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

Abstract—This paper presents a simple technique to forecast next-day electricity market prices based on the weighted nearest neighbors methodology. First, it is explained how the relevant parameters defining the adopted model are obtained. Such parameters have to do with the window length of the time series and with the number of neighbors chosen for the prediction. Then, results corresponding to the Spanish electricity market during 2002 are presented and discussed. Finally, the performance of the proposed method is compared with that of recently published techniques.

Index Terms—Electricity market prices, forecasting, time series, weighted nearest neighbors.

I. INTRODUCTION

IN competitive electricity markets, prediction tools have become important for participating agents to be able to develop their bidding strategies in order to maximize the profit obtained by trading energy. Such techniques, traditionally applied to load forecasting, have recently focused on the problem of predicting the hourly energy prices of pool-based electricity markets.

Electricity price models can be broadly classified in two sets [14], namely production cost models and statistical models. Production cost models try to simulate the operation of the system taking into account the strategic behavior of the involved agents. The main drawback of simulation methods is the large amount of information required, which is difficult to obtain in liberalized markets.

On the other hand, statistical models predict price evolution based on historically observed relationships, without explicitly modeling underlying physical processes. This category comprises a diversity of methods ranging from the simplest "black box" time-series methods, using only previous price as input data, to more complex structural forecasting models that include explanatory (causal) variables such as load demand, fuel prices, and generation availability.

In turn, time-series methods can be grouped as follows: classical time-series methods [2]–[4], [14] and automated learning techniques [5]–[7], [1], [16] (the reader is referred to [14] for an excellent taxonomy of electricity price models).

The main advantage of classical statistical methods, including regression models, ARIMA, transfer function models, etc., is their relative simplicity. However, owing to the nonlinear nature of the price prediction problem, it is difficult to obtain accurate and realistic models for such methods.

In the last few years, machine learning techniques, such as artificial neural networks (ANN), have been applied to energy price prediction owing to their relatively good performance in load forecasting and load pattern recognition [1], [16]. ANNs are trained to learn the nonlinear relationships between the input variables (mainly past values of prices and other key variables affecting the prices) and historical patterns of energy prices. More complex arrangements, combining ANNs with fuzzy logic, have been recently proposed [17].

Weighted nearest neighbors (WNN) algorithms [8], [9] 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. This simple idea is the learning rule of the WNN classifier. Thus, the nearest neighbors decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points. Unlike most statistical methods, which elaborate a model from the information available in the data base, the WNN method considers the training set as the model itself.

Recently, classification techniques based on the WNN method have been successfully applied in new environments outside traditional pattern recognition such as medical diagnosis tools, game theory, expert systems or time series forecasting. Several papers have been published on the application of those techniques to forecast the next-day hourly energy consumption [10], [11], providing competitive results. In [16], the application of the WNN methodology to next-day energy price forecasting is preliminarily applied, promising results being reported. This paper further elaborates on the same idea, providing an improved way of tuning the WNN model and presenting an exhaustive set of test cases corresponding to the working days of an entire year. A comparison is also performed with recently published proposals, based on ANN [1], [17], making use of a GARCH model [2] or combining an ARIMA model with the Wavelet transform [13].

The paper is organized as follows. Section II summarizes the main characteristics of the day-ahead electric energy market of mainland Spain. In Section III, the proposed method based on WNN is presented. In Section IV, a methodology to determine the two parameters characterizing the adopted model

Fig. 1. Hourly average of the prices time series (March 2001).

Fig. 2. Histogram of prices during 2000 (top), 2001 (middle), and 2002 (bottom).

for price forecasting is provided. Section V presents numerical results corresponding to the electric energy market of Spain for the entire year 2002. In Section VI, a comparison between the proposed method and existing techniques is made. Finally, Section VII provides some concluding remarks.

II. CHARACTERIZATION OF SPANISH ELECTRICITY MARKET PRICES

The Spanish electricity market, started in 1998, relies basically on a pool where energy is traded through an auction process [15] (bilateral transactions, although possible, are still negligible). Generators, retailers and consumers submit hourly bids the day before to the market operator, containing at least the price assigned to each block of energy (more complex bids are possible). Then the resulting marginal price is used every hour to charge/remunerate all agents buying/selling energy, irrespective of their original bids. In an ideal market, submitted bids should reflect actual variable costs, but the Spanish market is still far from this perfect competitive environment, as over 70% of energy is produced by units belonging to two major generating holdings. In this context, it is crucial for all agents to have as accurate a prediction as possible of next-day energy prices in order to develop their optimal bidding strategies.

The hourly Spanish spot market prices arising from January 2000 to August 2002 have been recorded to visually illustrate the behavior and evolution of the energy prices. As weekends and holidays constitute separate cases, only data corresponding to working days have been retained and analyzed.

