COMPARISON OF LINEAR AND NEURAL PARALLEL TIME SERIES MODELS FOR SHORT TERM LOAD FORECASTING IN THE REPUBLIC OF IRELAND

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

† Dublin City University †† NUI Maynooth ††† Electricity Supply Board

ABSTRACT

This paper presents a comparison between parallel linear and parallel neural network models. Parallel models consist of 24 separate models, one for each hour of the day. Each parallel model decomposes the load into a linear Auto-Regressive (AR) part and a residual. Exogenous linear and neural network model performance is compared in predicting this residual. Three days or 72 hours of current and delayed weather variables are available as exogenous inputs for the residual models. Input selection comprises of testing the bootstrapped performance of a linear model. The inputs are ordered using 4 methods derived from a mix of the T-ratio of the linear coefficients and Principal Component Analysis (PCA). The neural network models are found to give superior results due to the non-linear AR nature of the residual.

1 INTRODUCTION

1.1 SYSTEM BACKGROUND

The Irish climate may be described as oceanic temperate with few weather extremes [1]. There is a strong correlation between weather and demand. In an Irish context, while extreme weather conditions rarely occur, week to week variations can be quite significant. The weather affects the load mainly due to heating requirements [1].

Irish demographics consist of a population of 3½ million with 2 million living in the greater Dublin area. The weather measurements used in this study are thus made at Dublin airport.

The national grid is essentially an isolated system with a lack of strong interconnection, with a yearly peak of the order of 4000 MWs [2].

1.2 RESEARCH OBJECTIVES

Individual linear models for each hour of the day, also known as linear parallel models, have been examined by [3] and [4] and found to be advantageous in representing differing time of day fluctuations. Non-linear techniques have been shown to provide superior results to linear techniques both in the Irish context and for other systems [1,5,6,7]. The purpose of this study is to examine the benefits of applying a parallel Neural Network (NN) model to Irish electricity demand data.

The modelling approach consists of decomposing the load into a linear AR part, $y(t)$, and a residual $r(t)$. [3,4,5] Have taken similar approaches. The residual is then modelled using exogenous weather inputs and in the case of the neural network additional lagged residual inputs (Figure 1).

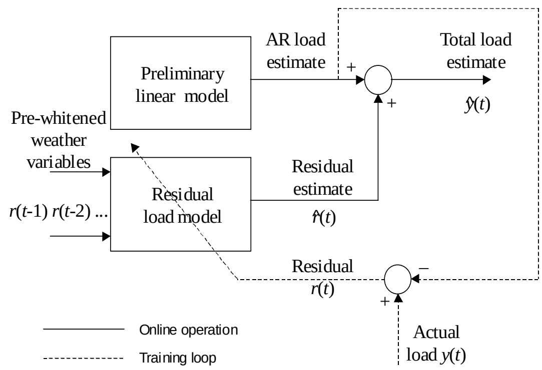

Figure 1 Load residual modelling approach

1.3 DATA SET DETAILS

A database of electricity demand from 1987 to 1998 on an hourly basis is available. The data is sifted for data between Tuesday and Thursday in the months January to March to avoid weekend, Christmas and daylight savings hour, changes in the data. This forms a late winter mid-week series of data, which is sifted by hour of the day to form 24 parallel series.

Three sets of data are used to tune and test the models (Table 1). The training set is used for training the linear and neural models. During training of the neural model the validation set is used to prevent overtraining (Section 5.1). Linear and neural model performance is compared over the novelty set.

Table 1 Segmentation of data set.

Set	Training	Validation	Novelty
Range	20th Jan '87 20th Mar '96	21st Mar '96 19th Mar '97	20th Mar '97 26th Mar '98
Size (Days)	300	30	30

Four exogenous inputs are available formed as the past 72 hours of weather relative to each load demand point, these are

- Temperature
- Humidity
- Wind speed
- Wind direction

As wind direction is a circular mapping it is transformed into 2 inputs using sine and cosine transformations.

2 A PRELIMINARY LINEAR MODEL

The preliminary linear model removes the linear AR information of the load data. As there is a high degree of correlation between the weather and the load this information must also be extracted from the weather variables by use of a pre-whitening filter.

