Electric energy demand forecasting with neural networks

Diego Carmona
Dpto. Ing. Química y Energética
Universidad de Extremadura
Avda. Elvas, 06071 Badajoz (Spain)
dcarmona@unex.es

Miguel A. Jaramillo
Dpto. Electrónica e Ing. Electromec.
Universidad de Extremadura
Avda. Elvas, 06071 Badajoz (Spain)
miguel@unex.es

Eva González
Dpto. Ing. Química y Energética
Universidad de Extremadura
Avda. Elvas, 06071 Badajoz (Spain)
evagzlez@unex.es

J. Antonio Álvarez
Dpto. Ing. Química y Energética
Universidad de Extremadura
Avda. Elvas, 06071 Badajoz (Spain)
jalvarez@unex.es

***Abstract* – Electric energy demand forecasting represents a fundamental information to plan the activities of the companies that generate and distribute it. So a good prediction of its demand will provide an invaluable tool to plan the production and purchase policies of both generation and distribution or reseller companies. This demand may be seen as a temporal series when its data are conveniently arranged. In this way the prediction of a future value may be performed studying the past ones. Neural networks have proved to be a very powerful tool to do this. They are mathematical structures that mimic that of the nervous system of living beings and are used extensively for system identification and prediction of their future evolution. In this work a neural network is presented to predict the evolution of the monthly demand of electric consumption. A Feedforward Multilayer Perceptron (MLP) has been used as neural model with Backpropagation as learning strategy. The network has three hidden layers with a 8-4-8 distribution. It takes twelve past values to predict the following one. Errors smaller than 5% have been obtained in most of the predictions.**

I. INTRODUCTION

Electric energy demand forecasting is a fundamental strategy for companies working on its production and distribution. As there is a growing consumption of electric energy in present world a precise prediction of its demand is fundamental for companies because this demand controls their economical evolution. A precise forecast of this demand will provide these companies with a valuable information to adapt production and distribution to the society needs. In this respect it is useful to differentiate between short and long term forecasting. The first one looks for a prediction of the electric demand with a horizon of hours, days or even weeks. It allows companies that generate and distribute electric energy to control their stocks and coordinate supply and demand. The second one works with monthly data and provides both producers and distributors with a forecast of the evolution of the demand, which allows the definition of strategies to augment the capability of the distribution lines or the construction of new production plants and policies to obtain new clients.

Short term prediction is very sensitive to specific distortions in the demand caused by the hour of the day, the day of the week, temperature falling or rise or another weather phenomena. A certain combination of all these factors may produce an unusual variation in the demand that must be forecasted. So it is necessary to consider all these data when performing a prediction. Sophisticated models are then needed to deal with this complex evolution. On the other hand the presence of outliers caused by unpredictable facts must be treated as perturbations and it will be necessary to define those models as robust as possible to reject their influence.

Long term prediction deals with data that rarely present large distortions in their temporal evolution. They are presented as the monthly total electric demand where specific peaks or valleys are diluted in the overall information considered, so their effect in the data is small. In this way any kind of isolated distortion in the demand will have little influence in its overall monthly value. Nevertheless, although their effect is not very important in the evolution of the time series, they generate fluctuations that are difficult to take into account because their causes are not provided. In any case long term demand forecasting needs less accurate predictions than the short time one since it influences global decisions regarding overall productions or purchases where small fluctuations have little influence. So a less robust prediction may be done avoiding the use of the sophisticated tools that are needed to deal with short term predictions. In the same way, as the influence of weather conditions or the kind of the day (holiday or weekday) is diluted in the monthly overall consumption, the forecasting model may use only past elements of the time series to define the time history of the electric demand. So an even more simplified structure may be used as forecasting tool.

As it has been pointed out previously the electric demand forecasting may be studied as a time series prediction problem, so tools which provide good results in solving this problem may be used to perform electric demand forecasting. Among them neural networks have shown to be an important tool because of their flexibility and easy configuration for solving the time series prediction problem, a fact that is hardly surprising if it is taken into account that recurrent neural networks may be considered as a special case of nonlinear autoregressive moving average models (NARMA), a very powerful tool for predicting the time series evolution [1]. So neural networks have become popular tools for the electric demand forecasting [2], [3], [4]. These works are mainly devoted to short term prediction, using the neural network capability of dealing with different kinds of data as the variables to be processed [4]. Nevertheless their use for long term forecasting have not received the same attention. In this work neural networks have been used to perform this work and several structures have been tested to find which of them performs the best.

