FORECASTING ELECTRICAL LOAD USING A MULTI-TIME-SCALE APPROACH

J.V. Ringwood[*] and F.T. Murray[**]

[*] School of Electronic Engineering
Dublin City University
Glasnevin, Dublin 9, Ireland

E-mail: ringwoodj@eeng.dcu.ie

[**] Xilinx Ltd.
Saggart
Co. Dublin, Ireland

E-mail: fiona.murray@xilinx.com

Keywords: Multi-time scale, load forecasting, state-space.

Abstract

This paper describes the application of a multi-time-scale technique to the modelling and forecasting of short-term electrical load. The multi-time-scale technique is based on adjusting the underlying short sampling period forecast time series with specific target points and possible aggregated demand. This allows not only improvement of the short sampling period forecast, but also focuses on weighting the accuracy of the forecast at certain critical points e.g. the overnight minimum and daily peaks. Various model types may be utilised at the upper level (forecating the aggregated consumption and target points at daily level), including intelligent models such as neural and fuzzy models, but the base model is currently restricted to a linear form. Results for the Irish national electrical grid demonstrate the effectiveness of the technique.

1 Introduction

In recent years, most of the work on short-term electrical load forecasting had concentrated on peak demand [1],[2]. This is reflected by the importance of accurately predicting the maximum demand over the day which determines the maximum generating capacity required. However, the complete daily load profile must also be forecast [3], in an attempt to provide inputs to the unit committment optimisation problem. Such forecasts are normally specified on an hourly or half-hourly time base. A danger associated with the development of an hourly (or half-hourly) *time series* model is that the model may produce a good forecast in terms of some overall indicator (e.g. Mean Square Error [MSE] over the 24-hour period), but have unacceptably large errors at the cardinal points e.g. midday peak, overnight minimum, evening peak.

This paper addresses the problem of producing complete 24-hour (and beyond) *time series* forecasts, while allowing separate forecasting models to concentrate exclusively on any required cardinal points (one model for each point). The time series forecast is then forced to fit the cardinal points, with the extra target of fitting to the aggregated load over the day (supplied by a further special model) being provided as a further option. In the current formulation, the hourly or half-hourly time series model must be a linear state-space model, including types such as AR, ARMA, ARIMA [4] and structural models [5].

One of the interesting features of the multi-time-scale technique is that a diverse range of inputs may be employed at the different time scales. For example, when a long term trend such as aggregated yearly demand is being imposed on aggregated weekly demand, inputs such as GDP and electricity price may be more appropriate at the annual level, while temperature inputs may be the dominant factor at the weekly level. In this way, various exogenous inputs, which are not available at the weekly level, can still be used to enhance the forecast. In the current application, both the (daily)target points and the hourly time series use the same (principally weather) input variables, so a univariate model is utilised at the hourly level, with the influence of the exogenous variables being taken into account in the target points.

2 The Multi-Time-Scale Technique

2.1 Problem Definition

Let the short time-scale model for the observations $y(k)$, $k = 1\ldots\ldots\ldots K$ be a general state-space model of the form:

$$\mathbf{x}(k) = \mathbf{F}\mathbf{x}(k-1) + \mathbf{G}\boldsymbol{\eta}(k-1) \qquad (1)$$

$$\mathrm{y}(\mathrm{k}) = \mathbf{H}\mathbf{x}(\mathrm{k}) + \boldsymbol{\varepsilon}(\mathrm{k}) \qquad (2)$$

where $\mathbf{x}(k) \in \mathrm{R}^n$ is the state vector, $\mathbf{F} \in R^{n\times n}$, $\mathbf{G} \in R^{n\times m}$ and $\mathbf{H}^T \in R^n$ are the system matrices which are assumed to be constant matrices, and $\boldsymbol{\eta}(k)$ and $\boldsymbol{\varepsilon}(k)$ are assumed to be zero mean independent and identically distributed normal random variables. Equation (1) is the state equation and (2) is the observation equation, together they make up the state space model for a system with *n* state variables, *m* system inputs and a single system output. An *l*-step-ahead forecast of the series is obtained through the following:

$$\hat{\mathbf{x}}(t+l/t) = \mathbf{F}^l \hat{\mathbf{x}}(t) \qquad (3)$$

$$\hat{y}(t+l/t) = \mathbf{H}\hat{\mathbf{x}}(t+l/t) \qquad (4)$$

where *t* represents the forecasting origin, *l* represents the forecasting lead time and $\hat{\mathbf{x}}(t+l/t)$ is the estimate of the state vector at time *t+l* given the state estimate $\hat{\mathbf{x}}(\ )$

