

Short-term load forecasting for demand side management

S.Sargunaraj
D.P.Sen Gupta
S.Devi

Indexing terms: Short-term loadforecasting, Demand side management

Abstract: A method for short-term load forecasting which would help demand side management is presented. This is particularly suitable for developing countries where the total load is not large, especially at substation levels, and the data available are grossly inadequate. It is based on the Kalman filtering algorithm with the incorporation of a 'fading memory'. A two-stage forecast is carried out, where the mean is first predicted and a correction is then incorporated in real time using an error feedback from the previous hours. This method has been used to predict the local load at 11kV and also the bulk load at 220kV. The results and the prediction errors are presented.

1 Introduction

Short-term load forecasting is the prediction of power demand with lead times ranging from an hour to a week. Short-term load forecasting is necessary so that the quantity of generation may be estimated and, on this basis, load management planned to meet the expected demand. In developing countries the power sector is often unable to meet peak demands. It seems essential that the scheduling of generation is to be planned carefully since one has to work within stringent limits. Hence, suitable strategies are necessary for generation control and load management. For this purpose, short-term load forecasting has to be carried out as accurately as possible.

In developing countries, load management usually takes the form of 'load shedding'. 'Power cuts' or 'voltage reduction' at substation levels are often resorted to. In the absence of centralised controls, decentralised load management needs to be carried out and reallocation of local peak loads can only be made if the local short-term load may be predicted. The prediction of local load at the 11kV level is significantly more difficult than the prediction of bulk load which exhibits a stronger statistical pattern due to

an averaging effect. Most papers on load forecasting usually deal with bulk load and can therefore, report fairly accurate predictions.

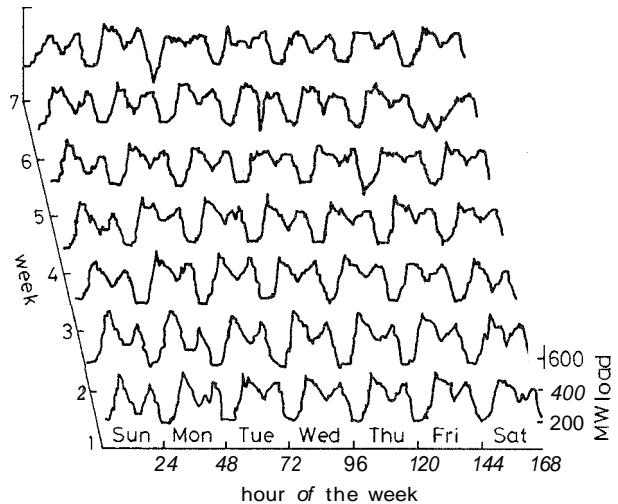


Fig.1 Load variation over period of several weeks 220kV level

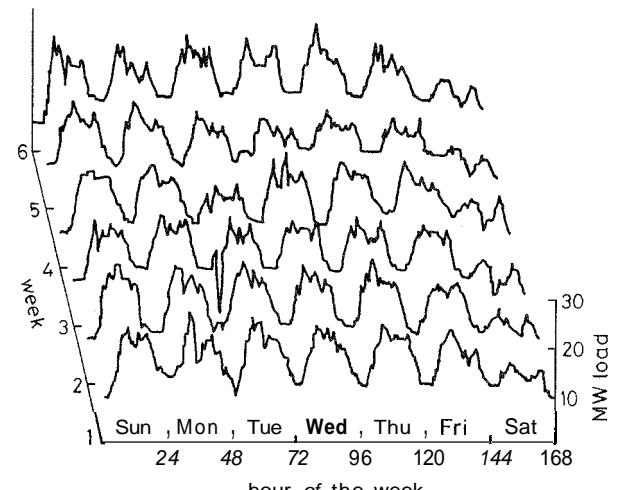


Fig.2 Load variation over period of several weeks 11kV level

Figs. 1 and 2 give a sample of the data used by us for the purpose of forecasting. Fig. 1 shows the load curve at 220kV over a period of several weeks. Fig. 2 gives the load curve for 11kV. It can be seen that the load shape more or less repeats every week, except for some fluctuations. It is also evident that the repetitive pattern is much stronger at the 220kV level than at the 11kV

level. The peak load at 220kV is less than 600MW and the peak load at 11kV is less than 30MW. Most papers on load forecasting usually deal with bulk load of the order of gigawatts, and it can be seen that the quantum of load for which prediction was carried out by us is comparatively very small. At this level, the accuracy of forecasts will not be as good as that for bulk load. Short-term load forecasting at the quantum of load we have dealt with is, however, essential for load management, especially if it needs to be decentralised as in countries with poor communication links.

