# 24-Hour Electrical Load Data - A Time Series or a Set of Independent Points ?

Damien Fay[†] John V. Ringwood[††] Marissa Condon[†] Michael Kelly[†††]

[†] Dublin City University [††] NUI Maynooth [†††] Electricity Supply Board
Ireland

**Keywords : Forecasting, electricity demand, time series, neural networks, principal component analysis.**

## Abstract

The paper investigates whether a time series or a set of independent points is a more appropriate description of 24-hour Irish electrical load data. A set of independent points means that the load at each hour of the day is independent from the load at any other hour. The data is first split into 24 series, one for each hour of the day i.e. a 1am 2am 3am series etc. These are called parallel series. The linear cross-correlation's of the parallel series are used to indicate independence. While the loads at 9am and 6pm to 8pm appear independent the remaining loads are highly inter-correlated. This suggests that 24-hour electrical load data has a dual nature. Two techniques are used to test this hypothesis. The first technique models each parallel series using neural networks. This technique is found to be computationally expensive. The second technique uses a hybrid technique called the Multi Time Scale (MTS) technique. This models 24-hour electrical load data as a time series that can be adjusted by 5 parallel forecasts and a daily cumulative model. The results show that the MTS forecasts are superior to the parallel forecasts except for 9am and 6pm to 8pm. A composite model using neural networks for 9am and 6pm to 8pm and the MTS model elsewhere takes advantage of the dual nature of the data reducing error and computational expense.

## 1. INTRODUCTION

Short term forecasting is required by electricity utilities for unit commitment and allocation of spinning reserve. Errors in short term forecasts cost the Irish Electricity Supply Board a significant amount of money per annum. In accordance with the Electricity Regulation Act of 1999, a deregulated market structure was set up which should lead to increased impetus to reducing forecast error and the associated costs.

Load forecasting models fall broadly into two categories, time series models and parallel models, examples of which can be found in [1,2] and [3-5] respectively. Time series models assume 24-hour electrical load to be a set of correlated points [1,2] while parallel models decompose electricity demand into 24 different time series, one for each hour of the day [3,4]. The different approaches are shown diagrammatically in Figure 1 below.

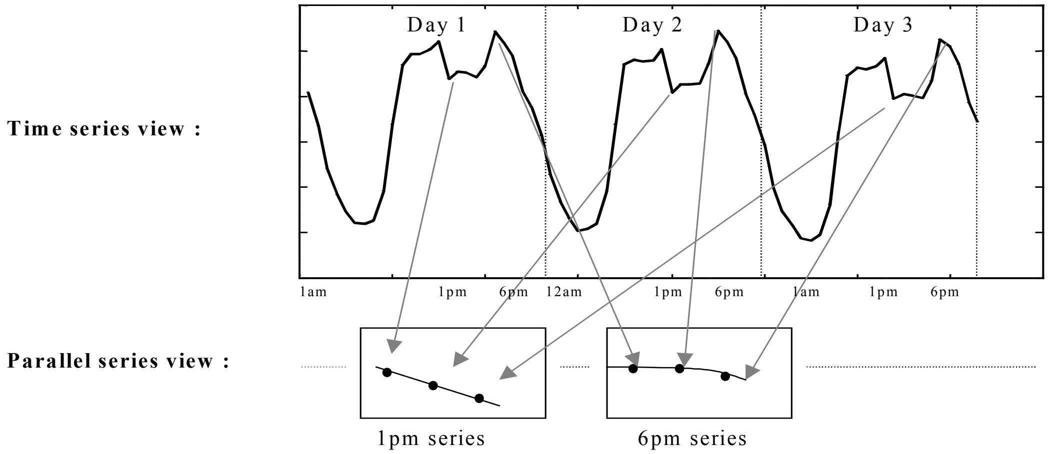


**Figure 1. Constructing parallel data series from electrical load data.**

The load for 3 sample days is shown in Figure 1. As can be seen the shape of the load curve for all three days is similar suggesting a time series model. However the 1pm series and 6pm series for example do not have the same shape and it may be advantageous to model them as parallel series.