0-7803-7474-6/02/$17.00 ©2002 IEEE

The work is organized as follows. Section II presents the data that will be used to train and validate the proposed models along with the network structure. In Section III some networks with different sizes and structures are tested and the results obtained with that performing the best are presented. Section IV describes the conclusions obtained from simulation.

II. NEURAL NETWORK STRUCTURE

As the time series describing the long term demand of electric energy is formed with the monthly overall consumption, a very sophisticated neural network is not needed to learn and then predict that electric demand. This is so because the differences in consumption that each day of the week presents have no influence in the monthly value since it is obtained as the sum of all the daily ones. In the same way the influence of the weather irregularities has a predictable effect in the time series elements because the variations from the usual monthly values of a certain meteorological parameter is usually embedded in a more general climate behavior whose effects may be detected from the analysis of the past values of the time series. So only the elements of the time series can be considered to perform the learning of the inherent monthly evolution of the demand, since it is reasonable to assume that the same month in different years will have similar consumptions whose differences will depend on the evolution of the demand in the past months. On the other hand, although outliers that may appear in short term consumption series impose the use of robust neural models to avoid their influence in the learning process, they produce a small effect in the monthly value of the electric consumption and the resulting time series will not present distortions of this kind. So it will not be necessary to use sophisticated models to reject the effects of this disturbances.

Taking into account all the aforementioned considerations the selected model is a Multilayer Feedforward Perceptron [5], one of the most popular neural structures because of its simplicity and easy configuration, but with enough prediction capability as to be used to predict time series of the kind considered.

A. Input data

In this work a simplified neural model is proposed to perform the long term demand forecasting based on the relaxation of the needs the short term prediction has. Nevertheless it has not been taken into account a very important factor that acts on that long term forecasting and was not considered for the short term one, where it had no effect: the influence of the economic and technological evolutions on the electric market. In fact, as the wealth of most nation presents a rising tendency and the technological development provides the society with more and more devices that need the electric energy to work, the electric demand suffers a constant rising tendency.

This general rising trend clearly appears in the example used to test the presented model. These data represent the Spanish overall monthly electric consumption from January 1959 to September 2000 (Fig. 1). They may be shared out into three periods when the general market trend is considered: the first from 1959 to 1973, the second from 1974 to 1991 and the last from 1992 to the end of the series.

The first one presents a strong rising in the monthly consumption. Increases higher than 10% appear every year except 1960 (8.46 %), 1963 (9.46%) and 1971 (7.10%). Indeed some months present very high increases with respect to the preceding year as September 1963 (22.7%), December 1964 (24.2%) and November 1966 (19.1%).

The second period begins with the well known 1973 energy crisis that provoked a decrease in the rising trend of the electric demand. So the annual consumption increases under 5% except in 1974 (7.35%), 1978 (5.62%) and 1979 (6.89 %). The smallest increments appear in 1981 (0.11%) and 1986 (1.85%). A new situation appears in the monthly demand: a decrease in the consumption in relation to the same month in the preceding year. This fact appears in August 1975 (-0.5%), December 1981 (-5.4%) and April 1987 (-5.5%).

In the third period the increase in the demand is even smaller than in the second, with values around 3%. In this way 1993 represents an unusual year because it suffers a decrease in the electric demand (-0.04%). Nevertheless in spite of this annual slow increase the series shows important monthly peaks (February 1996 with 11.88%, March 1998 with 11.11%, July 1998 with 10.38%, November 1999 with 10.47% or January 2000 with 10.21%) along with other very low increases.

As it can be seen in Fig. 1 the general rising trend of the electric demand have a very important influence in the evolution of the time series. So this effect must be considered in the definition of the neural network.

