Electric energy monthly demand forecasting by spectral analysis

M. A. Jaramillo[1], E. González[2], D. Carmona[2] and J. A. Álvarez[2]

[1] Department of Electronics and Electrical Engineering
E.I.I., University of Extremadura
Avda. de Elvas s/n, 06071 Badajoz (Spain)
phone:+34 924 289600, fax:+34 924 289601, e-mail: miguel@unex.es

[2] Department of Chemical and Energy Engineering
E.I.I., University of Extremadura
Avda. de Elvas s/n, 06071 Badajoz (Spain)
phone:+34 924 289600, fax:+34 924 289601, e-mail: evagzlez@unex.es, dcarmona@unex.es, jalvarez@unex.es

Abstract. Medium-term load forecasting is a useful tool for the maintenance planning of grids and as a market research of electric energy. In this work medium-term load forecasting methods are developed, the most forgotten time scaling process in the load forecasting bibliography. These methods will be applied to the peninsular Spanish monthly energy consumption. Methods traditionally employed with this objective are based on regression, statistical techniques (mainly Box-Jenkins ARIMA), and also with neural networks, fuzzy logic or expert systems. Most of them need the use of nonelectric variables, mainly climatic or economic ones, which strongly influence electric energy demand. These variables, of cyclic nature, provide a periodic behaviour to the energy consumption time series. This work presents a study of this periodic behaviour by means of spectral analysis, with the identification and interpretation of the dominant frequencies. A forecasting method for future values of electric energy demand will be then presented, which is based on a simple regression technique combined with neural networks. It does not take into account any climatic or economic variables, because only periodic behaviour of the time series is considered. Acceptable results are reached, with percentage errors lower than 5 % in most cases.

Key words

Electric energy forecasting, spectral analysis, time series, medium-term load forecasting, regression techniques.

1. Introduction

Electric energy demand forecasting is an essential strategy for companies that are responsible for producing, delivering and reselling it, as well as for power system controllers.

Three kinds of forecasting processes are distinguished according to their time scaling: short, medium and long-term. Short term forecasting tries to estimate load or energy demanded in the next hours or days. It is mainly useful as a help for buying and selling bidding to energy markets, for generation and distribution system operation, and for power system management, with the aim that energy supply to match the demand at every moment.

This time scaling has engaged the attention of most researchers.

Medium term forecasting consists of the estimation of the energy that consumers will demand next weeks or months. It is useful for the maintenance planning of grids and as a market research for electricity producers and resellers.

On the other hand, long term forecasting pursues the estimation of the annual peak load or global energy that consumers will demand up to twenty years ahead, with the objective of designing expansion planning strategies for both production and distribution systems.

This paper is organizes as follows. Section 2 presents the peninsular Spanish monthly electric energy consumption series. Section 3 proposes a smoothing method for the trend extraction and a neural network based model to forecast future values. Section 4 explains a curve fitting method for extrapolate future values of monthly consumption fluctuations and finally Section 5 presents conclusions.

2. Monthly electric energy consumption

As electric consumption data evolve along time they may be assumed to form a time series, where the prediction is done considering past values to forecast future ones. Nevertheless as electric energy consumption is highly influenced by factors different from past consumption, prediction needs to take into account, along with those past consumption values, other information as temperature, humidity, hour of the day or day of the week, because all of them have a remarkable influence on demand. All this information is essential when a short-term prediction is to be done because the electric energy consumption depends on all those factors. In this way a great amount of data is needed to perform the prediction and complex models must be used to obtain good results.

On the other hand, as medium term forecasting deals with monthly data, the influence of the aforementioned factors is diluted in an overall value. These values represent the total monthly electric consumption 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, medium term demand forecasting needs less accurate predictions than the short term one because 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 obtain a prediction of the electric demand.

