

# A Comparison of Two Techniques for Next-Day Electricity Price Forecasting

Alicia Troncoso Lora<sup>1</sup>, Jesús Riquelme Santos<sup>1</sup>, José Riquelme Santos<sup>2</sup>,  
Antonio Gómez Expósito<sup>1</sup>, and José Luíz Martínez Ramos<sup>1</sup>

<sup>1</sup> Department of Electrical Engineering, University of Sevilla, Sevilla, Spain

<sup>2</sup> Department of Languages and Systems, University of Sevilla, Sevilla, Spain  
ali@esi.us.es, jesus@us.es riquelme@lsi.us.es, {age, camel}@us.es

**Abstract.** In the framework of competitive markets, the market's participants need energy price forecasts in order to determine their optimal bidding strategies and maximize their benefits. Therefore, if generation companies have a good accuracy in forecasting hourly prices they can reduce the risk of over/underestimating the income obtained by selling energy. This paper presents and compares two energy price forecasting tools for day-ahead electricity market: a  $k$  Weighted Nearest Neighbours (kWNN) the weights being estimated by a genetic algorithm and a Dynamic Regression (DR). Results from realistic cases based on Spanish electricity market energy price forecasting are reported.

## 1 Introduction

The Spanish electric industry is moving from a centralized operational approach to a competitive one since 1998, and it is mainly based on two separated day-ahead markets [1]: the energy market, managed by the Market Operator and the market for regulation reserves.

Forecasting energy prices has become an essential tool in competitive electricity markets. In the short term, expected price profiles, both in terms of energy and reserve prices, help market participants to determine their optimal bidding strategies and, consequently, maximize their benefits. Therefore, if generation companies have a good accuracy in forecasting hourly prices they can reduce the risk of over/underestimating the income obtained by selling energy.

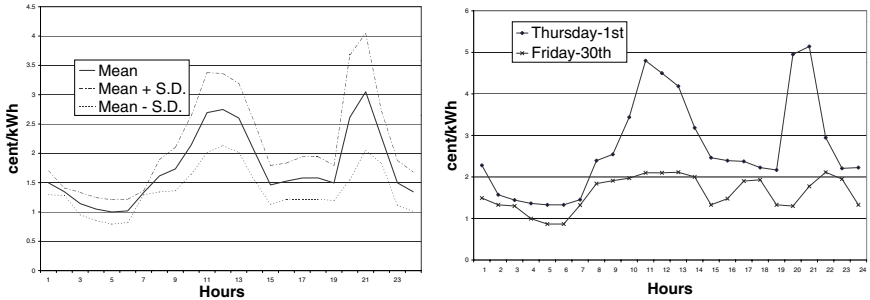
In a non-perfect oligopolistic market, the energy prices time series presents a high percentage of unusual prices. Due to this fact, nowadays, no approaches based on time series analysis that successfully forecast next-day Spanish electricity prices exist.

This paper presents and compares two energy price forecasting effective tools for day-ahead electricity market: a  $k$  Weighted Nearest Neighbours (kWNN), in which the weights are estimated by a genetic algorithm and a Dynamic Regression (DR). First, a  $k$  Weighted Nearest Neighbours algorithm is proposed using a Weighted-Euclidean distance. Second, two Dynamic Regression models are developed and applied to the 24-hour energy price forecasting problem, and the expected errors are quantified.

## 2 Structure of the Spanish Electricity Market

The study reported in this paper is based on the hourly Spanish spot market prices recorded from January 2001 to August 2001. As weekends and holidays constitute separate cases, only data corresponding to working days have been retained and analyzed.

Figure 1a shows the hourly averages and standard deviations of prices for the working days of March 2001, in cents of Euro per kWh.



**Fig. 1. a)** Hourly average of spot market prices for March 2001. **b)** Evolution of energy prices for two days of March 2001.

