Comparative Models for Electrical Load Forecasting
Edited by D. W. Bunn and E. D. Farmer

CHAPTER 12

Adaptive Forecasting of Daily Water Demand

M. J. H. Sterling and A. Bargiela

1 INTRODUCTION

The planning and operation of water supply systems has become a complex task which inevitably relies on a forecast of the likely consumer demand. The availability of a long-term forecast over several years enables a planned enhancement of the existing network with sufficient lead time to facilitate major construction work. In the short term the network resources can be regarded as fixed but the method of operation can be varied to minimize costs while satisfying all network constraints. The solution of the resultant optimization problem normally necessitates an on-line computer control system which provides for a comprehensive network monitoring scheme together with appropriate control actuations. Data is consequently available on the actual water consumption at various take-off points on every measurement scan which can then be assembled to provide an aggregated historical profile of the demand over the whole network or within selected areas. An examination of this data will normally reveal a cyclic pattern with similarities between successive days, weeks, and months which can form the basis of an on-line short-term forecasting technique. Such techniques have been investigated for prediction of consumer demand in other public utilities including the electricity supply industry,[1–6] but rather less attention has been paid to demand prediction for the water industry.

Essentially the available data on which to base a short-term forecast are a log of past consumptions with corresponding network configurations together with a meteorological record and weather forecast. Methods utilizing some of this information have been developed,[7,8] but normally require comprehensive on-line data collection of parameters other than those provided by conventional monitoring and control installations. Consequently methods which rely primarily on a log of past consumption only are likely to be more amenable for on-line demand prediction.[2]

This chapter considers the methodology of one such class of demand modelling and forecasting techniques, namely that of time series analysis based on a Box–Jenkins[9] form of autoregressive moving average demand model. The chapter considers the development of a suitable model from a nonstationary regular time series representation of consumption data and shows how such models may be evaluated, improved, and used for prediction. A new matrix formulation of the problem is developed and a novel application of the Fletcher–Powell Variable Metric Algorithm is shown to result in a unity rank updating formula which gives a very efficient, self-adaptive algorithm for on-line demand prediction.

2 DATA TRANSFORMATION

The water demand over a particular period can be characterized by a series of discrete values $[x_t, t=1,2,\ldots,n]$. The objective of demand prediction is consequently to establish the values of the series x_t for $t=n+1, n+2, \ldots$. The series x_t will in general be nonstationary since its mean value will not be constant over any period of time. Consequently the stochastic model used to represent x_t must either allow for nonstationarity or the data must be transformed to generate a new time series which is stationary.

The transformation of the demand data x_t should not only produce a stationary series but should also be devised such that it produces either a constant error variance or a normal error distribution. Transformed data differencing may also be required to ensure stationarity, and the period can be estimated from the sample autocorrelation function (SAF). Strong correlation between samples at some interval apart would indicate that differencing should be applied with this period. The data transformation may be expressed as

$$z_t = f(x_t) \tag{12.1}$$

where f is the particular transformation for the demand data.

The differencing operation is then represented by

$$w_t = (1-B)^d(1-B^s)^D z_t$$

where d, D, and s are integers and B is the backward difference operator such that

$$(1-B)z_t = z_t - z_{t-1}$$

and s is the seasonal periodicity of the time series z_t. Now defining

$$\nabla^d = (1-B)^d \quad \text{and} \quad \nabla s^D = (1-B^s)^D$$

then the stationary time series w_t is given by

$$w_t = \nabla^d \, \nabla s^D z_t. \tag{12.2}$$

It will rarely be necessary to use values d and D greater than three to achieve stationarity.

3 SELECTION OF THE MODEL STRUCTURE

The prediction problem has now been reduced to the determination of a class of models which will adequately represent stationary rather than nonstationary time series. One such class of models is composed of autoregressive (AR) and moving average (MA) components given by

$$\phi(B)\Phi(B^s)w_t = \theta(B)\Theta(B^s)a_t \tag{12.3}$$

where ϕ, Φ, θ, and Θ are polynomials in B such that

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{12.4}$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \ldots - \Phi_P B^{Ps} \tag{12.5}$$

are the autoregressive components and

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{12.6}$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \tag{12.7}$$

are the moving average components. If p, P_q, and Q are correctly determined then a_t will be a white noise sequence distributed as $N(0, \sigma_a^2)$ where σ_a^2 is the variance.