Time series of electric energy demand presents a rising tendency due to the influence of the economic and technological evolution on the electric market. In fact, as the wealth of most nations presents a rising tendency and the technological development provides the society with more and more devices that need electric energy to work, the electric demand suffers a constant rising tendency. Imbedded in this general trend a fluctuation caused by the difference in demand that each month has appears. The electric demand in the same month of different years must have very similar variations from the general rising trend because this demand is mainly controlled by climatological factors, which must be very similar in the same month of different years. Modifications in those factors are usually embedded in a more general climatological behaviour that will influence several months. In the same way specific political or economical factors different from those that govern the general rising tendency of the demand will influence several months. So variations in normal demand will affect several elements of the time series, providing a slight modification of the general rising trend. Those variations may be properly handled if enough past values are considered to perform a prediction. Then it may be assumed that the time series is formed with the combination of a general rising tendency and monthly variations that ought to be very similar for different years.

On the other hand that rising trend represents an important problem for the tool that is to be used to perform the demand forecasting, neural networks, because the use of non-linear saturating functions as neural output imposes boundaries limits that avoid the network providing an output higher that a certain upper limit. So the model will not be able to predict values of the time series falling out of those boundaries. The use of linear output functions in the output layer does not solve this problem because saturating functions in the inner

layers will still impose an upper limit to the network output. Removing all the saturating functions from the whole network is not a solution because in this way a linear network will be defined and it will not be able to follow the highly non-linear behaviour of time series that represent the electric energy demand. Therefore it is necessary to define a procedure that eliminates that rising tendency from the time series but retains the monthly fluctuations. In this work this procedure is defined as a trend extraction process which transforms the original time series into two ones, the trend and the monthly fluctuations.
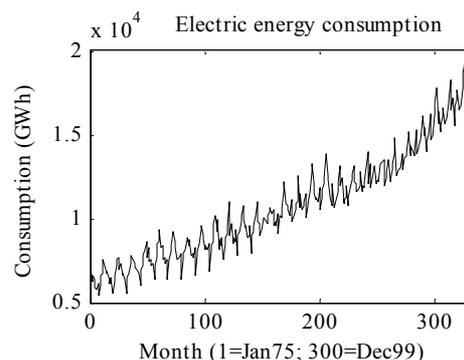

Fig. 1. Monthly Spanish electric consumption from January 1975 to December 2002.

The Spanish monthly consumption (Fig. 1) from January 1975 to December 2002 (a total of 336 values) has been used to validate the proposed model. All these data has been divided into two blocks: one for training (from January 1975 to December 1997, 276 months) and the other for validation (the remaining information, 60 months) [1].

3. Extraction and forecasting of the trend

As it has been pointed out previously, electric energy consumption forms a time series with a clear rising tendency.

This rising trend is the most interesting property of the series for the medium or long-term forecasting process, because it contains information about the magnitude of the energy amount that is consumed each month, about its rising rate and about how economic or politic events influences it.

Many works about smoothing techniques for time series have been published. A complete and updated summary of those methods is shown in [2]. Two kinds of series smoothing methods are presented in it: parametric and nonparametric ones.

Parametric methods try to fit data to a smoothing function or deterministic model. By means of any optimisation process, model parameters are obtained for the further extrapolation of future values.

On the other hand, nonparametric methods do not obtain a set of parameters capable of explaining the series behaviour, but they try to adapt a smoothing model to the series as new values appear. Local regression or filtering are examples of nonparametric techniques.

Parametric methods are simpler and allow an immediate extrapolation of future values of the series. However, an inadequate selection of the model parameters, or an unexpected change in the series tendency, will reduce the forecasting accuracy. Nonparametric techniques can adapt to local changes in the series trend, but they are not forecasting tools by themselves, because the smoothing model depends on the real values of the series and so it does not allow future values extrapolation.

It can be said that the objective of the trend extraction is to obtain a smooth series (without periodic components) faithful enough to the original series. In order to do it, nonparametric methods are employed to extract the series trend in this case, because of their best capability to adapt to the series changes.

Several factors will influence the decision about what is the most adequate smoothing technique to employ. Due to the mentioned objective, these factors are related to the goodness of fitting and the smoothness of the resulting series.

