IX Congreso de Ingeniería de Organización
Gijón, 8 y 9 de septiembre de 2005

Forecasting the Price of Energy in Spain's Electricity Production Market

Raúl Pino Diez[1], José Parreño Fernández[1]

[1] Dpto. de Admón. de Empresas y Contabilidad. Escuela Politécnica Superior de Ingeniería de Gijón. Campus de Viesques, 33204 Gijón (Asturias). pino@epsig.uniovi.es, parreno@epsig.uniovi.es

Resumen

In this job, short-term forecasts are calculated for the Energy Price in the Electricity Production Market of Spain. The methodology used to achieve these forecasts is based on Artificial Neural Networks, which have been used succesfully in recent years in many forecasting applications. To gauge the quality of forecasts, they have been compared with those obtained with the Box-Jenkins ARIMA models, another well-known forecasting methodology. Energy Price time series are usually composed of too many data, which can be a problem if we are looking for a short period of time to reach an adequate forecast. In this job, a training method for Neural Nets is proposed, which is based on making a previous selection for the Multilayer Perceptron (MLP) training samples, using an ART-type neural network. The MLP is then trained and finally used to calculate forecasts.

Palabras clave: Electricity Market, Forecasting Time Series, Neural Networks

1. Introduction

This study analyses the behaviour of Energy Price and Demand variables on the Spanish Electricity Production Market, and then proceeds to calculate Price forecasts. The methodology used to achieve these forecasts is based on Artificial Neural Networks, which have been used successfully in recent years in many forecasting applications. Forecasts obtained with this methodology are tested against forecasts calculated with Box-Jenkins ARIMA models. This paper focuses on short-term forecasting, as the time series to be forecast is the next-day Hourly Price for Energy on that market.

Nogales et al. (2002) and Contreras et al. (2003), among others, have analyzed the Spanish Electricity Market to achieve next-day Electricity Price Forecasts.

The paper is structured as follows: section 2 analyses the workings of the Electricity Market in Spain; section 3 provides an overview of Artificial Neural Nets forecasting methodology; section 4 describes the time series that are analysed in this study, and the structure of the Neural Nets used to forecast, as well as the results of forecasts compared with actual data. The study is rounded off by section 5, which draws relevant conclusions from the study.

2. Spain's Electricity Market

The Electricity Market in Spain is a stock exchange where purchase and sales operations of this particular merchandise are negotiated. Energy producers and energy purchasers both attend this stock exchange, just as any stakeholder interested in buying or selling financial assets would go to the Stock Market, and they present and match purchasing and selling prices for electricity, establish prices (once every hour), and settle the business of negotiating quantities to be bought and sold by the market's different agents.

The fact that electricity cannot be stored means that amounts sold, amounts bought and amounts consumed must all coincide each and every time. There must thus be a number of mechanisms and devices to link the commercial facet of the business (buying and selling) and the technical side of it (generating and consuming electricity), and to make them compatible.

The role of the Daily Market, an integral part of the Electricity Generation Market, is to handle electricity transactions for the day after calculations are made. Sellers' offers are presented to the market operator and are included in a 24-hour programmed matching process for the day following the one when the offer was made, which is divided into twenty-four consecutive programme periods.

Once the electricity purchase and sales offers have been received, checked and accepted (they are to have been received before 10:00 am every day), they are matched by the market operator, who calculates the marginal price and shares out production and demand amongst parties involved in the auction. The process can be either simple or complex, depending on whether the offers are all simple or, in contrast, a mix of simple and complex ones.

The Basic Working Daily Programme (BWDP), made public at 11:00 am, is the outcome of the above process of matching electrical supply and demand. Using the information in the BWDP, the offers presented by the agents and the bilateral contracts that have been signed, the market operator determines which technical restrictions are to be applied to the BWDP and how much electricity need be added or taken off the supply grid to meet the demands of the contracts. With this information, the market operator goes on to modify the result of the matching exercise by adding or eliminating electricity from the grid, following the order of economic precedence, until he recovers the balance between supply and demand for all the hourly periods of the programming. These adjustments are sent to the system operator who draws up the Provisional Feasible Daily Programme (PFDP) at 14.00 pm.

The transactions that transpire from the corrected matchings and the decisions on technical restrictions are forecasts that are unlikely to occur 100% of the time. Unforeseen circumstances may arise, such as breakdown of either the electricity generating unit, the distributing unit or the transmission lines, a higher or lower than expected local temperature, or simply a forecasting mistake.

Furthermore, the electricity system cannot be provided with intermediate storage, which might resolve the problem, so the system operator must have production units that can either up the load or reduce it in response to demand. This spectrum of available power is called the secondary regulation service, and it will be used to respond to unforeseen events like those that have already been mentioned. The result is called the Final Feasible Daily Programme (FFDP).

Finally, the Electricity Production Market offers stakeholders the opportunity to modify their buying and selling positions, by participating in the Inter-daily Market, at present made up of 6 sessions held between 16:00 and 12:00. Purchase and sales offers are matched each session, leading to the Final Timetable Programme (FTP) after analysis and the imposition of the relevant technical restrictions.