One of the major inputs in standard predictions is weather data. Although national weather forecasts are carried out nowadays in developing countries, it is seldom that reliable local weather data and their correlation with load demand are readily available at substation levels to be incorporated into load forecasting.

A simple method of error feedback has therefore, been devised to make corrections in the prediction, especially for the peak hours, by observing the deviations of the off-peak predictions from the measured values of the actual load at early hours of the day.

Kalman filter algorithm was found particularly suitable for predicting the average local hourly loads, and the dynamic corrections were carried out subsequently.

The study has been based on two years' hourly load data collected from 11kV and 220kV substations in Bangalore. Suitable corrections have been incorporated to take into account abrupt changes resulting from sudden load shedding or occasional wrong entries.

2 Methodology

The models used for load forecasting mainly belong to two categories. In one, the weather variables which affect the consumption appear explicitly in the model. In the other, the observed load data are treated purely as a time-series, the effect of weather being implicit in the data.

Weather variables affect load consumption significantly. This relationship may be expressed as

$$W(t) = f(T, H, P, \dots)$$

where $W(t)$ (which is a function of time t) is that part of the total load $L(t)$ dependent on the weather, and f denotes the functional relationship between the weather-sensitive part of the load and temperature (T), humidity (H), precipitation (P). Variables such as wind speed, cloud cover etc. are other factors. Davies [1] and Copernig *et al.* [2] have used this type of model.

The time series models include the **ARMA** model as used by Hagan and Klein [3], Hagan and Behr [4] and Natarajan and Bhagat [5]. The spectral decomposition model which is also a time series model has been used by Moutter *et al.* [6] and Bodger *et al.* [7]. The state space model has been used by Toyoda *et al.* [8]. Christiaanse [9], Sharma and Mahalanobis [10] and Sarma *et al.* [11] use a combination of the spectral decomposition model and the ARMA model.

In recent years, many of the papers on load forecasting use artificial neural networks (ANN). Generally a multilayered perceptron type neural network is used where the learning is through back-propagation. Two representative papers are [12, 13].

The forecasting procedure proposed by us is based on a Kalman filtering algorithm which is modified to give more emphasis on recent data as compared with

past data. The more recent the data, the greater the weightage. The weighting function is exponential with a negative exponent ($\exp[-1/M]$) so that the filter 'memory' may be said to 'fade' exponentially. The 'fading memory' has been incorporated in the usual Kalman filter algorithm and the following set of equations were derived (Appendix 6):

$$(a) I(T) = R(T) - D(T-1) \quad (1)$$

$$(b) V(T) = \frac{I(T)^2}{T} + \frac{T-1}{T} V(T-1) \quad (2)$$

$$(c) K(T) = \frac{P(T-1)}{P(T-1) + FV(T)} \quad (3)$$

$$(d) D(T) = D(T-1) + K(T)I(T) \quad (4)$$

$$(e) P(T) = (1/F)P(T-1)[1 - K(T)] \quad (5)$$

$$(f) R_f(T+1) = D(T) \quad (6)$$

where

$R(T)$ = measured value at time T

$D(T-1)$ = predicted value of $R(T)$

$D(T)$ = predicted value of $R(T+1)$

$I(T)$ = forecast error

$V(T)$ = variance of $R(T)$

F = $\exp(-1/M)$ where M is the 'time constant'

$K(T)$ = $P(T)/V(T)$ = Kalman gain

$R_f(T+1)$ = predicted value

It is observed that the load demand has a periodicity of a week (168h) within a season. The load demand at 10 o'clock on a Friday, for example, is different from the demand at 10 o'clock on other days of the week, but tends to repeat itself for a few consecutive Fridays, changing with the change in season. The load expected at a particular hour of the week can be predicted by filtering out the fluctuations from the load values observed at corresponding hours in previous weeks. The actual load observed fluctuates about these expected values. One of the factors contributing to these fluctuations is the weather. As has been mentioned, weather data are not easily available and therefore are not incorporated in this study. Their effects are however indirectly taken into account by a second stage correction of the forecast. It is assumed that the effects of the weather prevailing on a day are reflected in the load demand at early hours. The difference between the prediction and the actual measurement in the morning can be taken as an indication of how the predictions for the rest of the day need to be modified. The forecasting procedure thus comprises two stages.

In the first stage, 168 separate filters are applied in parallel to 168 separate hour-of-the-week load sequences. The load values $L(t)$ observed at time t is represented as

$$L(t) = B(t) + E_1(t) \quad (7)$$

where $L(t)$ is the sum of the average (or base) load $B(t)$ and a stochastic fluctuation $E_1(t)$. The first stage filters out $E_1(t)$ and the resulting estimate $B(t)$ gives the one-week ahead prediction of the expected load. The prediction errors for this lead time is equal to the value of $E_1(t)$.

