Integration of multi-time-scale models in time series forecasting

FIONA T. MURRAY†, JOHN V. RINGWOOD‡ and PAUL C. AUSTIN§

A solution to the problem of producing long-range forecasts on a short sampling interval is proposed. It involves the incorporation of information from a long sampling interval series, which could come from an independent source, into forecasts produced by a state-space model based on a short sampling interval. The solution is motivated by the desire to incorporate yearly electricity consumption information into weekly electricity consumption forecasts. The weekly electricity consumption forecasts are produced by a state-space structural time series model. It is shown that the forecasts produced by the forecasting model based on weekly data can be improved by the incorporation of longer-time-scale information, particularly when the forecast horizon is increased from 1 year to 3 years. A further example is used to demonstrate the approach, where yearly UK primary fuel consumption information is incorporated into quarterly fuel consumption forecasts.

1. Introduction

This paper deals with the problem of producing forecasts which are based on a short sampling interval over a relatively long period of time. The forecasting procedure applied to the problem should have the capability of producing long-range forecasts on a fine time interval, while allowing individual models to concentrate on natural time scales, or cycles, within the data. In addition, our proposed technique allows causal variables on the long sampling period to exert an influence over the short sampling period solution. The motivation for addressing this problem stems from forecasting problems in the electricity supply industry, including forecasting weekly consumption up to 10 years ahead and the forecasting of half-hourly consumption up to a year in advance.

One approach to the problem could be to use multirate analysis where a discrete-time multirate model (Berg *et al.* 1988) could conceptually contain both short and long sampling periods, for example weekly and yearly sampling rates. However, with this approach, all models used must fit within the same mathematical framework and longer-time-interval models are generally referred to the shortest sampling period, resulting in a structurally unwieldy model.

Our solution is to consider the problem in terms of a combination of *forecasts* (not models) produced by models which are based on different time scales, that is short and long time scales, which we refer to as multi-time-scale model integration. The intention is to provide the capability of producing forecasts on a relatively fine sampling period, while retaining the fidelity of longer trends and cycles in the data. Consider the example given in figure 1.

Model 1 is a short-time-scale model which is capable of making weekly predictions over 1 year, that is it is capable of predicting $y(t+l)$ for lead times $l = 1, 2, 3, \ldots, N$ weeks ahead where t is the forecasting origin. Models 2 and 3 are relatively longer-time-scale models which can make yearly predictions over a period of up to 10 years, that is they are capable of predicting $Y_{ep}(T+L)$ and $Y_s(T+L)$ for lead times $L = 1, 2, \ldots$ years ahead respectively, where T is the forecasting origin. This paper proposes to integrate the forecasts from these three models so that weekly predictions can be produced over a period of up to 10 years. The integration can be achieved by using the forecasts produced by models 2 and 3 to impose the longer-time-scale information on a long-term prediction produced by model 1. The forecasts obtained from model 2 will be termed

† Xilinx Ireland Limited. Saggart, Co. Dublin, Ireland. e-mail: fiona.murray@xilinx.com.

‡ School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland. e-mail: ringwoodj@eeng.dcu.ie.

§ Department of Electrical and Electronic Engineering, University of Auckland, Private Bag 92019, New Zealand.
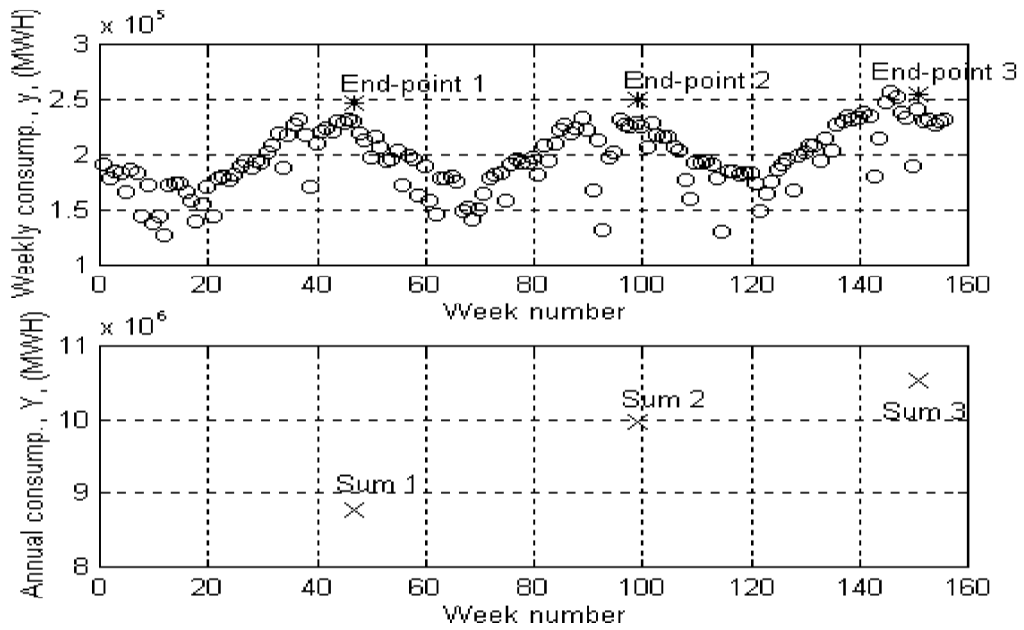e-mail: p.austin@auckland.ac.nz.

Figure 1. Multi-time-scale model integration for electricity consumption. Model 1 produces weekly predictions of the consumption (o). Model 2 forecasts for only a single week (end points) of each year (*). Model 3 produces yearly predictions of the annual aggregate (sum data) consumption (x).

'end-point' information and the forecasts obtained from model 3 will be termed 'sum' information. The longer-term models therefore provide intermediate 'targets' for the shorter interval forecast to follow. Such an approach of target achievement has been considered independently by Cholette (1982).

A suitable state-space model is required for the short-sampling-period time series and many forecasting studies are devoted to the use of such models. Examples include autoregressive moving-average (ARMA) models (Janacek and Swift 1993), autoregressive-integrated moving-average (ARIMA) models (Ansley and Kohn 1985) and structural time series models (Ng and Young 1990). The choice of the state-space model depends on the properties of the time series in question (e.g. stationary or non-stationary) and the various components present in the time series, such as trend, seasonal or special signal components.