Unfortunately these structures have prediction problems when dealing with time series that present a tendency of this kind because of the use of nonlinear saturating functions as neural outputs that impose boundaries to the network outputs. So the model will not be able to predict values of the time series falling out of that boundaries. Therefore it is necessary to define a procedure that eliminates that rising tendency from the time series. So a normalization process which transforms the original time series into another where the rising trend does not appear is defined and only the monthly evolution is retained.

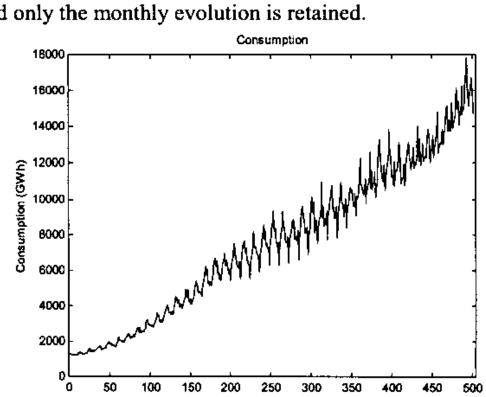

Fig. 1. Monthly Spanish electric consumption from January 1959 to September 2000.

B. Normalization of the input data

As the aim of the normalization process is to detect the monthly evolution throughout a year it is necessary to eliminate the general rising tendency of the time series. This may be done by representing every datum as a sort of variation from that tendency. So every datum will be substituted by another obtained by dividing each one by the sum of a group of the preceding data including itself. Two options have been tested to predict the following value of the time series based on the number of inputs presented to the network:

1) Twelve values are presented to the network. Every datum is divided by the sum of the preceding eleven along with itself. So a new time series is obtained that is to be used to perform both the learning and the forecasting processes (Fig. 2). This new series has lost its eleven first values during the normalization process because there was not enough data to perform their normalization. This normalization will be named in the following as "Option 1". Here a yearly data evolution is considered, because as the electric demand presents similar monthly patterns for different years it is reasonable to assume that a normalization that spans a year back will extract that monthly evolution from the overall yearly tendency.

2) Six values are presented to the network. The normalization constant is obtained in a similar way to that in "Option 1" with the only difference that five past values are considered. Here only five data are lost from the original time series. This will be named as "Option 2". Here the seasonal influence in the monthly demand is taken into account so that the network ought to predict the next value in the time series from the preceding six. These six past values may be enough to detect that seasonal influence.

On the other hand, when a prediction is done the value the network provides must be transformed into a value in the same scale the original time series was. Therefore a process opposite to the normalization must be performed. In this way each output was multiplied by the sum of the twelve actual data in the network input (the value of each input without normalization) for "Option 1" or by the six input actual data in "Option 2".

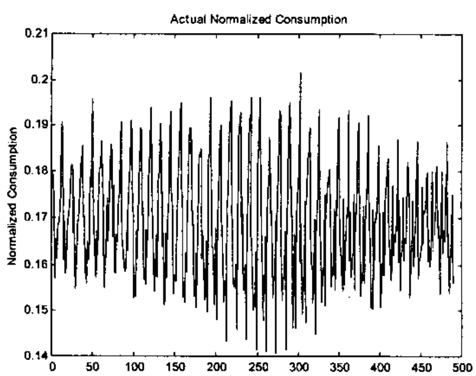

Fig. 2. Normalized monthly Spanish electric consumption with "Option 1" from January 1959 to September 2000.

C. Network structure

As it has been pointed out previously a Multilayer Perceptron has been selected as network model, one of the most popular neural network structures.

It is formed by an input layer whose elements are the data to be processed, an output layer that provides the output data of the network and one or several hidden layers that process the incoming information to obtain the network response at the output layer. As it can be seen the input layer is not actually a layer but rather the input data to the first hidden one. Every layer may be formed by a variable number of units named neurons. Each neuron computes the weighted sum of all its inputs and a bias constant. These inputs may come from both neurons in the preceding layer and in the same layer. The result is processed by an activation function that provides the neuron output.