In Stage 2, hourly corrections are carried out. As stated earlier, the deviations of the measured from the

expected load on a particular day may be taken as the general tendency for the rest of the day and therefore, these deviations are used to correct the forecast for the forthcoming hour. The deviation $E_1(t)$ obtained for every hour from

$$E_1(t) = L(t) - B(t) \quad (8)$$

constitutes the observed sequence for the hourly fluctuations. This sequence is modelled as

$$E_1(t) = W(t) - E_2(t) \quad (9)$$

based on the consideration that the deviation $E_1(t)$ consists of $W(t)$, the change in load demand due to weather change and $E_2(t)$, a residual random component. When the second stage of the filter is applied, $W(t)$ is abstracted from $E_1(t)$ based on the data at early hours of a particular day using the same filter. This is algebraically added to the expected load $B(t)$ obtained from the first stage filter with lead time of a week. The residual quantity $E_2(t)$ in eqn. 9 forms the prediction error with a lead time of 1h.

Forecast errors increase if there is excessive noise present in the data, and usually some smoothing technique is used to increase the signal to noise ratio for short-term load forecasting. With the Kalman filtering technique, smoothing is not really essential since the algorithm seeks the expected value despite the presence of noise.

Fig. 13 gives the flowchart describing the procedure used by us for load prediction.

3 Results

This method of short-term load forecasting based on the fading-memory Kalman filter algorithm has been applied to the short-term prediction of substation load (at 11kV) and bulk load (at 220kV). The data used is (i) that of Bangalore city with an average load of about 350MW at 220kV level and (ii) that of a local substation in an industrial suburb of Bangalore with an average load of 15MW at 11kV. Two years of data recorded every hour was used for both sets of data. Data screening was carried out before using the data for prediction.

The fading memory Kalman filter algorithm provides variable weightage to past data. This is done by the factor $\exp(-(T-t)/M)$ incorporated in the Kalman filter algorithm. If T is the instant of time at which the forecast is made, the weightage given to the data at time $T-1, T-2, \dots$ will be $\exp(-1/M), \exp(-2/M), \dots$, so that the weightage keeps on reducing as we go back in time (Appendix 6, eqn. 12). The selection of M is done empirically so as to reduce the error of prediction. This was done by using the first six months' data. It was found that the value of M remains almost constant for a substation. Once the value of M is fixed, it can be used in the algorithm and does not need to be re-evaluated. For various values of M , the forecasting was carried out and the error in prediction calculated in each case. Figs. 3 and 4 show the error in prediction for different values of M in two typical cases. It can be seen that, as M is increased, the error keeps reducing up to a point beyond which it increases again. The optimal value of M is chosen from Figs. 3 and 4. As M is increased the value of error seems to saturate and as M tends to infinity, the factor F will become one which gives the Kalman filter algorithm without the fading memory.

Data smoothing was carried out using Tukey's method [14] to see whether raw data (barring missing data) could be used for load prediction. Extensive studies have shown that marginal improvement is achieved by data smoothing, since we need to use data only from the recent past. Besides, Kalman's filter offers the distinct advantage that raw data may be directly utilised after preliminary screening for odd or missing data.

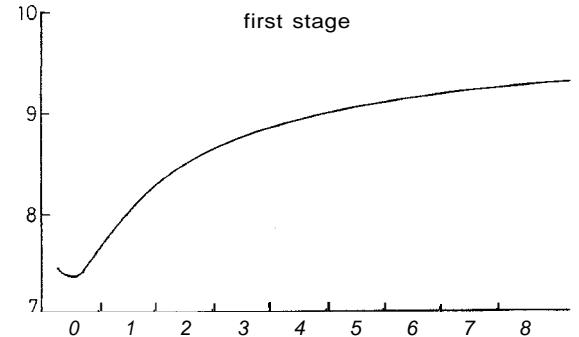


Fig.3 Forecast error for different memory depths at 220kV level
One week ahead forecast, $x = M$, $y = \text{percentage error}$

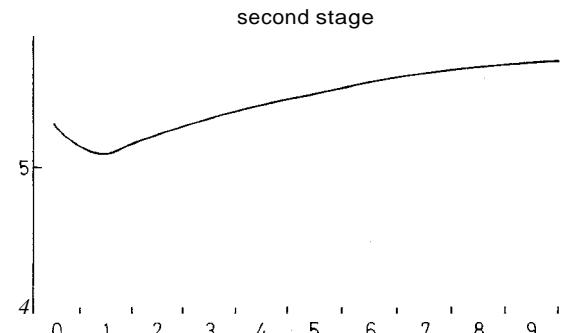


Fig.4 Forecast error for different memory depths at 220kV level
One hour ahead forecast, $x = M$, $y = \text{percentage error}$