The faithfulness of the smooth series to the original one will be measured by the R^2 index, the most used one for the evaluation of the goodness of fitting of curves. The expression below shows its definition:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(C[i] - T[i]\right)^2}{\sum_{i=1}^{N}\left(C[i] - \widehat{C}\right)^2} \tag{1}$$

where $C[i]$ y $T[i]$ are, respectively, the *ith* elements of the original and the smooth series, $\widehat{C}$ is the original series mean and N the number of data.

If the obtained curve fits accurately the original one, R^2 will be close to one. R^2 will decrease as the goodness of fitting gets worse.

The smoothness of the series resulting from the trend extraction is related to the behaviour of the first difference of the series, that is too say, the difference of each datum with regard to the proceeding. The more constant these differences are, the smoother the series is. First differences are similar to the first derivative of continuous functions. So the smoothness index S is defined as follows [3]:

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\Delta^2 T[i]\right)^2}$$

$$\Delta^2 T[i] = \left(T[i+1] - T[i]\right) - \left(T[i] - T[i-1]\right) = \\ = T[i-1] - 2\cdot T[i] + T[i+1] \tag{2}$$

$\Delta^2 T[i]$ is equivalent to the second derivative of the T series in point i, in the case of discrete series.

For this index, the smoother the series is, the lower value it will have, corresponding a zero value to the linear function.

The objective of pursuing the smoothness of the series is to avoid it to contain periodic components and to enhance its predictability. However, smoothness is not directly related to the periodic behaviour of the series, but to the absence of sharp tendency changes. So S will be used only as an indicative parameter and it will not be decisive for the election of the best technique.

Both indexes R^2 and S try to quantify the behaviour of the smooth series with regard to the original one, but they cannot substitute the visual observation of results.

A simple nonparametric technique is used for the trend extraction of the series, the moving average method.

A. Extracting the trend with moving average

Moving average consists of transforming values of a time series into other resulting of making a weighted mean of a set of data around that to modify.

Preceding values are always employed for the mean construction. In other case, some data at the end of the series could not be smoothed because no following data would be available for calculating the moving average. That would produce important forecasting errors.

For the series smoothing each datum is substituted by the mean of itself and the proceeding (n-1). So the weights applied to the n data are all equal to $1/n$. Equation (3) shows the process.

$$T[i] = \frac{1}{n}\left[C[i] + C[i-1] + ... + C[i-(n-1)]\right] \tag{3}$$

In this expression $C[i]$, $i = 1, 2, ..., N$ represents the original series, $T[i]$, $i = 1, 2, ..., N$ the smoothed series (trend) and N the number of data.

In this case, window lengths between 3 and 24 data have been considered for the moving average calculations, with the aim of taking into account every periodic component of the series.

Values of R^2 and S for the tested windows are presented in Table I.

TABLE I.- Indices R^2 and S obtained with moving average for different window lengths

Moving average size	Goodness of fitting (R^2)	Smoothness (S)
3 data	0.976667	417.706839
6 data	0.950851	130.581779
12 data	0.946279	45.268542
13 data	0.951823	97.762559
14 data	0.952430	73.421191
15 data	0.951438	84.190543
18 data	0.939880	46.968209
24 data	0.926141	29.236806

It can be proved that windows with 3, 6, 13, 14, 15 and 18 data produce trend series with a certain periodic behaviour. So, 12 or 24 months can be used for the moving average, because they produce the smoother series. However, Table I shows that the fitting accuracy is not sufficiently good when 24 months are employed for the moving average. So the most interesting option is the 12 data window, which is equivalent to an annual mean. Fig. 1 shows the result of applying moving average with windows of 12 months.
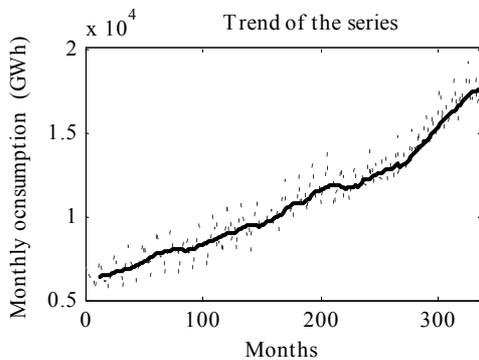

Fig. 1. Series obtained by moving average of 12 months.