3. Forecasting with Artificial Neural Networks

Before the early 20s, forecasts were made by simply extrapolating time series. What could be called the "modern way" of forecasting began in 1927, when Yule presented the auto-regressive techniques to forecast the annual number of sun spots (Yule, 1927). His model calculated forecasts as a weighted sum of previous data. If good performance was to be achieved from this linear system, the existence of an external factor called noise had to be assumed. This noise affects the linear system; this linear system with noise was widely used for the next 50 years, and research culminated in the ARIMA methodology proposed by Box and Jenkins (Box and Jenkins, 1970).

From this point onwards, with a strong theoretical base, other studies that focused on non-stationary and/or non-linear series were presented: bilinear, bi-spectral or threshold models are examples of this, to name but a few (Tong, 1983, 1990; Priestley, 1988; Tsay, 1991; Subba Rao, 1992).

During the 1980s, two crucial developments took place that affected the evolution of time series research. On the one hand, the increase in the capacity and features of personal computers means that much longer time series could be studied and more sophisticated algorithms could be used. This went hand in hand with a second aspect - the development of machine learning techniques, such as Artificial Neural Networks.

Artificial Neural Networks (ANNs) are mathematical models based upon the functioning of the human brain. The literature suggests certain characteristics of ANNs that make them particularly interesting for forecasting time series. Two might be mentioned here: the ability to approximate practically any function (even non-linear ones) and the possibility of *piece-wise* approximations of the functions.

From a mathematical point of view, ANNs can be considered universal functional approximators (Hornik, et al., 1989; Cybenko, 1989). This means that ANNs can approximate the best function to the data, which is more important when the functions are complex. Moreover, ANNs are non-linear by nature (Rumelhart and McClelland, 1986), which means that they can not only correctly estimate non linear functions, but also extract non linear elements from the data.

ANNs with one or more hidden layers can separate the space in different areas and build different functions for each of them. This means that ANNs have the capacity to build non linear piece-wise models. Collopy and Armstrong (1992) reviewed many forecast experts' opinions which agree with the importance of the existence of this kind of models, capable of identifying and treating abrupt changes in a time series pattern.

Some statistical methods for time series treatment have limitations due to the way they are estimated. This means that the forecaster has to supervise this phase, which leads to most statistical models having to be re-estimated periodically when new data are added. In contrast, estimation with ANNs can be automatized (Hoptroff, 1993) and model review is not necessary, given the fact that they learn automatically.

Any forecasting method that learns from past observations in order to predict the future has one problem: they cannot handle series that are not long enough. On the other hand, having too many data in a time series can also sometimes be a problem. For ANNs, having a set of learning examples that is long enough is important, but if this set is excessive a long time would quite possibly be required to achieve adequate solutions. Weigend and Gershenfeld (1993) reported that, in order to calculate forecasts for a time series composed of 100,000 observations, 100 hours on a Connection Machine CM-2 with 8,192 floating point processors were necessary. This problem will increase with the size of the net, and can be serious if forecasts are needed in a short period of time.

In order to solve this problem, we would like to highlight two methods. First of all, the use of faster algorithms such as *Backpropagation with Momentum* (Rumelhart and McClelland, 1986); this algorithm has some weaknesses, so that other algorithms were developed to accelerate the process by finding good local minimums if the absolute minimum cannot be reached. *Backpropagation with Weight Decay* (Werbos, 1988), figures amongst the methods; in this case a decrease of the weights for the connections takes place during the training process. The *Quickprop* algorithm (Fahlman, 1989) tries to accelerate the process by using information about the error surface curvature and attempts to find the optimal in just one step. The *RPROP* algorithm (Resilient backPROPagation, Riedmiller and Braun, 1993) estimates the change for every weight separately, using the topology of the error surface. Other algorithms and variations can be found in Parker, 1987; Jacobs, 1988; Tolleanere, 1990; Pack et al., 1991 (a, b); Fog et al., 1996; Cottrell et al., 1995.

Another possibility is to make a prior selection of the training examples in order to work with a training set composed of the fewer number of examples that ensure the correct learning of the net, which is what we do in this paper.

One of the main features of ANNs is its speed, so it is possible to achieve solutions immediately for some kind of ANNs, and in a very short time for most of them. Some conditions should exist for the latter to be true (in the Multilayer Perceptron case, MLP):

1. Network configuration cannot be too large, so that the number of connections whose weights have to be calculated is not large.
2. The training examples set cannot be too long. The smaller the set, the more number of times that every example will go through the network and the faster the solution will be obtained.
3. Composition of training examples has to be homogeneous. The more similar the examples, the faster the learning process.

According to this, a way to accelerate the process is to make a prior selection of the patterns or training examples, so that only the necessary patterns are chosen and all the underlying relationships have to be represented in these chosen patterns (Plutowski and White, 1991; Deco et al., 1997). In this case, the training process will be shorter, so that the ANN will be ready for forecasting in a short period of time.

In recent years, some work used algorithms based on the concept of continuous and selective training (Peng et al., 1992; Ho et al., 1992; Vermaak and Botha, 1998). These algorithms can be implemented thanks to the improvement in computer features during these years.