3.1 One week ahead prediction

Figs. 5 and 6 show the results of prediction after only the first stage forecast is carried out for the 220kV level and the 11kV level. The actual load and the predicted load have been plotted against time. Figs. 7 and 8 show the errors in prediction for the 220kV data and the 11kV data. The error of prediction for any hour of the week is the average of all the weeks of study and not just the error of prediction of one week taken at random. The average prediction error is 8.4% for the 220kV level load and is 11.5% for the 11kV load. The accuracy of prediction is, as expected, better for the 220kV load.

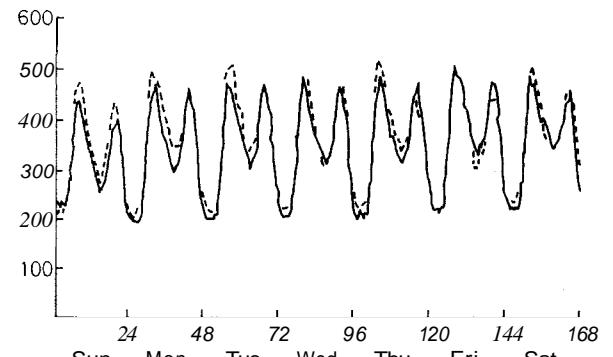


Fig.5 One week ahead forecast of actual demand at 220kV level
Sun 1 Dec to Sat 7 Dec 1985
— data - - - forecast
 $x = \text{hour of week}$
 $y = \text{load in MW}$

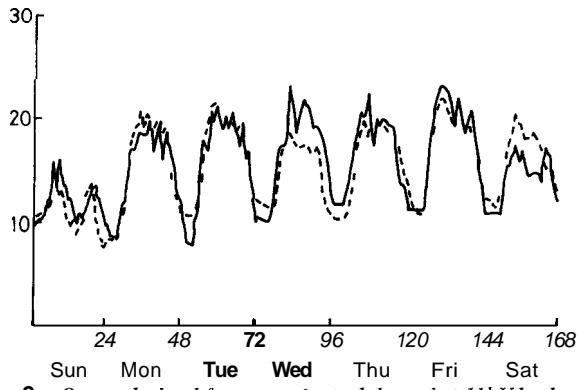


Fig.6 One week ahead forecast and actual demand at 11kV level
Sun 13 Oct to Sat 19 Oct 1985
— data —— forecast
x = hour of week
y = load in MW

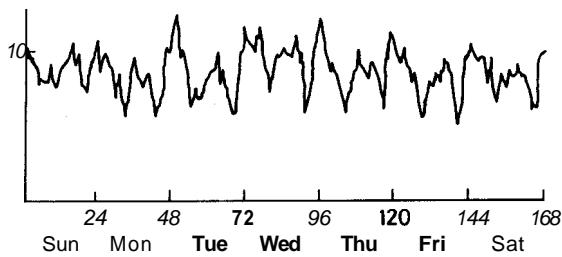


Fig.7 Hourly error in predicting total city load at 220kV level
Lead time of one week
x = hour of week
y = percentage error

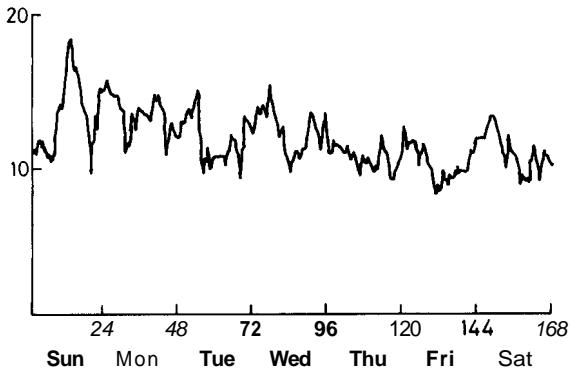


Fig.8 Hourly error in predicting local substation load at 11kV level
Lead time of one week
x = hour of week
y = percentage error

Two other methods, chosen particularly since weather data are not used here either, are based essentially on Fourier analysis [11, 7]. These were used to compare the relative accuracy of prediction. It was found (Table 1) that the Kalman filter gave better results and the computation time was about 25% of that of the other two methods. Data storage was significantly less.