$\hat{\mathbf{x}}(t)$ at forecasting origin *t* and back-solving for these freed states using the end-point and sum data. Define:

$$\boldsymbol{\Phi}(l) = \mathbf{F}^l \quad \in \quad R^{n\times n} \qquad (5)$$

The forecast from the short-time-scale model at the end-point *N* of the forecast horizon is given by:

$$\hat{y}(N) = \mathrm{H}\boldsymbol{\Phi}(N)\mathrm{x}(t) \qquad (6)$$

Let the value of the end-point predicted by the long-time-scale model be denoted $\hat{Y}_{ep}(L)$. Also, let the number of states which are fixed in the state vector at the forecasting origin $\hat{\mathbf{x}}(t)$ be *r*, and the number of states which are freed states be (*n-r*). The state vector is reconstructed as follows:

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}_1^T & \mathbf{x}_2^T\end{bmatrix}^T, \quad \mathbf{x}_1 \in R^{\mathrm{r}} \text{ and } \mathbf{x}_2 \in R^{(\mathrm{n-r})} \qquad (7)$$

and with $\mathbf{x_1}$ contains the fixed states and $\mathbf{x_2}$ contains the freed states which it is necessary to solve for using the long-time-scale end-point specification. The matrix $\boldsymbol{\Phi}$ is partitioned appropriately according to the above construction of **x**(*t*) (7) as follows:

$$\boldsymbol{\Phi} = \begin{bmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2\end{bmatrix}, \quad \boldsymbol{\Phi}_1 \in R^{n\times r} \text{ and } \boldsymbol{\Phi}_2 \in \mathrm{R}^{n\times(n-r)} \qquad (8)$$

then, using (6), (7) and (8),

$$\hat{Y}_{ep}(L) - \mathbf{H}\boldsymbol{\Phi}_1(N)\mathbf{x}_1(t) = \mathbf{H}\boldsymbol{\Phi}_2(N)\mathbf{x}_2(t) + e_1 \qquad (9)$$

where $e_1 = \hat{Y}_{ep}(L) - \hat{y}(N)$ represents the error on the end-point specification.

Similarly, if the forecast of the sum from the long-time-scale model is $\hat{Y}_s(L)$ then

$$\hat{Y}_s(L) - \sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_1(l)\mathbf{x}_1(t) = \sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_2(l)\mathbf{x}_2(t) + e_2 \qquad (10)$$

where $e_2 = \hat{Y}_s(L) - \sum_{l=1}^{N} \hat{y}(l)$ represents the error in the sum specification. Equations (9) and (10) represent the end-point and sum matching constraints imposed on the short-time-scale prediction.

2.2 Deviation From the Original Solution

When applying the end-point and sum information it is desirable that the prediction follows, to some degree, the original unaltered solution. Therefore, minimisation of the deviation from the unmodified prediction for *N-1* forecasts, i.e., all the forecasts minus the end-point, is sought. Let $\hat{y}(l)$ and $\hat{y}^*(l)$, *l*=1,2,.....*N*, be the unaltered and altered predictions respectively. Also, let **x**(*t*) be the original state vector at the forecasting origin and $\mathrm{x}^*(t)$ be the new altered state vector at the forecasting origin *t*, then it is required to minimise $\left|\hat{y}(l) - \hat{y}^*(l)\right|^2$ in

$$\mathbf{H}\boldsymbol{\Phi}(l)\mathbf{x}(t) = \mathbf{H}\boldsymbol{\Phi}(l)\mathbf{x}^*(t) + \left(\hat{y}(l) - \hat{y}^*(l)\right) \quad \text{for } l = 1,2,\ldots N-1 \qquad (11)$$