Fig. 1 shows the hourly averages and standard deviations of prices for the working days of March 2001 in cents of Euro per kWh (cE/kWh). It can be noted that, as expected, larger average spot prices occur during the morning and evening peak hours (10 A.M.–2 P.M. and 8–10 P.M., respectively). Except for a few valley hours, the standard deviation of hourly prices exceeds 20% of the mean value, reaching 40% at 8 P.M. and 9 P.M..

The way prices have historically evolved is summarized in the histograms of Fig. 2, corresponding to years 2000, 2001, and 2002. Apart from a clear trend for prices to spread and increase, it is evident that prices do not follow a normal distribution. Whereas over 50% of prices were lower than 1 cE/kWh
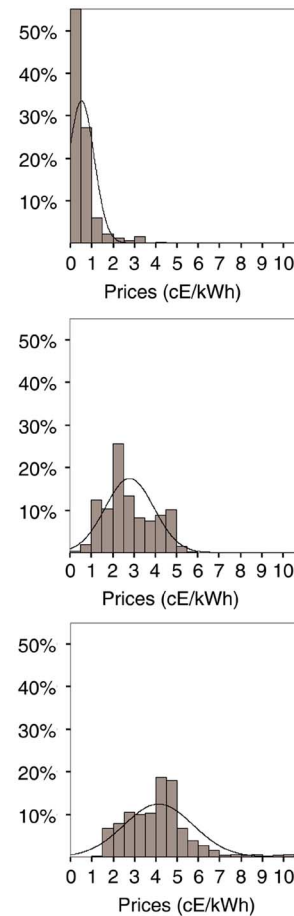
during 2000, a majority of prices lay between 2 and 3 cE/kWh in 2001, reflecting the fact that the Spanish electricity market was still rather unstable and subject to significant price changes.

To end this section, Fig. 3 represents hourly prices during 2002 (vertical axis) versus the respective energy demand (horizontal axis), along with the resulting regression line plus two parallel lines delimiting the ± 1 cE/kWh interval. When all data are considered, the correlation between price and energy is quite poor $(R = 0.63)$, particularly at peak hours. However, three different patterns (lower, central and upper) bounded by the two edge lines can be visually observed, the upper one being more diffuse. Results corresponding to each cluster of data will be separately analyzed in Section V. As the method proposed in this paper is intended to be of general applicability, no effort has been made to take advantage of such data clustering, that might be somehow useful in a future work to further improve the prediction accuracy.

III. WEIGHTED NEAREST NEIGHBORS METHODOLOGY

Given the hourly prices recorded in the past, up to day d, the problem consists of predicting the 24 hourly prices corresponding to day $d + 1$.

Let $P_i \in \mathbb{R}^{24}$ be a vector composed of the 24 hourly energy prices corresponding to an arbitrary day i

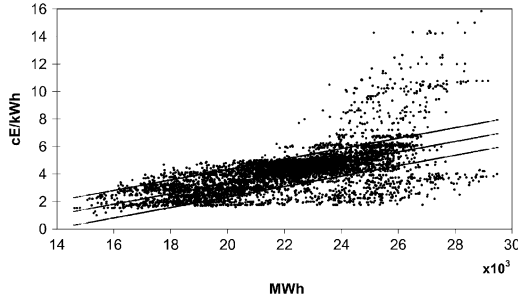$$P_i = [p_1, p_2, \dots, p_{24}]. \tag{1}$$

Fig. 3. Hourly price *versus* load during 2002.

Fig. 4. Illustration of the WNN approach.

Fig. 5. False nearest neighbors.

Then, the associated vector $PP_i \in \mathbb{R}^{24m}$ is defined by gathering the prices contained in a window composed of m consecutive days, from day i backward, as follows:

$$PP_i = [P_{i-m+1}, P_{i-m+2}, \ldots, P_{i-1}, P_i] \qquad (2)$$

where m is a parameter to be determined. For any couple of days, i and j, a distance can be defined

$$\text{dist}(i,j) = |PP_i - PP_j| \qquad (3)$$

where $|\cdot|$ represents a suitable vector norm (the Euclidean norm has been used in this work).

The WNN method first identifies the k nearest neighbors of day d, where k is a number to be determined and "neighborhood" in this context is measured according to (3) above. This leads to the neighbor set

$$NS = \{\text{set of } k \text{ days}, q_1, \ldots, q_k, \text{closest to day } d\} \qquad (4)$$

in which q_1 and q_k refer to the first and kth neighbors, respectively, in order of distance.

According to the WNN methodology, the 24 hourly prices of day $d+1$ are predicted by linearly combining the prices of the k days succeeding those in NS, that is

$$P_{d+1} = \frac{1}{\sum_{i \in NS} \alpha_i} \cdot \sum_{i \in NS} \alpha_i P_{i+1} \qquad (5)$$

where the weighting factors α_i are obtained from

$$\alpha_i = \frac{\text{dist}(q_k, d) - \text{dist}(i, d)}{\text{dist}(q_k, d) - \text{dist}(q_1, d)}. \qquad (6)$$

Obviously, α_i is null when $i = q_k$ (furthest neighbor) and is equal to one when $i = q_1$ (nearest neighbor), which means that only $k-1$ neighbors are actually used to determine the next-day prices. Note also that, although the 24 m prices contained in PP_i are used to determine if i is a good neighbor of d, only the 24 prices of P_{i+1} are relevant in determining P_{d+1}.