The output function represents an important element of the neural network paradigm, because it supplies a nonlinear element to the model that allows these structures to identify the nonlinear behavior inherent to complex dynamics. The most common function used as output is the so called sigmoid for its "S-shape" form. It may be the arctangent or hyperbolic tangent when it is needed an output included in the [-1,1] interval. When the desired interval is [0,1] the so-called logistic function is used. Nevertheless very simple algebraic manipulations may define any of the three functions with both outputs. The use of other kinds of output functions like linear, piecewise linear or step is also usual.

A combination of both linear and nonlinear functions is usually encountered in the literature, where a very common configuration is the use of nonlinear functions for the hidden layers and a linear one for the output one. This configuration is based on the fact that the information processing is performed by the hidden layers while the output one usually provides only an adaptation of the neural network response to the desired size (number of outputs) of that response, moreover, the use of a nonlinear saturating function in the output layer will eliminate the possibility of signals with higher than saturation values at the output, a fact that will diminish the precision of the predictions performed by the network. This is the configuration selected for the model proposed in this work.

Another important issue in the definition of the network structure is the connection scheme of every neuron. A fully connected feedforward network has been selected, where each neuron is connected to every output of the previous layer and no connection is allowed between neurons in the same layer. It has proved to have very good capabilities in the approximation of any nonlinear function, and also in time series prediction [6]. The inclusion of feedback will need the demonstration of the network stability and will not endow the model with further capabilities, therefore it is usually not included in neural network models used to identify time series.

D. Learning process

The learning capability of neural networks is provided by the adaptation of the input weights of every neuron. This process is performed by presenting an input pattern and the desired output to the network and then modifying every weight until an error function reaches a minimum or falls below a fixed value. This procedure is repeated for each pattern to be learned. The way of carrying out this adaptation process defines the learning strategy of the neural network. The well known Backpropagation algorithm has been selected to do it. It uses the mean squared error as error function. So, once an input pattern is presented to the network it provides an output whose value is compared with the desired one and an error function is obtained. With this function it is possible to calculate an expression that relates the gradient of the overall error function with every weight. Several algorithms may be used to perform this calculation. The selected one is the so-called Levenberg-Marquardt (a combination of the gradient descent and Newton methods for solving optimization problems) [7] which provides a very accurate weight adaptation with a moderate time consumption. Its only drawback is the need of a very large memory in the computer where the process will be simulated, an easy to accomplish requirement in modern computers.

III. COMPUTATIONAL RESULTS

As it has been previously stated the Spanish monthly consumption from January 1959 to September 2000 (a total of 501 values) has been used to validate the proposed model. All these data has been divided into two blocks: one for training (from January 1959 to December 1985, 324 months) and the other for validation (the remaining information, 177 months).

The learning algorithm had a maximum number of 5000 time steps to reach an adequate error (0.000001), a preventive measure to avoid the algorithm falling into a too long learning process, although in most cases the desired error was reached in less than 1500 time steps, which proves the speed of the selected algorithm.

A. Inputs to the network

As it has been previously stated two possibilities have been tested as network inputs. The first one took twelve values of the time series spanning from months *(t-11)* to *(t)*. Then the output will be the network prediction. In the second option six values were considered, from *(t-5)* to *(t)*. The simulation process has proved that better results were obtained with the first option, which represents a hardly surprising result because, as it considers twelve months, the whole information about the monthly evolution of the demand during the last year is presented to the network while only a part of it is taken into account when only six inputs were considered.

B. Network structure

The number of neurons in the hidden layer is an essential issue in the network design strategy. So if a network is provided with too few neurons it will not be able to reproduce the system dynamics accurately and therefore it could not provide a reliable forecasting. On the other hand, too many hidden neurons will define a network that, in the best case, will provide an appropriate behavior but with an excessive computing time and a high memory need while, in the worst case, will only learn the presented pattern and will not be able to generalize the acquired knowledge to predict non-learned patterns. Therefore the selection of the appropriate network size is not an easy task. Some algorithms have been developed to look for the best dimension of the network, the so-called pruning algorithms [8]. Their working strategy is very simple: it starts with the definition of a larger than necessary network and then proceeds to simulate its behavior to detect redundant neurons or links that will be next removed. The process is iteratively repeated while an error function is kept under a certain value. These algorithms usually consume too much time and not always provide the smallest network size.