The process for calculating the forecast for a given period of time is the following: with the N observations of the time series ($Y_1, Y_2, ..., Y_N$), using an ANN with n input nodes, we have a training data base composed of N-n training patterns (the first will be composed of $Y_1, Y_2, ...,$

Y_n as inputs and Y_{n+1} as the target output; the second will contain $Y_2, Y_3, ..., Y_{n+1}$ as inputs and Y_{n+2} as output, and the last pattern will be $Y_{N-n}, Y_{N-n+1}, ..., Y_{N-1}$ as inputs and Y_N as desired output). Now, the observation to be forecast (e.g. $\hat{Y}_{N+1}$) will have a known input vector, composed of the n past values ($Y_{N-n+1}, Y_{N-n+2}, ..., Y_N$); then, a comparison can be made between this vector with all the input vectors of the examples contained in the training database (that is composed by all the vectors that represent past situations whose solutions are already know). This way, those examples which are more similar to the one to be forecast, can be selected. Then, the ANN is trained with these examples and, when this process is over, the forecast is calculated. Finally, the forecast example is added to the training data base in order to be used in other forecasts. All this process is depicted in figure 1.

The similarity between the input vector $\hat{Y}$ and the stored patterns, $Y^{(k)}$ ($k = 1, 2, ..., N-n$), can be measured by (1). This formula is not very appropriate when one of the vector parts is very different and the rest of the parts are very similar, because such vectors would be too different for the formula to choose the pattern, whereas the reality is that they differ in just one part.

$$d(\hat{Y}, Y^{(k)}) = \sqrt{\sum_{i=1}^{n} \left(Y_i^{(k)} - Y_i\right)^2} \tag{1}$$

In order to overcome this drawback, we propose a pattern selection process based on ART networks (Carpenter and Grossberg, 1987). These networks are able to classify the vectors that are being introduced as inputs to the net, based on their similarity with the vectors already classified. To achieve this, a similarity test is carried out. In this test, the similarity of a vector with the vectors belonging to a given category is quantified, and then, based on a previously set similarity coefficient, it is decided whether the new vector belongs to that category. The similarity coefficient can take values between 0 and 100 (value 100 means that the vector needs to be identical to the one representing the group).

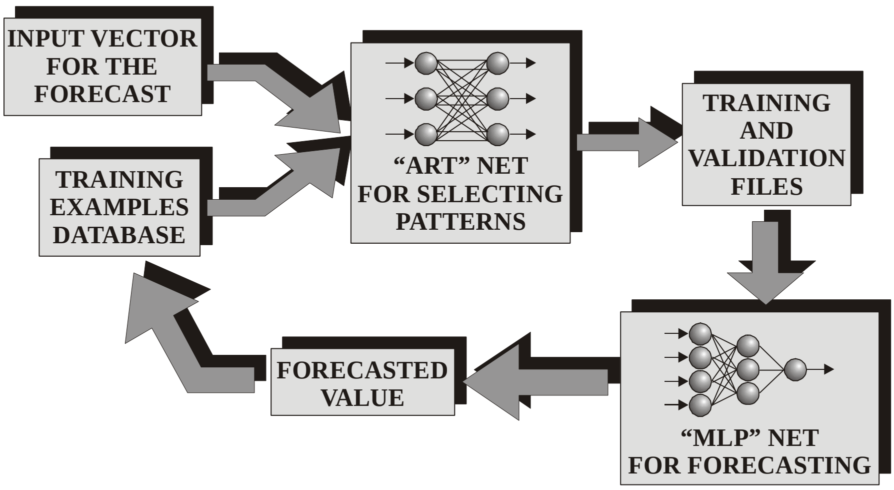


Figure 1. Selective and continuous method to neural networks training.

In our case, the input vector of the case to be forecasted is the first to be incorporated to the net, so that it will be classified in the first category. Then all the patterns included in the training data base, are put into the ART network, and the net will classify in the same first category the examples whose inputs are similar to the one that is going to be forecast, based on the previously set similarity coefficient. All this process is very fast, so that in a few seconds, we have selected the examples more similar to the one to be forecast.

All these selected examples are stored in a file from which the training and the validation sets for the MLP will be obtained. These two sets (training and validation) are composed of a small number of examples (up to 200 and 40 for training and validation respectively). The examples conforming these two sets are also very similar to the example to be forecasted. This fact makes the MLP training process very fast, so the required forecast can be made in a few minutes.

4. Forecasting the Electricity Market in Spain

As it has been pointed out in paragraph 2, the Electricity Production Market in Spain is the set of transactions of certain agents in the Daily and Intra-Daily Market sessions, along with the application of the Procedures of Technical Operations of the System.

The Daily Market is a part of the Electricity Market, and its aim is to carry out the transactions of electric power for the following day when the calculations are made. To do so, the market agents have to present their offers to sell and buy electricity. The Market Operator match up supply and demand, with the aim of assigning the marginal price (€/kW) and the production and demand among the participating agents. The result of this process is the Basic Working Daily Program (BWDP), in which prices and electric power quantities to be sold and bought for the next 24 hours are set.