One of the features of the proposed technique is the ability to use any model type, or data from an independent source, to produce the long-sampling-period series, since the focus is on the forecast that it produces and not the model itself. The versatility of our approach is demonstrated by the use of a neural network model as the long-sampling-period series model. Such model types have been shown to be effective in forecasting annual electricity consumption, although much care is required in order to apply them effectively. In other applications, different (e.g. state-space) models may be more appropriate. Further flexibility is provided by the

ability to utilize either causal or univariate models at the different sampling periods. In the paper, however, a univariate model is assumed at the shortest sampling period. The numerical examples presented demonstrate the use of both causal and univariate models at the long sampling period.

A possible solution to the current problem is also provided by smoothing analysis in conjunction with an appropriate model structure (Jones 1980, Harvey 1984, Kohn and Ansley 1989, Ng and Young 1994). However, while such an approach allows the fitting of the forecast to particular points of the series, there is no mechanism for the inclusion of 'sum' or cumulative data.

There are a number of other studies which deal with the subject of combining forecasts produced on different time scales. In contrast with the current technique, some of these aggregate forecasts obtained with a higher sampling frequency and then combine them with the lower-sampling-frequency forecast (Corrado and Greene 1988, Howrey et al. 1991), the final 'pooled' forecast being produced at the lower frequency level. One exception is the paper by Fuhrer and Haltmaier (1988), which extends the approach put forward by Corrado and Greene (1988), where the pooled forecast may be obtained at the disaggregated higher frequency level.

Harvey (1989) also examines the subject of forecasting using different timing intervals. However, this work formulates state-space forecasting models at timing intervals which are shorter than the observation interval.

2. Multi-time-scale model integration

The multi-time-scale model integration procedure involves modifying the prediction obtained from the short-time-scale model so that it matches, in a least-squares sense, the end-point and sum information obtained from longer-time-scale model predictions. This is achieved by adjusting the solution of the state-space short-time-scale model through the relaxation of some of the states in the state vector at the forecasting origin and then back solving for a new set of states using end-point and sum specifications. The modified solution is now found using this new state vector where a weighted least-squares solution is sought to create selective alteration of the raw unmodified solution.

2.1. *The multi-time-scale models*

Depending on the nature of the time series in question there are a number of possible state-space models available for the short sampling period. For a stationary time series an ARMA model could be employed. The state-space form, identification and estimation of such a model have been dealt with by Janacek and Swift (1993). For some non-stationary time series, detrending or differencing can be applied to achieve stationarity. ARIMA and seasonal autoregressive-integrated moving-average (SARIMA) models use differencing operators to deal with non-stationary non-seasonal time series and non-stationary seasonal time series respectively. However, it has been shown that differencing or detrending analysis may not be suitable for a time series with a time trend (Kang 1990). A structural time series model which uses separate models for trend, seasonal and signal components present in the time series can handle non-stationarity where there is an underlying trend component which is changing randomly over time. The state-space form, identification and estimation of ARIMA and SARIMA models have been dealt with by Ansley and Kohn (1985) and Kohn and Ansley (1986) and structural time series models by Young (1988, 1994), Harvey (1989), Young *et al.* (1989, 1991) and Ng and Young (1990). Each of these studies use state-space Kalman filtering (or variants of it) and fixed interval or fixed-point smoothing. The Kalman filter provides the minimum mean squared estimate of the state vector at any given point in time based only on the observations available at that point in time and it can be used to forecast future values of the time series. Smoothing allows the estimation of the state vector at any point in time given all the available data and it can be used to interpolate missing observations of the time series.

2.2. *Adjustment of the short-time-scale model prediction*

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

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

$$y(k) = \mathbf{H}x(k) + \varepsilon(k), \qquad (1\,b)$$

where $x(k) \in R^n$ is the state vector, $\mathbf{F} \in R^{n \times n}$, $\mathbf{G} \in R^{n \times m}$ and $\mathbf{H}^{\mathrm{T}} \in R^n$ are the system matrices which are assumed to be constant matrices, and $\eta(k)$ and $\varepsilon(k)$ are assumed to be zero mean independent and identically distributed normal random variables. Equation $(1\,a)$ is the state equation and equation $(1\,b)$ 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:

$$x(t+l/t) = \mathbf{F}^l x(t), \qquad (2\,a)$$

$$y(t+l/t) = \mathbf{H}x(t+l/t), \qquad (2\,b)$$

where t represents the forecasting origin, l represents the forecasting lead time and $x(t+l/t)$ is the estimate of the state vector at time $t+l$ given the state estimate $x(t)$ at time t. Let the forecast be required for lead times $l = 1,2,3,\ldots,N$. It is required to modify the prediction so that the forecast obtained at the end point N matches the end-point value obtained from the prediction of the longer-time-scale model. It is also required that the sum of the forecasts, aggregated over the forecast horizon, matches the sum value obtained from the longer-time-scale model. The adjustment is achieved by relaxing some of the states in the state vector at forecasting origin t and back solving for these freed states using the end-point and sum data. Define

$$\mathbf{\Phi}(l) = \mathbf{F}^l \in R^{n \times n}. \qquad (3)$$

The forecast at the end point N is given by

$$y(N) = \mathbf{H}\mathbf{\Phi}(N)x(t). \qquad (4)$$

Let the value of the end point obtained from the longer-time-scale model be $Y_{\mathrm{ep}}(L)$. Also let the number of fixed states in the state vector $x(t)$ be r; therefore $n-r$ states are freed. Split the state vector into

$$x = [x_1 \quad x_2]^{\mathrm{T}}, \qquad (5)$$

where

$$x_1 \in R^r \quad \text{and} \quad x_2 \in R^{n-r}.$$

Therefore x_1 are the original fixed states and x_2 are the freed states which it is necessary to solve for using the end-point and sum specifications. Partition $\mathbf{\Phi}$ as

$$\mathbf{\Phi} = [\mathbf{\Phi}_1 \mathbf{\Phi}_2], \qquad (6)$$

where

$$\boldsymbol{\Phi}_1 \in R^{n \times r} \quad \text{and} \quad \boldsymbol{\Phi}_2 \in R^{n \times (n-r)};$$

then

$$\mathbf{Y}_{\text{ep}}(L) - \mathbf{H}\boldsymbol{\Phi}_1(N)\boldsymbol{x}_1(t) = [\mathbf{H}\boldsymbol{\Phi}_2(N)]\boldsymbol{x}_2(t) + \boldsymbol{e}_1, \quad (7)$$

where $\boldsymbol{e}_1 = \mathbf{Y}_{\text{ep}}(L) - \boldsymbol{y}(N)$ represents the error in the end-point specification. If the sum value obtained from the longer-time-scale model is $\mathbf{Y}_s(L)$, then

$$\mathbf{Y}_s(L) - \sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_1(l)\boldsymbol{x}_1(t) = \left(\sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_2(l) \right)\boldsymbol{x}_2(t) + \boldsymbol{e}_2,$$
$$(8)$$

where $\boldsymbol{e}_2 = \mathbf{Y}_s(L) - \sum_{l=1}^{N} \boldsymbol{y}(l)$ represents the error in the sum specification.