An easier approach may be done with a trial and error strategy, where several networks with different sizes are simulated and that with the best performance is selected. If a more precise response is needed variations about this size will be proved until the desired precision is obtained.

Along with the definition of the network size it is also important to determine the number of hidden layers it has, that is to say the way the neurons are arranged. Sometimes the use of a multilayer network provides better results than using an only one, others an only layer will be enough. Here there is no algorithmic procedure to obtain the best solution and a trial and error strategy is to be used.

We have tested several layer numbers and sizes to find the best one. For one layer networks sizes of 4, 6, 8 and 12 neurons have been tested. The one performing the best was that with 8 neurons. The network with 12 units gave a result equivalent to that provided by the selected one. On the other hand some multilayer structures have been tested with two and three hidden units. The most significant ones were (4,2), (6,2), (6,4), (8,4) and (12,6) for a two layers network and (4,2,4), (8,4,8) and (12,6,12) for a three layers structure. The best ones are listed in the Table I. It is important to notice that, as it has been previously pointed out, a higher number of neurons does not mean a better performance of the network. Indeed a higher number of layers neither means a better forecast capability. These results prove what has been previously stated: a higher number of neurons not only does not always mean a higher precision in the forecast tasks of the network but also in some cases the resulting network will perform worse.

It can be seen in Table I that the best results were obtained with a three layers network with the distribution (8,4,8), so this structure has been used to compare the results of taking six or twelve inputs to the network.

TABLE I. LIST OF THE TESTED STRUCTURES THAT PROVIDE BETTER RESULTS IN SIMULATION.

NET-OPTION	DESCRIPTION
8-4-8-1. 1	13 values out of the ±5% band. No one higher than ±10%.
8-4-8-1. 1 (3th layer linear)	10 values out of the ±5%. Higher overall error.
8-1. 1	Many values out of the ±5% band. Little computing time.
8-4-8-1. 2	Worse behavior than Option 1.
8-4-1. 1	Many values out of the ±5% band. The overall error has increased
8-8-1. 1	Worse behavior than the preceding structure

It was also removed the nonlinear output from one of the layers of this structure (specifically from the third layer) to test the effect of using linear functions as neural output. The overall behavior without nonlinear outputs worked worse as it might be expected.

C. Simulation Results

It is shown in Fig. 3 and Fig. 4 the simulation results obtained with the structure that performs the best from all the tested ones: three hidden layers with a (8,4,8) neuron distribution. As it can be seen in Fig. 5 there is a very good relation between the data used to train the network (the first 324 months) and those obtained as forecasting. This is not surprising because these are the directly learned data. On the other hand a greater error appears when the forecast is performed to obtain data that have not been previously learned (the last 177 months). Nevertheless in most cases the error is less than 5% and only 13 values are higher than that, although none of them exceed 10%. These results show the good capability of the proposed model to forecast the electric demand as only monthly consumption data have been presented to the network to predict a future value. This implies that no data different from them have been taken into account, although information like unusual weather variations, economic evolution or other social factors may slightly influence the long term electric demand. Nevertheless as the training and forecasting processes have been performed with monthly data, and they represent an overall magnitude, the influence of sporadic variation in the expected demand caused by those factors is diluted in that overall value. Moreover if those factors influence the general monthly evolution this influence will be taken into account by the normalization process.

IV. CONCLUSIONS

A neural network has been presented to perform electric demand forecasting that takes into account only past monthly demand data to obtain a prediction of the following month.

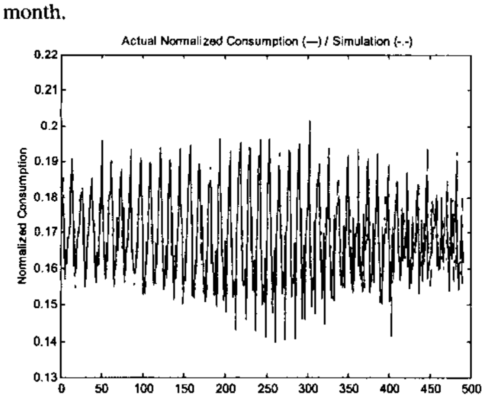

Fig. 3. Comparison between the prediction performed by the neural network with the (8,4,8) configuration and the actual data. Normalized values.

