Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks

Article *in* Journal of Intelligent and Robotic Systems · May 2001
DOI: 10.1023/A:1012046824237 · Source: dx.doi.org

John Ringwood
National University of Ireland, Maynooth

Journal of Intelligent and Robotic Systems **31:** 129–147, 2001.

Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks

J. V. RINGWOOD and D. BOFELLI
Department of Electronic Engineering, National University of Ireland, Maynooth, County Kildare, Ireland

F. T. MURRAY
Xilinx Ireland Limited, Saggart, County Dublin, Ireland

(Received: 27 October 2000; accepted: 20 June 2000)

Abstract. This paper examines the application of artificial neural networks (ANNs) to the modelling and forecasting of electricity demand experienced by an electricity supplier. The data used in the application examples relates to the national electricity demand in the Republic of Ireland, generously supplied by the Electricity Supply Board (ESB). The paper focusses on three different time scales of interest to power boards: yearly (up to fifteen years in advance), weekly (up to three years in advance) and hourly (up to 24 h ahead). Electricity demand exhibits considerably different characteristics on these different time scales, both in terms of the underlying autoregressive processes and the causal inputs appropriate to each time scale. Where possible, the ANN-based models draw on the applications experience gained with linear modelling techniques and in one particular case, manual forecasting methods.

Key words: artificial neural networks, electrical load, electricity demand, load forecasting, Box–Jenkins model.

1. Introduction

Electricity demand, accumulated on different time scales, presents considerably different characteristics to the time series modeller. Figure 1 shows graphs of demand accumulated on hourly, weekly and yearly bases. Note, in particular, that detailed variations in the daily profile are lost as demand is accumulated up to the weekly level, the seasonal variations of which are subsequently lost through accumulation up to the annual level.

With ever increasing drives towards efficiency in the supply of electical energy, both for economic and environmental reasons, accurate forecasting of electricity demand is vital if generating capacity is to be closely matched to that of demand. A good example is that quoted by Bunn and Farmer [7] of the UK CEGB, where in 1984 an improvement in forecasting accuracy of 1% was estimated to yield a saving in operating costs of approximately £10 million per year. For the daily profile, the requirement is for optimal scheduling of generating sets, in order to minimise the total cost of supply, which can only be obtained through advance knowledge of the

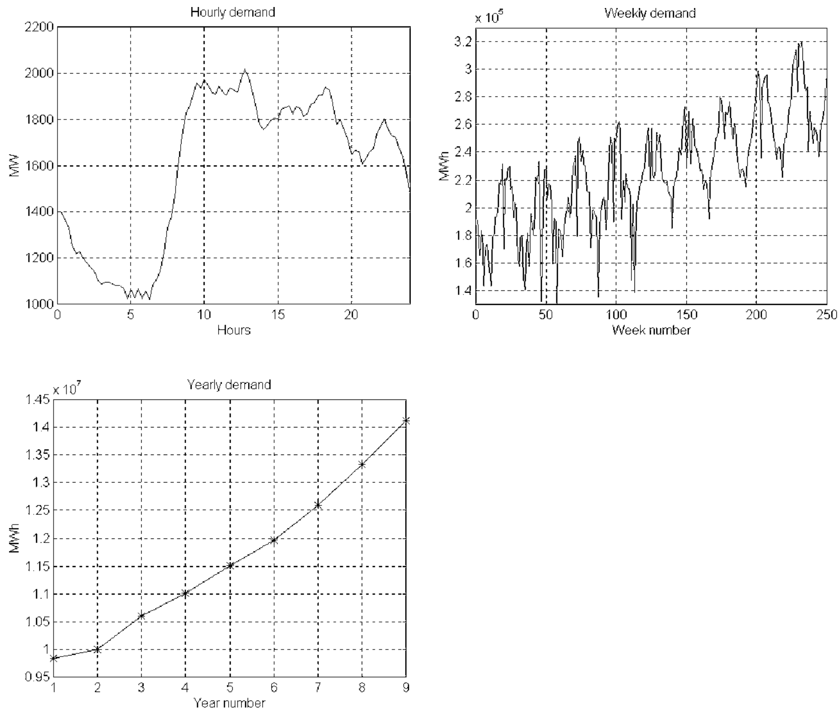

Figure 1. Electricity demand profiles on different time scales.

daily demand profile. In the medium term, weekly forecasts are used to implement maintenance schedules and fuel purchasing policies, while in the long term, annual forecasts are used for strategic planning of generating plant. A useful overview of the range of requirements is given in [1].

Many researchers have considered the forecasting of electricity demand using a variety of modelling techniques. These range from manual methods which rely on operator experience [19] to formal mathematical approaches, such as structural techniques [6], Bayesian methods [12], Box–Jenkins methods [25] and intelligent techniques, such as fuzzy logic [2] and neural networks [24]. In the case of neural networks, most studies concentrate on forecasting daily demand profiles or peak daily demand. In [3], neural networks are used to forecast peak demand, [17] and [15] forecast the entire 24 hour vector using a single network and [11] generates a separate ANN model for each hour. In *this* study, a time series modelling approach is adopted for each of the three timescales, with a single model being used recursively to forecast the series of future points required.

Recently, the power of neural networks in modelling nonlinear time series has been shown in many applications, with the documented results of a recent forecasting competition giving some comparative results [26]. However, it is important that, in applying neural networks to time series modelling, as much information as possible about the application area and experience from earlier (possibly linear) approaches be utilised where available, so that an intuitive and parsimonious solution can be determined. Most of all, it is important that appropriate tools are chosen for the job – it makes little sense to employ neural networks with their inherent problems (sensitivity to initial conditions, difficulty in structure determination and global optimisation) when the problem does not warrant a nonlinear modelling tool.