When applying the end-point and sum information it is desirable that the prediction follows, to some degree, the original unaltered solution. Therefore minimization of the deviation from the unmodified prediction for $N-1$ forecasts, that is all the forecasts minus the end point, is required. If $\boldsymbol{y}^*(l)$ and $\boldsymbol{y}(l)$ are the altered and unaltered predictions respectively, and $\boldsymbol{x}^*(t)$ and $\boldsymbol{x}(t)$ are the altered and unaltered state vectors at the forecast origin t respectively, then it is required to minimize $|\boldsymbol{y}(l) - \boldsymbol{y}^*(l)|^2$ in

$$\mathbf{H}\boldsymbol{\Phi}(l)\boldsymbol{x}(t) = \mathbf{H}\boldsymbol{\Phi}(l)\boldsymbol{x}^*(t) + [\boldsymbol{y}(l) - \boldsymbol{y}^*(l)]$$
$$\text{for } l = 1, 2, \ldots, N-1. \quad (9)$$

Partitioning $\boldsymbol{\Phi}$ and $\boldsymbol{x}^*(t)$ according to (5) and (6), (9) can be written as

$$\mathbf{H}\boldsymbol{\Phi}(l)\boldsymbol{x}(t) - \mathbf{H}\boldsymbol{\Phi}_1(l)\boldsymbol{x}_1^*(t) = \mathbf{H}\boldsymbol{\Phi}_2(l)\boldsymbol{x}_2^*(t) + [\boldsymbol{y}(l) - \boldsymbol{y}^*(l)]$$
$$\text{for } l = 1, 2, \ldots, N-1. \quad (10)$$

A weighted least-squares formulation of the problem is sought which allows for selective adjustment of the original prediction. Combining (7), (8) and (10)

$$\begin{bmatrix} \mathbf{Y}(L) - \mathbf{H}\boldsymbol{\Phi}_1(N)\boldsymbol{x}_1(t) \\ \mathbf{Y}_s(L) - \sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_1(l)\boldsymbol{x}_1(t) \\ \mathbf{H}\boldsymbol{\Phi}(1)\boldsymbol{x}(t) - \mathbf{H}\boldsymbol{\Phi}_1(1)\boldsymbol{x}_1^*(t) \\ \vdots \\ \mathbf{H}\boldsymbol{\Phi}(N-1)\boldsymbol{x}(t) - \mathbf{H}\boldsymbol{\Phi}_1(N-1)\boldsymbol{x}_1^*(t) \end{bmatrix}$$

$$= \begin{bmatrix} [\mathbf{H}\boldsymbol{\Phi}_2(N)] \\ \sum_{l=1}^{N} \mathbf{H}\boldsymbol{\Phi}_2(l) \\ \mathbf{H}\boldsymbol{\Phi}_2(1) \\ \vdots \\ \mathbf{H}\boldsymbol{\Phi}_2(N-1) \end{bmatrix} \boldsymbol{x}_2(t) + \mathbf{E}, (11)$$

where

$$\mathbf{E} = [\boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \boldsymbol{e}_3 \quad \cdots \quad \boldsymbol{e}_{N+1}]^{\mathrm{T}}$$

$\boldsymbol{e}_3, \ldots, \boldsymbol{e}_{N+1}$ represent the deviation of the modified prediction from the original prediction, where $\boldsymbol{e}_{l+2} = \boldsymbol{y}(l) - \boldsymbol{y}^*(l)$ for $l = 1, \ldots, N-1$. Equation (11) is of the form

$$\mathbf{z} = \mathbf{A}\boldsymbol{x}_2(t) + E. \quad (12)$$

A weighted least-squares solution of the freed states $\boldsymbol{x}_2(t)$ which minimizes $\boldsymbol{E}^{\mathrm{T}} \cdot \mathbf{W}\boldsymbol{E}$ is determined using

$$\boldsymbol{x}_2(t) = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{z}. \quad (13)$$

The relative importance of the error minimization can be specified in the weighting matrix $\mathbf{W}$ which is chosen to be of the form

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & w_2 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & w_i & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & w_{N+1} \end{bmatrix}, \quad (14)$$

where the weight w_i corresponds to the error term $\boldsymbol{e}_i^2$ in $\boldsymbol{E}^{\mathrm{T}} \cdot \mathbf{W}\boldsymbol{E}$. The errors may be weighted differently, allowing selective control of the modified solution. The modified initial state vector which consits of r original states and $n - r$ modified states can now be used to obtain a modified short-time-scale prediction from (2).

The multi-time-scale integration approach has been described for imposing a single end point and sum value on a short-time-interval prediction. It is straightforward to extend the approach to deal with more than one end point and sum value and thus the integration approach can be used in an attempt to produce short-time-scale predictions over a long-range forecast horizon. This may be achieved either by augmenting the least-squares problem in (11) with extra end points and sum targets or by reapplying the adjustment after each cycle.

2.3. *Solution parameters*

The objective of the method is to provide enough freedom in the state vector at the forecasting origin,

while maintaining reasonable fidelity with the original short-sampling period forecast. In particular, the modified forecast should not deviate significantly from the original forecast in the area close to the forecasting origin but should be allowed more freedom later, in order to achieve the end point. To this end, the weighting on the deviation from the original forecast is heavier near the forecasting origin and lighter towards the end; in practice, good results have been obtained from a linear weighting profile. Following experimental analysis, larger values are assigned to the weights on the end-point error e_1 and the sum error e_2, reflecting the importance of minimizing these errors. For cases involving more than one end-point and sum specification, end-point errors are heavily penalized, since the end point of each prediction cycle provides the starting point for the next cycle.