We will use ANNs to forecast this BWDP time series. We will try to forecast the price in Euros/kW for each one of the 24 hours of a given day. To do so, the Hourly Electricity Price data for the Daily Market Price (Hourly Price, HP) and the Electricity Demand (DEM) time series are used, from January 1st, 1998, to September 30th, 2004. There are more than 60,000 observations in each time series. Given that the time series are so long, the ANN training process needs to be accelerated, by using a selection method like the one proposed before.

In order to test the forecasting performance of the proposed method, forecasts for the Hourly Price time series for the 24 hours in 12 different days and months were carried out. To forecast the 24 hours of any given day, observations from January 1st, 1998, until the day before the one to be forecast are used. The days to be forecast and the number of observations are shown in table 1.

Table 1. Days to forecast and available data.

Day to forecast (month/day/year)	***From (hour-day)***	***Until (hour-day)***	***No. of observations***
02/12/99 (Friday)	*01:00 - 01/01/98*	*24:00 - 02/11/99*	*9,769*
07/07/99 (Wednesday)	*01:00 - 01/01/98*	*24:00 – 07/06/99*	*13,251*
01/29/00 (Saturday)	*01:00 - 01/01/98*	*24:00 – 01/28/00*	*18,193*
09/20/00 (Wednesday)	*01:00 - 01/01/98*	*24:00 – 09/19/00*	*23,833*
11/01/01 (Thursday)*	*01:00 - 01/01/98*	*24:00 – 10/31/01*	*34,129*
12/03/01 (Monday)	*01:00 - 01/01/98*	*24:00 – 12/02/01*	*34,369*
03/18/02 (Monday)	*01:00 - 01/01/98*	*24:00 – 03/17/02*	*36,889*
08/08/02 (Thursday)	*01:00 - 01/01/98*	*24:00 – 08/07/02*	*40,321*
05/13/03 (Tuesday)	*01:00 - 01/01/98*	*24:00 – 05/12/03*	*46,993*
10/12/03 (Sunday)	*01:00 - 01/01/98*	*24:00 – 10/11/03*	*50,641*
04/17/04 (Saturday)	*01:00 - 01/01/98*	*24:00 – 04/16/04*	*55,153*
06/01/04 (Tuesday)	*01:00 - 01/01/98*	*24:00 – 05/31/04*	*56,233*

(*) National holiday.

The quality of forecasts will be tested against forecasts calculated with the well-known Box-Jenkins (BJ) methodology.

Two different methods are used to calculate 24 Hourly Price forecasts for each day subject to analysis:

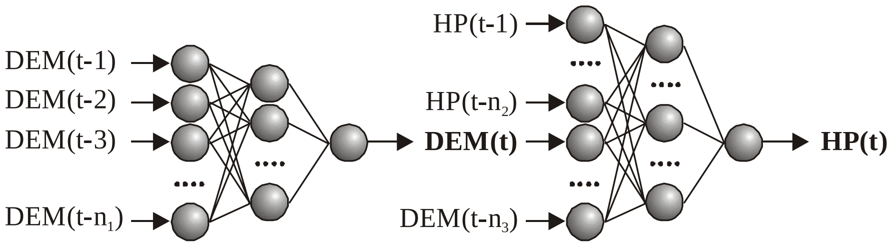
- 24 "*one-step ahead*" forecasts; for each case, real data up to the time period prior to forecast are used:

$$s_t = NN(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, \ldots)$$
$$s_{t+1} = NN(x_t, x_{t-1}, x_{t-2}, x_{t-3}, \ldots)$$
$$\ldots\ldots \quad \ldots\ldots$$
$$s_{t+23} = NN(x_{t+22}, x_{t+21}, x_{t+20}, \ldots)$$

- "*24-step ahead*", which have been calculated in an iterative way (Zhang et al, 1998); obtained forecasts are used to calculate the next ones:

$$s_t = NN(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, \ldots)$$
$$s_{t+1} = NN(s_t, x_{t-1}, x_{t-2}, x_{t-3}, \ldots)$$
$$\ldots\ldots \quad \ldots\ldots$$
$$s_{t+23} = NN(s_{t+22}, s_{t+21}, s_{t+20}, \ldots)$$

For both cases, the process has been the same: due to the strong relationship between the Hourly Price and Hourly Demand series, forecast process is composed of two phases (figure 2).



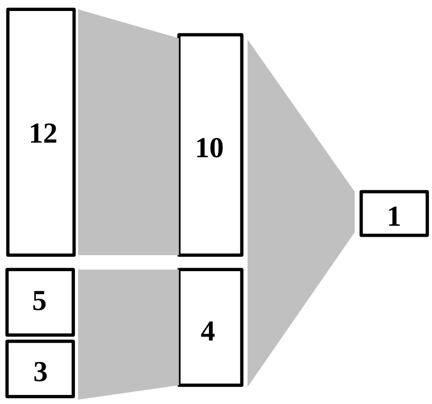


Figure 3. Multilayer Perceptron NN1 (shaded areas represent the connections from all left nodes to all the right nodes).