B. Neural network used for forecasting

Future values of the trend series are to be forecasted from past values. A smoothing process have already been applied to the original series, so that it only remains to find a reliable forecasting tool which is capable of adapting to local trend changes.

Neural networks have been widely used for time series forecasting and for electric load forecasting specifically. The non-linear behaviour of this algorithm provides a great generalisation capability, a very useful quality for forecasting processes.

Among the different neural network models that bibliography presents, feedforward networks (FF) have been selected. They implement a non-linear autorregresive process, while recurrent networks (networks where feedback between neurons have been included) are equivalent to a non-linear ARMA model. The first kind of networks are simpler than the second, and they assure a better convergence, while recurrent networks require a delicate selection of initial conditions and network parameters to avoid stability problems.

The FF neural network structure used in this work is the well-known Multilayer Perceptron (MLP). It has been widely used in time series prediction because of its ability to identify the time evolution of a dynamic system. Actually it may be assumed as a universal approximator [4,5].

The network 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. The result is processed by an activation function that provides the neuron output:

$$x_j^i = \sum_k w_{jk}^i\, y_k^{i-1} + \theta_j^i \qquad (4)$$

$$y_j^i = \sigma\left(x_j^i\right) \qquad (5)$$

In this expression x_j^i represents the activity of neuron j in layer i, w_{jk}^i the strength of the connections between this neuron and all those that are in the previous layer, y_j^i the neuron output while θ_j^i is a bias constant.

The output function $\sigma\left(x_j^i\right)$ represents an important element of the neural network paradigm, because it supplies a non-linear element to the model that allows these structures to identify the non-linear behaviour inherent to complex dynamics. The most frequently used as output function is the so called sigmoid for its "S-shape" form. It may be the so-called logistic function when it is needed an output included in the [0,1] interval. When the desired interval is [-1,1] hyperbolic tangent is used.

Nevertheless very simple algebraic manipulations allow the use of any of the two sigmoids with both output intervals. The hyperbolic tangent obtained from the logistic function is represented in the following expression:

$$y_j^i = \sigma\left(x_j^i\right) = \frac{2}{1+\exp(-2\,x_j^i)} - 1 \qquad (6)$$

The use of other kinds of output functions like linear, piecewise linear or step is also usual.

A combination of both linear and non-linear functions is usually found in the literature, where a very common configuration is the use of non-linear functions for the hidden layers and linear functions 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 non-linear 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.

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 this adaptation process is carried out defines the learning strategy of the neural network. The well-known Backpropagation algorithm [6] has been selected to do it. It uses the mean squared error as performance 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 to 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 optimisation problems) [6] 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.

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 behaviour but with an excessive computing time and a high memory need while, in the worst case, will only learn the presented patterns and will not be able to generalise 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 [7]. Their working strategy is very simple: it starts with the definition of a larger than necessary network and then proceeds to simulate its behaviour 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 tested 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. 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.

C. Trend forecasting

Input data ought to be approximately enclosed within the interval marked by neurons output function. Nonstationary behaviour of the trend series opposes this requirement, so a scaling process is necessary to confer stationarity to the input data. Differentiation is the classic method to do it. Box and Jenkins introduced it in their early works [8-9] and other authors employed it for electric load forecasting approaches.

A first order differentiation (each datum minus the previous one) provides a new series with a quite stationary behaviour. Dividing its elements by a constant will enclose them within the desired interval.

Consumption has proved to present an increasing trend in the considered time period, so first differences will be positive, allowing the use of logistic output functions for the neurons. However, many values in that series are close to zero and so close to one of the saturation limits of the function. The hyperbolic tangent has proved to be more adequate for using in this case as close-to-zero input data are far enough from the saturation zones of the function.