To ensure that equation (12.3) represents a stationary time series the coefficients of (12.4) and (12.5) must all lie outside the unit circle, hence

$$-1 < \phi_i < 1 \qquad i = 1, 2, \ldots, p$$

$$-1 < \Phi_j < 1 \qquad j = 1, 2, \ldots, P.$$

Furthermore, to ensure invertibility the coefficients of equations (12.6) and (12.7) must also all lie outside the unit circle, hence

$$-1 < \theta_k < 1 \qquad k = 1, 2, \ldots, q$$

$$-1 < \Theta_i < 1 \qquad l = 1, 2, \ldots, Q.$$

The autoregressive moving average model class defined by equation (12.3) is extensive and it is consequently necessary to establish which parameters are significant for the data to be modelled. The sample autocorrelation function can be used as an indication of the subset of models likely to be relevant, and is defined for a time series w_t as

$$r_k(w_t) = c_k(w_t)/c_0(w_t) \tag{12.8}$$

where

$$c_k(w_t) = \frac{1}{n}\sum_{t=1}^{n-k}(w_t - \bar{w})(w_{t+k} - \bar{w}) \tag{12.9}$$

and

$$\bar{w} = \frac{1}{n} \sum_{t=1}^{n} w_t. \tag{12.10}$$

The value of r_k can thus be calculated from the data w_t for $k = 0, 1, \ldots, n-1$ and will indicate the relationship between samples k periods apart. Values of $r_k(w_t)$ which are close to zero imply that samples k periods apart are not significantly related, whereas a SAF approaching unity indicates a strong interdependence of samples at that frequency. The threshold for parameter significance is normnally defined in terms of the standard error (SE) given by

$$\mathrm{SE}(r_k) \simeq \frac{1}{n}\left(1 + 2\sum_{j=1}^{q} r_j^2\right) \qquad k > q \tag{12.11}$$

where q is the value of k beyond which the SAF is small, such that if $r_k(w) > 2\mathrm{SE}(r_k)$ for multiples of k, then a differencing parameter with period k should be included in the model. Further aids in the parameter identification process include the partial autocorrelation function and fast Fourier transforms.

4 MODEL PARAMETER FITTING

Having established the form of the autoregressive moving average model required for the data, the coefficients associated with each backshift operator must be determined. The objective function normally associated with this fitting procedure is a minimization of the sum of squared errors for each sample point. Thus the coefficients ϕ, Φ, θ, and Θ which minimize

$$S = \sum_t a_t^2(\phi, \Phi, \theta, \Theta) \tag{12.12}$$

complete the time series model.

In order to produce a computationally efficient algorithm a special matrix formulation of the problem has been developed. Writing equation (12.3) at each time step $t = 1, \ldots, N$ the model can be represented as

$$FD\mathbf{w} = RT\mathbf{a} \tag{12.13}$$

where

$$D = \begin{bmatrix} 1 & & & & \\ \phi_1 & 1 & & & \\ \phi_2 & \phi_1 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & \phi_2 & \phi_1 & 1 \end{bmatrix} \qquad F = \begin{matrix} \\ s\left\{\begin{matrix} \\ \\ \\ \end{matrix}\right. \\ \\ \\ \\ \\ \end{matrix}\begin{bmatrix} 1 & & & & \\ \vdots & \ddots & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \Phi_1 & \ldots & 1 & & \\ & \ddots & & \ddots & \\ & & \ddots & & \ddots \\ & & & \Phi_1 & \ldots \quad 1 \end{bmatrix}$$

$$T=\begin{bmatrix} 1 & & & & & \\ \theta_1 & 1 & & & & \\ \theta_2 & \theta_1 & 1 & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & \theta_2 & \theta_1 & 1 \end{bmatrix} \qquad R=\begin{array}{l} \left.\begin{array}{c} \\ s \\ \\ \end{array}\right\{ \\ \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ \cdot & \cdot & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \Theta_1 & \dots & & 1 & & \\ & \cdot & & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot\ \cdot & & \cdot \\ & & & \Theta_1 & \dots & 1 \end{bmatrix}$$

$$\mathbf{w}^{\mathrm{T}}=[w_1, w_2, \dots, w_N]$$

$$\mathbf{a}^{\mathrm{T}}=[a_1, a_2, \dots, a_N].$$

The function S to be minimized is now

$$S=\mathbf{w}^{\mathrm{T}}F^{\mathrm{T}}D^{\mathrm{T}}(R^{-1})^{\mathrm{T}}(T^{-1})^{\mathrm{T}}T^{-1}R^{-1}DF\mathbf{w}. \tag{12.14}$$