Fig. 4 illustrates the basic idea behind the WNN methodology. It considers that, if PP_i is close to PP_d, then P_{i+1}, already known, should be also similar to P_{d+1}.

In order to find candidate neighbors, a window of m days is simply slid along the entire list of hourly prices contained in the data base [15].

IV. TUNING THE MODEL

Before applying the WNN method, a training phase is necessary in order to find suitable values for m and k. Generally, after using the resulting model for a certain period, prediction errors tend 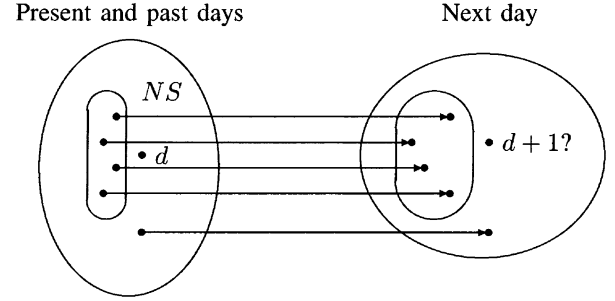to increase slightly, particularly when applied t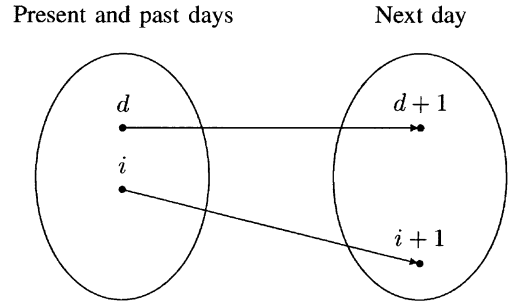o a very dynamic context, like the Spanish one (see Fig. 2), which may call for new training processes. As will be seen below, the training phase may be relatively costly and, in any case, even if the model was trained in a daily basis, the prediction accuracy would be always limited by the intrinsic uncertainty of prices. Therefore, a compromise should be found between cost and accuracy. In this paper, unless otherwise indicated, monthly training periods will be considered, but yearly training provides almost as accurate results.

A. Optimal Window Length

The number of days contained in the window that will be used to find candidate neighbors (parameter m) is determined in advance by resorting to the so-called *false nearest neighbors* (FNN) method [12]. This method compares the distance between a day d and a candidate neighbor i with that between $d+1$ and $i+1$. If the second distance is larger, as illustrated in Fig. 5, it is said that d and i are false neighbors, because the trajectory of the associated prices tend to diverge. Note that, according to (3), such distances depend on m in an implicit manner. By trying all days contained in the training set, m can be chosen so as to minimize the number of false neighbors.

In practice, as m increases, the cost of the training process also increases and the number of candidate neighbors gets significantly reduced (in the limit, if m approached the size of the training set, a single candidate would remain). Hence, the suboptimal but cheaper scheme adopted in this paper consists of choosing the minimum value of m leading to a percentage of false neighbors not exceeding a given threshold (e.g., 10%). Frequently, but not necessarily, $m = 1$ leads to forecasting errors that are close to that of the optimal value.

B. Optimal Number of Nearest Neighbors

The optimal number of nearest neighbors (parameter k) is the one that minimizes the forecasting error when the WNN methodology is applied to the training set.

2002 month	Training Set	$m \geq 1$		$m = 1$
		k	m	k
Jan	Jan 2001 - Dec 2001	8	5	10
Feb	Feb 2001 - Jan 2002	3	9	8
Mar	Mar 2001 - Feb 2002	3	14	6
Apr	Apr 2001 - Mar 2002	2	15	5
May	May 2001 - Apr 2002	6	15	8
Jun	Jun 2001 - May 2002	4	1	4
Jul	Jul 2001 - Jun 2002	4	9	9
Aug	Aug 2001 - Jul 2002	3	15	10
Sep	Sep 2001 - Aug 2002	3	14	5
Oct	Oct 2001 - Sep 2002	9	5	9
Nov	Nov 2001 - Oct 2002	3	10	3
Dec	Dec 2001 - Nov 2002	5	1	5

Mathematically, this is equivalent to finding the value of k that minimizes the following quadratic function:

$$\sum_{d \in \text{TS}} |\widehat{P}_{d+1} - P_{d+1}| \tag{7}$$

where $\widehat{P}_{d+1}$ are forecasted prices for day $d+1$, according to the WNN method, P_{d+1} are actual recorded prices and TS refers to the training set. Note that, according to (5), P_{d+1} is an implicit function of the discrete variable k, which prevents application of standard mathematical programming methods when searching for k. In practice, k is assigned successive integer numbers ($k = 2, 3, \ldots$) until a local minimum is found.