A time series approach can be less complex than a parallel approach as only one set of parameters needs to be calculated. Also as the data set is not divided into 24 separate time series there are more input-output pairs for training of the model. Training 24 separate models can be computationally expensive. As found in [5] calculating the topology of 24 separate neural networks was prohibitive and just one topology was calculated and used in all the models.

[6,7] have suggested hybrid approaches where parallel forecasts are combined with time-series forecasts. Both approaches have the advantage that only a few parallel series are required as opposed to 24.

The purpose of this paper is to determine which approach is superior for predicting Irish daily load profiles.

## 2. DATA SET DETAILS

A database of electricity demand, actual temperature, wind speed and cloud cover from 1987 to 1998 on an hourly basis is available. Data between Tuesday and Thursday in the months January to March is selected so as to avoid the exceptions associated with weekend, Christmas and changes due to the daylight saving hour.

Three sets of data are used to train and test the models (Table 1). The training set is used to calculate model parameters. In the case of neural network models the validation set is used for early stopping and topology determination. As the validation and training sets have significantly influenced the model, a novelty set is used to evaluate model performance.

**Table 1 Segmentation of data set.**

| Set | Training | Validation | Novelty |
|---|---|---|---|
| Range | 20th Jan 1987<br>20th Mar 1996 | 21st Mar 1996<br>19th Mar 1997 | 20th Mar 1997<br>26th Mar 1998 |
| Size (Days) | 300 | 30 | 30 |

## 3. TESTING IRISH LOAD DATA FOR INDEPENDENCE

If electricity demand is a time series then by definition the load at hour $i$+1 on day $k$, i.e. $y_{i+1}(k)$ is related to the load at hour $i$ on day $k$ via a function $f$ (i.e. a time series model) and a white noise component ε

$$y_{i+1}(k) = f(y_i(k)) + \varepsilon_{i+1}(k) \quad (1)$$

The load for hour $i$+2 can be calculated from (1) as

$$y_{i+2}(k) = f(f(y_i(k) + \varepsilon_{i+1}(k)) + \varepsilon_{i+2}(k) \quad (2)$$

Assuming that electrical load is a time series then (1) also holds if the data is organised into parallel series $y_i$ with $i$=1…24. If $f()$ is a linear function then this hypothesis can be tested using the linear cross-correlation coefficient $r_{i,j}$ between parallel series $i$ and $j$ defined as [8]:

$$r_{i,j} = \frac{\mathrm{E}[(\bar{y}_i - y_i)(\bar{y}_j - y_j)]}{\sqrt{\mathrm{E}[(\bar{y}_i - y_i)^2]}\sqrt{\mathrm{E}[(\bar{y}_j - y_j)^2]}} \quad (3)$$

where E[] denotes the expectation operator and $\bar{y}_i$ is the average of $y_i$. As an example the cross correlation matrix between the 1pm, 2pm and 3pm series is shown in Table 2 below.

**Table 2 Cross-correlation matrix of $y_{13}$, $y_{14}$ and $y_{15}$**

| Hour | 1pm (13:00) | 2pm(14:00) | 3pm (15:00) |
|---|---|---|---|
| 1pm(13:00) | 1 | .9958 | .9924 |
| 2pm(14:00) | .9958 | 1 | .9934 |
| 3pm(15:00) | .9924 | .9934 | 1 |

As can be seen the cross-correlation's are very high (Table 2). This is not surprising, the load profiles for 3 sample days shown in Figure 1 are very similar which is typical and thus a large component of the data is highly correlated. The uncorrelated component between $y_{13}$ and $y_{14}$ is due to noise $\varepsilon_{14}$, the fact that $f()$ may be non-linear and possibly an independent component.

Also note that the correlation coefficient between $y_{13}$ and $y_{15}$ , i.e. $r_{13,15}$ is less than $r_{14,15}$ as the noise term in (2) is larger than (1).

The cross-correlation matrix between all the parallel series is calculated and the contour for $r_{i,j}$ =0.99 is shown in Figure 2.