Average spot prices larger than 2 cent/kWh take place during the morning and evening peak hours (10am-2pm and 8-10pm respectively). Except for a few valley hours, the s.d. of this price exceeds 20% of the mean value, reaching even 40% at 8pm and 9pm.

Figure 1b represents the energy prices for two selected days of March 2001. The significant differences in the prices of the peak hours can not be explained by a change in the demand profile, probably revealing market power mechanisms.

## 3 k Weighted Nearest Neighbours

In this section, a combined kWNN [2] and GA [3] for hourly market energy prices forecasting is described. The kWNN finds the daily market energy prices that are “similar to” the prices of previous days. GA are computational techniques based on the mechanics of natural selection in which each individual in the population represents a possible solution to the problem, and a fitness factor is assigned to each individual, with the best ones receiving higher fitness factors. A particular GA is characterized by issues such as population size, mutation rates, and selection and new population creation mechanisms. In this paper, the GA (with float codification) is used to compute the optimal weights to outweigh the most significant hours.

The prediction aims at estimating the prices for a certain day from a linear combination of the energy prices of the days that follow the nearest neighbours days.

The prediction of a time series requires the knowledge of some parameters: the time delay, the embedding dimension and the value of  $k$ . The optimal determination of these parameters has been studied in the last years [4].

Energy prices are collected every hour and, in consequence, the time delay is an hour. For short-term forecasting, it has been decided to organize the temporal data with a shifting window of prices comprising 24 hours and, consequently, the embedding dimension is 24.

As far as the number of steps to predict is concerned, only one possibility has been evaluated: twenty four hours corresponding to the prices of a whole day, whose values are determined by those of the previous day. This possibility implies that the window is shifted 24 hours each time.

In this case, the time series is the electricity market energy prices, and the Weighted-Euclidean distance is preferred because not all the hours of a day have the same influence on the prices of the following day. This distance is defined by

$$d_w^2(q, z) = \sum_{i=1}^r w_i \cdot (q_i - z_i)^2 \tag{1}$$

where  $r = \text{length}(q)$  and  $w_i \in [0, 1]$ .

The prediction of stock energy prices for one day  $d + 1$  is computed through two steps taking into account the weights of all the hours of a day:

1. Calculate the distances between the prices of the day  $d$ ,  $P_d$ , and the preceding points  $\{P_{d-1}, P_{d-2}, \dots\}$  using (1). Let be  $v_1, \dots, v_k$  the  $k$  nearest days to the day  $d$ , sorted by closeness.
2. The prediction is:

$$\hat{P}_{d+1} = \frac{1}{\alpha_1 + \dots + \alpha_k} \sum_{l=1}^k \alpha_l \cdot P_{v_l+1} \tag{2}$$

where

$$\alpha_i = \frac{d_w(P_d, P_{v_k}) - d_w(P_d, P_{v_i})}{d_w(P_d, P_{v_k}) - d_w(P_d, P_{v_1})} \tag{3}$$

Notice that if the  $k$  nearest neighbours of a vector  $P_d$  are  $[P_{v_1}, \dots, P_{v_k}]$ , where  $v_i$  is the  $i^{th}$  nearest neighbour, the set of points  $[P_{v_1+1}, \dots, P_{v_k+1}]$  will usually be the nearest to  $P_{d+1}$ , at least if the dynamic system is stable.

### 3.1 Test Results

The kNN+GA has been applied in several experiments to obtain the forecast of Spanish electricity market energy prices. February 2001 has been used to determine the weight for every hour of the day, using a GA whose main parameters

have been: the population size 100, the probability of crossover and mutation 1 and 0.1 respectively and the maximum number of generations 5000. The available period of March-August 2001 has been chosen as a test set to check the forecasting errors. The number of neighbours has been considered equal to one in all experiments due to the low influence on the forecasted errors.