The application examples used in this paper are based on an island utility with a peak demand of about 2500 MW (used for hourly, weekly and yearly examples) operated by the Electricity Supply Board (ESB) in Ireland. For each particular utility, there are particular inputs which affect demand on the different time scales. One of the chief difficulties in accurate forecasting of electricity demand is the determination of effective inputs to use and the future prediction of them.

2. Linear Modelling Approaches

In order to provide a foundation for the neural network techniques employed, two linear modelling approaches are presented which are appropriate for seasonal data, such as weekly and hourly load *and* non-seasonal data, such as annual load. The first (more traditional) approach involves modelling using the Box–Jenkins [5] methodolody. This procedure involves the application of transformations which eases the subsequent modelling exercise, which is performed using a model with seasonal and non-seasonal sections. The second method has similarities to the first, but segments the model into three distinct parts, each of which contributes to the overall model output. Such a model is termed a structural model [14]. Models developed using these procedures can be either purely autoregressive (depend only on previous model outputs) or can be causal (driven by appropriate inputs).

2.1. BOX–JENKINS METHODOLODY

This general linear modelling approach follows the following procedure:

1. Determination of seasonality of time series and application of seasonal differencing.
2. Application of further differencing transformations to make the time series stationary.
3. Investigation of significant inputs to use as causal variables with the model.
4. Determination of orders of seasonal and non-seasonal regressors.
5. Identification of model parameters.

The univariate Box–Jenkins model is derived from the general SARI(p, d) (P, D) (seasonal autoregressive integrated) model which can be written (with B as the delay operator) as:

$$\Phi_p(B)\Phi_P(B^L)\nabla_L^D\nabla^d Y_t = a_t, \tag{1}$$

where:

- Y_t is the time series to be modelled,
- $\nabla_L^D \nabla^d = (1 - B^L)^D (1 - B)^d$ is a differencing transformation required if the data is nonstationary,
- d is the degree of non-seasonal differencing,
- D is the degree of seasonal differencing, and
- L is the season length,
- a_t is the forecast error,
- $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ is the non-seasonal autoregressive operator of order p,
- $\Phi_p(B^L) = (1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \cdots - \phi_{P,L} B^{PL})$ is the seasonal autoregressive operator of order P.

The lags p and P are determined using correlation analysis, as are the degree of the differencing operators, d and D. The seasonality of the data, L, is usually known *a priori*, or may also be determined using correlation analysis. A variety of methods may be used to determine the model parameters in the $\Phi(B)$ polynomials, iterative least squares proving a popular approach. Following model construction, t-ratio tests may be used to assess the significance of the model.

For input-driven models, the general Box–Jenkins transfer function model is of the form

$$z^{(y)}(t) = \frac{\Omega(B)}{\Delta(B)} z^{(x)}(t - b) + \eta(t), \tag{2}$$

where $z^{(y)}(t)$, $z^{(x)}(t)$ represents the stationary output and stationary input respectively and b is a pure delay parameter. $\Omega(B) = (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)$, where s is the number of past input values influencing current output values and $\Delta(B) = (1 - \delta_1 B - \cdots - \delta_r B^r)$, where r is the number of past output values influencing current output values. $\eta(t)$ is a coloured noise series which is usually represented by an ARIMA model. A plot of the sample cross correlation between the input series and the output series is used to evaluate the response lag time b, and the orders r and s of the polynomials $\Omega(B)$ and $\Delta(B)$. Once the b, r and s values are determined a preliminary transfer function model is estimated. The sample autocorrelation (SAC) and sample partial autocorrelation (SPAC) of the residuals of this model are examined in order to identify an ARIMA model for the noise series $\eta(t)$. The adequacy of the final transfer function model can now be tested using the techniques outlined in [5, 4].

2.2. STRUCTURAL STATE-SPACE MODELS

Structural models adopt a different methodology than in Section 2.1 by modelling the trend and seasonal components, rather than removing their effect prior to modelling using transformations. A structural time series model consisting of a trend and a seasonal component may described by

$$y(k) = t(k) + p(k) + \varepsilon(k), \tag{3}$$

where $t(k)$ is a linear trend, $p(k)$ a seasonal component and $\varepsilon(k)$ a zero mean, serially uncorrelated white noise component. A *generalised random walk* (GRW) model [23] can be used to model the trend behaviour $t(k)$. The state-space form of the GRW model is defined by:

$$\begin{aligned} \begin{bmatrix} t(k) \\ d(k) \end{bmatrix} &= \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} t(k-1) \\ d(k-1) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1(k) \\ \eta_2(k) \end{bmatrix}, \\ t(k) &= [1 \quad 0]\mathbf{x}(k), \end{aligned} \tag{4}$$

where α, β and γ are constant parameters; $t(k)$ is the trend at sample k, $d(k)$ is a second state variable and $\eta_1(k)$ and $\eta_2(k)$ are zero mean, serially uncorrelated discrete white noise inputs. An *integrated random walk* (IRW) is obtained with $\alpha = \beta = \gamma = 1$; $\eta_1(k) = 0$. If the seasonal component is well defined and stationary it can be modelled by a *periodic random walk* (PRW) or a *differenced periodic random walk* (DPRW) model [23]. The DPRW model is defined by:

$$p(k) = -\sum_{i=1}^{s-1} p(k-i) + \eta_p(k-1), \tag{5}$$

where s is the seasonal period and $\eta_p(k)$ is a zero mean white noise disturbance input. If the trend component is represented by an IRW model and the seasonal component is represented by a DPRW model, then the complete state-space model is defined by the following:

$$\begin{bmatrix} t(k) \\ d(k) \\ ----- \\ p(k) \\ p(k-1) \\ \vdots \\ p(k-(s-2)) \end{bmatrix} = \left[\begin{array}{cc|cccccc} 1 & 1 & 0 & 0 & . & . & 0 \\ 0 & 1 & 0 & 0 & . & . & 0 \\ - & - & - & - & - & - & - \\ 0 & 0 & -1 & -1 & . & . & -1 \\ 0 & 0 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & 1 & 0 \end{array}\right] \times$$

$$\begin{bmatrix} t(k-1) \\ d(k-1) \\ ----- \\ p(k-1) \\ p(k-2) \\ \vdots \\ p(k-(s-1)) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_2(k-1) \\ --- \\ \eta_p(k-1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

$$y(k) = [1 \quad 0 \quad 1 \quad 0 \quad \ldots \quad 0 \quad 0 \quad 0]x(k) + \varepsilon(k). \tag{7}$$