A first artificial neural network (NN1; 20-14-1 represented in figure 3) is used to forecast Electricity Demand for the hour "t" (table 2 shows the structure of the input layer for this neural network). As it can be seen in the net structure, hidden layer has been divided into two sets: a first set composed of ten neurones does the processing of the previous twelve data from the Hourly Price series. A second set composed of four neurones does the processing of the day and hour to be forecast. The use of these separated nets usually has some advantages in the case of multivariate forecasting (Chen et al., 1992; Choi et al., 1997). In these nets, the hidden layer is separated in sections, each of them processing one kind of information from the input layer. The final configuration was chosen after several tests that included different numbers of neurones in the input and hidden layers.

Table 2. Composition of NN1 input layer.

Node	***Description***
1	***DEM(t-1)***
2	***DEM(t-2)***
3	***DEM(t-3)***
4	***DEM(t-4)***
5	***DEM(t-5)***
6	***DEM(t-6)***
7	***DEM(t-7)***
8	***DEM(t-8)***
9	***DEM(t-9)***
10	***DEM(t-10)***
11	***DEM(t-11)***
12	***DEM(t-12)***

Node	***Description***					
13	***0***	***0***	***0***	***0***	***.....***	***1***
14	***0***	***0***	***0***	***0***	***.....***	***0***
15	***0***	***0***	***0***	***0***	***.....***	***1***
16	***0***	***0***	***1***	***1***	***.....***	***1***
17	***0***	***1***	***0***	***1***	***.....***	***1***
	1	*2*	*3*	*4*	*.....*	*24*

Hour of Day

18	***0***	***0***	***0***	***0***	***1***	***1***	***1***
19	***0***	***0***	***1***	***1***	***0***	***0***	***1***
20	***0***	***1***	***0***	***1***	***0***	***1***	***0***
	M	*T*	*W*	*T*	*F*	*Sa*	*Su*

Day of Week

In second place, the forecast value of DEM series, is used as an input for the input layer of a second neural network (named NN2), that will be used to forecast the Hourly Price. As can be seen in figure 4, NN2 is a partially separated MLP; there is a main net, 18-10-1, which is supported by another net, 6-3-1, that processes the information from the DEM series. Moreover, there are other two separated nets: the 5-3-1 net, which processes the information from the hour to be forecast; and the 3-2-1 net, which processes the information from the day of the week that is going to be forecast. Table 3 shows the composition of the input layer.

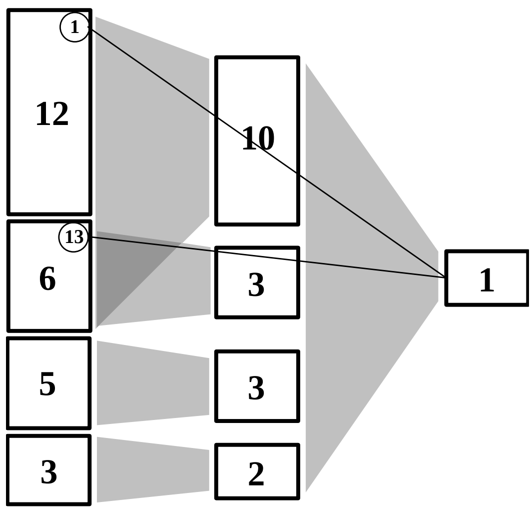


Figure 4. Multilayer Perceptron NN2 (shaded areas represent the connections from all nodes on the left to all nodes on the right).

Table 3. Composition of NN2 input layer.

Node	Description
1	HP(t-1)
2	HP(t-2)
3	HP(t-3)
4	HP(t-4)
5	HP(t-5)
6	HP(t-6)
7	HP(t-24)
8	HP(t-25)
9	HP(t-26)
10	HP(t-27)
11	HP(t-28)
12	HP(t-29)
13	DEM(t)
14	DEM(t-1)
15	DEM(t-2)
16	DEM(t-3)
17	DEM(t-4)
18	DEM(t-5)

Node	Description					
19	0	0	0	0	.....	1
20	0	0	0	0	.....	0
21	0	0	0	0	.....	1
22	0	0	1	1	.....	1
23	0	1	0	1	.....	1
	1	2	3	4	.....	24

Hour of Day

24	0	0	0	0	1	1	1
25	0	0	1	1	0	0	1
26	0	1	0	1	0	1	0
	M	T	W	T	F	Sa	Su

Day of Week

In some cases, *shortcuts* or direct links, are made between any neuron from the input and output layers; this is to enforce the effect of some variable, or to model some linear relationship among input variables (Duliba, 1991; Chen et al., 1992). Sometimes it can be proved that the using of this shortcut enhances the ANN capacity to forecast, although in other cases the use of shortcuts only leads to a longer training time due to the existence of more links. In our case, after some trials, we decided to use 2 shortcuts, between neuron 1 and 13, which belong to the value immediately prior from Hourly Price and the value from period “t” of DEM time series.

As a sample of the results obtained, Figures 5 and 6 show 24 one-step ahead ANN and BJ forecasts, compared with real data for two of the twelve days analyzed: Monday, March 18th, 2002 and Sunday, October 12th, 2002. Figures 7 and 8 show 24-step ahead ANN and BJ forecasts compared with real data for these two days.