Partitioning $\boldsymbol{\Phi}$ and $\mathrm{x}^*(t)$ according to equations (7) and (8), equation (11) can be written as:

$$\mathbf{H}\boldsymbol{\Phi}(l)\mathbf{x}(t) - \mathbf{H}\boldsymbol{\Phi}_1(l)\mathbf{x}_1^*(t) = \mathbf{H}\boldsymbol{\Phi}_2(l)\mathbf{x}_2^*(t) + e_l \quad \text{for } l = 1,2........, N-1 \qquad (12)$$

where $e_l = \left(\hat{y}(l) - \hat{y}^*(l)\right)$ represents the error on the deviation from the original solution for the forecast at lead time *l*. Therefore, equation (12) represents the deviation from the original unadjusted solution constraint. The errors $e_l = \left(\hat{y}(l) - \hat{y}^*(l)\right)$, *l*=1,2,.....*N*, are referred to as the *deviation errors*.

2.3 Weighted Least Squares Solution

A weighted least squares formulation (Franklin *et al*, 1990) of the problem is sought which allows for selective adjustment of the original prediction of the short sampling period time series. The combination of equations (9), (10) and (12) gives:

$$\begin{bmatrix} \hat{Y}(L) - H\mathbf{\Phi}_1(N)x_1(t) \\ \hat{Y}_s(L) - \sum_{l=1}^{N} H\mathbf{\Phi}_1(l)x_1(t) \\ H\mathbf{\Phi}(1)x(t) - H\mathbf{\Phi}_1(1)x_1^*(t) \\ . \\ . \\ . \\ H\mathbf{\Phi}(N-1)x(t) - H\mathbf{\Phi}_1(N-1)x_1^*(t) \end{bmatrix} = \begin{bmatrix} [H\mathbf{\Phi}_2(N)] \\ \sum_{l=1}^{N} H\mathbf{\Phi}_2(l) \\ H\mathbf{\Phi}_2(1) \\ . \\ . \\ . \\ H\mathbf{\Phi}_2(N-1) \end{bmatrix} x_2(t) + E \tag{13}$$

where

$$\mathbf{E} = \begin{bmatrix} e_1 & e_2 & e_3 & \ldots\ldots & e_{N+1} \end{bmatrix}^T$$

Equation (13) may be rewritten as:

$$\mathbf{b} = \mathbf{A}\mathbf{x}_2 + \mathbf{E} \tag{14}$$

with the obvious identification of **b** and **A** from (13). A weighted least squares solution for the freed states $x_2(t)$ which minimises $E^T\mathbf{W}E$ is given by:

$$\mathbf{x}_2(t) = \left(\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{W}\ \mathbf{b} \tag{15}$$

If the forecast of an additional point from the long-time-scale model is denoted by $\hat{Y}_a(I)$ then the constraint at this point is given by:

$$\hat{Y}_a(I) - \mathbf{H}\mathbf{\Phi}_1(i)\mathbf{x}_1(t) = [\mathbf{H}\mathbf{\Phi}_2(i)]\mathbf{x}_2(t) + e_i \tag{16}$$

These additional constraints are straightforwardly added to the formulation in (13).

3 Application to Short-Term Forecasting

3.1 Long-term Models

The long term (daily) models comprise the aggregated daily consumption (sum) and the selected cardinal points as follows:

- Peak load of the day at 1800 hrs.
- Overnight min. At 0500hrs.
- Lunch-time peak at 1300 hrs.
- Afternoon valley at 1400 hrs.

The peak load of the day was chosen as the end-point and the others as additional target points.

Daily data recorded over 9 years is used to determine the sum and cardinal point models, giving a total of 245 data points. The daily data is seasonal, with a season length of 7, and basic structural models (BSM) with a dummy seasonal component [6] were used for each individual model. The parameters of such models are determined using a combination of the Kalman filter and maximum liklihood optimisation [5].

3.2 Short-term models

The hourly model is also a BSM with a dummy seasonal component [6], with the identification data used being taken from 17 days of hourly load, giving 408 data points.

The adequacy of both long and short-term models was checked using the Ljung-Box statistic [7].