A further design parameter is available in the number of states to free $n - r$ at the forecasting origin. Considerations here include the model type and the desire to maintain the values of the most recent states intact. For autoregressive (AR) or ARMA models, the state vector is made up of present and past values of the model output, and it is therefore desirable to leave a proportion of the states which represent the most recent values of the model output unchanged since they provide the starting point and higher-order derivatives for the solution. In practice, fixing 25% of states for AR and 20% for ARMA models has been found to give good results. AR models may be overparametrized, thus allowing a greater number of the states that represent the most recent values of the model output to remain fixed while allowing approximately the same number of states to be freed. In the case of structural state-space models, some of the states may represent components such as trend. In such cases, it is beneficial to maintain the trend state at its original value and to free other states.

To date, selection of r and the weighting matrix $\mathbf{W}$ has been largely based on trial and error. In spite of this, reasonable initial results have been obtained, and current efforts focus on analytical and numerical techniques for optimization of these parameters. Preliminary results suggest that r is dependent on the model structure and order, while $\mathbf{W}$ depends on the characteristics of the particular time series.

3. Practical examples

Two examples are used to demonstrate the effectiveness of the multi-time-scale technique. Example 1 concentrates on the weekly electricity demand, which exhibits a slowly rising trend and seasonality of regular amplitude, while example 2 focuses on fuel series data (Janacek and Swift 1993), which is also seasonal, but with a more irregular long-term trend. Both actual end-point and sum information and the corresponding predicted values are used to reshape the forecast, giving an indication of the ideal and realistic improvement obtained.

3.1. *Example 1: weekly electricity demand*

In this example, the short-time-scale series is made up of weekly electricity demand (in megawatt hours) from July 1982 to July 1990, as shown in figure 2. The data are divided into three identification sets, where these data sets and their corresponding forecast horizons are described in table 1.

A state-space structural time series model consisting of a trend and seasonal component is used for the short-time-interval series. The smooth changes in the trend are represented by an integrated random walk (IRW) model (Young 1988, 1994, Young *et al.* 1989, 1991, Ng and Young 1990). The seasonal component with sustained amplitude will be represented by a differenced periodic
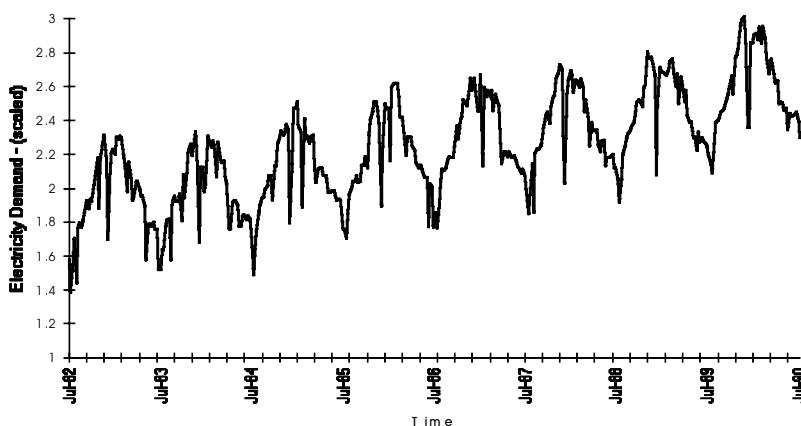

Figure 2. Weekly electricity demand.

Table 1. Identification data sets and corresponding forecast horizons.

Identification data set	Forecast horizon
Week 30 of July 1982–Week 29 of July 1987	Week 30 of July 1987–Week 29 of July 1988
Week 30 of July 1982–Week 29 of July 1988	Week 30 of July 1988–Week 29 of July 1989
Week 30 of July 1982–Week 29 of July 1989	Week 30 of July 1989–Week 30 of July 1990
Week 30 of July 1982–Week 29 of July 1987	Week 30 of July 1987–Week 30 of July 1990

random walk (DPRW) model (Harvey 1984, Kitagawa and Gersch 1984, Young 1988, Ng and Young 1990). A description and a detailed discussion of the estimation of such a model have been given by Harvey (1989) and Harvey and Peters (1990) where prediction error decomposition is employed and also by Young (1988, 1994), Young *et al.* (1989, 1991), Ng and Young (1990) where sequential spectral decomposition (SSD) is carried out. The estimation of the short-time-scale model in this example is carried out using SSD. The noise variance ratio (NVR) selected for the IRW model for each of the identification data sets is 0.000 01 (using the NVR $= 1650(F_{50})^4$ relationship (Young 1994) and 0.1 for the DPRW model, estimated using prediction error decomposition. The unaltered forecast results for this model are shown later in table 4 and also later in figures 3–6. The criteria, namely the mean squared emr (MSE) and the mean absolute error (MAE) used in assessing the quality of the forecast are defined as follows:

$$\text{MSE} = \frac{1}{n_l} \sum_i \left(\text{actual}_i - \text{predicted}_i\right)^2$$
$$\text{MAE} = \frac{1}{n_l} \sum_i |\text{actual}_i - \text{predicted}_i| \qquad (15)$$

where n_l is the number of samples in the forecasting horizon.

An attempt is now made to improve the forecast using annual data. To obtain an indication of the maximum achievable improvement, actual end-point and sum values are initially employed. The end point, for example, in the 1987–1988 prediction is taken as the weekly electricity demand in week 29 of 1988 and the sum value is taken as the sum of the weekly electricity demand from week 30 of 1987 to week 29 of 1988. For the structural state-space models used in this example the first state in the state vector is fixed and the remaining states are freed. In the weighting matrix **W**, w_1 and w_2 are assigned large values (1×10^5) relative to the other weights $w_3, \ldots, w_{N+1}$ (1×10^0) for the one-year-ahead forecasts. For the three-year-ahead forecast, the end-point weighting is increased to 1×10^6. The three-year-ahead forecast is carried out through the augmentation of the least-squares problem given by (11). The results for this case are shown later in table 4, with a mean improvement of approximately 9% for

the one-year-ahead forecasts and approximately 21% for the three-year-ahead forecast.