First order differences, divided by an enough high constant, have been used for scaling the trend series. The resulting normalized series $Tn[i]$ is formed with values close to zero that have shown a good behaviour in the forecasting process.

$$Tn[i] = \frac{T[i] - T[i-1]}{1000} \qquad (7)$$

Before applying the neural network, the input vector length has to be selected. For autorregresive models, these vectors must be constituted by past values of the series. The number of data is usually stated from periodic behaviour of the series. The widely used ARIMA method [8] advises it. But periodic components have already been eliminated from the series. However, meteorological and socio-politic factors can influence the trend behaviour, and those factors usually repeat annually. So 12 past values of the trend series will form the input vectors.

Once the trend series has been extracted, its values conveniently rescales and the neural network structure defined the training process may be carried out. As it has

been previously stated the first 276 values will be used for this purpose.

The neural network will be trained with input vectors (formed by 12 values of the normalized series $Tn[i]$) and the expected output for each one (the next value).

After the network has been trained the validation of the network will be carried out predicting one by one each new value of $Tn[i]$ of the remaining 60 data of the series.

The network outputs (those obtained both from training and from validation) must be rescaled applying a process opposite to the previously by (7) and then compared with real values of the smooth trend series.

Forecasting results are shown in error plots, where errors are obtained as the relation between real and forecasted values, in percentage. Some statistical indexes have been used to explain results:

- *MAPE*: mean absolute percentage errors, in validation period.
- Percentile 95 of the same absolute percentage errors.
- Percentile 90 of the same absolute percentage errors.
- Minimum and maximum errors, retaining sign.

Different numbers of neurons for one hidden layer have been tested for the neural network. Fig. 5 shows resulting *MAPE*.
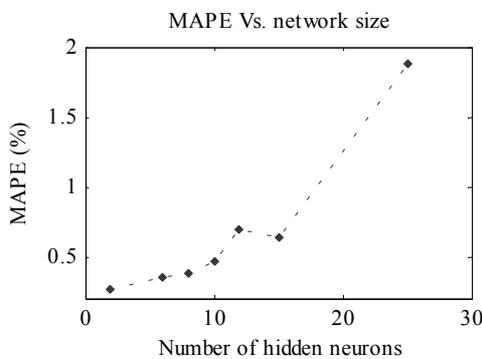
Fig. 5. *MAPE* obtained by neural networks with different number of neurons on hidden layer

Results shown in Fig. 5 correspond to networks with only one hidden layer. Two layers have been tested too, but they have not been capable of enhancing results, and computational time spent increases significantly.

In this case, the best MAPE has been reached with a two neuron MLP. Forecasting errors it produced are presented bellow:

- *MAPE*: 0.439712 %
- Percentile 95: 1.159996 %
- Percentile 90: 1.087768 %
- Minimum error: -0.866968 %
- Maximum error: 1.589953 %

4. Curve fitting of the fluctuation series

Section 3 presented the smoothing process of the consumption series and forecasting of the resulting trend series. Extracting this trend from the original consumption values, a new series appears which represents monthly fluctuation of consumption. Fig. 7 shows that new series.
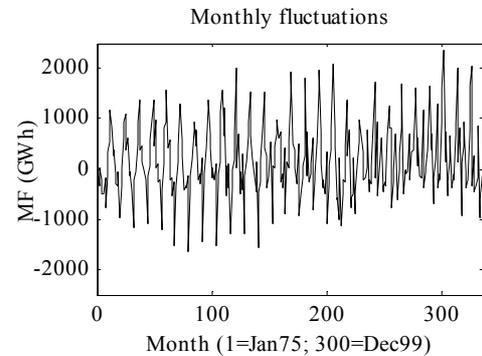
Fig. 7. Monthly fluctuation of the electric energy consumption