In order to perform a prediction, a Kalman filter is used over the identification data set to provide initial state estimates for the model. Covariances for the (process) noise sources $\eta_1(k)$ and $\eta_2(k)$ and the measurement noise, $\varepsilon(k)$, are determined using maximum liklihood optimisation.

3. Time Series Modelling Using Neural Networks

3.1. NETWORK INPUT STRUCTURE

A total black-box approach to neural network modelling of dynamical systems or time series would be to utilise a model of the form shown in Figure 2, with tapped delay lines for input and output variables forming the input to the neural network. Such an approach is common in a variety of time series and model-based control system applications [8, 16]. However, such an approach may disregard structural information about the dynamical model available from linear analysis. In the current study, an effort is made to incorporate information on an effective input and model structure suggested by linear time-series modelling techniques.

3.1.1. *Box–Jenkins Model*

For the Box–Jenkins methodology in Section 2.1, two nonlinear options are possible. Expansion of Equation (1) gives:

$$\begin{aligned}&\big(1-\phi_1 B-\cdots-\phi_p B^p-\phi_{1,L}B^L+\phi_1\phi_{1,L}B^{L+1}+\cdots+\phi_p\phi_{1,L}B^{L+p}-\\&\quad-\cdots-\phi_{P,L}B^{PL}+\phi_1\phi_{P,L}B^{PL+1}+\cdots+\phi_p\phi_{P,L}B^{PL+p}\big)\big(1-B^L\big)^D\times\\&\quad\times(1-B)^d Y_t=a_t.\end{aligned}\tag{8}$$

A corresponding nonlinear model for the structure in Equation (8) could now be defined as:

$$\begin{aligned}&f(Y_y,\ldots,Y_{t-p-1},Y_{t-L},\ldots,Y_{t-L-p-1},\ldots,Y_{t-PL},\ldots,Y_{t-PL-p-1},\ldots,\\&\quad Y_{t-PL-D},\ldots,Y_{t-PL-p-D-1})=a_t.\end{aligned}\tag{9}$$

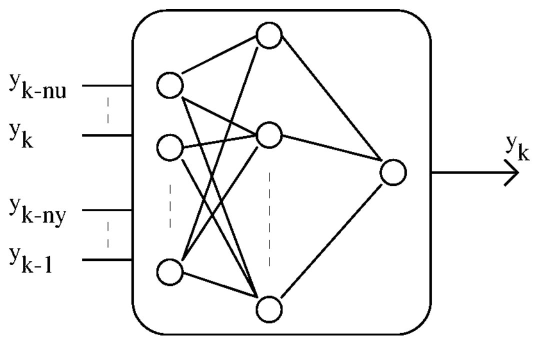

Figure 2. Network with classical input structure.

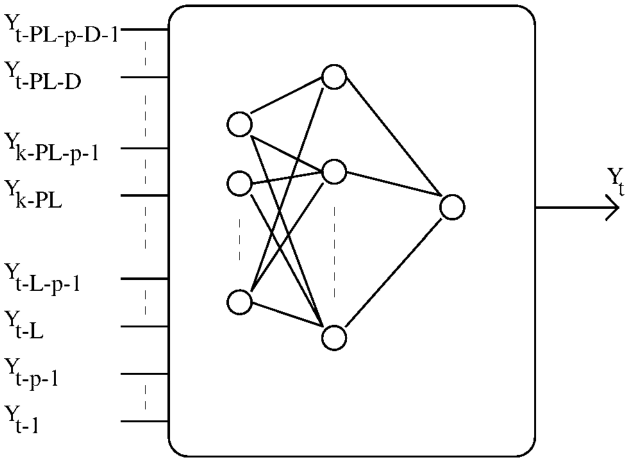

Figure 3. Network with Box–Jenkins input structure.

This model will be termed a NNBJ type 'A' model. Alternatively, defining Z_t as:

$$Z_t = \left(1 - B^L\right)^D (1 - B)^d Y_t \tag{10}$$

a model (NNBJ type 'B') of the form:

$$g(Z_y, \dots, Z_{t-p}, Z_{t-L}, \dots, Z_{t-L-p}, \dots, Z_{t-PL}, \dots, Z_{t-PL-p}) = a_t \tag{11}$$

results, where the neural network is used to model data which has already been subject to seasonal and one-step differencing. To obtain the final forecast, the output from the neural network must be appropriately integrated, using seasonal and one-step integration. As an example, Figure 3 shows a neural network forecasting model using a form corresponding to Equation (9). Compared to Figure 2, the input structure had been modified so that the network is focussed on the most effective inputs. This generally also results in fewer inputs overall, resulting in reductions in training times. For a model of the form of Equation (9), the total number of inputs is $(P + D)(p + d + 1) + p + d$. For a 'standard' autoregressive (AR) model of the form of Figure 2, it would be usual to choose inputs which span a season, i.e., L inputs. In the weekly example, this would yield 52 inputs for the AR model, with only 31 and 21 inputs respectively for models based on Equations (9) and (11), while the hourly example would yield 24 inputs for the AR model as against 17 and 9 inputs, respectively.