Using matrix E_i defined as

$$E_i=\left[\begin{array}{c:c} \multicolumn{2}{c}{0_{iN}} \\ \hdashline I_{N-1} & 0_{N-i,i} \end{array}\right] \qquad 1\leqslant i\leqslant N-1$$

the necessary conditions for this optimization problem can be written in a concise form:

$$-2\mathbf{a}^{\mathrm{T}}E_iR^{-1}T^{-1}F\mathbf{w}=0 \qquad 1\leqslant i\leqslant p \tag{12.15}$$

$$-2\mathbf{a}^{\mathrm{T}}E_{si}R^{-1}T^{-1}D\mathbf{w}=0 \qquad 1\leqslant i\leqslant P \tag{12.16}$$

$$-2\mathbf{a}^{\mathrm{T}}E_iT^{-1}\mathbf{a}=0 \qquad 1\leqslant i\leqslant q \tag{12.17}$$

$$-2\mathbf{a}^{\mathrm{T}}E_{si}R^{-1}\mathbf{a}=0 \qquad 1\leqslant i\leqslant Q. \tag{12.18}$$

These equations can be solved with respect to model parameters in several different ways, for example by using any appropriate hill-climbing method. However, in order to take full advantage of the problem structure the Newton-Raphson iterative procedure, modified to overcome numerical problems, is shown below to result in a highly efficient solution procedure.

In a standard Newton–Raphson process the corrections of the parameter estimates are calculated according to the formula

$$\Delta\Psi=H^{-1}h(\Psi) \tag{12.19}$$

where

$\Psi^{\mathrm{T}}=[\phi_1, \dots, \phi_p, \Phi_1, \dots, \Phi_P, \theta_1, \dots, \theta_q, \Theta_1, \dots, \Theta_Q]$

H^{-1} is the inverse of the Hessian matrix

$h(\Psi)$ is the gradient of the function S

the biggest computational effort being associated with the calculations of the inverse of the Hessian matrix. Additionally, if the initial estimates of parameters Ψ are far from their optimal values, equation (12.19) may prove to be numerically unstable because of an approximate singularity of the computed Hessian matrix.

These difficulties can be remedied by calculation of the corrections of the parameter estimates using an approximation of the inverse of the Hessian according to the Fletcher–Powell method. Taking Ψ_i as the current parameter estimates, the algorithm for updating model parameters may be expressed as:

(i) $\Psi_{i+1} = \Psi_i + \lambda_i v_i$ where $v_i = Q_i g_i$, and λ_i is chosen to minimize $S(\Psi_i + \lambda_i v_i)$

(ii) $u_i = \lambda_i v_i$ and $y_i = g_{i+1} - g_i$ where g_{i+1} is the new gradient vector (evaluated at Ψ_{i+1})

(iii) $$Q_{i+1} = Q_i + \frac{u_i u_i^{\mathrm{T}}}{u_i^{\mathrm{T}} y_i} - \frac{Q_i y_i y_i^{\mathrm{T}} Q_i^{\mathrm{T}}}{y_i^{\mathrm{T}} Q_i y_i}.$$

The above recurrence relationships provide a continuous revision of the model parameters with minimal computational effort. The parameters are updated in a single-step operation at each successive time interval by re-evaluating the gradient so as to incorporate the latest measurements. The resulting model will be the best, in a least-squares sense, that can be fitted to the data w_t up to the current time N. If now the sample autocorrelation function of the residual error between the model and actual data at each point given by

$$a_t = \theta^{-1}(B)\Theta^{-1}(B)\phi(B)\Phi(B) \qquad (12.20)$$

is computed then $r_k(a_t)$ can be used as an indication of the success of the modelling.