2.1 MODEL STRUCTURE

Twenty-four Basic Structural Models (BSM) [1,8] are used to model the AR component of the parallel series. The BSM consists of modelling the load y_k at time k, as the sum of a trend component t_k, a seasonal component s_k with seasonal length N and a white noise error or residual r_k defined by [7]:

$$y_k = t_k + s_k + r_k \tag{1}$$

With

$$\sum_{i=1}^{N} s_k = 0 \tag{2}$$

And

$$\begin{bmatrix} t_k \\ \dot{t}_k \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t_{k-1} \\ \dot{t}_{k-1} \end{bmatrix} \tag{3}$$

The BSM requires that the variances of t_k, (σ^2_t) and s_k (σ^2_s) be tuned relative to the variance of r_k (σ^2_e). This may be achieved by use of prediction error decomposition [8] or by use of the IRWSMOOTH algorithm [9]. The models in this paper use the latter method to tune σ^2_t and σ^2_e. σ^2_s is then tuned independently using prediction error decomposition [8].

Note on boundary conditions

As the data is formed from appending data from Jan-March and Tuesday-Thursday there is a time jump of 9 months or 3 days between the boundaries. The trend component t_k is updated at these boundaries using a Kalman filter run over all the data.

2.2 MODEL RESIDUAL

The residual r_k for the 6pm series is shown in figure 2[1] together with its Sample Auto-Correlation Function (SACF) (Figure 3).

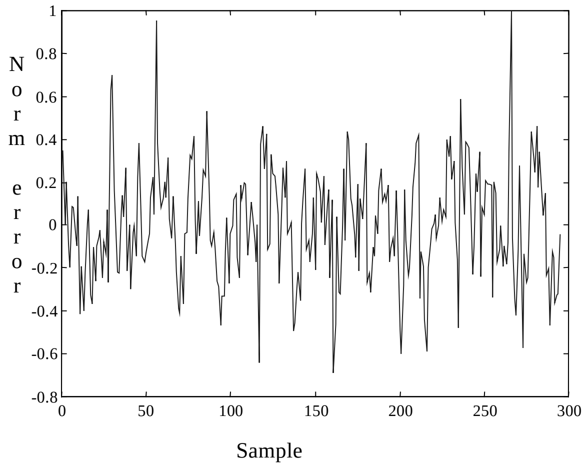

Figure 2 AR load residual (6pm series)

Examination of the residual shows that r_k is stationary in the sample mean. However, it is not stationary in the variance.

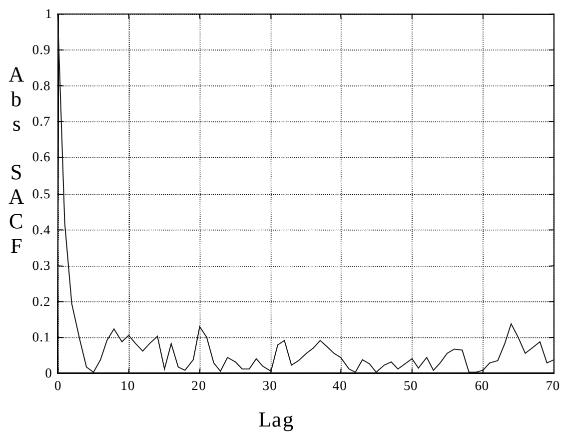

Figure 3 AR load residual SACF (6pm series)

The SACF of the residual shows no significant lags as the AR information has been removed.

2.3 EXOGENOUS VARIABLE PRE-WHITENING

[1] Data has been normalised for confidentiality reasons.

The weather variables are cross-correlated with s_k as they are also seasonal. Thus they require pre-whitening to remove the correlated AR component. The pre-whitening filter used is the BSM in (1) with the same parameters achieved by tuning with load data.

3 INPUT SELECTION

Input selection forms perhaps the most important step in model building [7]. Inclusion of non-causal variables leads to model performance degradation outside of the training set.

The selection of the most significant weather inputs is examined here.

3.1 PRE-PROCESSING TECHNIQUES

A mixture of two techniques where examined for input selection. The first technique involves ordering the input variables relative to the variance/amplitude or T-ratio [10] of each input w_i with coefficient a_i using a simple linear model of the form

$$r_k = a_1 w_1 + a_2 w_2 + ... + a_{72} w_{72} \quad (4)$$

Where r_k is the residual to be modelled. The linear coefficients are calculated by means of least squares.

The second technique employs PCA to transform w_i into a set of orthogonal variables such that each variable represents the coefficient along a basis vector in the original data set [11]. The transformed variables or components are organised in descending order of percentage of variance explained in the original data set, thus the first component contains the highest level of information [11]. The target data r_k is not used in the transformation and thus an assumption made when using PCA is that all the information in the input set is correlated to the target data. This is not always the case as in the trivial example where the target data is simply the second or higher order component.