3.1.2. *Structural Model*

For the structural state-space model presented in Section 2.2, the approach is to let the linear sections (IRW + DPRW) model the trend and seasonal components, with the neural network used to model the remaining residuals in ε_k. This concurs roughly with the model presented in (11). However, in order to ensure a good training set for the network, a state-space smoothing algorithm [13] is employed to back-smooth the state estimates resulting from the Kalman filter, since the estimates from the filter will be poor during initial convergence. The output $[t(k) + p(k)]$ from the smoothed state estimates are then subtracted from the identification data, with this difference providing the network training set. Since the network is now dealing with (approximately) data which has been detrended and deseasonalised, an input structure similar to that in Equation (11) can be used.

3.2. NEURAL NETWORK DESIGN

3.2.1. *Network Architecture and Structure*

Network architecture requirements are for a network which can operate recurrently (since the time series is autoregressive) and produce a continuous output. In addition the size of the network should not be intractable. The latter condition excludes the utilisation of local approximators, due to the dimension of the input space encountered in the current application and feedforward Multi-Layer Perceptrons (MLPs) were adopted as a suitable network structure [10, 20]. Feedforward networks were preferred over recurrent networks due to ease of training. In terms of configuration, a three layer structure was adopted with a linear output neuron (effectively removing any restriction on output range), while the number of neurons in each layer was determined using optimisation across Monte-Carlo runs, with the mean-squared error across a validation set as a criteria.

3.2.2. *Neural Network Training*

Two important aspects of network training which must be considered here are the choice of training algorithm and the training cessation point. A standard LMS gradient technique with backpropagation was employed for training, which also included a momentum term and adaptive learning rate. Faster techniques, such as the Lervenberg–Marquardt algorithm [18] were also examined, but found to be extremely sensitive to initial conditions and local minima. This can be overcome, to some extent, by utilising sufficient Monte-Carlo runs, but this extra computation, combined with the slower computational speed of such algorithms was found to more than offset any gains in convergence speed. Neural networks trained for time series applications are typically trained using single step prediction criteria. However, this does not always determine the weight set which optimises the multi-step prediction performance of the network. One reason for this is that backpropagation training with multi-step criteria are difficult to design and can be computationally intensive, particularly when the prediction horizon is long. A compromise is to train the network for single-step performance, but examine the multi-step performance during training. Figure 5 shows, for a weekly example, the variation in single step sum-squared error (SSE) and the multi-step mean absolute error (MAE) over

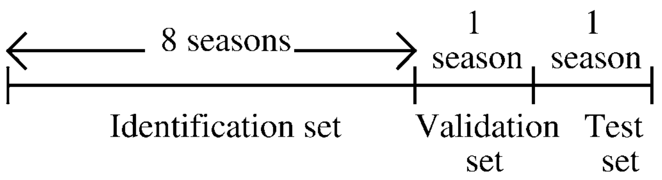

Figure 4. Time series data segmentation (weekly data).

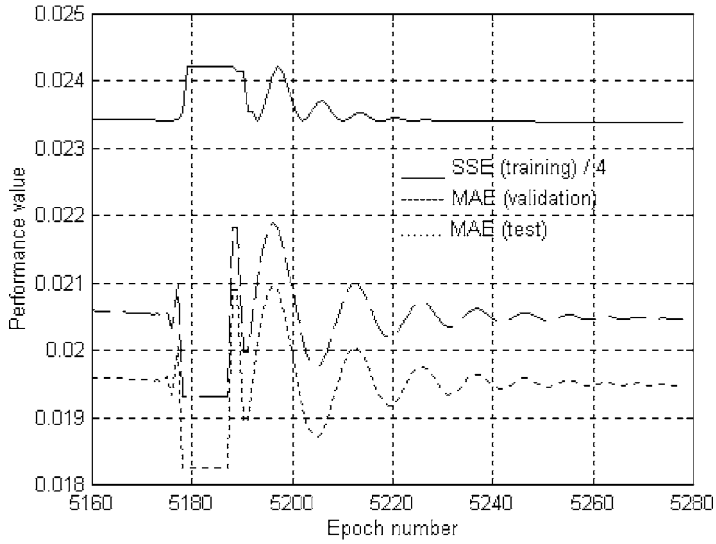

Figure 5. Variation in single- and multi-step criteria during training.

the validation set and test set (over which the prediction is to be performed, see Figure 4). Note that the variation in the multi-step criteria for the validation and test portions are *consistent*, allowing a stopping point (weight vector) to be chosen based on the validation set which will give good multi-step performance when doing the actual forecast. For example, a choice of weights at epoch 5180 gives a multi-step performance value 0.01825, while the corresponding value at epoch 5188 is 0.02090 (approximately 14% worse), in spite of the fact that the single-step SSE suggests that a choice of weights at epoch 5180.

3.3. PERFORMANCE CRITERIA

The performance criteria used in this study are the mean absolute error and mean squared error defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_p(i) - y(i)\right|, \qquad \mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_p(i) - y(i)\right|^2. \tag{12}$$

The MAE penalises all errors uniformly, whereas the MSE penalises larger errors more heavily.

4. Modelling of Hourly Electricity Demand

4.1. PRELIMINARY DATA ANALYSIS

Figure 6 shows the characteristic curve for a full week of daily load profiles for Ireland. The profiles generally follows the same basic shape (other than the presence of an evening peak in Summer only) for each weekday, with variations in the basic profile for Saturdays and Sundays (see Figure 1 also). In addition, the peak consumption values vary considerably between Summer and Winter causing a vertical translation of the profile. This can be accounted for, to some degree at least, by using a temperature input, but it was found necessary to use separate sets of models for Summer and Winter, with separate models for each day of the week, indicating the unique nature of each daily pattern.