V. RESULTS

The methodology described above has been applied to the working days of 2002 (four atypical days of the last week of December, much closer to a holiday than to a working day, have also been excluded from the analysis). Table I shows the parameters m and k that result when each monthly model corresponding to 2002 is trained with the 12 preceding months. For instance, the model corresponding to April 2002 is trained with data corresponding to the period April 2001-March 2002. However, the data base in which neighbors are searched for comprises all working days from 2000 up to the day of 2002 right before the day whose prices are to be forecasted.

Note that if m is not equal to 1, then it is, or is very close to, a multiple of 5, revealing the weekly repetitive pattern of energy prices (not so regular anyway than that of the load). The fact that m sometimes equals 9 or 14, instead of the more intuitive values 10 or 15, is explained either by some isolated holidays being removed occasionally from the data base or by the slightly different behavior of Monday valleys compared to the remaining working days.

For comparison, the rightmost column collects the values of k that result when $m = 1$, instead of the value provided by the FNN methodology, is adopted. Note the larger number of neighbors involved when the sliding window comprises only the current day ($m = 1$).

The following prediction errors have been computed to assess the adequacy of the WNN methodology:

- *mean relative error (MRE)*

$$\text{MRE} = 100 \cdot \frac{1}{n} \sum_{h=1}^{n} \frac{|p_h - \widehat{p}_h|}{p_h} \tag{8}$$

TABLE II
MONTHLY PREDICTION ERRORS OBTAINED WITH THE
WNN METHODOLOGY IN 2002

Working days 2002	$m = 1$	$m \geq 1$			$\bar{p}_{\text{month}}$
	MRE (%)	MRE (%)	MAE (cE/kWh)	MMRE (%)	(cE/kWh)
January	13.72	11.18	0.93	12.54	7.45
February	8.12	6.48	0.27	5.96	4.00
March	6.20	5.10	0.23	5.33	3.68
April	5.82	5.55	0.24	5.28	4.00
May	6.64	6.12	0.27	5.56	4.18
June	9.67	9.67	0.52	10.07	4.38
July	7.99	7.13	0.37	6.59	4.86
August	9.79	8.98	0.37	8.73	3.37
September	11.54	9.70	0.35	8.75	4.00
October	8.93	8.75	0.29	7.83	3.76
November	16.69	13.34	0.41	12.78	3.18
December	12.05	12.05	0.32	12.02	2.68
Average	9.76	8.67	0.38	8.45	4.07

- *mean absolute error (MAE)*

$$\text{MAE} = \frac{1}{n} \sum_{h=1}^{n} |p_h - \widehat{p}_h| \tag{9}$$

- *mean error relative to $\bar{p}_{\text{month}}$ (MMRE)*

$$\text{MMRE} = \frac{1}{n} \sum_{h=1}^{n} \frac{|p_h - \widehat{p}_h|}{\bar{p}_{\text{month}}} \tag{10}$$

where

$$\bar{p}_{\text{month}} = \frac{1}{n} \sum_{h=1}^{n} p_h. \tag{11}$$

$\widehat{p}_h$ and p_h are the predicted and actual hourly prices, respectively, and n is the number of hours for the period of interest (usually a month).

Table II presents the resulting errors, along with $\bar{p}_{\text{month}}$, for each month of 2002. When the parameter m is determined by the FNN ($m \geq 1$), the MRE ranges from 5.10% in March to 13.34% in November, yielding an average value of 8.67% in 2002. In a majority of cases, the MMRE is smaller than the MRE, which is logical considering that, unlike $\bar{p}_{\text{month}}$, p_h is sometimes relatively close to zero. Note that the MRE corresponding to $m = 1$, also included for comparison in the second column, is larger than the one provided by the value of m determined through the FNN method.

Apart from December, which is quite an atypical month in Spain, unusually large prediction errors can be observed in January, June and November. Unlike demand, whose behavior is rather predictable, there are no definite clues about prices behaving so irregularly in those months. Nevertheless, the market and system operators monthly reports reveal the following.

- In January, extremely low temperatures gave rise to sustained demand peaks. In addition, important blackouts that had recently affected Madrid and other large cities, along with reduced levels of hydraulic production, led to expensive fuel units to be resorted to for security of supply rea-

sons. This probably explains the fact that the average price in January was 197% higher than that of the same month in 2001, and also significantly higher than those of the remaining months.

- After several dry years, hydraulic generation beat minimum historic records in June. Furthermore, a general strike on June 20 (Thursday) significantly affected the demand and prices for several consecutive days, as many people took advantage of the resulting long weekend for a break. The average price on June 20 was 30% lower than that of the day before, yielding an average prediction error of about 31.2% for that day and 23% for the next (the peak error on Thursday exceeded 80% and took place at 8 A.M.). In spite of that, the monthly average price was 15% larger than that of June 2001, clearly showing that June was also a rather special month.