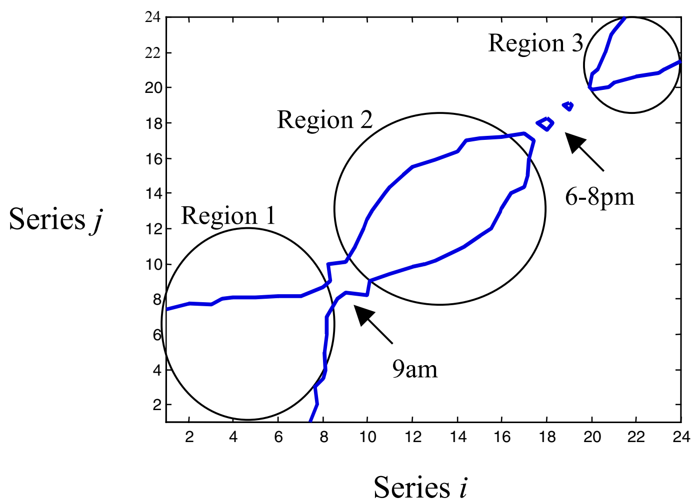


**Figure 2. Contour plot of $r_{i,j}$=0.99 for parallel series $i$ with $j$.**

Inside the contour the cross-correlation is higher than 0.99 and outside it is lower. Examination of (2) shows that the cross-correlation component is expected to drop as $i$-$j$ increases i.e. with distance from the diagonal (Figure 2). However the narrowness of the contour at 9am and 6-8pm show that these hours of the day may have an independent component or a larger noise component than the others. Excluding these hours there are 3 regions where the cross-correlation's are typical of a time series. This suggests that 24-hour Irish electrical load data may have a dual nature. The data is a time series with the exception of the hours of 9am and 6pm to 8pm, which are independent points. The next step is to model the data as both a set of independent points (Section 4) and using hybrid model (Section 5) and examine the results.

## 4. PARALLEL MODELS

### 4.1 Preliminary AR linear model

The parallel series for hour $j$ on day $k$, $y_j(k)$, has a low frequency trend $d_j(k)$ due to year on year changes in usage of electricity and a seasonal component $s_j(k)$ due to more electricity being used in winter for heating and less in summer. These components, are first removed using a Basic Structural Model (BSM) leaving a residual $x_j(k)$ (Figure 3) which is composed of weather, non-linear AR and white noise components.

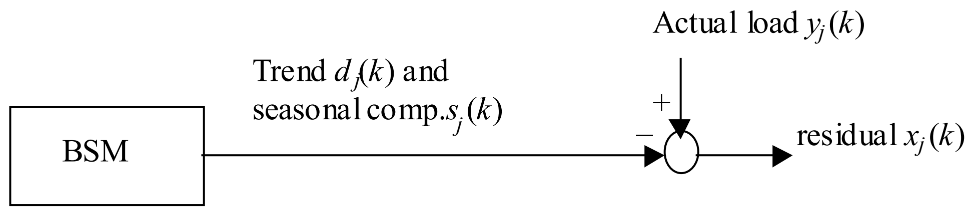


**Figure 3. Preliminary AR linear model overview**

There is a high degree of correlation between the weather and the seasonal component of the load. The seasonal component of the weather is related to $y_j(k)$ but not to $x_j(k)$ and so must be removed if the non-seasonal part of the weather variables is to be used to model $x_j(k)$. The seasonal component in the weather variables is removed by use of a pre-whitening filter. The pre-whitening filter is the same as the preliminary AR filter (Figure 1) except that the load is replaced by the weather variables and the residual is now the pre-whitened weather variables.

### 4.2 Input selection

Input selection forms perhaps the most important step in model building [9]. Inclusion of non-causal variables leads to model performance degradation outside of the training set. In addition, reducing the dimensionality of the inputs aids training of neural networks [9].