The methods examined for variable selection employing both T-ratio and PCA are:

1. Identify most significant w_i according to T-ratio alone.
2. Use the T-ratio to identify 50 most significant variables. Transform using PCA and order according to variance explained.
3. Transform w_i using PCA and order according to variance explained.
4. Transform w_i using PCA and order using T-ratio.

3.2 RESULTS

When the exogenous or transformed variables have been ordered by significance (4) is evaluated using just the first component and then subsequent components until an optimal Mean Absolute Error (MAE) is found[2]. A bootstrap is used to increase the confidence of the resulting model performance.

Bootstrapping involves training N identical models using the entire data set with each model using non-overlapping test sets. In this case the test set size is 45 points or one eighth of the data points. N is eight. The training set consists of all the data less the test set. Figure 4 shows the performance achieved with method 3. The test set MAE's reach a minimum at 10 components. Table 2 summarises the performance over all hours of the day for each four methods. The optimum MAE achieved is given. Method 3 gives the best result.

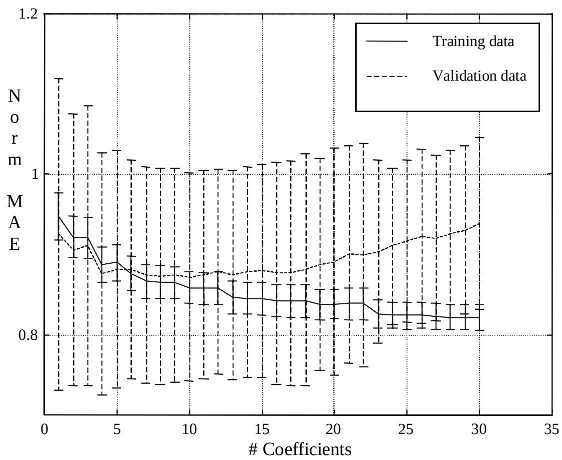

Figure 4 Bootstrapped MAE for method 3 (6pm series)

Table 2 Input selection MAE (normalised)[3]

Summary	Method 1	Method 2	Method 3	Method 4
Mean MAE Test	0.90	0.90	0.86	1.00
Mean STD Test	0.13	0.13	0.13	0.15
Mean MAE Trn	0.83	0.87	0.83	0.96
Mean STD Trn	0.03	0.03	0.02	0.06

The optimum number of components used in method 3 is found to vary from 7 to 15 depending on the hour of the day. The mode is 10 and due to the computational expense of calculating the topology of neural networks 10 components are used for all hours of the day.

[2] The MAE is used in place of the Mean Absolute Percentage Error (MAPE) as the trend component t_k of the data is increasing. It should be noted that the residual in not strictly stationary and this exaggerates the standard deviations of the bootstrapped MAE's.

[3] Data has been normalised for confidentiality reasons.

4 LINEAR MODELLING

The linear models are trained using the training set defined in Table 1. The input structure consists of the first 10 PCA components ordered using percentage of variance explained. The model is of the form

$$r_k = a_1 c_1 + a_2 c_2 + ... + a_{10} c_{10} \tag{5}$$

Where c_i is principal component i and a_i is the coefficient applied to that component. The model coefficients are tuned via least squares.

5 NEURAL NETWORK MODELLING

5.1 MODEL STRUCTURE

A multi-layer perceptron network using the back-propagation learning algorithm [7,11] is used. Each network consists of 2 hidden layers with tansig activation functions as the data is normalised between ± 1 and a linear activation function in the output layer.

Each model is trained until a minimum is found over the validation data set (Table 1) or for a maximum of ten thousand epochs. Cessation of training at the validation set minimum prevents over-training of the NN [11].

5.2 INPUT STRUCTURE

As in the case of the linear model (Section 4) PCA is used to transform the pre-whitened weather variables and the first 10 components are used. From the SACF of the residual (Figure 3) identification of significant lags in r_k is difficult to identify as the linearly AR element of the load has been removed. It was found that using r_{k-1} and r_{k-2} as additional inputs gave the best results.

5.3 TOPOLOGY DETERMINATION

The topology of a NN determines the degrees of freedom available to model the data [11]. If the NN is too simple then the network will generalise, while an over-complex NN will learn the noise in the data and performance over the validation set will degrade [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 NN's were trained for each topology with random initial weights to ensure reliable results.

Table 3 shows the achieved 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 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 minimum MAPE's (Table 3). Topology 4x4 is chosen using the second criteria.

24 NN's with a topology of 4x4 are trained for each parallel series. Ten NN's are again trained for each parallel series with random initial weights to ensure reliable results.