- Somewhat the opposite happened late in 2002. Owing to unusually mild temperatures and, for the first time in the year, hydraulic productions well over the average, prices in November were abnormally lower than expected (20.1% lower in average than those of the same month in 2001). For unknown reasons, the demand forecasted by the system operator on Wednesday 6 was about 10% lower than the actual demand, yielding an exceptionally low daily average price of 1.4 cE/kWh (in fact, the price was 0.0 cE/kWh for four consecutive hours). Note, however, that the absolute prediction error in November is not far from the average corresponding to 2002.

- In December mild temperatures and rain continued, leading to low demand levels and twice as much hydraulic production as that of November.

In summary, extreme weather factors (very cold and dry season early in 2002, rather mild and rainy fall) along with the peculiarities of the generation mix in Spain, where the hydraulic production in 2002 ranged from virtually null to nearly 27% in December, may partly explain the irregular price behavior and associated prediction errors. Furthermore, market agents are very sensitive to energy-related government policies, among which the so-called "transition-to-competition costs" were quite relevant those years. Such costs, intended to compensate existing Gencos for potential profit losses originated by the regulatory change, were somewhat conditioned to the "reasonable" evolution of the energy pool prices, constituting in this way an economic signal whose real influence on prices is very difficult to analyze.

Table III shows the standard deviation of MRE and MAE corresponding to the variable m case, which is useful to assess the variability of prediction errors.

Table IV collects, for each month, the maximum and minimum average daily prediction errors, in absolute value, corresponding to the variable m case. Interestingly enough, the largest prediction errors corresponding to July, August and September took place respectively on Monday 15, Monday 5, and Monday 16, typical days in which a majority of Spaniards start their summer holiday season.

Figs. 6 and 7 show the evolution of actual and forecasted prices corresponding respectively to the best and worst day of the third week of May (Monday 20 to Friday 24), in terms of

TABLE III
STANDARD DEVIATION OF MRE AND MAE

Working days 2002	σ_{MRE} (%)	σ_{MAE} (cE/kWh)
January	6.28	0.87
February	4.98	0.17
March	2.64	0.11
April	3.51	0.15
May	2.36	0.08
June	7.47	0.45
July	3.99	0.19
August	4.77	0.19
September	5.86	0.18
October	7.61	0.17
November	6.26	0.19
December	7.83	0.18
Average	5.30	0.24

TABLE IV
MAXIMUM AND MINIMUM MEAN DAILY ABSOLUTE ERRORS

Working days 2002	maximum daily error (cE/kWh)	minimum daily error (cE/kWh)
January	3.21	0.25
February	0.83	0.07
March	0.45	0.08
April	0.65	0.08
May	0.43	0.11
June	1.81	0.11
July	0.82	0.12
August	0.80	0.08
September	0.81	0.15
October	0.92	0.11
November	0.82	0.13
December	0.74	0.15

prediction accuracy. The resulting MMRE error for those days is 1.54% (May 23) and 5.60% (May 24), respectively.

To end this section, prediction errors corresponding to 2002 will be separately analyzed for the three clusters visually identified in Fig. 3. Table V presents, along with the number of hours comprising each cluster, the absolute and relative mean prediction errors. Figures in the rightmost column, referring to the hourly averages for the entire period considered, slightly differ from those in the last row of Table II, because in that case the average is computed by months. Note that the performance of the WNN method is better for the central cluster, where correlation between price and demand is larger. As expected, the largest absolute prediction errors arise in the upper cluster. It is also interesting to split the data in Table V by months, as in Table VI. A careful examination of this table shows that a majority of hours corresponding to November and December lie in the lower cluster, where prices are statistically small for the associated demand, while somewhat the opposite happens in January. Such findings support the explanations provided above regarding the anomalous behavior of prices for certain months in 2002.

For simplicity, holidays and weekends have been removed from the records to obtain the above results, but they could be gathered in a separate data base to be processed in the same way as labor days.
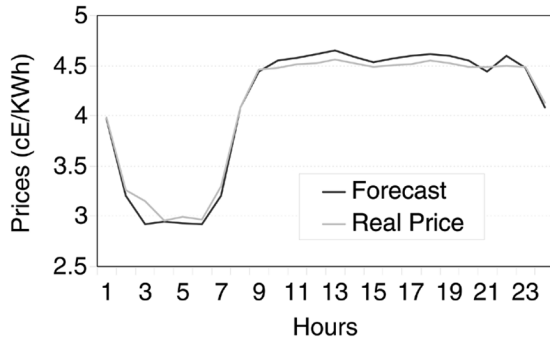
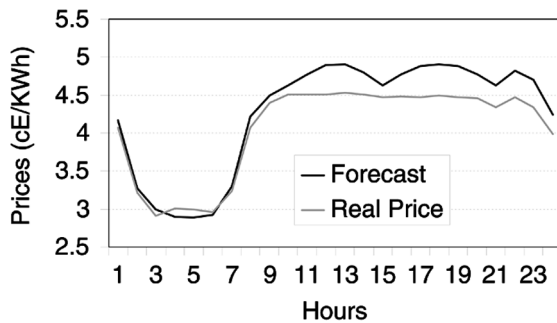Fig. 6.   Best daily prediction in a week of May 2002.