The pre-whitened weather variables are organised into:

- $\boldsymbol{t}_j(k)$, a vector of pre-whitened temperature from hour $j$ to hour $j$-23 on day $k$,
- $\boldsymbol{ws}_j$,$(k)$, a vector of pre-whitened wind speed from hour $j$ to hour $j$-23 on day $k$,
- $\boldsymbol{h}_j(k)$, a vector of pre-whitened humidity from hour $j$ to hour $j$-23 on day $k$.

These inputs are transformed using Principal Component Analysis (PCA) [9], to produce a vector of principal components. As found in [5] calculating the topology of 24 different neural networks is too computationally expensive and a single topology is used. The number of principal components retained is thus also constant [5]. The number of components retained is determined by the performance of a linear model, using these components, over the validation set (see [5] for more details).

It was found in [5] that there is a non-linear relationship between $x_j(k)$ and $x_j(k\text{-}1)$, $x_j(k\text{-}2)$… . The optimum lag for including AR residuals was found to be 2. Figure 4 shows an overview of the neural network models which produce an estimate of $x_j(k)$ called $x_j^{nn}(k)$ .

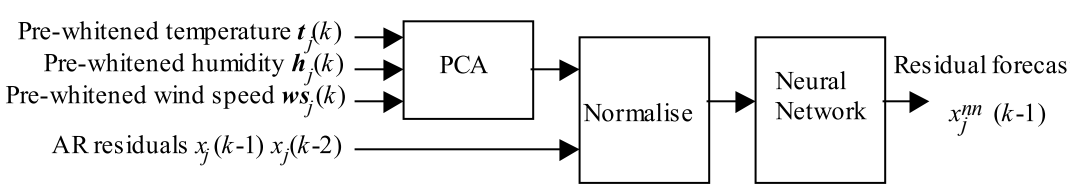


**Figure 4. Overview of neural network model.**

### 4.3 Neural network model

A multi-layer perceptron network using the back-propagation learning algorithm [1] is used. Each network consists of 2 hidden layers with *tan sigmoid* activation functions and a linear activation function in the output layer. The input data is normalised between ± 1 so that the *tan sigmoid* activation functions are not driven into saturation [10].

Each model is trained using early stopping [10] in which training ceases when the sum squared error of predictions in the validation set reaches a minimum. If a minimum is not found the training stops after ten thousand epochs. Cessation of training at the validation set minimum prevents over-training of the NN [10].

The topology of a neural network determines the degrees of freedom available to model the data [11]. If the neural network is too simple then the network will not be able to learn the function relating the input to output [11]. An over-complex network will learn the noise in the data and will not be able to generalise [11].

In order to determine the correct topology 50 NN architectures were examined using 1-5 and 1-10 nodes in the first and second hidden layers consecutively. Ten neural networks were trained for each topology with random initial weights to ensure reliable results. To perform this for each parallel series would require training 1,200 neural networks which is too expensive computationally so the network topology is refined for the 6pm series and applied to the others. There are other techniques for determining topology such as network pruning, constructive methods and weight sharing methods [11,12], however these are also computationally expensive.

Table 3 shows the average validation MAPE for each network topology. Topology selection is based on two criteria:

- Networks, which failed to reach a minimum MAPE over the validation data, are deemed to be too elementary.
- A subjective bias is shown towards networks with less complexity but similar validation MAPE's.

Using the first criteria networks with topologies of 1 and 2 nodes in the either the first or second hidden layer are eliminated. Of the remaining networks topologies 4x4, 5x7 and 4x9 achieved the lowest MAPE's (Table 3). Topology 4x4 is chosen using the second criteria.