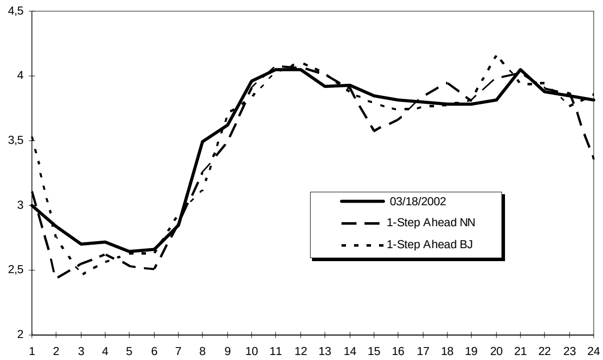


Figure 5. One-Step Ahead, Monday March 18th

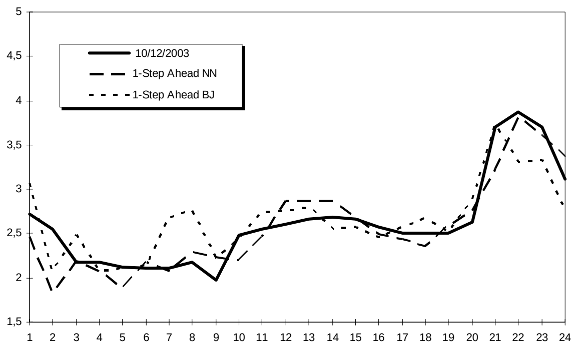


Figure 6. One-step ahead, Sunday, October 12th

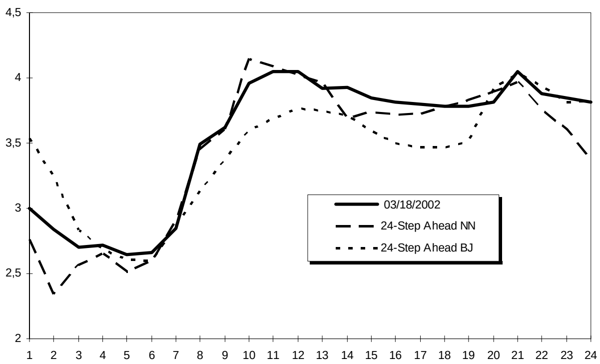


Figure 7. 24-Step Ahead, Monday March 18th

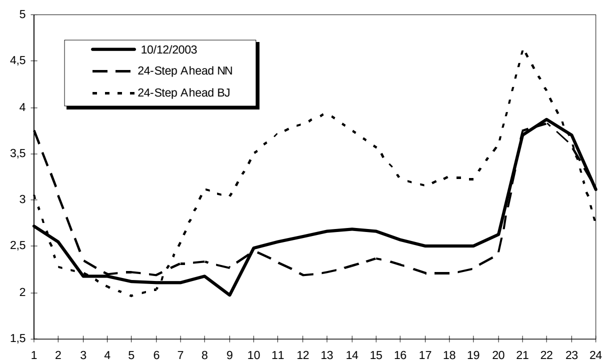


Figure 8. 24-step ahead, Sunday, October 12th

Table 4 shows the Mean Absolute Percentage Error (MAPE, introduced by Makridakis, 1993), for every case studied.

$$MAPE = \frac{1}{p}\sum \frac{|e_t|}{\frac{1}{2}|y_t + \hat{y}_t|} \quad (2)$$

where e_t is the forecast error for period t; y_t and $\hat{y}_t$ are the actual and forecast values, and p is the number of forecasts.

Table 4. Results in terms of MAPE.

Day to Forecast (month/day/year)	*Hourly Price (One-Step Ahead)*		*Hourly Price (24-Step Ahead)*	
	Neural Networks	*Box-Jenkins*	*Neural Networks*	*Box-Jenkins*
02/12/99 (Friday)	*5.77*	***4.78***	*8.48*	*5.04*
07/07/99 (Wednesday)	*4.37*	***2.74***	*6.90*	*4.49*
01/29/00 (Saturday)	***8.18***	*12.12*	*10.18*	*17.25*
09/20/00 (Wednesday)	***3.27***	*3.81*	*4.15*	*7.59*
11/01/01 (Thursday)*	***6.66***	*12.91*	*9.90*	*27.18*
12/03/01 (Monday)	*5.31*	*5.63*	***4.79***	*19.73*
03/18/02 (Monday)	*3.80*	***3.51***	*3.96*	*5.88*
08/08/02 (Thursday)	*5.74*	***3.60***	*5.01*	*4.34*
05/13/03 (Tuesday)	***5.10***	*5.31*	*5.54*	*14.82*
10/12/03 (Sunday)	***6.91***	*8.41*	*9.79*	*21.61*
04/17/04 (Saturday)	***2.53***	*4.67*	*6.30*	*7.41*
06/01/04 (Tuesday)	***5.43***	*8.49*	*7.63*	*13.34*

(*) National holiday.

From data shown in Table 4, some interesting conclusions can be drawn: first of all, for eight out of the twelve days analyzed, better forecasts are obtained with ANNs. Second of all, for every day but just one, better results are obtained with one-step ahead forecasting rather than 24-step ahead forecasting. This is due to the fact that for the one-step ahead method, ANNs always use real data, and BJ re-estimates the parameters for each forecast. When calculating 24-step ahead forecasts, ANNs obtain each forecast using the previously calculated ones, and BJ uses the same model to calculate every forecast.