Fig. 7.   Worst daily prediction in a week of May 2002.

TABLE V
ABSOLUTE AND RELATIVE PREDICTION ERRORS BY CLUSTERS OF DATA

	Lower	Central	Upper	Total
MRE (%)	13.57	7.7	8.96	8.6
MAE (cE/kWh)	0.27	0.29	0.75	0.33
No. hours	767	4473	592	5832

VI.   COMPARISON WITH OTHER TECHNIQUES

In this section the performance of the WNN method is compared, whenever possible, with that of other approaches recently published, namely the ANN approach described in [1], the Neuro-Fuzzy system described in [17], the GARCH model described in [2] and the proposal of [13] combining the Wavelet transform with an ARIMA model.

A.   Comparison With a Plain ANN

The ANN described in [1] has been used for comparison purposes, as both the ANN and the proposed WNN approach resort to nonlinear models.

For this application, the ANN is fed with a shifting window of prices comprising 24 hours, which means that the input layer is composed of 24 perceptrons. As far as the number of output perceptrons is concerned, 24 outputs corresponding to the prices of a whole day are used, whose values are thus determined by those of the previous day. This implies that the window is shifted 24 hours each time. Finally, the intermediate layer is also composed of 24 perceptrons, this number being determined in an optimal way after several tests. The period used in [1] to train and tune the ANN is January–February 2001, while the period March-August is devoted to check the forecasting errors.

Table VII presents the resulting errors when market clearing prices of working days are forecasted using both the ANN tested

in [1] and the proposed WNN approach, along with $\bar{p}_{\mathrm{month}}$, for each tested month. As can be observed, the WNN method clearly outperforms this particular ANN implementation.

B.   Comparison With A Neuro-Fuzzy System

In order to compare the proposed method with a more sophisticated Neural Network arrangement, the work reported in [17] has been considered, for which Ontario market data has been downloaded from [18]. For this comparison, four WNN monthly models corresponding to the period June-September 2002 have been developed and tested, additional data corresponding to May being included in the data base (otherwise, there would be no price patterns to compare with during the first days of June). Like in [17], the MRE has been obtained and tabulated (note, however, that this error is termed MAPE in [17]). Table VIII shows the prediction errors provided by the WNN method when all days are considered (case A) and when several extreme days (June 11, July 2, August 13, and September 3, according to [17]) are omitted (case B). The WNN average errors for the considered period are 24.58% and 22.34% for cases A and B, respectively. Such results could be significantly improved if some other critical days, like September 15 for which the prediction error reaches 115%, were omitted in addition to those already ignored in case B.

Several prediction results provided by an ANN are presented in [17], according to different variables used as inputs by the model (demand, capacity shortfall and both), number of neurons at the intermediate layer and learning techniques (LM and momentum algorithms). Table IX summarizes the best results given in [17] for cases A and B. Note that, in spite of its simplicity and reduced size of the data base, the average results provided by the WNN method are better than those of Table IX.

A more complex model, based on a Neuro-Fuzzy system, is also reported in [17]. In addition to those cited above, forecasted energy imports and expected generator outages are used as inputs to the model, several membership functions being tested. The best results obtained by this approach are collected in Table X. Comparing with Table VIII, it can be concluded that this sophisticated model, that needs to be periodically trained, outperforms the proposed WNN method only when the predicted demand is considered.

C.   Comparison With Garch Model

Following [2], the data set comprises in this case the period ranging from January 1999 to November 2000 (both included), and the resulting monthly models are only used to predict the last week of every month in 2000 (except for December, whose last week refers to 1999).

For each of the twelve weeks under study the MRE error is computed. Table XI compares the results provided by the WNN methodology with those taken from [2]. Both the plain GARCH model and the GARCH model using as additional input the forecasted load are included.

Note that, in average, the WNN method outperforms both GARCH models. The average MRE is 8.57% for the WNN method and 9.55% for the plain GARCH model, neither of them using information about the expected load (better indices are obtained when December is excluded).