**Table 3 Average validation MAPE for differing NN topologies (6pm series)**

| #Nodes | Layer 2 : 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Layer 1 : 1 | 1.42 | 1.43 | 1.60 | 1.68 | 1.65 | 1.62 | 1.69 | 1.73 | 1.81 | 1.83 |
| 2 | 1.90 | 1.92 | 1.53 | 1.55 | 1.56 | 1.58 | 1.52 | 1.54 | 1.77 | 1.64 |
| 3 | 1.44 | 1.47 | 1.53 | 1.48 | 1.50 | 1.50 | 1.51 | 1.60 | 1.59 | 1.68 |
| 4 | 1.42 | 1.49 | 1.48 | **1.45** | 1.50 | 1.48 | 1.63 | 1.48 | **1.43** | 1.51 |
| 5 | 1.47 | 1.46 | 1.48 | 1.52 | 1.53 | 1.47 | **1.44** | 1.47 | 1.62 | 1.53 |

Finally the neural network forecast of the load $y_j^{nn}(k)$ is forecast by reintroducing the trend and seasonal component:

$$y_j^{nn}(k) = d_j(k) + s_j(k) + x_j^{nn}(k) \tag{4}$$

## 5. MULTI TIME SCALE MODEL

In Section 3 it is shown that 24-hour Irish electrical load data may have a dual nature. The hybrid approach of Murray [7] can be used to take advantage of this nature. The approach combines three *different* types of forecast:

1. A load curve forecast from a Load Curve Model (LCM).
2. Forecasts of the load at individual hours (or cardinal points).

3. A forecast of the sum of the load for the day to be forecast, day+1 and day+2.

The daily peaks and troughs in 24-hour electrical load, which occur at 5am, 1pm, 2pm, 6pm and 11pm, are of prime importance to electricity companies in scheduling plant operation. Thus these points are chosen as the cardinal points. The forecasts for these points are generated using *parallel models*. The sum forecasts are generated using a separate sum model based on a neural network. The model is trained as in Section 4 but with a sum series instead of a parallel series. This is shown diagrammatically in figure 5 below where the original LCM forecast is combined with cardinal point forecasts and cumulative forecast to give a MTS forecast.

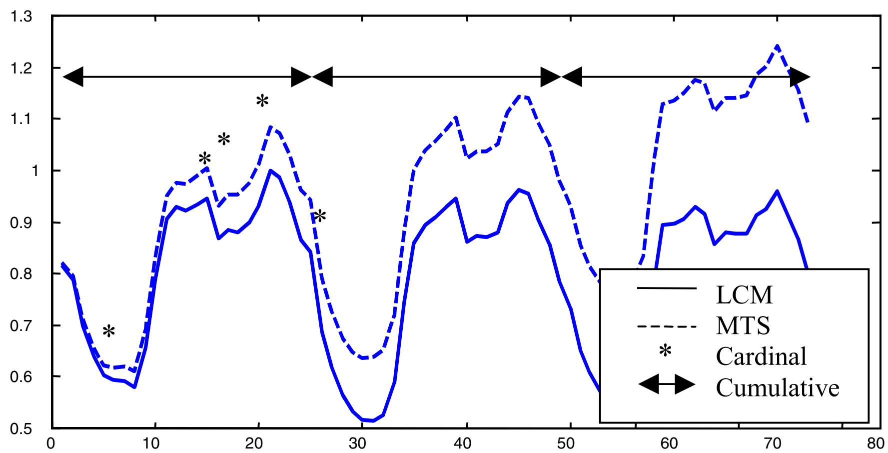


**Figure 5. Diagram showing MTS technique.**

The LCM is restricted to a state space model of the form:

$$y^{LCM}(n) = Hz(n) + \varepsilon(n) \tag{5}$$

where $n$ is the forecast origin, z($n$) is the state vector and $y^{LCM}(n)$ is the LCM forecast at the forecasting origin time $n$. The state vector z($n$) is updated via:

$$z(n) = \Phi z(n-1) + G\eta(n) \tag{6}$$

where $H$, $\Phi$, $G$ are called the observation, transition and input matrices respectively and $\varepsilon(k)$ and $\eta(k)$ are known as the channel and process noise respectively. Forecasts are generated $p$ steps ahead using:

$$y^{LCM}(n+p) = Hz(n)\Phi^{p} \tag{7}$$

The state space model used is a BSM with a dummy seasonal component ( see [10] for more details). The MTS forecast is generated by adjusting the state vector at the forecast horizon based on the differences between the LCM forecast and the cardinal and cumulative forecasts (See [10] for more details). This results in a forecast that is based on an adjusted state vector $z^{*}(n)$ at the forecasting horizon such that the MTS forecast $y^{MTS}(n+p)$, $p$ steps ahead is:

$$y^{MTS}(n+p) = Hz^{*}(n)\Phi^{p} \tag{8}$$

## 6. RESULTS

The Mean Absolute Percentage Error (MAPE) achieved by the MTS and parallel models in the novelty set are shown in Figure 6. The Preliminary AR linear model results are included as a baseline. The most significant differences between the MTS and parallel model performances are at midnight, 9am and 6-8pm. 9am and 6-8pm are also the times at which load showed possibly independent behaviour in Section 3. The 6pm parallel forecasts are however also included in the MTS model, as a cardinal point. An independent point will lead to a degraded result if used to adjust a load curve forecast as is the case in the MTS model. This can be seen in Figure 6, where the MTS MAPE for 6pm is above the preliminary AR linear model MAPE. The large MAPE for the MTS forecasts for midnight is unexplained.

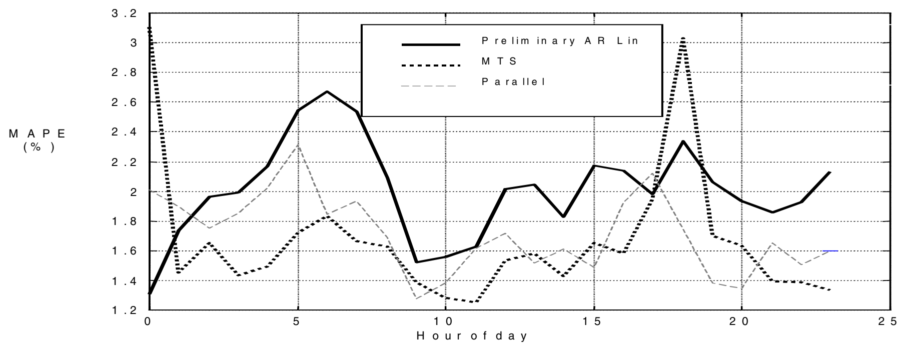


**Figure 6. MAPE as a function of the hour of day in the novelty set**

Given that the MTS model is not superior for forecasting independent points (9am, 6pm-8pm) and generates a large error for midnight, the forecasts at these times need not be used. Parallel models for these hours of the day can be used. A composite model using forecasts from the MTS model except at midnight, 9am and 6-8pm where parallel forecasts are used gives a better result than either (Table 4).

Table 4 shows the MAPE for all hours, in both the training and novelty data sets. Although the parallel models are superior in the training set the performance degrades relative to the MTS in the novelty set. This shows the parallel models are failing to generalise well and is caused by the chosen network topologies used in the parallel models.

**Table 4. Summary of results.**

| Model | Parallel | MTS | Preliminary AR | Composite model |
|---|---|---|---|---|
| Training set MAPE | 1.64% | 1.90% | 2.31% | 1.73 |
| Novelty set MAPE | 1.71% | 1.67% | 2.00% | 1.57 |

The composite model shows a consistent result in both data sets and a lower MAPE than both the MTS or parallel models in the novelty set. A sample forecast is shown in Figure 7.