3.3 Multi-Time-Scale Modelling

In the multi-time-scale formulation, states 1 (level of trend component) and states 3 to 25 were freed i.e. the $\mathbf{x}_2$ states in (15) and state 2, corresponding to the slope of the trend component, fixed. This gives sufficient freedom to adjust the level and the shape of the time series forecast according to the intermediate points obtained. The weighting matrix in (15) is given as:

$$W = diag\begin{bmatrix} w_{ep} & w_s & w_{addpt} & w_{dev\_1900\ hrs} & w_{dev\_2000\ hrs} & . & . & . & w_{dev\_1700\ hrs} \end{bmatrix} \tag{17}$$

4 Results

The accuracy of the forecasts from the long-term models is given in the table below:

Table 1: Accuracy (%) of long-time-scale information for 24-hour-ahead prediction

Long-Time-Scale Forecast	**% Error**
Σ(1900 hrs Tues. to 1800 hrs Wed.	1.56
1800 hrs Wed	2.90
1400 hrs Wed	1.11
1300 hrs Wed	3.67
0500 hrs Wed	2.13

In order to give a basis for comparison, the unadjusted short-term time series forecast obtained the following accuracy:

Table 2: Unadjusted 24-hour-ahead forecast

Forecast Horizon	MAE	MSE x 10^{-3}	MAPE
1900 hrs Tues to 1800 hrs Wed	45.54	2.8074	2.19

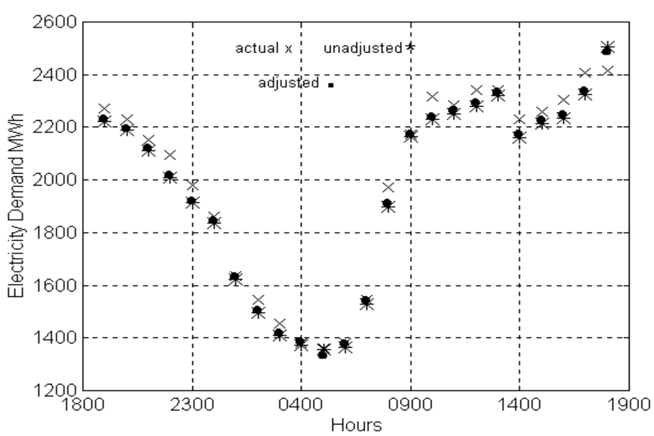
where MAPE represents the Mean Absolute Percentage Error. A number of combinations of cardinal points were tried, with the best combination being sum (aggredated daily consumption), end-point (peak daily load) and 0500 hrs load (overnight minimum). The results for this case are:

Table 3: Adjusted 24-hour-ahead forecast

Type of adjustment	MAE	MSE x 10^{-3}	MAPE
Adjusted using actual daily values	19.69	0.5663	1.05
Adjusted using predicted daily values	37.21	1.9880	1.80

Fig.1 indicates (graphically) the effect of the adjustment:

Figure 1: Actual and predicted load values



5 Conclusions

The results show a significant improvement in the hourly forecast values when adjusted by the specific cardinal point models. Dramatically improved results using actual cardinal point inputs demonstrate that considerable effort should be expended in modelling the cardinal points. This technique has also been successfully applied to forecasting a number of 24-hour periods ahead and also the combination of weekly/annual forecasts [8].

References

1. Azzam-ul-Azzar and McDonald, J.R. A Specification of Neural Network Applications in the Load Forecasting Problem, *IEEE Trans. on Control Systems Tech.*, Vol.2, 1994, pp 135-141.
2. Haido, T. and Muto, S. Regression-based Peak Load Forecasting Using a Transformation Technique, *IEEE Trans. on Power Systems*, Vol.9, 1994, pp 437-455.
3. Park, D.C., El-Sharkawi, M.A., Marks II, R.J., Atlas, L.E. and Damborg, M.J. Electric Load Forecasting Using an Artificial Neural Network, *IEEE Trans. on Power Systems*, Vol.6, No.2, May 1991, pp 442-449.
4. Ljung, L. *System Identification - Theory for the User*, Prentice-Hall, 1987.
5. Harvey, A.C. A Unified View of Statistical Forecasting Procedures, *Journal of Forecasting*, Vol. 3, 1984, pp 245-275.
6. Harvey, A.C. *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1989.
7. Bowermann, B. and O'Connell, R.T. *Time Series Forecasting: Unified Concepts and Computer Implementations*, Duxbury Press, 1987.
8. Murray, F.T. *Time Series Forecasting Methodologies for Electricity Supply Systems*, PhD Thesis, Dublin City University, Ireland, 1997.