TABLE VI
MONTHLY RELATIVE PREDICTION ERRORS BY CLUSTERS OF DATA

	Lower		Central		Upper	
	No. hours	MRE (%)	No. hours	MRE (%)	No. hours	MRE (%)
January	4	24.56	218	9.66	234	12.30
February	29	15.69	448	5.86	3	10.11
March	27	2.81	423	5.38	6	15.94
April	0	0.00	440	5.31	88	6.73
May	0	0.00	437	5.99	91	6.73
June	0	0.00	444	9.80	36	8.19
July	8	25.68	430	7.93	114	4.28
August	25	19.57	474	8.31	5	19.90
September	22	19.29	476	9.75	6	8.87
October	41	16.67	509	8.10	2	12.65
November	289	14.16	160	11.62	7	19.24
December	322	12.07	14	11.59	0	0.00

TABLE VII
MONTHLY PREDICTION ERRORS PROVIDED BY THE ANN AND WNN METHODOLOGIES IN 2001

Working days 2001	ANN			WNN			$\bar{p}_{month}$
	MRE (%)	MAE (cE/kWh)	MMRE (%)	MRE (%)	MAE (cE/kWh)	MMRE (%)	(cE/kWh)
March	21.45	0.37	20.75	14.24	0.25	14.23	1.77
April	16.92	0.35	16.34	11.06	0.22	10.33	2.15
May	10.71	0.32	11.29	6.08	0.18	6.36	2.84
June	10.82	0.42	10.99	5.54	0.21	5.44	3.84
July	9.25	0.34	9.05	7.11	0.25	6.64	3.74
August	18.57	0.52	16.97	10.06	0.30	9.78	3.08

TABLE VIII
MRE PROVIDED BY THE WNN METHODOLOGY ON THE
ONTARIO CASE STUDIES

	Case A	Case B
June	19.99	17.08
July	23.82	22.42
August	22.25	20.85
September	35.17	31.48
Average	24.58	22.34

TABLE IX
MRE PROVIDED BY THE ANN ON THE ONTARIO CASE STUDIES

Input	Case A	Case B
Demand	25.88	23.30
Cap. shortfall	27.37	24.28
Both	27.04	24.03

TABLE X
MRE PROVIDED BY THE NEURO-FUZZY SYSTEM ON THE
ONTARIO CASE STUDIES

Input	Case A	Case B
Demand	23.03	20.17
Cap. shortfall	27.11	22.96
Import	30.52	28.09
Demand+Cap. shortfall	24.48	21.86
Demand+Outages	21.85	19.83
Demand+Import	21.67	20.4

TABLE XI
COMPARISON OF MEAN WEEKLY PREDICTION ERRORS, MRE (%),
PROVIDED BY WNN AND GARCH METHODS

	WNN	GARCH	GARCH with load
January	6.00	9.25	8.62
February	4.49	7.24	6.64
March	9.34	9.94	9.75
April	9.70	12	11.91
May	5.58	5.19	4.62
June	9.70	8.92	8.67
July	7.60	8.49	8.23
August	7.52	7.28	7.2
September	5.55	9.46	9.08
October	9.26	8.99	8.83
November	9.28	10.92	10.24
December	18.89	16.96	15.41
Average	8.57	9.55	9.10
Average (without December)	7.64	8.88	8.53

D. Comparison With a Hybrid Wavelet-Arima Model

According to [13], the data set in this case comprises the working days of years 2001 and 2002, and the following four weeks of 2002 are selected to perform the experiments: February 18 to 22, May 20 to 24, August 19 to 23 and November 18 to 22. For the sake of comparison, in this case the WNN approach has been tuned specifically for each week, separate models being developed for working days and weekends.

Table XII compares, for each week, the error provided by the WNN method with those corresponding to the plain ARIMA

and the Wavelet-ARIMA models, both taken from [13]. Results provided by the "naive" approach, by which prices of the present week are directly taken as estimates for the next week, are also included.

In spite of its simplicity, the proposed method outperforms in all cases the plain ARIMA model and, in average, also the enhanced Wavelet-ARIMA model. The average MMRE for the whole test period is 8.05% when the WNN method is applied, whereas the Wavelet-ARIMA model yields an average of 8.11% for the same period.

VII. CONCLUSION

In this paper a simple methodology based on the WNN technique is proposed to forecast hourly prices in deregulated electricity markets. Prediction results corresponding to the market of mainland Spain for the entire year 2002 are reported, yielding an average monthly error which is close to 8%. The performance

TABLE XII
COMPARISON OF THE MMRE PROVIDED BY WNN AND ARIMA METHODS FOR THE FOUR WEEKS ANALYZED

Selected week of	WNN (weekly model)			ARIMA	WAVELET and ARIMA	Naïve
	working days	week-end	average			
February	5.25	4.89	5.15	6.32	4.78	7.68
May	3.98	5.24	4.34	6.36	5.69	7.27
August	9.88	13.41	10.89	13.39	10.7	27.30
November	11.27	13.21	11.83	13.78	11.27	19.98
Average	7.60	9.19	8.05	9.96	8.11	15.56

of the proposed method is also compared with that of other techniques such as ANN, Neuro-Fuzzy systems, GARCH, and ARIMA (with and without wavelet transform). In view of the results previously reported in the literature, it can be stated that the proposed method performs satisfactorily, particularly when the uncertainty of prices associated with the Spanish system are taken into account.