Causal artificial neural networks (ANNs) are now used to forecast the yearly end-point and sum values. Neural networks has been shown to be useful in forecasting daily (Azzam-ul-Azar and McDonald 1994) and yearly (Murray *et al.* 1996) electricity demand. Only brief details of the neural network modelling methodology are given here; the interested reader is referred to the books by Azoff (1994) and Vemuri and Rogers (1994). The neural networks considered are multilayer perceptrons (MLPs) consisting of an input layer, two hidden layers and an output layer. The weekly average temperature is used as an exogenous input for the end-point model, while yearly economic data such as the gross domestic product, the average unit price of electricity, the average industrial wage and the number of customers; the period 1960–1990 is used for the sum model. The output $Y_{\text{ep}}(k)$ of the end-point network is given by:

$$Y_{\text{ep}}(k) = h(\text{at}(k), Y_{\text{ep}}(k-1)), \qquad (16)$$

where h is the nonlinear function defined by the trained end-point network, $Y_{\text{ep}}(k)$ is the electricity demand at the end-point week in year k and at(k) is the average temperature over the end-point week in year k. Similarly the output of the network which predicts the elements of the sum series is given by

$$Y_{\text{s}}(k) = q(\text{gdp}(k), \text{pr}(k), \text{aiw}(k), \text{noc}(k), Y_{\text{s}}(k-1)), \qquad (17)$$

where q is the nonlinear function defined by the trained sum network, $Y_{\text{s}}(k)$ is the sum of the weekly electricity demand for year k. gdp(k) (£) is the gross domestic product for year k. pr(k) (p kW^{-1} l^{-1}) is the average unit price of electricity in year k, aiw(k) (£ week^{-1} is the average industrial wage in year k and noc(k) is the number of customers in year k.

The networks are trained using the error back-propagation algorithm with momentum and an adaptive learning rate (Hertz *et al.* 1991). Using numerical optimization techniques (Murray *et al.* 1996), the structure selected for the end-point series is a 3–2–6–1 network (three inputs in the input layer, two neurons in the first hidden layer and six neurons in the second hidden layer, with one linear neuron in the output layer). Similarly,

Table 2. Models used to predict inputs to the ANNs: seasonal auto-regressive integrated (SARI), autoregressive integrated (ARI)

Input	Model
at	SARI(4, 1, 0)(1, 52, 0)
gdp	ANN 2–1–3–1
pr	ARI(6, 2, 0)
aiw	ARI(1, 2, 0)
noc	ARI(3, 2, 0)

the structure selected for the sum series is a 5–1–3–1 network. The activation function used in all the non-linear neurons is a log-sigmoid function, since electricity demand always has a positive value. Although a small number of annual data points are available, the use of causal models and training using multiple epochs helps to ensure an effective forecasting model for the end-point and sum values, which are reasonably straightforward to predict. Inputs to the causal models are provided by the AR models given in table 2.

Table 3 shows the accuracy of the end-point and sum predictions obtained from the neural network forecasting models. The improvement in the weekly forecast using predicted end-point and sum information is given in table 4, with a mean improvement of approximately 7% for the one-year-ahead forecasts and approximately 21% for the three-year-ahead forecast. Figures 3–5 show graphs of the one-year-ahead predictions, while figure 6 shows graphs of the three-year-ahead predictions. 95% confidence intervals are given for the adjusted forecast

using predicted end-point and sum information. For clarity, plots of actual and adjusted profiles only are given.

3.2. *Example 2: UK primary fuel consumption*

In this example the multi-time-scale integration technique is applied to the quarterly UK primary fuel consumption series 1965–1985 (Janacek and Swift 1993), shown in figure 7.

The time series exhibits strong seasonality, following the UK seasonal variation in temperature, and a long term trend with a small slope. The identification sets and forecast horizons are given in table 5.

Following Janacek and Swift (1993, chapter 6) a structural state-space model with a trend and quarterly trigonometric seasonal component is fitted to the fuel series. They used maximum-likelihood estimation techniques to estimate the hyperparameters $(\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2)$ of the structural state-space model, where the estimates obtained are $\sigma_\varepsilon^2 = 2746.40$, $\sigma_\eta^2 = 204.43$, $\sigma_\zeta^2 = 0$ and $\sigma_\omega^2 = 3.83$. The unadjusted forecasting results for this model are given later in table 7 and also later in figures 8–10. The results show a higher degree of error in the 1984 forecast compared with the 1985 forecast, because in the year 1984, for quarters 3 and 4, the time series behaves differently from previous years or in 1985. This difference in behaviour of the time series has been dealt with in more detail by Murray (1996).

The multi-time-scale integration technique is now applied to the quarterly forecast using annual fuel con-

Table 3. Results for the predicted end-point and sum electricity demand values using ANNs.

Prediction	Year	End-point error (%)	Sum error (%)
Week 30 of July 1987–Week 29 of July 1988	1987–1988	0.2	0.9
Week 30 of July 1988–Week 29 of July 1989	1988–1989	5.3	0.2
Week 30 of July 1989–Week 30 of July 1990	1989–1990	0.2	0.1
Week 30 of July 1987–Week 30 of July 1990	1987–1988	0.2	0.9
	1988–1989	4.1	0.2
	1989–1990	4.3	4.5

Table 4. Results for the electricity demand example.

| Year	Unaltered forecast		Altered forecast: using *actual* data for end-point and sum values		Altered forecast: using *predicted* data for end-point and sum values	
	MAE	MSE	MAE	MSE	MAE	MSE
1987–1988	0.8108×10^4	1.4455×10^8	0.7381×10^4	1.3197×10^8	0.7372×10^4	1.3198×10^8
1988–1989	0.7652×10^4	1.0122×10^8	0.6723×10^4	0.9158×10^8	0.7083×10^4	0.9485×10^8
1989–1990	0.5942×10^4	0.7740×10^8	0.5574×10^4	0.7143×10^8	0.5575×10^4	0.7156×10^8
1987–1990	1.0191×10^4	1.9439×10^8	0.8043×10^4	1.5080×10^8	0.7998×10^4	1.4948×10^8
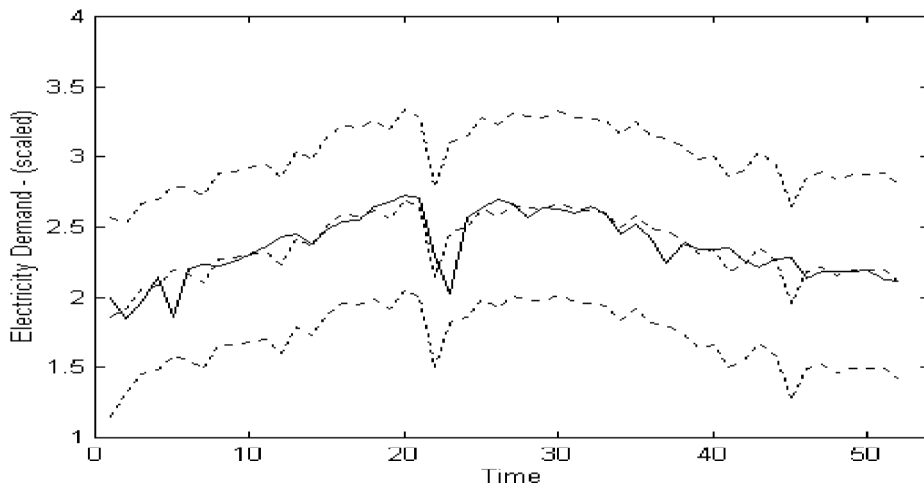
Figure 3.   Actual versus predicted profiles for the 1987–1988 forecast: (——), actual data; (- - - -), data adjusted using predicted information and confidence limits.
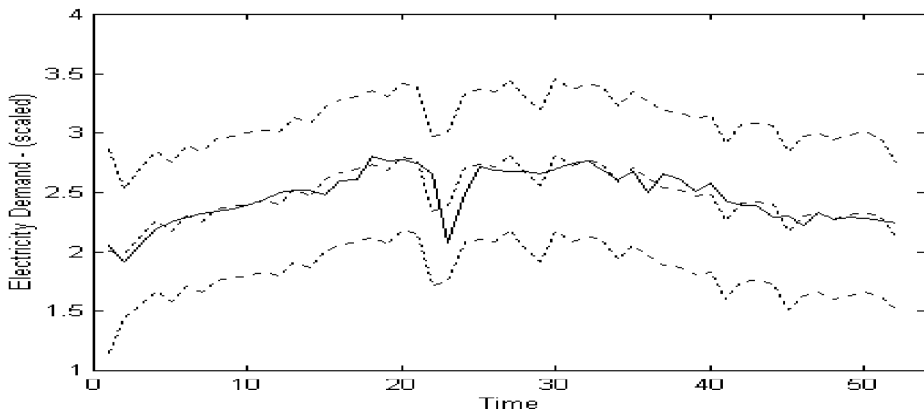