5 DEMAND PREDICTION

Having established a model of the time series w_t derived from the demand data over the period $t = 1, \ldots, N$ in the form of equation (12.13). this model may now be used for prediction by expanding it forward in time $t = N+1, \ldots, N+k$ with the assumption that the white noise error function a_t is zero for $t > n$. Thus

$$FD\mathbf{w}' = RT\mathbf{a}' \qquad (12.21)$$

where

$$\mathbf{w}'^{\mathrm{T}} = [\mathbf{w}^{\mathrm{T}} : w_{N+1}, \ldots, w_{N+k}]$$

$$\mathbf{a}'^{\mathrm{T}} = [\mathbf{a}^{\mathrm{T}} : 0, \ldots, 0]$$

and all the matrices are now $(N+k) \times (N+k)$

Introducing new variables

$$\mathbf{a}'' = T\mathbf{a}' \tag{12.22}$$

$$\mathbf{a}^* = R\mathbf{a}'' \tag{12.23}$$

it is easy to see that computations of the values $a_t^*, t = N+1, \ldots, N+k$, requires only a_t'', $t = (N+1-sQ), \ldots, N+k$, which in turn requires only a_t', $t = (N+1-sQ-q), \ldots, N+k$. Similarly by introducing

$$\mathbf{w}'' = D\mathbf{w}' \tag{12.24}$$

$$\mathbf{w}^* = F\mathbf{w}'' \tag{12.25}$$

it can be shown that to solve (12.24) for w_t', $t = N+1, \ldots, N+k$, requires w_t'', $t = N+1, \ldots, N+k$. These in turn can be found from equation (12.25), given w_t'', $t = (N+1-sP), \ldots, N$, which are calculated from (12.25) using known values w_t', $t = (N+1-sP-p), \ldots, N$.

The probability limits for the forecast can be determined by re-writing equations (12.2) and (12.3) as

$$z_t = (1 + \Psi_1 B + \Psi_2 B^2 + \ldots) a_t. \tag{12.26}$$

The variance of the forecast errors $[z_{t+1} - z_t(l)]$, where $z_t(l)$ represents predicted transformed data, is given by

$$V(z_{t+l} \mid a_t, a_{t-1}, \ldots) = \left(1 + \sum_{j=1}^{l-1} \Psi_j^2\right)\sigma_a^2 \tag{12.27}$$

where

$$\sigma_a^2 = \frac{1}{n}\sum_{t}^{n} a_t^2.$$

The 95% probability limits on the forecast are consequently given by

$$z_t(1) \pm 2\left(1 + \sum_{j=1}^{l-1} \Psi_j^2\right)\sigma_a^2. \tag{12.28}$$

6 COMPUTATIONAL RESULTS

A software package has been developed in the Department of Engineering at the University of Durham to implement the above ARMA modelling and prediction technique together with on-line parameter tracking. The software includes colour graphical display of logged and predicted demand, together with an error analysis enabling an interactive selection of the model structure before the prediction is run. The method has been successfully applied both to power and water demand data giving prediction errors of a few percent over an extended time period.

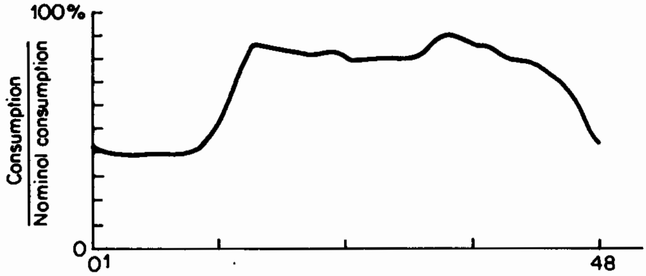

FIGURE 12.1 Daily consumption pattern

The performance of the above technique has been investigated for half-hourly water demand data shown in Figure 12.1 and an appropriate model structure found to be

$$(1+\phi_1 B)(1+\Phi_1 B^{48})w_t = (1+\theta_1 B)(1+\Theta_1 B^{48})a_t \qquad (12.29)$$

which gives an almost exact fit to the data.