The following comments, taken from [19], related with the load forecasting problem, are fully applicable to the price forecasting case: "...because of particular and often heuristic nature of short-term load forecasting (STLF), it is not always possible to assume portability of an STLF system from one utility to another. General models and algorithms have wider applicability, but must be used cautiously and should be experimentally tested with sufficiently lengthy data records". Therefore, general models, like the one proposed in this paper, should be preferably chosen if they prove to be competitive enough with specific methods, developed and tuned for a very particular case.

REFERENCES

[1] J. L. M. Ramos, A. G. Expósito, J. M. R. Santos, A. T. Lora, and A. R. M. Guerra, "Influence of ANN-based market price forecasting uncertainty on optimal bidding," in *Proc. PSCC Power System Computation Conf.*, Sevilla, Spain, Nov. 2002.

[2] R. C. García, J. Contreras, M. van Akkeren, J. Batista, and C. García, "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, May 2005.

[3] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003.

[4] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, "Forecasting next-day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.

[5] B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, "Electricity price short-term forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 851–857, Aug. 1999.

[6] B. Ramsay and A. J. Wang, "A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays," *Neurocomputing*, vol. 23, pp. 47–57, 1998.

[7] L. Zhang, P. B. Luh, and K. Kasiviswanathan, "Energy clearing price prediction and confidence interval estimation with cascaded neural networks," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 99–105, Feb. 2003.

[8] B. V. Dasarathy, *Nearest Neighbour (NN) Norms: NN Pattern Classification Techniques*. New York: IEEE Computer Society Press, 1991.

[9] S. Dudani, "The distance-weighted K-nearest-neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, no. 1, pp. 325–327, 1975.

[10] A. Troncoso, J. M. Riquelme, J. C. Riquelme, A. Gómez, and J. L. Martínez, "Time-series prediction: Application to the short-term electric energy demand," *Lecture Notes in Artific. Intell.*, vol. 3040, pp. 577–586, 2004.

[11] A. Troncoso, J. C. Riquelme, J. L. Martínez, J. M. Riquelme, and A. Gómez, "Influence of kNN-based load forecasting errors on optimal energy production," *Lecture Notes in Artific. Intell.*, vol. 2902, pp. 189–203, 2003.

[12] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A.*, vol. 45, no. 6, pp. 3403–3411, 1992.

[13] A. J. Conejo, M. A. Plazas, R. Espínola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1035–1042, May 2005.

[14] A. M. González, A. M. San Roque, and J. García-González, "Modeling and forecasting electricity prices with input/output hidden markov models," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 13–24, Feb. 2005.

[15] *Compañía Operadora del Mercado Español de Electricidad (Spanish Market Operator)*, [Online]. Available: http://www.omel.es

[16] A. Troncoso, J. Riquelme, J. Riquelme, J. L. Martínez, and A. Gómez, "Electricity market price forecasting: Neural networks versus weighted-distance k nearest neighbours," *Lecture Notes in Comput. Sci.*, vol. 2453, pp. 321–330, 2002.

[17] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the Ontario competitive power system market," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 366–374, Feb. 2004.

[18] *IESO. Power to Ontario. On Demand*, [Online]. Available: http://www.ieso.ca/imoweb/pubs/marketReports/download/HOEP_20051//209.csv

[19] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.

Alicia Troncoso Lora was born in Carmona, Spain, in 1974. She received the Ph.D. degree in computer science from the University of Sevilla, Sevilla, Spain, in 2005.

From 2002 to 2005, she was with the Department of Computer Science, University of Seville. Presently, she is an Assistant Professor at the University Pablo de Olavide, Seville. Her primary areas of interest are time series analysis, control and forecasting, and optimization techniques.

Jesús M. Riquelme Santos was born in Las Palmas de Gran Canarias, Spain, in 1967. He received the Ph.D. degree in electrical engineering from the University of Sevilla, Sevilla, Spain.

Since 1994, he has been with the Department of Electrical Engineering, University of Seville, where he is currently an Associate Professor. His primary areas of interest are active and reactive power optimization and control, power system analysis, and power quality and forecasting techniques.

Antonio Gómez Expósito was born in Spain in 1957. He received the electrical and Dr. Eng. degrees from the University of Sevilla, Sevilla, Spain.

Since 1982, he has been with the Department of Electrical Engineering, University of Seville, where he is currently a Professor and Chairman of the Department. His primary areas of interest are power system optimization, state estimation, and digital signal processing.

José Luis Martínez Ramos was born in Dos Hermanas, Spain, in 1964. He received the Ph.D. degree in electrical engineering from the University of Sevilla, Sevilla, Spain.

Since 1990, he has been with the Department of Electrical Engineering, University of Sevilla, where he is currently a Professor. His primary areas of interest are active and reactive power optimization and control and power system analysis.

José C. Riquelme Santos was born in Las Palmas de Gran Canarias, Spain, in 1962. He received the Ph.D. degree in computer science from the University of Sevilla, Sevilla, Spain.

Since 1987 he has been with the Department of Computer Science, University of Sevilla, where he is currently an Associate Professor. His primary areas of interest are data mining, knowledge discovery in databases, and machine learning techniques.