Besides, the four days for which BJ obtain better forecasts correspond to labour days, while ANNs always calculate better forecasts than BJ for weekends and/or holidays. When forecasting labour days, the performance of both methods is similar in terms of MAPE, as can bee seen from data in Table 4. But when weekends or holidays are to be forecast, then ANNs clearly outperform BJ.

A relevant situation occurs on Thursday, November 1st, 2001, which is a national holiday. ANNs training samples selection method uses the most similar samples to the day to be forecast. And in this case, even though it is not a weekend day, the most similar samples correspond to weekends. So that, from ANNs point of view, forecasting November 1st is like forecasting a weekend day, and MAPE value is similar to the ones obtained when forecasting weekends.

Average values for MAPE are calculated for the eight labour days, and for the four holidays and/or weekend days considered in this study (see Table 5).

Table 5. MAPE Average values.

Day to Forecast (month/day/year)	*Hourly Price (One-Step Ahead)*		*Hourly Price (24-Step Ahead)*	
	Neural Networks	*Box-Jenkins*	*Neural Networks*	*Box-Jenkins*
Labour Days	*4.84*	***4.73***	*5.80 (+19.8%)*	*9.40 (+98.7%)*
Weekend and/or Holidays	***6.07***	*9.52*	*9.04 (+48.9%)*	*18.36 (+92.8%)*

Data in this table shows that better forecasts are obtained when forecasting Hourly Price for labour days than weekends or holidays. In particular, ANNs achieve better forecasts than BJ except for the one-step ahead forecasting for labour days. When forecasting 24-step ahead, MAPE values increase for both forecasting methods, from almost 100% increase for BJ in weekends and labour days, to 49% and 20% for ANNs in weekends and labour days respectively.

The reason for these MAPE increases when forecasting 24-step ahead resides in the fact that more updated information is used when forecasting one-step ahead.

Table 5 also shows that ANNs perform better for labour days than weekends and/or holidays. This is due to the fact that the database of examples is composed of fewer examples corresponding to weekends than to labour days. So that, the selected set of training examples is more heterogeneous for weekends, which leads to a worse quality of the training process, and then poorer forecasts are calculated.

5. Conclusions

In this work, forecasts for a time series from the Spanish Electricity Production Market were calculated. This market arose as a result of the liberalisation process in the Electricity sector in Spain.

Although it is a similar market to the Stock Market, the inherent characteristics of the product that is sold and bought gives this market special features that make it very interesting for forecasting purposes. For instance, the Electricity Market is not influenced by as many exogenous factors as the Stock Market. For these reasons, forecasting time series from this market should be easier, and the quality of results could be related to the forecasting method used to obtain them. Thus, it would be easy to evaluate the predictive quality of the method used.

When forecasts for the Hourly Price series are calculated using ANNs, we come up against the problem of having too many observations. As this Market has been working for a few years, the Electricity Hourly Price time series is composed of thousands of observations. This abundance of data makes ANN training phase very slow, taking hours or even days, which is against the initial purpose of making forecasts for the next 24 hours of a given day.

This problem led us to design a training method in which it is possible to calculate forecasts in a few minutes. There are many other methods that try to accelerate the training process, but the advantage of the proposed method resides in the use of an ART Neural Net to make the selection of training examples. This kind of net has been broadly used with great success in the solving of complex classification problems, and this is the reason why we integrated them into our system. This process ends with the selection of a few samples (around 200) very similar among them and to the one to be forecast. This makes the MLP training process much faster.

The final results suggest that ANN forecasts with the proposed training method outperform those calculated with the BJ methodology, particularly when the days to be forecast are weekend days or national holidays.

References

Box, G.E. and Jenkins, G.M. (1970): *Time Series Analysis*, Holden-Day, San Francisco.

Carpenter, G.A. and Grossberg, S. (1987): "A Massively Parallel Architecture for a Selforganizing Neural Pattern Recognition Machine". *Computer Vision, Graphics and Image Processing*, 37, pp. 54-115.

Chen, S.T.; Yu, D.C. and Moghaddamjo, A.R. (1992): "Weather Sensitive Short-Term Load Forecasting Using Nonfully Connected Artificial Neural Networks". *IEEE Transactions on Power Systems*, 7 (3), pp. 1098-1105.

Choi, H.R.; Kim, W. and An, S.Y. (1997): "Recurrent and decomposed neural network-based hotel occupancy forecasting". *The New Review of Applied Expert Systems*, 1997, pp. 121-136.

Collopy, F. and Armstrong, J.S. (1992): "Expert opinions about extrapolation and the mystery of the overlooked discontinuities". *International Journal of Forecasting*, 8, pp. 575-582.

Contreras, J.; Espinola, R.; Nogales, F.J. and Conejo, A.J. (2003): "ARIMA models to predict next-day electricity prices". *IEEE Transactions on Power Systems*, 18 (3), pp. 1014-1020.

Cottrell, M.; Girard, B.; Girard, Y.; Mangeas, M. and Muller, C. (1995): "Neural modeling for time series: a statistical stepwise method for weight elimination". *IEEE Transactions on Neural Networks*, 6 (6), pp. 1355-1364.