Figure 4.   Actual versus predicted profiles for the 1988–1989 forecast: (——), actual data; (- - - -), data adjusted using predicted information and confidence limits.
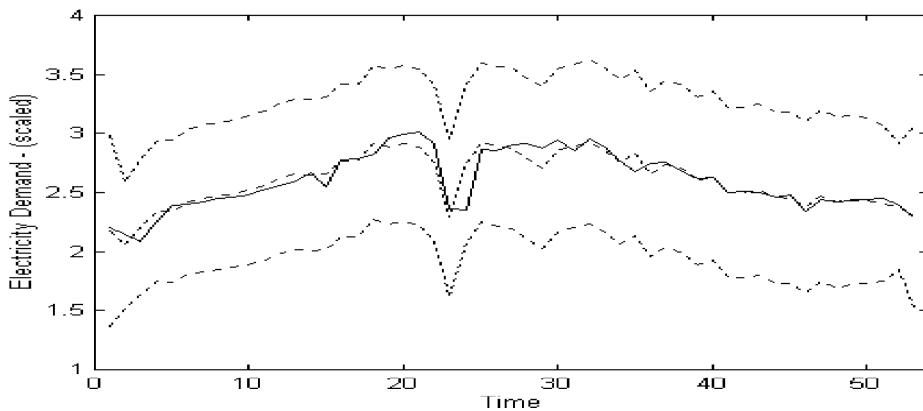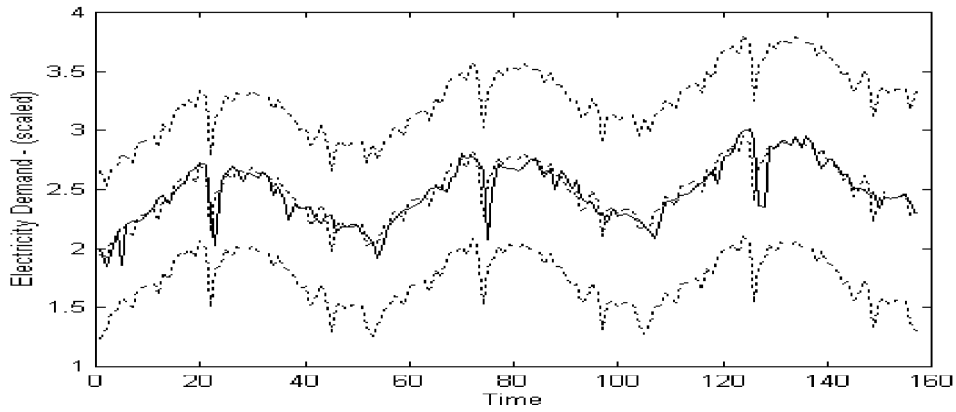

Figure 5.   Actual versus predicted profiles for the 1989–1990 forecast: (——), actual data; (- - - -), data adjusted using predicted information and confidence limits.

Figure 6. Actual versus predicted profiles for the 1987−1990 forecast: (——), actual data; (- - - -), data adjusted using predicted information and confidence limits.
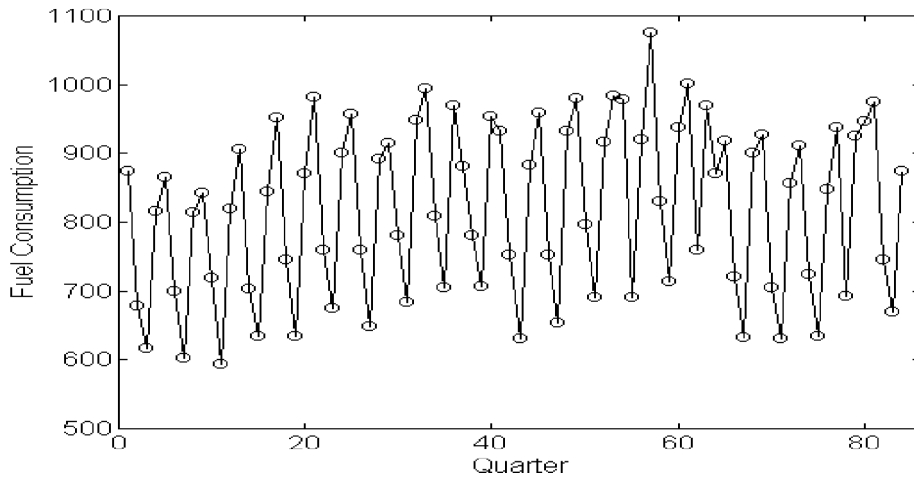

Figure 7. UK primary fuel consumption for 1965−1985.

sumption data. The end-point data are given as the fuel consumption for last quarter in each year, and the sum data as the aggregate of the quarterly fuel consumption over each year. The use of actual data provides a measure of the maximum achievable improvement. For the quarterly state-space model the most accurate results are obtained when the first state, corresponding to the trend component, is fixed and the remaining states are freed. In the weighting matrix $\mathbf{W}, [w_1, w_2, w_3, w_4, w_5] = [10, 1, 1, 0.67, 0.34]$ for the one-year-ahead predictions and $[w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}] = [10, 10, 1, 1, 1, 0.67, 0.34, 1, 0.67, 0.34]$ for the two-year-ahead prediction (Murray 1996). The results using actual end-point and sum data are given later in table 7, with a mean improvement of approximately 68% for the one-year-ahead forecasts and approximately 13% for the two-year-ahead forecast.