Although the proposed Kalman filter method seems to be a reasonably accurate method for short-term load forecasting in a statistical sense, it is also true that the maximum error in the bulk load forecast may be as high as 13 (Fig. 7). Although this is better than what may be obtained from the other two methods in which

the maximum error in prediction may be as high as 16% and 30%, a 13% error in the total load of around 450MW means an error of 60MW in the prediction of load. This may not present a problem where there is sufficient spinning reserve but in a country like India, where spinning reserve is inadequate, a shortfall of 60MW in prediction is difficult to cope with. This however is the worst case and Fig. 5 shows generally good agreement between the predicted and the actual load.

At present, there is hardly any forecasting or generation planning carried out in India, or for that matter in any developing country, except that a rough estimate is made based on experience. When the load exceeds generation, the system frequency drops. Generation is increased if possible or load shedding is resorted to as stated earlier.

Table 1: Performance of the proposed method vis-a-vis two other methods for a lead time of one week

	Overall mean absolute percentage error	Standard deviation of percentage error	Relative frequency of absolute percentage error less than 10%	Relative computer time required for prediction
Proposed method				
11kV	11.5	15.3	56	1
220kV	8.4	11.4	69	1
Method of [11]				
11kV	13.2	17.2	48	4
220kV	10.1	12.6	57	4
Method of [7]				
11kV	12.8	17.4	52	4
220kV	17.9	21.9	34	4

The one week ahead prediction as shown in Fig. 5 is a reasonably accurate one to plan generation and power import/export ahead of time.

The prediction at 11kV level, although less accurate, would help demand side management. If the predicted peak demands from all 11kV stations cannot be met, bulk consumers may be prevailed upon to shift a part of their loads away from peak hours, care being taken to see that a larger peak may not appear elsewhere. The second stage correction provides a distinct improvement.

3.2 Second stage correction

The second stage correction as described earlier gives distinctly better predictions. This estimate has to be carried out in real time. The one-week-ahead forecasts are already computed and stored in the computer. The load measured at early hours of the morning are compared with the predictions already stored in the computer. The error is used, as described in Section 3 to improve the prediction for the next hour. The process is continued, and the correction is invariably found to bring the week-ahead prediction closer to the actual measured values.

Figs. 9 and 10 show the results of prediction at the 220kV and the 11kV voltage levels. The actual load and the predicted load are shown as a function of time. Figs. 11 and 12 show the prediction error for the

220kV and the 11kV voltage level when hourly correction has been carried out. It can be seen that there is a significant improvement in the predicted results after the second stage is carried out. Table 2 shows the results obtained both for one-week-ahead forecasts and for one-hour-ahead forecasts at both levels of load.

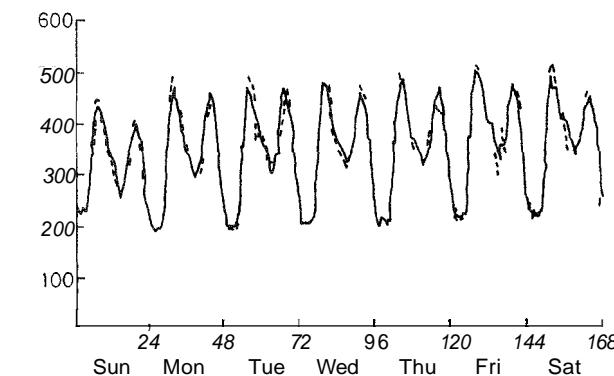


Fig.9 One hour ahead forecast and actual demand at 220kV level
Sun 1 Dec to Sat 7 Dec 1985
— data —— forecast
 x = hour of week
 y = load in MW

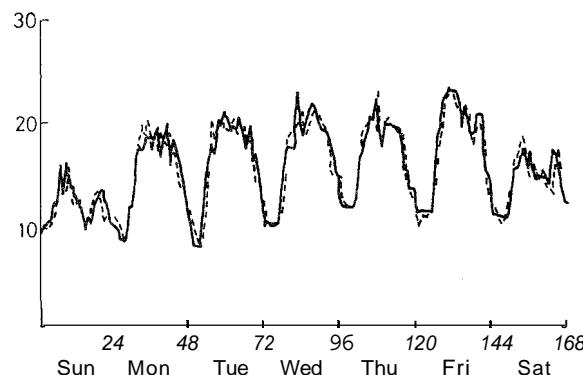


Fig.10 One hour ahead forecast and actual demand at 11kV level
Sun 13 Oct to Sat 19 Oct 1985
— data —— forecast
 x = hour of week
 y = load in MW

