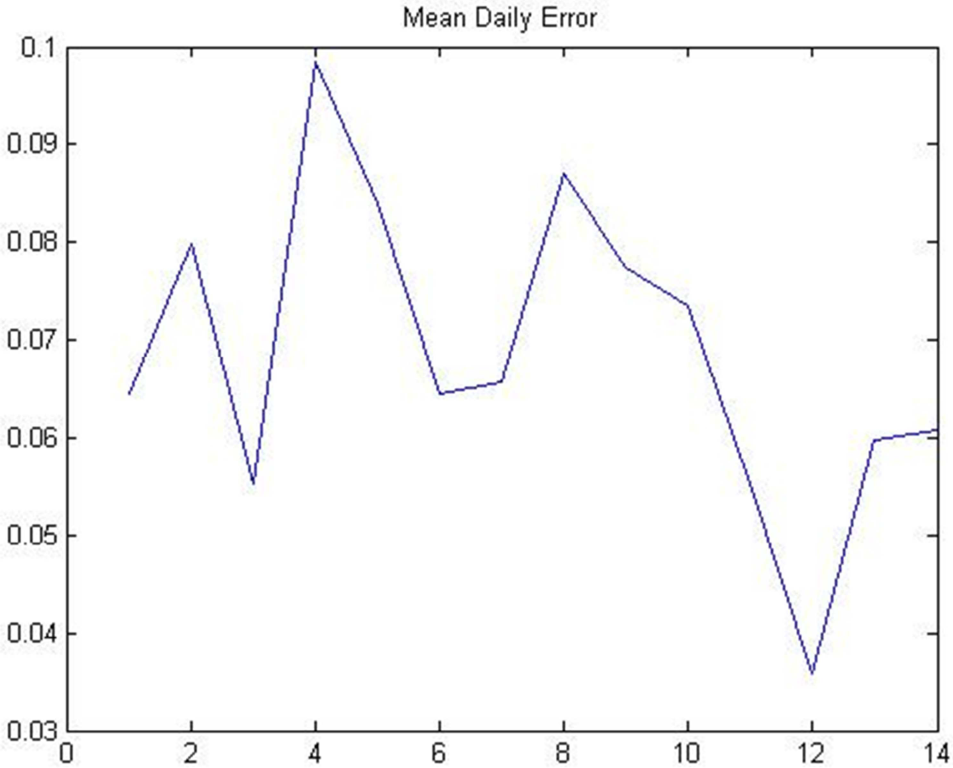
The WNN method first identifies the k nearest neighbors of day , where k is a number to be determined and “neighborhood” in this context is measured according to distances.

The algorithm is coded in MATLAB and to find the optimum value of k can be selected after analyzing mean daily error and mean weekly error.

In this work the forecasting done for next 2 weeks using day by day approach. Following figure shows the forecasted prices and actual prices.



The means daily error is plotted as follows.



4. Conclusions.

The price forecasting is done using different methods. Following table shows the mean weekly errors observed in different approaches.

	ARIMA	ARIMA + Demand	ARIMA + Demand , by days	KNN	KNN by demand
MWE 1	2.05%	1.8%	5%	7.3%	14%
MWE 2	2.1%	1.8%	7%	6.4%	8.4%

Although the kNN method does not give the best results, it is a very simple and time efficient method for forecasting the future that holds lots of promise. We have only touched upon the possible ways this method could be applied. It should be investigated for a variety of norms. For example, when searching past demand a Euclidean norm may not be the best norm to use. It may be better to use a norm that takes into account the variation of the demand, e.g. a gradient norm. This may explain why the “kNN by demand” results are not as good as the normal kNN results. Another way to improve the kNN by demand results would be to use a seasonality factor where days with similar demand from the same season have a higher weighting in the linear combination for tomorrow´s price. The kNN method could also be investigated using a by day approach where, for example, if tomorrow is Monday we only use Mondays as potential nearest neighbours. Another possible avenue for investigation would be to combine all these different methods for finding the nearest neighbours.

It is clear that the “ARIMA + demand” gives the best results. It should be noted, however, that this is for one test set in summer. Each method should be re-investigated for a variety of test sets and seasons. Another future avenue for investigation would be to try and combine the kNN and ARIMA methods.

Future Extensions :

Study the volatility with Garch or wavelet transform.
Different norms can be used to improve kNN..
All the methods can be combined e.g kNN could be used for developing time series and ARIMA can be used for predictions.
Test data can be separated in different seasons.
Market´s strategic behavior can be taken in to account.
Forecasting can be done with one hour time series.
In the ARIMA model by days , previous days can be taken as exogenous variables.

5. **Bibliography.**

R. Weron (2008) Forecasting wholesale electricity prices: A review of time series models, in "Financial Markets: Principles of Modelling, Forecasting and Decision-Making", eds. W. Milo, P. Wdowiński, FindEcon Monograph Series, WUŁ, Łódź.

Electricity Market Price Forecasting Based on Weighted Nearest Neighbors Techniques .Alicia Troncoso Lora, Jesús M. Riquelme Santos, Antonio Gómez Expósito, Fellow, IEEE, José Luis Martínez Ramos, Senior Member, IEEE, and José C. Riquelme Santos . IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 22, NO. 3, AUGUST 2007