The autocorrelation function of the residuals between the actual data values and the values calculated from the model is given in Figure 12.2. Although expansion of the model (12.29) of second-order nonseasonal components improves the autocorrelation function, it has been found that it has only a marginal effect on prediction accuracy. Figure 12.3 shows that the prediction error over short times is of order 1.0% with an increase to approximately 5.0% for longer term forecasts. This significant worsening of the longer term prediction error is caused by the absense of a model component reflecting weekly periodicity which becomes more important as the prediction time expands. However, an improvement in longer term

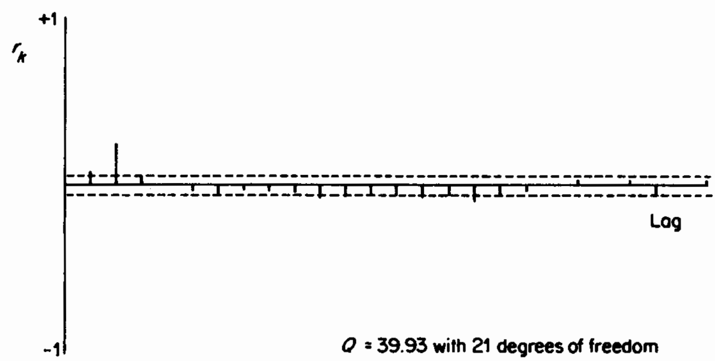

FIGURE 12.2 Autocorrelation function of residuals

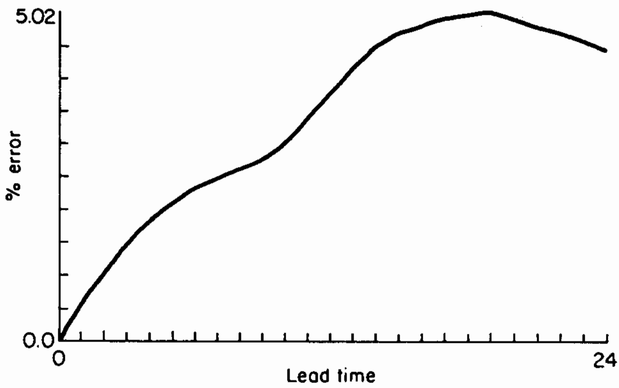

FIGURE 12.3 Mean absolute prediction error for up to 24 steps ahead forecast after 48 prediction steps

accuracy can be achieved by introducing an additional model, based on average daily consumption data, which would facilitate predictions for several days ahead.

The daily demand model would then have the form

$$(1+\phi_1 B)w_t = (1+\theta_1 B)(1+\Theta_1 B^7)a_t \tag{12.30}$$

and produces the autocorrelation function of residuals shown in Figure 12.4. The prediction error is now reduced to approximately 1.7% (Figure 12.5). Introduction of a hierarchy of models according to their time scales allows a simple structure to be retained and consequently enhances the computational efficiency of the method.

The significance of the tracking of model parameters has also been studied. Figures 12.6–12.9 illustrate the variation of parameters of the

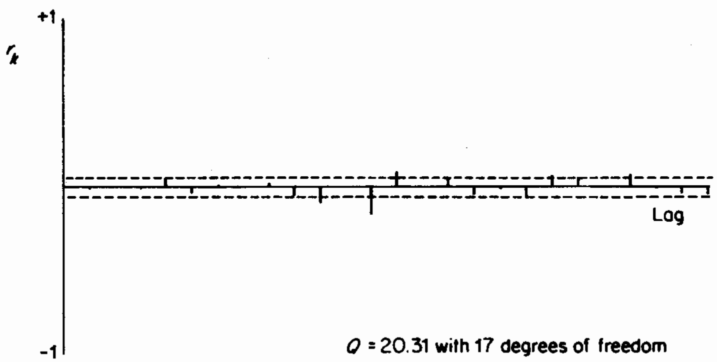

FIGURE 12.4 Autocorrelation function of residuals

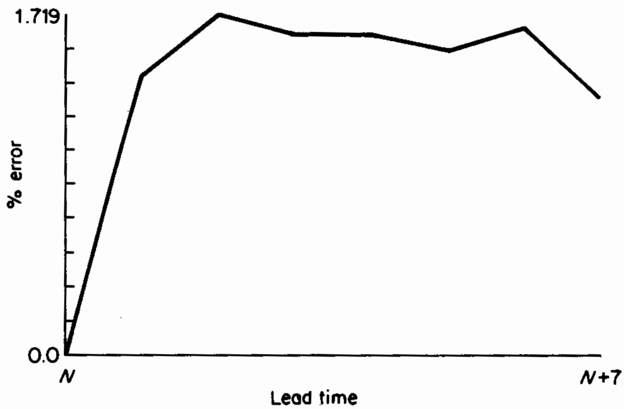

FIGURE 12.5 Mean absolute prediction error for up to 7 steps ahead forecast after 14 prediction steps