6 RESULTS

The MAPE as a function of the time of day (Figure 5) shows that the NN model achieves better results over the novelty set for most hours of the day.

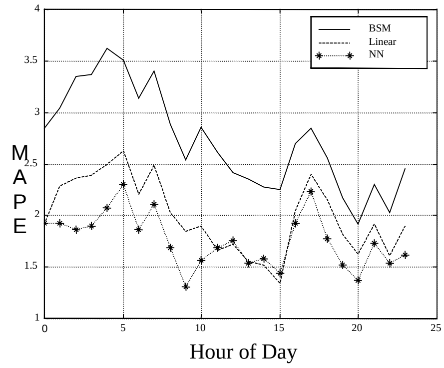

Figure 5 MAPE over the novelty set.

Table 4 shows the overall average MAPE achieved using the BSM, the linear model and the NN model.

Table 4 Average daily MAPE

Data set	Training	Validation	Novelty
BSM	2.60	2.13	2.73
Linear Model	1.91	1.68	1.99
Neural Network	1.65	1.67	1.76

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

	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

The NN model achieved the best results in all three data sets with an improvement of 0.2% over the linear model in the novelty data set. Both the linear and NN model show a significant improvement over the BSM.

7 DISCUSSION

The residual modelling approach allows the analysis of the linear AR and residual part of the data. The MAPE achieved utilising a NN model is superior to a linear model (Table 4). However the NN performance over the hours 11am-3pm is comparable to the linear model performance (Figure 5). The topology of the NN's used assumes that the optimum topology for the 6pm series (i.e. 4x4) also applies to the series for these hours. Input selection (Section 3) also utilises only one input structure, based on the mode of the optimum number of components. The calculation of 24 individual input structures and network topologies should improve the performance and is proposed for each parallel model.

Similar NN's using the structure in Section 5 but without the inclusion of lagged residual inputs exhibit a near identical performance to the linear models. This suggests that the superior performance of the NN presented is due to the inclusion of the lagged residual elements in the NN input data. Examination of the residual for evidence of a non-linear AR relationship is proposed using techniques such as [12][13].

8 CONCLUSION

The purpose of this paper has been to compare linear and neural parallel models and determine which is superior. For the models and Irish data used the neural network models were found to give superior results.

REFERENCES

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

[2] D.A. Fay, J.V. Ringwood, M.V. Kelly, M. Condon, "New research developments in half hourly forecasting methodologies and technologies applied to the Irish electricity market", (*Proceedings of the IIR conference on Accurate Energy Demand Forecasting*), *IIR Conferences*, June 1999, London, UK.

[3] 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.

[4] D.A. Fay, J.V. Ringwood, M. Condon, M.V. Kelly, "Linear parallel and multi-time series forecasting techniques for short term load forecasting.", to be published.

[5] P.K.Dash , H.P. Satpathy, A.C. Liew, S. Rahman, "A real-time short-term load forecasting system using functional link network", *IEEE Transactions on Power Systems*, Vol.12, pp 675-680, May 1997.

[6] K.Liu, S. Subbarayan, R.R. Shoults, M.T.Manry, C. Kwan, F.L. Lewis, J.Naccarino, "Comparison of very short term load forecasting techniques", *IEEE Transactions on Power Systems*, Vol.11, pp 877-882, May 1996.

[7] A.D. Papalexopoulos, S. Hao, T.M. Peng, "An implementation of a neural network based load forecasting model for the EMS", *IEEE Transactions on Power Systems*, Vol. 9, pp 1956-1962, November 1994.

[8] A.C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1994.

[9] C.N. Ng, P.C. Young , "Recursive estimation and forecasting of non-stationary time series", *Journal of Forecasting*, Vol. 9 , 1990, pp 1730-204.

[10] L.J. Kazmier, N.F. Pohl, *Basic Statistics for Business and Economics*, McGraw Hill, 1987.

[11] 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.

[12] T.S. Rao, M.M.Gabr, "A test for linearity of stationary time series", *Journal of Time Series Analysis*, Vol. 1, pp 145 – 158, 1980.

[13] J.K. Tugnait, "Testing for linearity of noisy stationary signals", *IEEE Transactions on Signal Processing*, Vol. 42, pp 2742-2748, October 1994.

AUTHOR'S ADDRESS

The first author can be contacted at

School of electronic engineering,
Dublin City University,
Glasnevin,
Dublin 9,
Republic of Ireland.

E-mail : fayd@eeng.dcu.ie