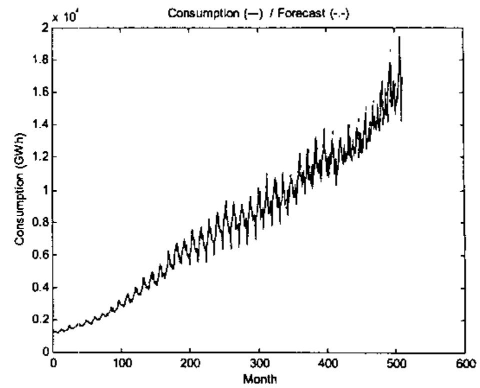

Fig. 4. Comparison of the prediction performed by the neural network with the (8,4,8) configuration with the actual data. Actual values.

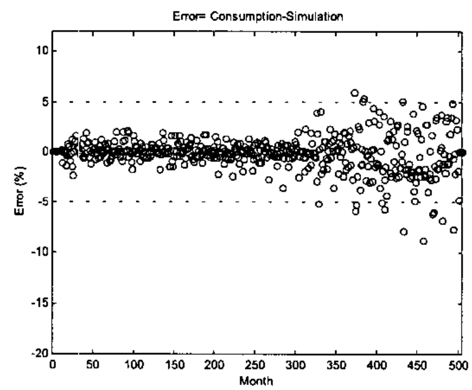

Fig. 5. Error obtained in the prediction of the electric demand by the neural network with a (8,4,8) structure.

This model defines a structure different from those usually used to predict short or medium-term data consumption where, along with that information, another data as weather parameters (temperature, humidity...) or the day of the week to be forecasted must be considered. This fact means that not only less data are presented to the network but also a more simplified information structure is needed. As a matter of fact these data have different nature and are represented in different units, and therefore they must be previously coded to be appropriately processed by the network, which implies a more complex definition of the network. On the other hand the increase in the size of the overall structure (a higher number of inputs needs a higher number of neurons and then a higher number of weights) implies a longer training time along with higher time and memory demands during simulation. So the proposed structure represents a very simple although effective option for the monthly electric demand forecasting, an issue that have been reached by taking into account the cumulative nature of the data to be processed.

V. REFERENCES

[1] J. T. Connor, R. D. Martin, L. E. Atlas. "Recurrent Neural Networks and Time Series Prediction". *IEEE Trans. on Neural Networks*, Vol. 5, No. 2, pp. 240-254, March 1994.

[2] W. Charytoniuk, E. Don Box, W.J. Lee, M.-S. Chen, P. Kostas, P. Van Olinda. "Neural-Network-Based Demand Forecasting in a Deregulated Environment". *IEEE Trans. on Industry Applications*, Vol. 36, No. 3, pp. 893-898, May/June 2000.

[3] F. Doveh, P. Fegin, D. Greig, L. Hyams. "Experience with FNN Models for Medium Term Power Demand Predictions". *IEEE Trans. on Power Systems*, Vol. 14, No. 2, pp. 538-546, May 1999.

[4] I. Derzga, S. Rahman. "Input Variable Selection for ANN-Based Short-Term Load Forecasting". *IEEE Trans. on Power Systems*, Vol. 13, No. 4, pp. 1238-1244, November 1998.

[5] S. Haykin. "Neural Networks. A Comprehensive Foundation". *IEEE Press*, 1994.

[6] K. Hornik, M. Stinchcombe, H. White. "Multilayer Feedforward Networks Are Universal Approximators". *Neural Networks*, Vol. 2, pp- 359-366, 1989.

[7] C. M. Bishop. "Neural Networks for Pattern Recognition". *Oxford University Press*, 1995.

[8] R. Reed. "Pruning Algorithms. A Survey". *IEEE Trans. on Neural Networks*, Vol. 4, No. 5, pp. 740-747, September 1993.