In order to optimise the performance of the neural network model, the very obvious regular daily profile is removed from the data, leaving the NN model to forecast the *differences* between the chosen day and the standard profile. This accords with the practice of operators in the ESB [19], where a particular 'standard day' is chosen as a template upon which the variations are forecast. Typical choices for 'standard' day include the same day last week or same day last year. Variations are typically forecast using the difference of input variables (e.g., weather variables) between the standard day and the forecast day. Forecasts for these input variables are usually obtained from national meteorological services. A set of candidate weather inputs are chosen [9] according to correlation analysis as temperature, wind speed, solar intensity and humidity.

4.2. MODELLING

In order to ensure consistency between training and forecasting, the 'standard' day was chosen as the same day last year, with the model being used to forecast

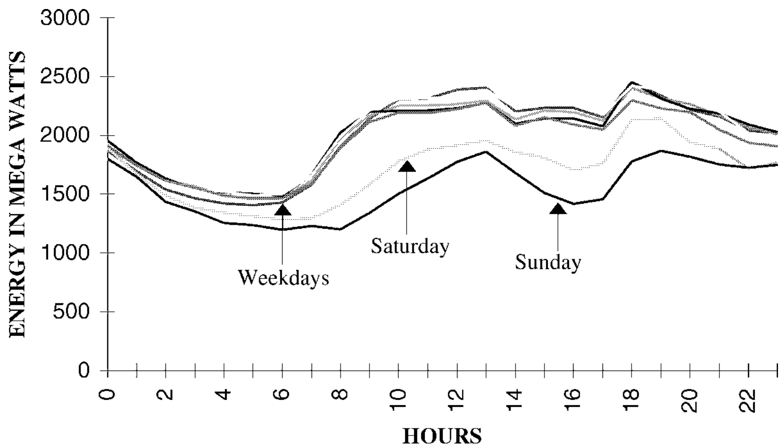

Figure 6. Daily profile for each day of week 2 February 1991.

the differences between this day and the chosen day, based on the autoregressive behaviour of the data and the input variables. Such a choice is confirmed by performing pattern matching (with dc removed) with a range of standard day candidates. From correlation analysis, stationarity transformations corresponding to $d = 1 = D$ (as in (1)) were suggested. However, performance tests would indicate that stationarity transformations do not add greatly to model performance, which may be due to the fact that the seasonal, dc and trend components of the series is removed by subtracting off the 'standard' day profile. Further correlation tests revealed candidate B–J model structures as $p = 4$ and $P = 1$ or $P = 0$, with the same shape correlation functions for *all* days of the week. Model-based evaluation is used to determine:

- Final B–J model structure ($P = 1$ or 0).
- Final selection of effective inputs.
- Network size (can be different for each model).

An MLP network with tan-sigmoid neurons are employed, with a linear output neuron. The data, including the causal inputs, are scaled using a common factor to simplify rescaling. Data for training was taken from the 10 previous weeks, giving 240 time-series training points for each model. From initial experiences with the neural network models, the following choices were made:

Effective input:	Temperature
Max training duration:	8000 epochs

4.3. RESULTS

Note that Peak 1 is at 1200 h (the midday peak) and Peak 2 at 1900 h (the evening peak). Results are presented here for the Summer only and it is likely that larger errors could occur in Winter, given the larger sensitivity to temperature variations.

Table I. Summary of typical results from hourly forecasting

Day of week	Model BJ(p, P)	Network size	MAE	$\sqrt{\text{MSE}}$	Peak 1 % error	Peak 2 % error	Average % error
Monday	BJ(4, 0)	3-5-1	32.7	40.0	1.08	1.31	2.43
Tues.	BJ(4, 1)	3-5-1	20.9	25.9	1.01	2.43	1.48
Wed.	BJ(4, 1)	2-4-1	27.0	33.5	0.69	1.49	1.90
Thurs.	BJ(4, 1)	5-10-1	25.0	32.1	1.79	2.95	1.76
Friday	BJ(4, 0)	3-5-1	22.7	26.9	1.67	0.63	1.59
Sat.	BJ(4, 1)	3-5-1	22.6	26.3	2.0	1.35	1.7
Sunday	BJ(4, 1)	3-5-1	28.4	35.4	0.19	0.12	2.62

Table II. Comparative hourly results from other researchers

Model	Error type	Max. % error		Min. % error		Aver. % error	
Azzam-ul-Asar [3]	Peak load	5.71		0.02		1.96	
Park [24]	Peak load	6.64		0.13		2.06	
Ours	Peak load	Peak 1 4.6	Peak 2 3.7	Peak 1 0.19	Peak 2 0.12	Peak 1 1.69	Peak 2 1.73
Park [24]	Total load	5.64		0.03		1.68	
Ours	Total load	3.85		1.48		2.09	

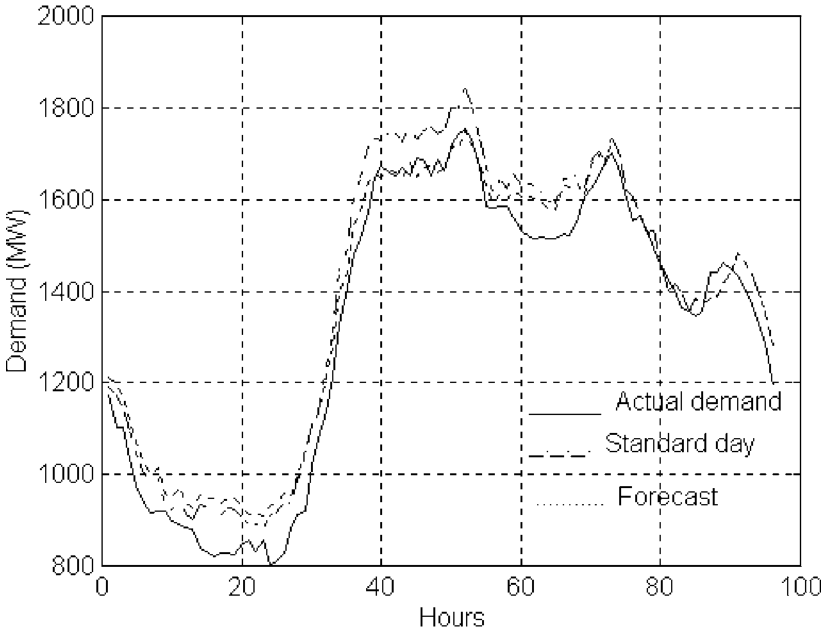

Figure 7. Typical hourly forecasting result.