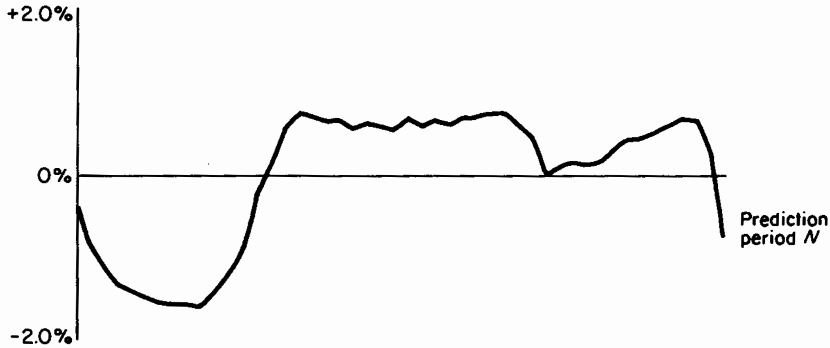

FIGURE 12.6 Percentage change in ϕ_1 from its mean value of 0.966

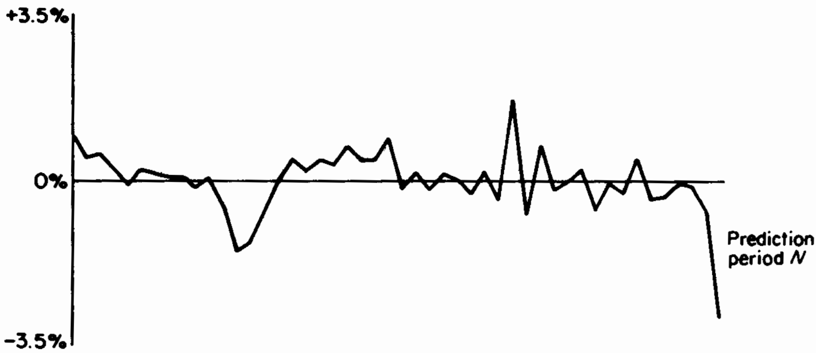

FIGURE 12.7 Percentage change in Φ_1 from its mean value of 0.981

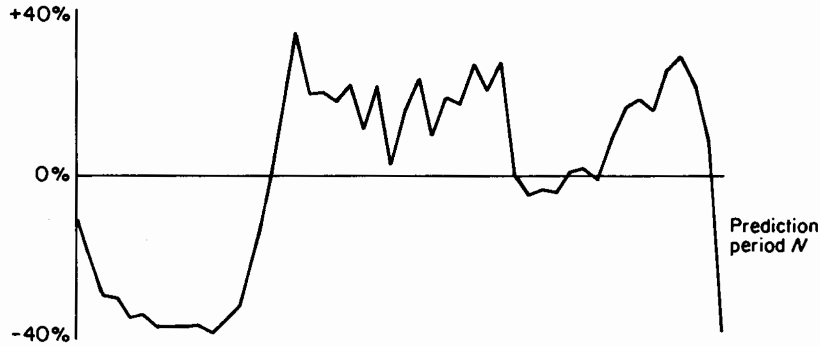

FIGURE 12.8 Percentage change in θ_1 from its mean value of 0.564

model (12.29) during a 24-hour period. As expected, the seasonal autoregressive parameter Φ_1 shows only some random fluctuations around its mean value (Figure 12.7) since each new data point remains in agreement with the periodicity of the demand. The parameters ϕ_1, θ_1, and Θ_1 are modified in line with the changes of the demand giving a good justification for the adaptive procedure. As can be seen from Figures 12.8 and 12.9 the changes of magnitude of the moving average parameters θ_1 and Θ_1 are particularly high and alter by as much as +35% to −39% and +42% to −28% respectively of the mean value of the parameters. This lends support to the conclusion that the adaptive structure of the model is highly desirable.