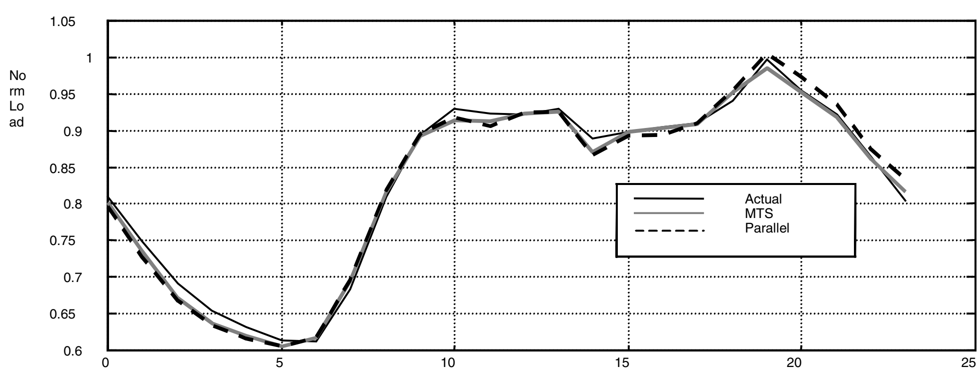


**Figure 7. Sample 1-day ahead forecast**

As can be both MTS and parallel models produce a good forecast.

## 7. CONCLUSION

24-Hour electrical load data has been shown to have a dual nature with the load at 9am and 6-8pm. The load at these hours is sufficiently independent from the load at other hours to warrant separate models. Constructing a parallel model with neural networks for every hour of the day is too computationally expensive. However with a composite model approach, far fewer parallel models are used and the computational expense is greatly reduced. The inclusion of the 6pm parallel model forecasts as a cardinal point in the MTS model influenced the load curve model not only at 6pm but also at the other hours. As the 6pm parallel series is independent it should be excluded as a cardinal point.

By taking advantage of the dual nature of the data a composite model was proposed which improved results and reduced computational expense.

## REFERENCES


[1] M.H. Choueiki, C.A. Mount-Campbell, S.C. Ahalt, "Building a quasi-optimal neural network to solve the short-term load forecasting problem", *IEEE Transactions on Power Systems*, Vol. 12, pp 1432-1439, November 1997.

[2] S.E. Papadakis, J.B. Theocharis, A.G. Bakirtzis, "Fuzzy short-term load forecasting based on load curve –shaped prototype fuzzy clustering", *European Control Conference*, Karlsruhe, Germany, (Not paginated on CD-ROM), 31st August –3rd September, 1999.

[3] R.Ramanthan, R.Engle, C.W.J. Granger, F. V. Araghi, "Short-run forecasts of electricity loads and peaks", *International Journal of Forecasting*, Vol. 13, pp 161-174, 1997.

[4] D.G. Infield, D.C. Hill, "Optimal smoothing for trend removal in short term electricity demand forecasting", *IEEE Transactions on Power Systems*, Vol.13, pp 1115-1120, August 1998.

[5] D.Fay, J.V.Ringwood, M.Condon, M.Kelly, "Comparison of linear and neural parallel time series models for short term load forecasting in the Republic of Ireland", *35rd Universities Power Engineering Conference*, Vol 1, Belfast, UK., , (Not paginated on CD-ROM), September 2000 6-8th.

[6] P.C. Gupta, "Adaptive short-term forecasting of hourly loads using weather information*", Comparative Models for Electrical Load Forecasting*, D.W. Bunn, E.D. Farmer (Eds), J. Wiley & Sons, 1985.

[7] F. Murray, J.V. Ringwood, P.C. Austin., "Integration of multi-time-scale models in time series forecasting", *International Journal of Systems Science*, Vol. 31, No. 10, 2000.

[8] A. Papoulis, Probability, random variables, and stochastic processes, 3rd ed. McGraw Hill, 1991.

[9] D.W. Long, M.Brown, C.Harris, "Principal component analysis in time-series modelling",*35rd Universities Power Engineering Conference*, Vol 1, Belfast, UK., , (Not paginated on CD-ROM), September 2000 6-8th.

[10] F.T. Murray, "Forecasting methodologies for electricity supply systems", *Ph.D. Thesis, School of Electronic engineering*, Dublin City University, 1997.

[11] G.Bebis,M.Georgiopoulos, T. Kasparis,"Coupling weight elimination with genetic algorithms to reduce network size and preserve generalisation", Neurocomputing, 17, pp 167-194, 1997.

[12] A.S. Weigend , D.E. Rumelhart, B.A. Hubermann, "Generalisation by weight elimination with application to forecasting", *Advances in Neural Information Processing Systems*, Vol. 3, pp 875-882, 1991.