However, the results are encouraging, given that there is completely random (unpredictable) load on the system of 50 MW corresponding to the switching on and off of a steel furnace. In most of the cases considered, the maximum error was less than 80 MW. Peak load is of particular interest to utilities, since it determines the maximum generating capacity required. Although dealing with different utilities, we present some comparative results from other researchers (see Table II).

Figure 7 shows a typical plot of the adjustment carried out on the ‘standard’ day profile for a Thursday, as an example.

5. Modelling of Weekly Electricity Demand

5.1. PRELIMINARY DATA ANALYSIS

From Figure 1, it is seen that the weekly data is seasonal with a rising trend. This seasonality is generally due to temperature variations, with extra heating requirements in Winter, but may be due to other factors, such as agricultural activity, where this is the dominant factor [22]. In some cases, weekly consumption does not exhibit such a strong harmonic form, for example in cases where significant air conditioning is used in the Summer months. This suggests the use of a temperature input, but for the current application, the harmonic variation can be adequately captured by an autoregressive model, due to the regularity of the variation.

5.2. MODELLING

Both Box–Jenkins and structural models may be used to model this seasonal demand profile. Although the cyclical variations depend heavily on weather inputs, it is assumed that these harmonic variations are adequately captured by the seasonal differencing and regressor of the Box–Jenkins model and the harmonic components of the structural model. In addition to the Box–Jenkins and structural models and for comparison purposes, a traditional AR neural net model was also evaluated which utilised 52 lagged inputs, spanning a complete season. A network architecture of 3-5-1 was used for this model. For each model, the data was normalised to ensure an input range in the region $[-1, +1]$ and approximately 30,000 epochs used for training, where the actual training cessation point determined as outlined in Section 3.2.2.

5.2.1. *Box–Jenkins ANN Model*

For the case of the Box–Jenkins model, correlation tests were used to determine the system input structure as in Figure 3. From correlation analysis (SAC and SPAC), the appropriate model structure is specified by the following parameters:

$$p = 6, \quad P = 2, \quad d = 1, \quad D = 1, \quad L = 52.$$

Neural network Box–Jenkins models of both type ‘A’ and ‘B’ (see Section 3.1.1) were identified and a network architecture of 3-5-1 arrived at, based on examination of the multi-step errors for various architectures over a set of Monte-Carlo trials. Log-sigmoid neurons were utilised in the hidden layers (employing 3 and 5 neurons, respectively), with a linear output neuron.

5.2.2. *Structural Model*

As shown in Section 3.1.2, the structural model contains components which allow effective modelling of the seasonal and trend components. Following the determination of the structural model components, the residual is formed, which is now modelled by the neural network, following backward smoothing to generate quality state estimates near the forecasting origin. Correlation analysis (SACF) was used to determine the appropriately lagged inputs to use, returning $p = 3$ and $P = 3$. Again, a multi-step error-based performance analysis was used to determine an appropriate network architecture as 20-60-1, based on a set of Monte-Carlo runs.

5.3. RESULTS

Table III compares the multi-step (52 step ahead) forecasting results for linear and neural models. Figures 8 and 9 demonstrate qualitatively the improvement achieved. While it is interesting to note that there is significant improvement in the use of neural networks with Box–Jenkins and Structural models, it is perhaps more dramatic the effect that the choice of input structure has on performance. In particular, note the poor performance of the 'classical' autoregressive (with inputs formed from the past 52 week's demand) model. The neural version of this, which incidentally is inferior to its linear counterpart, has a much poorer performance than (either linear or neural) Box–Jenkins or structural models.

Table III. Comparative results for linear and neural models

Model type	MAE linear	MAE neural network
Autoregressive (52 inputs)	1.1101×10^4	1.2481×10^4
Box–Jenkins	1.0691×10^4	0.9040×10^4 NNBJ type 'A'
		0.6587×10^4 NNBJ type 'B'
Structural state space	0.7687×10^4	0.7198×10^4

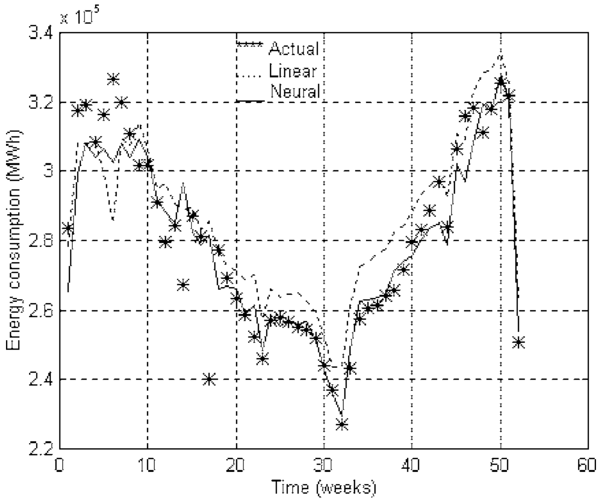

Figure 8. Forecasts using Box–Jenkins models.