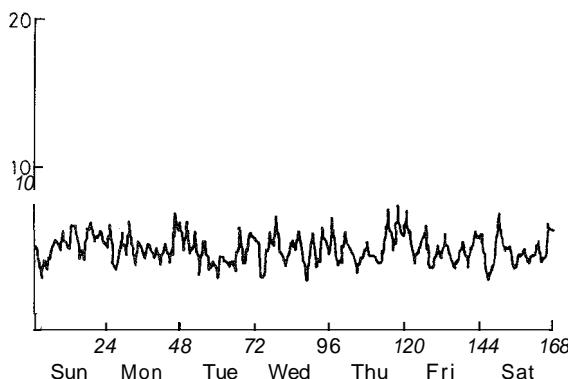


Fig.11 Hourly error in predicting total city load at 220kV level
Lead time of one hour
 x = hour of week
 y = percentage error

The one-hour-ahead prediction, even for local load (11kV level) is good and reliable. This will enable us to carry out local load management with confidence. In the event of a threatened emergency, bulk industrial load may be reduced by making requests to local industries to shed load.

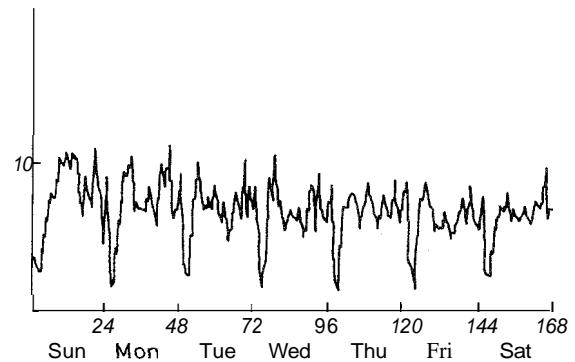


Fig.12 Hourly error in predicting local substation load at 11kV level
Lead time of one hour
 x = hour of week
 y = percentage error

Table 2 : Mean absolute percentage errors in prediction

	220kV voltage level	11kV voltage level
	%	%
One-week-ahead forecast error (Stage 1)	8.4	11.5
One-hour-ahead forecast error (Stage 2)	5.3	6.6

4 Conclusion

In this paper the fading-memory Kalman filter has been proposed to assign variable weightage to past data. This has not only resulted in reduction of dependence on data far back into the past, but also improved the accuracy of prediction to a certain extent. Space for data storage and the time taken for computation are both significantly low and makes this method highly suitable for use in small computers.

The load despatch centre can undertake the generation planning when the anticipated load is known one week ahead. This forecast can then be compared with the sum of the forecasts made at the 11kV level. The expected generation, import and export of power, are then checked for adequacy. In the event of an expected shortage, remedial measures may be taken at the local levels to reduce the demand at critical hours. By using the one-week-ahead predictions, gross planning is carried out and predictions made one hour ahead would enable final corrections to be made. This will provide a lead time during which additional generation or import of power may be attempted, failing which voluntary load shedding is tried at local levels. In this way arbitrary power cuts may be avoided.

5 References

- DAVIES, M.: 'The relationship between weather and electricity demand', *IEE Proc. C*, 1958, **106**, pp. 27-37
- CORPENING, S.L., REPPEN, N.D., and RINGLEE, R.J.: 'Experience with weather sensitive load models for short and long term forecasting', *IEEE Trans.*, 1973, **PAS-92**, (6), pp. 1966-1972
- HAGAN, M., and KLEIN, R.: 'On-line maximum likelihood estimation for load forecasting', *IEEE Trans.*, 1978, **SMC-8**, (9), pp. 711-715
- HAGAN, M.T., and BEHR, S.M.: 'The time-series approach to short term load forecasting', *IEEE Trans.*, 1987, **PWRS-2**, (3), pp. 785-791

- 5 NATARAJAN, B., and BHAGAT, G.: 'Methodology for short-term load forecasting – a case study of Delhi (India)'. The IASTED international conference on *High technology in the power industry*, Phoenix, Arizona, USA, 1988
- 6 MOUTTER, S.P., BODGER, P.S., and GOUGH, P.T.: 'Spectral decomposition and extrapolation of variations in electricity loading', *IEE Proc. C*, 1986, **133**, (5), pp. 247–255
- 7 BÖDGER, P.S., BROOKS, D.R.D., and MOUTTER, S.P.: 'Spectral decomposition of variations in electricity loading using mixed radix fast fourier transform', *IEE Proc. C*, 1987, **134**, (3), pp. 197–202
- 8 TOYODA, J., CHEN, M.J., and INOUE, Y.: 'An application of state estimation to short-term load forecasting, Part 1: Forecast modelling, Part 2: Implementation', *IEEE Trans.*, 1970, **PAS-89**, (7), pp. 1678–1688
- 9 CHRÍSTIAANSE, W.R.: 'Short-term load forecasting using general exponential smoothing', *IEEE Trans.*, 1971, **PAS-90**, (2), pp. 900–910
- 10 SHARMA, K.L.S., and MAHALANABIS, A.K.: 'Recursive short-term load forecasting algorithm', *IEE Proc.*, 1974, **121**, (1), pp. 59–62
- 11 SARMA, V.K., BASU, T.K., and SINHA, S.: 'Medium range forecasting for power systems load by fast fourier transform', Platinum jubilee conference on *Systems and signal processing*, Bangalore, 1986, pp. 418–421 (Department of Electrical Engineering, IISc)
- 12 MOHAMMED, O., PARK, D., MERCHANT, R., DINH, T., TONG, C., AZEEM, A., FARAH, J., and DRAKE, C.: 'Practical experiences with an adaptive neural network short-term load forecasting system', *IEEE Trans.*, 1995, **PWRS-10**, (1)
- 13 PENG, T.M., HUBELE, N.F., and KARADY, G.: 'Advancement in the application of neural networks for short-term load forecasting', *IEEE Trans.*, 1992, **PS-7**, (1), pp. 250–257
- 14 MCNEIL, D.R.: 'Interactive data analysis: a practical primer' (John Wiley, New York, 1983)