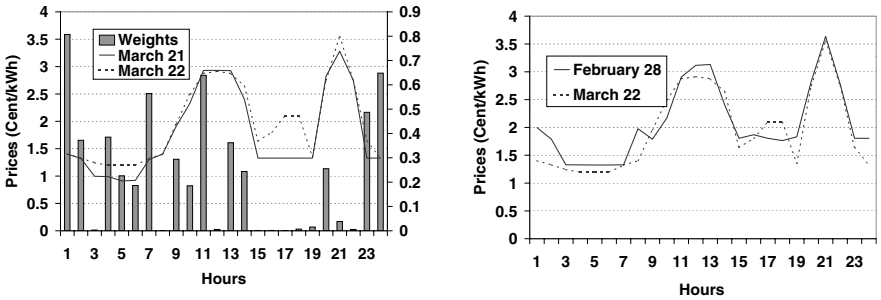


Fig. 2. Nearest neighbour using: a) the Weighted-Euclidean, b) the Euclidean distance.

Figure 2a represents the relative weights for every hour, along with the energy prices for a selected day, March 22, and its nearest neighbour, calculated using the Weighted-Euclidean distance (1). The nearest neighbour corresponds to March 21.

It can be noticed that the hours with greater weights, (1am, 7am, 11am, 11pm and 12pm), have almost the same energy prices. However, at 3pm, 4pm, 5pm, 6pm and 7pm, when the differences in prices are larger, the weights are approximately equal to zero.

Figure 2b represents the energy prices of March 22, along with its nearest neighbour calculated using the Euclidean distance. In this case, the nearest neighbour corresponds to February 28. Notice the influence of the chosen distance in the calculation of the nearest neighbour.

#### 4 Dynamic Regression

In this section, several models based on DR approach are described and tested [5,6]. Basically, this approach aims at estimating the price at hour  $t$  from the values of past prices at hours  $t-1, t-2, \dots$ , etc. A study of the correlation between the variables  $P_t, P_{t-1} \dots$  has been first made to determine the underlying model of the price time series.

Figure 3a shows the average correlation coefficient between the past prices and the present price for the period March-August 2001. Notice that this coefficient presents a periodicity corresponding to a day. Moreover, the correlation decreases as the number of past hours increases.

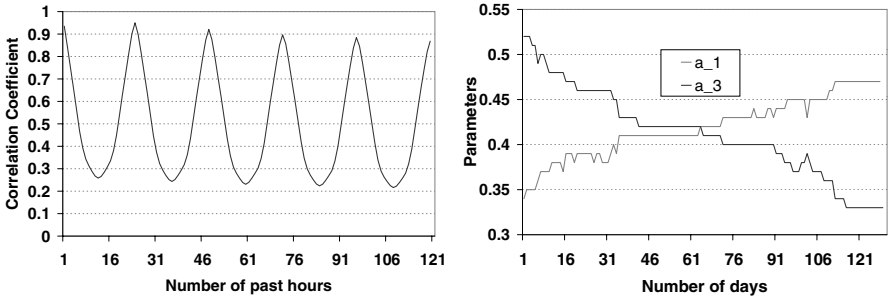


Fig. 3. a) Correlation Coefficient. b) Evolution of the parameters  $a_1$  and  $a_3$

The two proposed models based on the former correlation study are the following:

$$\hat{P}_t = a_0 P_{t-1} + a_1 P_{t-23} + a_2 P_{t-24} + a_3 P_{t-25} + a_4 P_{t-48} \tag{4}$$

$$\hat{P}_t = a_0 P_{t-23} + a_1 P_{t-24} + a_2 P_{t-25} + a_3 P_{t-48} \tag{5}$$

The difference lies in the presence of  $P_{t-1}$  which, except for the first time, is also an estimated quantity. Experience has shown that, including other terms in (4) and (5), like  $P_{t-47}$ ,  $P_{t-49}$ , makes no difference as far as accuracy is concerned. In turn, the parameters  $a_i$  are estimated by minimizing the least-squares objective function that follows:

$$\sum_t (P_t - \hat{P}_t)^2 \text{ where } \hat{P}_t \text{ is defined by (4) or (5)} \tag{6}$$

These parameters can be calculated at once on a given training set or updated every day of the selected period to forecast and validate the performance of both models.