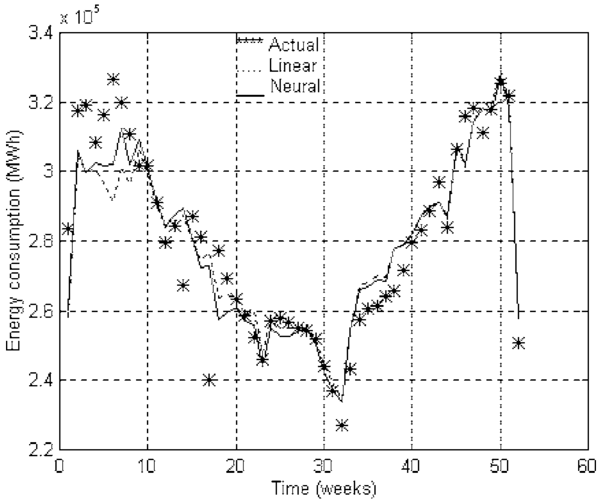

Figure 9. Forecasts using structural models.

6. Modelling of Yearly Electricity Demand

6.1. PRELIMINARY DATA ANALYSIS

Yearly load demand is only available from 1972–1994, which gives only 23 data points for the load and input variables. The input candidates which are appropriate on this time scale are average industrial wage (AIW), gross domestic product (GDP), average electricity unit price (AUP), number of customers (NOC) and a temperature input. The temperature input which was found to be most appropriate for this application is annual heating degree-days recorded below 15.5 °C. For a description of heating degree–days, see [22]. Disaggregated data is also available, where load is broken down into industrial, domestic and commercial sectors with similarly disaggregated data available for number of customers and price of electricity units. This offers the potential to consider both aggregated and disaggregated approaches to the problem. Finally, it is seen from Figure 1 that the yearly data does not contain any obvious seasonal components, and so seasonal differencing is not appropriate.

6.2. MODELLING

The Box–Jenkins transfer function (as given in (2)) modelling approach is used to determine the appropriate inputs for the neural net model, as in Section 4. Both type ‘A’ and ‘B’ models can be considered, and predifferencing the data (in accordance with type ‘B’) would remove the strong trend prior to modelling, but at the expense of the loss of a data point. Since stationarity is not a requirement with neural network modelling [10] and there is already a shortage of data, one-step differencing will be omitted. Following the Box–Jenkins modelling procedure, model structures

Table IV. Structural parameters for transfer function models

TF model	Order of $\eta(t)$	Input	b	r	s
Total	δ and $p = (1)$	AIW	0	0	2
		GDP	0	0	0
		TAUP	0	1, 8	0
Industrial sector	δ and $p = (2, 4, 6, 8)$	AIW	0	0	0
		GDP	0	0	0
		IAUP	1	0	0
Domestic sector	$p = (1)$	AIW	0	1	2
		GDP	0	0	0
		DAUP	0	0, 3	0
Commercial sector	$p = (3, 4)$	GDP	0	0	0
		TAUP	0	1	0
		CNOC	0	0	0

Table V. Autoregressive lags for aggregated and disaggregated load

Electricity sales	Autoregressive lags	p
Total	1, 2, 3, 4	4
Industrial sector	1, 2	2
Domestic sector	1	1
Commercial sector	1	1

were determined for each of the disaggregated load variables and Table IV shows the structural parameters along with the appropriate inputs for each load sector. The sample cross-correlation (SCC) function is used to determine the orders b, r and s for each of the inputs associated with the output and the residuals of this initial model are then examined to identify a suitable noise model $\eta(t)$. The resulting lags suggested using the SPAC are given in Table V.

A selection of neural networks were now trained with a maximum training duration of 20,000 epochs. The optimal network size was determined using performance analysis MAE across the validations set (see Figure 4), with a summary of the results for total load given in Table VI.

6.3. RESULTS

The results for the neural network models are shown in Table VII, with linear results shown for comparative purposes. It is interesting to note that the linear models

Table VI. Selection of network architecture

Sector	Minimum MAE
Total	2-6-1
Industrial	3-9-1
Domestic	3-5-1
Commercial	1-3-1

Table VII. Comparative performance of linear and ANN yearly models

Sector	MAE ANN models	MAE linear models
Aggregated load	153.98	218.82
Industrial	41.15	81.52
Domestic	98.31	78.39
Commercial	84.95	38.05
Σ disaggregated	224.41	197.96

outperform the neural models for the disaggregated case, although the ANN results are considerably better than the linear models for the aggregated case.

The main difficulty with the disaggregated case seems to be the modelling of domestic and (particularly) commercial load. Plots of actual and predicted load variations cannot be shown for confidentiality reasons.

7. Conclusions

This study of the use of neural networks as a time series modelling tool for forecasting electricity demand has shown some of the benefits and drawbacks of the application of neural networks. Possibly one of the greatest difficulties encountered is the choice of MLP network architecture. A network growing approach has been used, with multi-step performance across the validation sequence used as a selection criterion, but the sensitivity to initial conditions, requiring the evalualtion of Monte-Carlo runs, makes the procedure somewhat unsatisfactory.

Another general comment relates to the performance of the ANN models against linear models. In general, from the results presented in the paper, it can be said that the ANN models perform better. However, the result in the case of the yearly demand, where the disaggregated modelling results were best for the linear models, creates some cause for concern. Some comfort can be drawn from the fact that the Industrial ANN model is clearly superior to its linear counterpart, and a combination of linear and neural disaggregated models would result in an MAE of 157.59, which, although marginally inferior, compares well with the aggregated models. However, the result, in general, mitigates against the use of disaggregation, which is generally accepted to be beneficial in exploiting the detailed characteristics present in the disaggregated data. The fact that the neural models do not include the linear models as a particular case [10] is clear and indicates a serious potential pitfall in the application of neural networks. This problem could be somewhat avoided with the availability of a clear test for nonlinearity, which would provide an indication of the appropriateness of neural networks for particular time series applications. It is, however, probable that the difficulties encountered in the yearly demand application are exacerbated by the brevity of the data set and the application of *linear* correlation techniques for input structure determination.