Visual observation of this series reveals a certain periodic behaviour. This idea leads to make a spectral analysis of the series. Fig. 8 shows the Fast Fourier Transform of the series, and Fig. 9, the power spectrum. X-axe of plot in Fig. 9 represents the normalized frequency, where value 1 is the so-called Nyquist frequency, defined as half the sampling frequency. It is the highest frequency in the series, and corresponds to an angular frequency of π rad./sample [10].
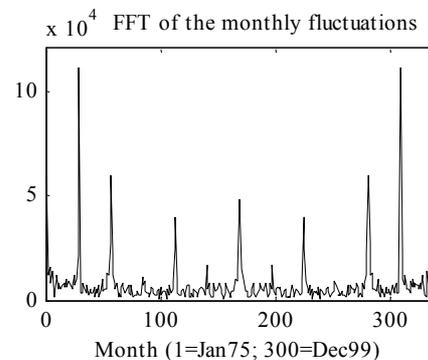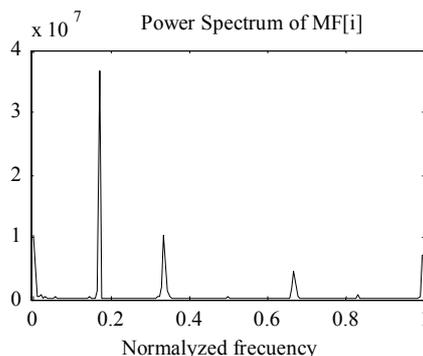
Fig. 8. FFT of the monthly fluctuation series.

Fig 9. Power Spectrum of monthly fluctuation series

Figures 8 and 9 show important peaks in frequencies corresponding to periods of 12, 6, 3 and 2 months and lower peaks in those corresponding to periods of 4 and 2.4 months. These frequencies are integral multiples of the fundamental one (corresponding to a 12-month period), so a six-term-Fourier series could fit the fluctuation series with arbitrary accuracy:

$$F(n) = a_0 + \sum_{k=1}^{6} \left[a_k \cos(k\omega n) + b_k \, \text{sen}(k\omega n) \right] \quad (8)$$

being ω the fundamental angular frequency of the series, and a_0, a_k and b_k the fitting coefficients.

Curve fitting is a widely employed tool in different scientific disciplines for explaining process dynamics. Many works propose this technique for electric load forecasting [11]. However, most of them employ regression techniques instead of using deterministic models like polynomial or Fourier series. They try to find a consumer behaviour pattern from past values of energy consumption.

The approach proposed in this work consists of fitting the monthly fluctuation series to a six-term-Fourier series, presented in equation (8). Two options have been tested:

- Selection of a training subset of the series for the curve fitting. Then every future values of the series will be extrapolated from the obtained curve.

- Every past values are employed to make the curve fitting for the extrapolation of each series element. Many fitting processes are necessary, so this option will spend more time but it could adapt better to changes in consumer behaviour.

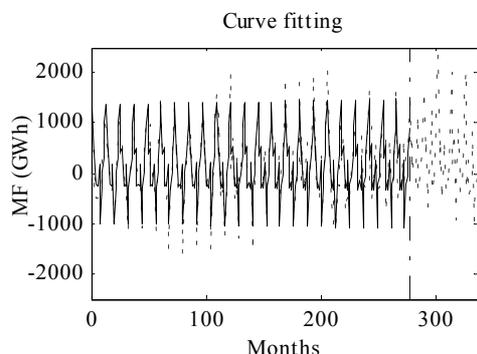Fig. 10 compares the curve obtained with the first option with the original series.

Fig. 10. Curve fitting of training subset (solid line) vs. original series (dashed line)

Future values (corresponding to the validation period) were then extrapolated from the curve. Forecasted trend

values are added to the results and then compared to the real values of consumption. Forecasting errors are shown in Fig. 11.
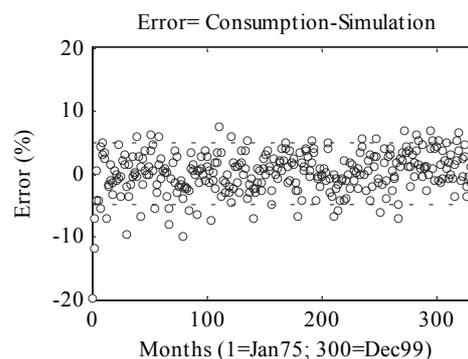

Fig. 11. Errors obtained by trend forecasting and monthly fluctuation fitting (in training period)

In the validation period, the next statistical results were obtained:

- *MAPE*: 2.373855 %
- Percentile 95: 5.706764 %
- Percentile 90: 5.125731 %
- Minimum error: -3.758720 %
- Maximum error: 6.694291 %

This forecasting technique has two clear disadvantages:

- On the first hand, a great amount of data has been employed for the fitting process. So the curve describes the global behaviour of the series along 23 years (276 months) and provides the curve with certain obsolescence.