Computation times obtained for a FORTRAN 77 implementation of the algorithm using a 32-bit minicomputer (Perkin–Elmer 3220) are presented in Table 12.1.

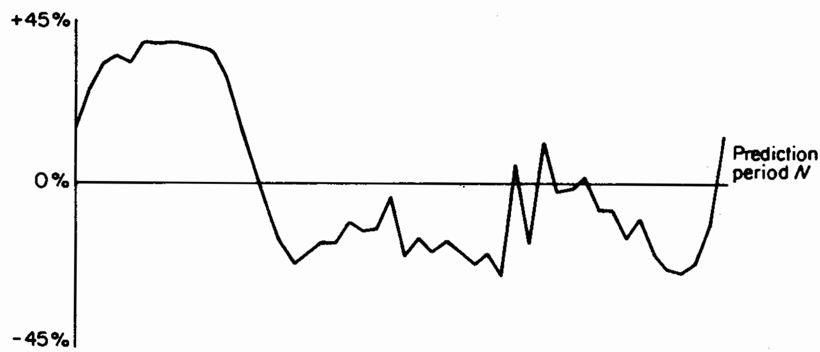

FIGURE 12.9 Percentage change in Θ_1 from its mean value of −0.377

TABLE 12.1

	Computation time (s) Model 1 –336 data points –4 model parameters –48 time-step season –24 time-atep prediction	Computation time (s) Model 2 –100 data points –3 model parameters –7 time-step season –7 time-step prediction
Initial model fitting	8.0	1.6
Model parameters update	0.720	0.150
Prediction	0.085	0.030

7 CONCLUSIONS

A technique for the modelling of time series has been presented and shown to be of a form suitable for on-line prediction of water demand but the method has also been found to be suitable for electricity demand prediction. A novel application of the Fletcher–Powell variable metric algorithm for function minimization in the Newton–Raphson method has been shown to result in a unity rank updating formula which enables the tracking of variation in the model parameters. The significance of the adaptive structure of the algorithm has been illustrated above and found to be a highly desirable feature for an on-line computer-based demand prediction scheme.

REFERENCES

1. E. D. Farmer (1963). 'A method of prediction for nonstationary processes and its application to the problem of load estimation.' *Proc. IFAC Congress, Basle, 1963.*
2. M. J. H. Sterling and D. J. Antcliffe (1974). 'A technique fór the prediction of water demand from past consumption data.' *J. Inst. of Water Engs.*, **28**, No. 8, 413–420.
3. G. T. Heineman, D. A. Nordman, and E. C. Plant (1966). 'The relationship between summer weather and summer load—A regression analysis.' *IEEE Trans.*, **PAS-85**, No. 11, 1144–1154.
4. W. R. Christiansee (1971). 'Short term load forecasting using general exponential smoothing.' *IEEE Trans.*, **PAS-90**, 909–910.
5. J. Toyoda, M. Chen, and W. S. Inoue (1970). 'An application of state estimation to short term forecasting.' *IEE Trans.*, **PAS-89**, 1678–1682.
6. M. J. H. Sterling and M. R. Irving, 'Integrated software for power system control.' In M. Singh (Ed.), *Encyclopaedia of systems and control*, Pergamon Press.
7. F. Fallside and P. F. Perry (1975). 'On-line prediction of consumption for water supply network control.' *Proc. IFAC, Boston, Mass., August 1975.*
8. R. DeMoyer and L. B. Horwitz (1975). *A system approach to water distribution modelling and control*, Lexington Books, D. C. Heath & Co.

9. G. E. P. Box and G. M. Jenkins (1970). *Time series analysis: Forecasting and control*, Holden-Day, San Francisco.
10. M. J. H. Sterling (1978). *Power systems control*, Peter Peregrinus, London.
11. W. Gilchrist (1976). *Statistical forecasting*, Wiley, London.
12. R. Fletcher and M. J. D. Powell (1963). 'A rapidly convergent descent method for minimisation.' *Comput. J.*, **6**, 163–168.