Univariate AR neural network models are now used to forecast the yearly end-point and sum values (Murray 1996). As in example 1, MLP networks with an input layer, two hidden layers and an output layer, using a log-sigmoid activation function in the nonlinear neurons and trained using the error back-propagation algorithm with momentum and an adaptive learning rate, are utilized. The output $Y_{ep}(k)$ of the end-point network is given by

Table 5. Identification data sets and corresponding forecast horizons.

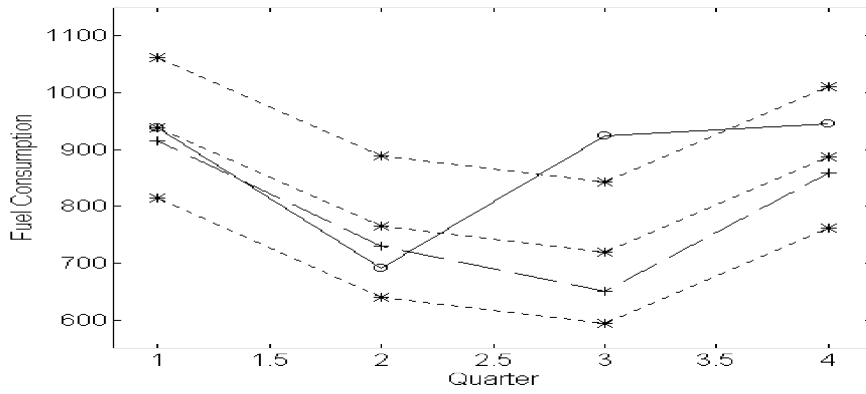| Identification data set	Forecast horizon
Quarter 1 of 1965−quarter 4 of 1983	1984
Quarter 1 of 1965−quarter 4 of 1984	1985
Quarter 1 of 1965−quarter 4 of 1983	1984−1985

Figure 8. Actual versus predicted profiles for the 1984 forecast: (——O——), actual data; (− − + − −), adjusted data; (- - -*- - -), data adjusted using predicted information and confidence limits.
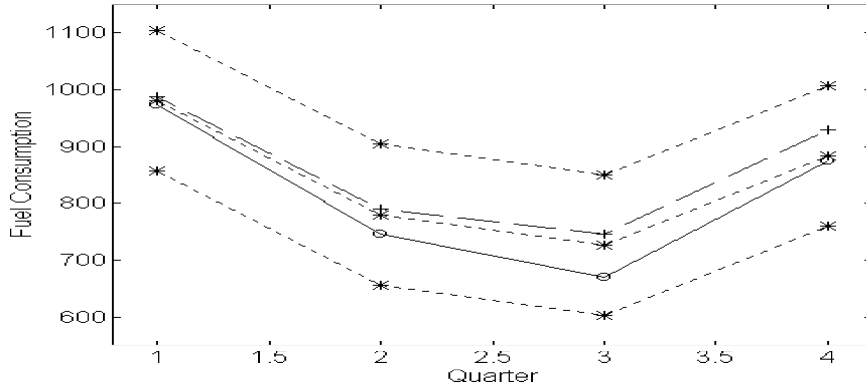

Figure 9. Actual versus predicted profiles for the 1985 forecast: (——O——) actual data; (− − + − −), adjusted data; (- - - * - - -), data adjusted using predicted information and confidence limits.
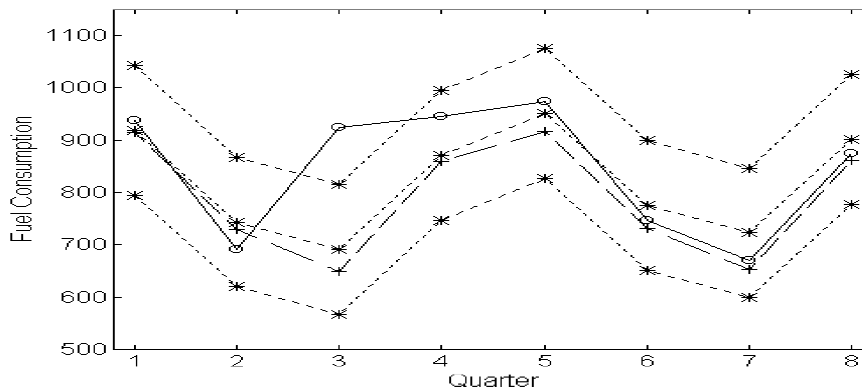

Figure 10. Actual versus predicted profiles for the 1984−1985 forecast: (——O——), actual data; (− − + − −), adjusted data; (- - - * - - -), data adjusted using predicted information and confidence limits.

$$Y_{\mathrm{ep}}(k) = f\left(Y_{\mathrm{ep}}(k-1), Y_{\mathrm{ep}}(k-2), Y_{\mathrm{ep}}(k-3),\right.$$
$$\left. Y_{\mathrm{ep}}(k-4)\right) \quad (18)$$

where f is the nonlinear function defined by the trained end-point network and $Y_{\mathrm{ep}}(k)$ is the quarterly fuel consumption for the last quarter in year k. Similarly the output of the network which predicts the elements of the sum series is given by

$$Y_{\mathrm{s}}(k) = g\left(Y_{\mathrm{s}}(k-1), Y_{\mathrm{s}}(k-2), Y_{\mathrm{s}}(k-3), Y_{\mathrm{s}}(k-4)\right),$$
$$(19)$$

where g is the nonlinear function defined by the trained sum network $Y_{\mathrm{s}}(k)$ is the sum of the quarterly fuel consumption for year k.

Using numerical optimization techniques, (Murray *et al.* 1996), the structure selected for the end-point series and sum series is a 4–3–5–1 network. Table 6 shows the accuracy of the end-point and sum predictions obtained from the neural network forecasting models. The improvement in the weekly forecast using predicted end-point and sum information is given in table 7, with a mean improvement of approximately 40% for the one-year-ahead forecasts and approximately 13% for the two-year-ahead forecast.