Cybenko, G. (1989): "Approximation by Superpositions of a Sygmoidal Function", *Mathematics of Control, Signals and Systems*, 2, pp. 303-314.

Deco, G.; Neuneier, R. and Shürmann, B. (1997): "Non-parametric Data Selection for Neural Learning in Non-stationary Time Series". *Neural Networks*, 10 (3), pp. 401-407.

Duliba, K.A. (1991): "Contrasting neural nets with regression in predicting performance in the transportation industry". *Proceedings of the Annual IEEE International Conference on Systems Sciences*, 25, pp. 163-170.

Fahlman, S.E. (1989): "Faster-learning variations of back-propagation: An empirical study". *Proceedings of the 1988 Connectionist Models Summer School*, pp. 38-51.

Fog, T.L.; Larsen, J. and Hansen, L.K. (1996): "Training and Evaluation of Neural Networks for Multivariate Time Series Processing". *Proceedings of IEEE International Conference on Neural Networks*. Perth, Australia.

Ho, K.L.; Hsu, Y.Y. and Yang, C.C. (1992): "Short-term load forecasting using a multilayer neural network with an adaptive learning algorithm". *IEEE Transactions on Power Systems*, 7 (1), pp. 141-149.

Hoptroff, R.G. (1993): "The principles and practice of time series forecasting and business modeling using neural nets". *Neural Computing and Applications*, 1, pp. 59-66.

Hornik, K.; Stinchcombe, M. and White, H. (1989): "Multilayer feedforward networks are universal approximators". *Neural Networks*, 2, pp. 359-366.

Jacobs, R.A. (1988): "Increased rates of convergence through learning rate adaptation", *Neural Netwoks*, 1 (4), pp. 295-308.

Makridakis, S. (1993): "Accuracy Measures: Theoretical and Practical Concerns". *International Journal of Forecasting*, 9, pp. 527-529.

Nogales, F.J.; Contreras, J.; Conejo, A.J. and Espinola, R. (2002): "Forecasting next-day electricity prices by time series models". *IEEE Transactions on Power Systems*, 17 (2), pp. 342-348.

Pack, D.C.; El-Sharkawi, M.A. and Marks II, R.J. (1991a): "An adaptively trained neural network". *IEEE Transactions on Neural Networks*, 2 (3), pp. 334-345.

Pack, D.C.; El-Sharkawi, M.A.; Marks II, R.J.; Atlas, L.E. and Damborg, M.J. (1991b): "Electric load forecasting using an artificial neural network". *IEEE Transactions on Power Systems*, 6 (2), pp. 442-449.

Parker, D.B. (1987): "Optimal algorithm for adaptive network: second order back propagation, second order direct propagation, and second order Hebbian learning". *Proceedings of the IEEE International Conference on Neural Networks 2*, IEEE Press, pp. 593-600.

Peng, T.M.; Hubele, N.F. and Karady, G.G. (1992): "Advancement in the application of neural networks for short-term load forecasting". *IEEE Transactions on Power Systems*, 7 (1), pp. 250-257.

Plutowski, M. and White, H. (1991): "Active selection of training examples for network learning in noiseless environments". *Technical Report No. CS91-180*. University of California, San Diego.

Priestley, M.B. (1988): *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, London.

Riedmiller, M. and Braun, H. (1993): "RPROP – A Fast Adaptive Learning Algorithm". *Proceeding of ISCIS VII*.

Rumelhart, D.E. and McClelland, J. (1986): *Parallel Distributed Processing*. MIT Press, Cambridge.

Subba Rao, T. (1992): "Analysis of Nonlinear Time Series (and Chaos) by Bispectral Methods". In *Nonlinear Modeling and Forecasting*, (Casdagli, M. and Eubank, S., ed.). Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol XII, pp. 199-226.

Tolleanere, T. (1990): "SuperSAB: Fast Adaptive Backpropagation with good Scaling Properties". *Neural Networks*, 3, pp. 561-573.

Tong, H. (1983): *Threshold Models in Non-Linear Time Series Analysis*. Springer-Verlag, New York.

Tong, H. (1990): *Non-Linear Time Series: A Dynamic System Approach*. Oxford University Press, Oxford.

Tsay, R.S. (1991): "Non-Linear Time Series Analysis: Diagnostics and Modelling". *Stat. Sinica*, 1, pp. 432-451.

Vermaak, J. and Botha, E.C. (1998): "Recurrent Neural Networks for Short-Term Load Forecasting". *IEEE Transactions on Power Systems*, 13 (1), pp. 126-132.

Weigend, A.S. and Gershenfeld, N.A. (1993): *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison-Wesley, Reading, MA.

Werbos, P. (1988): "Backpropagation: Past and Future". *Proceedings of the IEEE International Conference on Neural Networks*, IEEE Press, pp. 343-353.

Yule, G. (1927): "On a Method of Investigating Periodicity in Disturbed Series with special reference to Wolfer's Sunspot Numbers". *Phil. Trans. Royal Society London*, 226, pp. 267-298.

Zhang, G.; Patuwo, B.E. and Hu, M.Y. (1998): "Forecasting with artificial neural networks: The state of the art". *International Journal of Forecasting*, 14 pp. 35-62.