- On the other hand, any periodic components of the series whose frequencies differ from those implicit on the curve are eliminated. Power spectrum (Fig. 9) shows that frequencies near the main ones are not always negligible, and model removes them.

With the aim of avoiding the curve fitting become obsolete, new data can be added to the training set, which contain updated information about the series behaviour. This approach consists of using every past values for the curve fitting when a new value is to be predicted. The extrapolation process will produce only one forecasted value and subsequent fitting processes are necessary for predicting the rest of the series.

Trend forecasting has been done in the same way than before, using 276 data for a first network training and then predicting every value in the validation period.

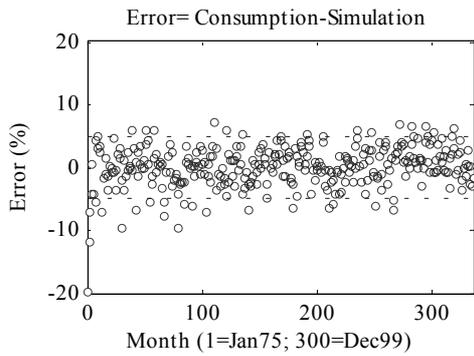Forecasting errors obtained with this approach are shown in Fig. 12.

Fig. 12. Forecasting error obtained by trend prediction and fitting of all past values of the datum to forecast.

The indices bellow summarises the results:

- *MAPE*: 2.233223 %
- Percentile 95: 5.854312 %
- Percentile 90: 4.984499 %
- Minimum error: -3.715556 %
- Maximum error: 6.579173 %

As could be expected, adaptive fitting outperforms the results obtained with a fixed set of training data.

It can be seen that positive errors are quite higher than negative ones. The reason is in the trend forecasting process. The smoothing model used was the 12 data moving average, so each original data is substituted by the mean of itself and eleven past values. The rising tendency of the series causes the smoothed data are lower than the original and so, the forecasted values are lower than the real.

5. Conclusion

A new approach has been proposed for trend extraction and prediction in medium-term load forecasting. It has been completed with the prediction of the periodic components of the time series rejected during the trend extraction by means of a Fourier series adjustment. The combination of both predictions provides a very satisfactory forecast of the electric demand.

References

[1] D. Carmona, M.A. Jaramillo, E. González, J.A. Álvarez, "Electric energy demand forecasting with neural networks", in Proc. of the 28th Annual Conference of the IEEE Industrial Electronics Society - IECON 2002, Vol. 3, pp. 1860-1865.

[2] S. Zhao, G.W. Wei, "Jump process for the trend estimation of time series", Computational Statistics & Data Analysis, no. 42, pp. 219-241, 2003

[3] T.M. Peng, N.F. Hubele, G.G. Karadi, "An adaptative neural network approach to one-week ahead load forecasting", IEEE Trans. on Power Systems, Vol. 8, No. 3, pp. 1195-1203, 1993

[4] K. Hornik, M. Stinchcombe, H. While, "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, Vol. 2, pp. 359-366, 1989

[5] G. Cybenco, "Approximation by superpositions of a sigmoidal function", Mathematics of Control, Signals and Systems, Vol. 2, pp.303-314, 1989

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

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

[8] G.E.P. Box, G.M. Jenkins, "Time Series Analysis: Forecasting & Control", Holden-Day Inc., Oakland, CA, U.S.A., 1976

[9] T.W.S. Chow, C.T. Leung, "Neural Network based Short-Term Load Forecasting Using Weather Compensation", IEEE Trans. on Power Systems, Vol. 11, No. 4, pp. 1736-1742, 1996

[10] S.K. Mitra, J.F. Kaiser, "Handbook for digital signal processing", John Wiley and sons, 1993

[11] IEEE Committee Report, "Load Forecasting Bibliography Phase I", IEEE Trans. On Power Systems, Vol. 99, pp 53-58, 1980