### 4.1 Results

The two models described in the previous section have been applied to predict the energy prices of the Spanish electricity market. January-February 2001 has been used to determine the parameters of the models, by the least-squares method, providing:  $a_0 = 0.6760$ ,  $a_1 = -0.4815$ ,  $a_2 = 0.5213$ ,  $a_3 = 0.077$  and  $a_4 = 0.2072$  for the first model, and  $a_0 = 0.0848$ ,  $a_1 = 0.3369$ ,  $a_2 = 0.0416$  and  $a_3 = 0.5236$  for the second model. Table 1 presents the s.d. of actual prices, the average absolute value of forecasting errors and the maximum errors for Spring and Summer seasons for every model with the previous parameters. It can be concluded that the presence of  $P_{t-1}$  in (4) has a negative influence on the forecasting errors, because errors get accumulated during the last hours

**Table 1.** Comparison of predicted prices for two models.

	Model 1		Model 2	
	March-May	June-August	March-May	June-August
s.d.	0.184	0.242	0.173	0.209
Average absolute errors	0.275	0.332	0.239	0.288
Maximum errors	2.22	2.81	2.28	2.55
Average Relative errors (%)	12.2	9.4	10.6	8.1

of the prediction horizon. Figure 3b shows the evolution of the parameters corresponding to the second model when they are calculated for every day of the period March-August 2001 (excluding weekends).

**Table 2.** Comparison of predicted daily prices for both methods.

	Daily Prices			
	March-May		June-August	
	kWNN	DR	kWNN	DR
s.d.	0.2186	0.170	0.26	0.210
Average absolute errors	0.256	0.228	0.33	0.265
Maximum errors	2.13	2.11	2.36	2.41
Average Relative errors (%)	11.4	10.1	9.3	7.5

Finally, table 2 presents the s.d. of actual prices, the average absolute value of forecasting errors and the maximum errors for Spring and Summer seasons. Note that, the average error, when using the second DR model with daily parameter updating, ranges from 7.5% (Summer) to 10% (Spring) of the hourly average price, while the obtained average error, using a kNN combined with a GA ranges from 9% (Summer) to 11% (Spring).

## 5 Conclusions

Optimal bidding strategies are relevant in a competitive market, the computation of a good forecasted price profiles being crucial for the generation companies. In this sense, this paper presents two energy price forecasting tools for day-ahead electricity market: a *k* Weighted Nearest Neighbours and a Dynamic Regression. Both algorithms have been applied to the 24-hour energy price forecasting problems, using real data of the Spanish energy markets and their performance have been compared.

**Acknowledgments.** The authors would like to acknowledge the financial support of the Spanish Government under grants PB97-0719, DPI2001-2612 and TIC2001-1143-C03-02.

## References

1. A. Canoyra, C. Illán, A. Landa, J.M. Moreno, J.I. Pérez Arriaga, C. Sallé and C. Solé: The Hierarchical Market Approach to the Economic and Secure Operation of the Spanish Power System. Bulk Power System Dynamic and Control IV, August 24-28, Santorini, Greece.
2. B.V. Dasarathy (Ed): Nearest neighbour (NN) Norms: NN pattern classification techniques. IEEE Computer Society Press, 1991.
3. D. E. Goldberg: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, 1989.
4. A. W. Jayawardena, W. K. Li and P. Xu: Neighbourhood selection for local modelling and prediction of hydrological time series. *Journal of Hydrology* 258, 40-57.
5. A. D. Papalexopoulos and T. C. Hesterberg: A Regression-Based Approach to Short-Term System Load Forecasting. *IEEE Trans. on Power System*, Vol. 5, pp. 1535-1547. 1990.
6. F. J. Nogales, J. Contreras, A. J. Conejo and R. Spínola: Forecasting Next-Day Electricity Prices by Time Series Models. *IEEE Trans. on Power System*, to appear in 2002.