6 Appendix: Fading memory Kalman filter algorithm

The measured value at time t can be modelled as

$$R(t) = C(t) + N(t) \quad (10)$$

where $C(t)$ = expected value of $R(t)$ and $N(t)$ = zero-mean noise with variance $V(t)$. Let $t = T$ be the instant of time when a forecast is to be made of $R(T+1)$. Let the predicted value of $R(T+1)$ be $D(T)$, which is the minimum variance estimate of $C(T)$ computed with the recorded measurements available at time T . $D(T)$ can be evaluated by minimising

$$f(D(T)) = \sum_{t=T-n}^T \frac{[R(t) - D(t)]^2}{V(t)} \quad (11)$$

where $n = 1, 2, 3, \dots$ indicates time units going backward in time. To give a greater weight to more recent data, a fading memory is incorporated into eqn. 11 to give

$$f_m(D(T)) = \sum_{t=T-n}^T \frac{[R(t) - D(t)]^2 F^{(T-t)}}{V(t)} \quad (12)$$

where $0 < F < 1$. F can also be expressed as

$$F = \exp(-1/M) \quad M > 0 \quad (13)$$

Minimising $f_m(D(T))$ by differentiating with respect to $D(T)$ and setting it to zero we obtain

$$D(T) = \frac{\sum_{t=T-n}^T \frac{F^{(T-t)} R(t)}{V(t)}}{\sum_{t=T-n}^T \frac{F^{(T-t)}}{V(t)}} \quad (14)$$

if

$$\frac{1}{\sum_{t=T-n}^T \frac{F^{(T-t)}}{V(t)}} = P(T) \quad (15)$$

say. Eqn. 14 becomes:

$$D(T) = P(T) \sum \frac{F^{(T-t)} R(t)}{V(t)} \quad (16)$$

$$D(T) = P(T) \left[\frac{R(T)}{V(T)} + F \sum_{t=T-1-n}^{T-1} \frac{F^{(T-1-t)} R(t)}{V(t)} \right] \quad (17)$$

$$D(T) = P(T) \left[\frac{R(T)}{V(T)} + F \frac{D(T-1)}{P(T-1)} \right] \quad (18)$$

since

$$\begin{aligned} \frac{1}{P(T)} &= \sum_{t=T-n}^T \frac{F^{(T-t)}}{V(t)} = \frac{1}{V(T)} + F \sum_{t=T-1-n}^{T-1-t} \frac{F^{(T-1-t)}}{V(t)} \\ &= \frac{1}{V(T)} + \frac{F}{P(T-1)} \end{aligned} \quad (19)$$

i.e.

$$\frac{FP(T)}{P(T-1)} = 1 - \frac{P(T)}{V(T)} \quad (20)$$

So eqn. 18 becomes

$$D(T) = \frac{P(T)R(T)}{V(T)} + \left[1 - \frac{P(T)}{V(T)} \right] D(T-1) \quad (21)$$

Rearranging, we get

$$D(T) = D(T-1) + \frac{P(T)}{V(T)} [R(T) - D(T-1)] \quad (22)$$

which can be written as

$$D(T) = D(T-1) + K(T)I(T) \quad (23)$$

where

$$K(T) = \frac{P(T)}{V(T)} \quad (24)$$

is the Kalman gain for this filter and the forecast error

$$I(T) = R(T) - D(T-1) \quad (25)$$

is the filter innovation.

Using eqns. 19 and 24 we get

$$K(T) = \frac{P(T-1)}{P(T-1) + FV(T)} \quad (26)$$

which is convenient for computing the gain.