The results obtained for the hourly load application are also noteworthy. In particular, although a time series approach was taken, the accuracy of the peak prediction is good, which is usually the focus of short-timescale demand forecasting exercises [19]. If further accuracy in peak prediction is required, a separate set of models, which predict the cardinal points on the daily profile (such as peaks and valleys), can be generated and integrated with a multi-time scale method [21], in order to provide particularly good fitting of the time series at the cardinal points.

Acknowledgements

The authors are grateful to the ESB for the supply of data and their input to the electricity demand forecasting project. In particular, wish to thank Dermot Byrne, Pat Mangan and Paddy McEvoy of ESB National Grid and Keelin O'Brien of the Cystomer Supply and Marketing Department.

References

1. Al-Alawi, S. M. and Islam, S. M.: Principles of electricity demand forecasting – Part 1: Methodologies, *Power Engineering J.* (June 1996).
2. Al-Anbuky, A., Bataineh, S., and Al-Aqtash, S.: Power demand prediction using fuzzy logic, *Control Engrg Practice* **3**(9) (1995).
3. Azzam-ul-Asar and McDonald, J. R.: A specification of neural network applications in the load forecasting problem, *IEEE Trans. Control Systems Techn.* **2**(2) (June 1994).
4. Bowerman, B. and O'Connell, R.: *Time Series Forecastig – Unified Concepts and Computer Implementations*, Duxbury Press, 1987.
5. Box, G. and Jenkins, G.: *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1976.
6. Bruce, A., Jurke, S., and Thomson, P.: Forecasting load-duration curves, *J. Forecasting* **13** (1994).
7. Bunn, D. W. and Farmer, E. D. (eds): *Comparative Models for Electrical Load Forecasting*, Wiley, New York, 1985.
8. Chakraborty, K., Mehrotra, K., Mohan, C. K., and Ranka, S.: Forecasting the behaviour of multivariate time series using neural networks, *Neural Networks* **5** (1992) 961–970.

9. Commarmond, P. and Ringwood, J. V.: Quantitative fuzzy modelling of short-time-scale electricity consumption, in: *Proc. of Irish Signals and Systems Conf. (ISSC'97)*, Derry, June 1997.
10. Dorffner, G.: Neural networks for time series processing, *Neural Network World* **6** (1996).
11. El-Sharkawi, M. A., Marks II, R. J., Oh, S., and Brace, C. M.: Data partitioning for training a layered perceptron to forecast electric load, in: *Proc. of the 2nd IEEE Internat. Forum on Appl. of Neural Networks to Power Systems*, Yokohama, 1993.
12. Gunel, I. Forecasting system energy demand, *J. Forecasting* **6** (1987).
13. Harvey, A. C.: A unified view of statistical forecasting proceedures, *J. Forecasting* **3** (1984) 245–275.
14. Harvey, A. C.: *Structural Time Series Models and the Kalman Filter*, Cambridge Univ. Press, Cambridge, 1989.
15. Kermanshahi, B. S., Poskar, C. H., Swift, G., McLaren, P., Pedrycz, W., Buhr, W., and Silk, A.: Artificial neural network for forecasting daily loads of a Canadian electric utility, in: *Proc. of the 2nd IEEE Internat. Forum on Appl. of Neural Networks to Power Systems*, Yokohama, 1993.
16. Kuschewski, J. G., Hui, S., and Zak, S.: Application of feedforward neural networks to dynamical system identification and control, *IEEE Trans. Control Systems Techn.* **1**(1) (March 1993).
17. Lee, K. Y., Choi, T. I., Ku, C. C., and Park, J. H.: Short-term load forecasting using diagonal recurrent neural network, in: *Proc. of the 2nd IEEE Internat. Forum on Appl. of Neural Networks to Power Systems*, Yokohama, 1993.
18. Ljung, L.: *System Identification: Theory for the User*, Prentice-Hall, Englewood Clifs, NJ, 1987.
19. Lonergan, T. and Ringwood, J. V.: Linguistic modelling of short-timescale electricity comsumption using fuzzy modelling techniques, in: *Proc. of Irish DSP and Control Colloquium (IDSPCC'95)*, Belfast, June 1995.
20. Mitchell, J.: Comparing feedforward neural network models for time series prediction, in: *Proc. of Conf. on Neural Computing Research and Applications*, Belfast, June 1992.
21. Murray, F. T., Ringwood, J. V., and Austin, P. C.: Integration of multi-time-scale models in time series forecasting, *Internat. J. of Systems Sci.* **31**(10) (2000).
22. Murray, F. T. and Ringwood, J. V.: Improvement of electricity consumption forecasts using temperature inputs, *Simulation Practice Theory* **2** (1994).
23. Ng, N. C. and Young, P. C.: *Recursive estimation and forecasting of non-stationary time series*, *J. Forecasting* **9** (1990).
24. Park, D. C., El-Sharkawi, M. A., Marks II, R. J., Atlas, L. E., and Damborg, M. J.: Electricity load forecasting using an artificial neural network, *IEEE Trans. Power Systems* **6**(2) (May 1991).
25. Watson, M. W., Pastuszek, L. M., and Cody, E.: Forecasting commercial electricity sales, *J. Forecasting* **6** (1987).
26. Weigend, A. S. and Gerschenfeld, N. A.: *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, Reading, MA, 1992.