Figures 8 and 9 show graphs of the one-year-ahead predictions, while figure 10 shows graphs of the two-year-ahead predictions. As in example 1, 95% confidence intervals are given for the adjusted forecast using predicted end-point and sum information. The original unadjusted solution is also included in the graphs.

4. Conclusions

This paper develops a technique which allows the integration of forecasts produced at different sampling periods. Specifically, the method provides for utilization of long-sampling-period forecast information to force a short-sampling-period forecast in order to maintain the long-term characteristics of the time series with reasonable fidelity. The main use is where long-range forecasts are required on a short sampling interval, as in the electricity supply industry, for example. Based on the experience to date, the proposed technique has shown merit (e.g. for the two practical examples considered in the paper, the improvement in forecast accuracy is of the order of 10–20%) although a rigorous multi-time-scale modelling procedure is not yet available. Current research is concentrated on optimum procedures for the determination of the weighting matrix **W** and the number of free initial states. The use of the relatively unconventional neural network models for a long sampling interval demonstrates the diversity of model types which may be used at this level, although models with a state-space structure must be used for a short sampling interval. For the examples considered, the technique compares favourably with state-space smoothing, which does not allow the incorporation of cumulative sum information. It is anticipated that the method may be straightforwardly extended to include more than two levels of integration, when state-space models will be required for all but the longest sampling period.

References

ANSLEY, C. F., and KOHN, R., 1985, A structured state space approach to computing the likelihood of an ARIMA process and its derivatives. *Journal of Statistical Computing Simulation,* **21**, 135–169.

AZOFF, M. E., 1994, *Neural Network Time Series Forecasting of Financial Markets* (New York: Wiley).

AZZAM-UL-AZAR, and MCDONALD, J. R., 1994, A specification of neural network applications in the load forecasting problem. *IEEE Transactions on Control Systems Technology,* **2**, 135–141.

BERG, M. C., AMIT, N., and POWELL, D. J., 1988, Multirate digital control system design. *IEEE Transactions on Automatic Control,* **33**, 1139–1150.

CHOLETTE, P., 1982, Prior information and ARIMA forecasting. *Journal of Forecasting,* **1**, 375–383.

Table 6. **Results for the predicted end-point and sum fuel values using ANNs.**

Prediction	End-point error (%)	Sum error (%)
1984	6.5	4.8
1985	1.1	3.1

Table 7. **Results for the fuel consumption example.**

Year	Unaltered forecast		Altered forecast: using *actual* data for end-point and sum values		Altered forecast: using *predicted* data for end-point and sum values	
	MAE	MSE	MAE	MSE	MAE	MSE
1984	105.85	21 322	68.85	8651	84.63	12 846
1985	47.29	2762	9.78	126	26.64	1128
1984–1985	65.54	11 148	60.41	9852	64.06	8527

CORRADO, C., and GREENE, M., 1988, Reducing uncertainty in short-term projections: linkage of monthly and quarterly models. *Journal of Forecasting, 7*, 77−102.

FURHER, J., and HALTMAIER, J., 1988, Minimum variance pooling of forecasts at different levels of aggregation. *Journal of Forecasting, 7*, 63−73.

HARVEY, A. C., 1984, A unified view of statistical forecasting procedures. *Journal of Forecasting, 3*, 245-275; 1989, *Forecasting Structural Time Series Models and the Kalman Filter* (Cambridge University Press).

HARVEY, A. C., and PIERSE, R. G., 1984, Estimating missing observations in economic time series. *Journal of American Statistical Association, 79*, 125−131.

HARVEY, A. C., and PETERS, S., 1990, Estimation procedures for structural time series models. *Journal of Forecasting, 9*, 89−108.

HERTZ, J., KROGH, A., and PALMER, R. G., 1991, *Introduction to the Theory of Neural Computations* (Reading, MA: Addison-Wesley).

HOWREY, P., HYMANS, S., and DONIHUE, M., 1991, Merging monthly and quarterly forecasts: experience with MQME. *Journal of Forecasting, 10*, 225−268.

JANACEK, S., and SWIFT, L., 1993, *Time Series Forecasting, Simulation and Applications.* (Chichester, West Sussex: Ellis Horwood).

JONES, R. H., 1980, Maximum likelihood fitting of ARMA models to time series with missing data. *Technometrics, 22*, 389−395.

KANG, H., 1990, A composite model for deterministic and stochastic trends. *International Journal of Forecasting, 6*, 175−186.

KITAGAWA, G., and GERCH, W., 1984, A smoothness priors—state space modelling of time series with trend and seasonality. *Journal of American Statistical Association, 79*, 378−389.

KOHN, R., and ANSLEY, C. F., 1986, Estimation, prediction and interpolation for ARIMA models with missing data. *Journal of American Statistical Association, 81*, 751−761; 1989, Filtering and smoothing algorithms for state space models. *Computers and Mathematics with Applications, 18*, 515−528.

MURRAY, F., 1996, Forecasting UK primary fuel consumption using multi-time-scale techniques. Report EE/FTM/961, School of Electronic Engineering, Dublin City University.

MURRAY, F., RINGWOOD, J. V., and O'BRIEN, K., 1996, Yearly electricity sales forecasting using artificial neural networks. *Neural Computing and Applications* (submitted).

NG, C. N., and YOUNG, P. C., 1990, Recursive estimation and forecasting of non-stationary time series. *Journal of Forecasting, 9*, 173−204.

VEMURI, V. R., and ROGERS, R. D., 1994, *Artificial Neural Networks: Forecasting Time Series* (New York: IEEE).

YOUNG, P. C., 1984, *Recursive Estimation and Time-Series Analysis* (Berlin: Springer-Verlag); 1988, Recursive extrapolation, interpolation and smoothing of nonstationary time series. *Proceedings of the IFAC Symposium on Identification and System Parameter Estimation*, Beijing, PR China, 35–46; 1994, Time-variable parameter and trend estimation in non-stationary economic time series. *Journal of Forecasting, 13*, 179−210.

YOUNG, P. C., NG, C. N., and ARMITAGE, P., 1989, A systems approach to recursive economic forecasting and seasonal adjustment. *International Journal on Computers and Mathematics with Applications, 18*, 481−501.

YOUNG, P. C., NG, C. N., LANE, K., and PARKER, D., 1991, Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data. *Journal of Forecasting, 10*, 57−89.