Using eqns. 20 and 24,

$$P(T) = \frac{P(T-1)}{F} [1 - K(T)] \quad (27)$$

This equation can be used to update the quantity $P(T)$.

Since the expected value of $N(t)$ is zero using the equation

$$V(T) = \frac{1}{T} \sum_{t=1}^T N(t)^2 \quad (28)$$

we get

$$V(T) = (1/T)[N(T)^2 + (T-1)V(T-1)] \quad (29)$$

Here, $N(T) = R(T) - C(T)$ and since the best estimate of $C(T)$ is $D(T-1)$, $N(T)$ will be $R(T) - D(T-1)$ which is the filter innovation $I(T)$.

Thus eqn. 29 becomes

$$V(T) = (1/T)[I(T)^2 - (T-1)V(T-1)] \quad (30)$$

Grouping eqns. 25, 30, 26, 23 and 27 derived above, the fading memory Kalman filter algorithm can be described using the set of eqns. 1–6 as given in the paper.

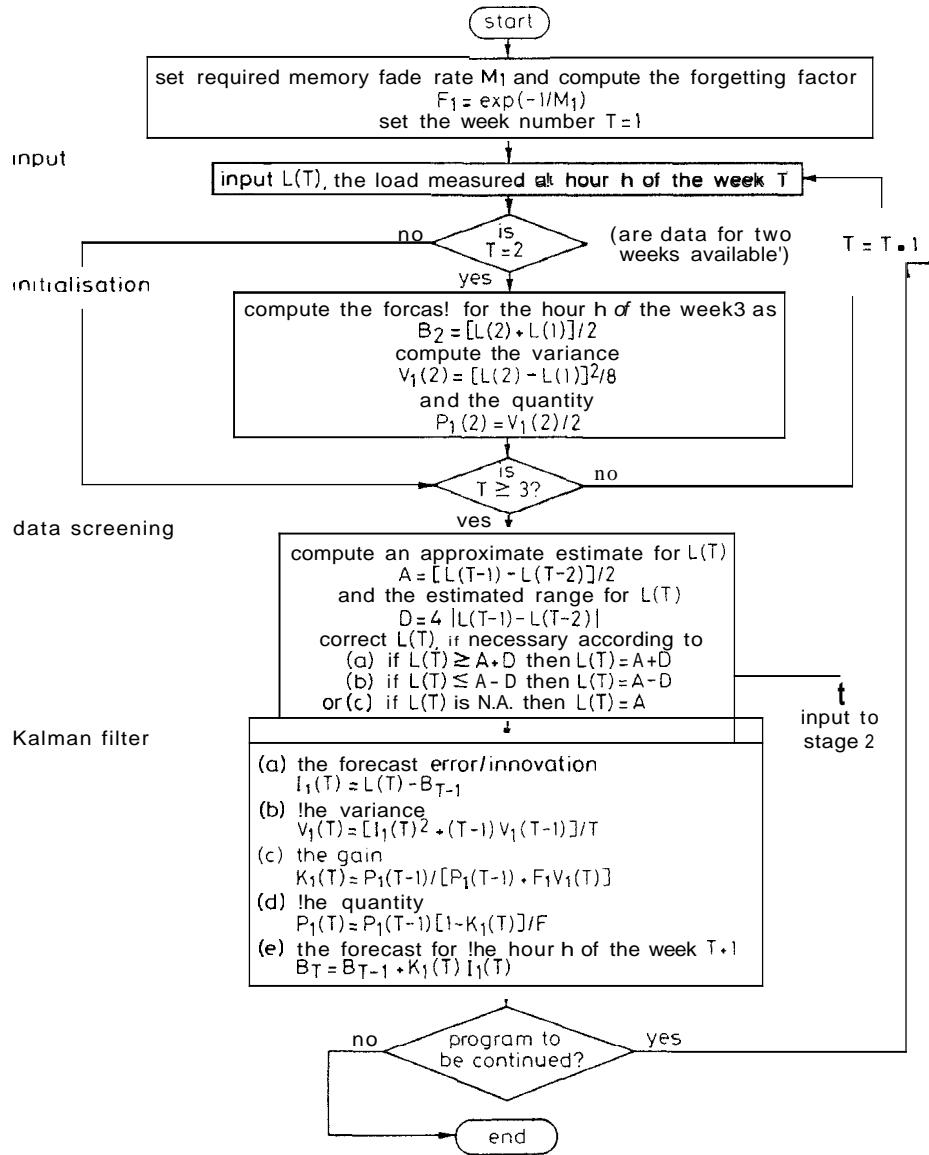


Fig. 13 Simplified flowchart of proposed short-term loadforecasting method