

## Medium range forecasting for power system load by fast Fourier transform

U. K. SARMA, T. K. BASU AND S. SINHA

Department of Electrical Engineering, Indian Institute of Technology, Kharagpur 721 302, West Bengal.

### Abstract

Medium range forecasts of daily electric power demand, spanning one to two weeks are required for preparation of short time maintenance schedule of unit auxiliaries and peaking stations. Forecast of daily load for seven days by the available multiplicative SARIMA model, suffers from divergent error levels of multistep forecasts and to take all the specific features of any particular day, the order of such a model is prohibitively large. The present paper demonstrates an altogether different approach to the time-series analysis of stochastic processes; the periodic nature of electric power system load demand spanning 24 and 168 hours has been gainfully exploited by grouping the data in seven subgroups characterising each week-day separately and then each day's data is transformed into frequency spectrum. The amplitudes of the spectrum have then been time-series modelled and forecasts are made by inverse Fourier transform.

**Key words:** Unit auxiliaries, peaking stations, maintenance schedule, medium range forecasting.

### 1. Introduction

Estimation theory and forecasting techniques have been in use in power system planning and control since early seventies<sup>1-6</sup>. Power systems all over the world are expanding at a rapid rate and with dwindling reserves of fossil fuels, economy and reliability of generation and also operation have assumed greater importance. To meet this challenge operational engineers have to judiciously coordinate generations of different thermal units, gas turbines and hydro units. For this task, *a priori* knowledge of load demand patterns at various important nodes as well as the total demand along with some meteorological data like temperatures, humidity, rainfall, etc., for various ranges of time intervals are very essential. Modern techniques of forecasting are, therefore, effective tools for working out any control strategy with the least amount of uncertainty. The demand patterns required may be of hourly or half-hourly MW-load for a few hours to week, weekly or monthly energy (MWh) demand or a yearly profile of monthly peak (MW) and energy (MWh) demand, etc. Once a tentative model is selected, following some criterion, the parameters need be optimised, diagnostic checks are to be made and then the model is ready for forecasting. Short-time maintenance involving smaller generating units like peaking stations (hydro/gas turbines) and unit auxiliaries can be planned if the multistep forecasts of hourly demand can be made for the next 24 to 168

\*First presented at the Platinum Jubilee Conference on Systems and Signal Processing held at the Indian Institute of Science, Bangalore, India, during December 11-13, 1986.

hours (1 week). Such a long penetration into the future is quite an involved task. For programming the schedule of preventive maintenance of larger generating sets and boiler turbine units, prediction of load with a few weeks' or even months' time in advance is desirable. For forecasting the demand profile for the next 24 hours, we need past data to determine the different elements of periodicity and trends. In general, forecasts can be made either through causal models<sup>3</sup> entailing functional relationship between load demand (output) and variables such as temperature, humidity or economic factors (input), or time-series model necessitating only a set of a sequence of discrete data. Since the various input variables are not always available in finer details and quantification of certain factors and their relationship with the load demand are difficult to express mathematically, the casual models are much less in use for such purposes.

## 2. Statement of the problem

Power system data is marked by periodicities of 24 hours, one week and/or one year depending on the interval and length of data points. For medium range forecasting of hourly data spanning one day to one week, a non-multiplicative time-series model of the SARIMA (seasonal auto-regressive moving average) type will be rather inadequate in representing all the patterns in a day and the specific features of a particular week day. Also with such a model multistep-ahead forecast will converge either to the mean or the last value with increasing variance of the forecast error, for a given confidence level and quite often, the forecasts beyond one or two steps are unsuitable for any practical use. Thus, the problem is to make a forecast of the load for 168 hours (one week), in advance with uniform level of accuracy.

## 3. Sarma-Sinha-Basu model approach

The present work is aimed at circumventing multiplicativeness of models based on the following three new ideas:

(a) Hourly load data of a particular week-day for 24 hours are grouped together, say, for total *NW* weeks. Thus, seven groups are formed like, Sunday group, Monday group, etc. Each of these groups will be represented by a data matrix of  $(24 \times NW)$ . Each group is modelled separately and forecasts are made for each week-day separately. Thus seven days' forecasts are made separately without going for the multistep-ahead forecasts.

It is to be understood here, that in the Indian context and in a large metropolitan power system, the industrial loads are distributed in such a way that their holidays are staggered and each week-day has its own characteristic features regarding the occurrence

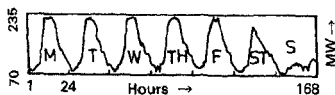


FIG 1. Hourly loads (MW) over a week.

of different peaks and troughs. (fig. 1) and the duration of different constant level blocks in the entire load profile for 24 hours. Thus, each of the seven subgroups will reflect the characteristic trends in the data set.

(b) For each subgroup the same single model cannot be used for two reasons:

- i) To accommodate all load patterns, the order of the model to be used will be very large.
- (ii) Forecasting with 24 steps will be less useful due to larger limits of variances.

To bring the 24-hourly periodicities into focus the data is resolved into frequency spectrum with the help of discrete Fourier transforms (DFT)<sup>7</sup>. Each complex component is then modelled as a suitable ARMA with the help of Box-Jenkins method and the parameters are determined following least squares algorithm. The real and imaginary parts are modelled separately. Based on the estimated parameter values, forecasts are made for each component and finally, by taking inverse discrete Fourier transforms (IDFT), the total forecast for 24 hours can be obtained.

Model order and parameter values for each DFT component are selected so that the model residuals give minimum variance and pass a 'whiteness test'.

At this juncture two important aspects of the proposed method need further elaboration.

(1) A particular day group is assumed to follow a well defined stochastic process and treated as stationary over a year and each DFT component in that day-group is modelled separately. Growth rate is appreciable only when yearly data over a span of nine to ten years are handled. Further, the stationarity of any day-group in any power system is justified if forecasts are needed for one or two weeks.

In the continuous domain we can explain the load in terms of the histogram. Now the average of the daily load of, say, the Monday group may exhibit an ARMA (1,1) process whereas, the amplitude of a component with a time period of 24 hours (fundamental) may exhibit an AR (3) process. Except for contingencies, the frequency components follow certain definite patterns depending on how different components of loads (office load, traction load, textile mill load, shipyard load and so on) vary with seasons over the whole year.

(2) While modelling a particular series (formed by a certain DFT component over a number of weeks) a tacit assumption is being made that it is being driven by a white noise. In the stricter sense, the DFT component series is not a time-series of the normal kind and the input is not 'white noise'. It is the amplitude of a particular frequency measured on (say) every Monday for the Monday group, that is being modelled. Thus let us suppose that the second harmonic corresponding to a period of 12 hours, has a magnitude (in MW) as 50, 52, 51, 47 and so on in a sequence of Mondays. When we talk of modelling this time-series, we presume that it is driven by an input noise of corresponding frequency, but the noise amplitude varies randomly with an ACF pattern similar to that of a 'white noise' in the time domain. Extrapolation of such a concept of 'white noise' may be justified as long as the load curves are taken for the same span (24 hours) for all the week days and the days are evenly spaced with an interval of seven days.

This is a departure from the current trend where periodicities are taken into account by multiplicative Box-Jenkins model<sup>4,5</sup> giving rise to a nonlinear estimation algorithm, whereas the proposed model will have a fairly straight forward algorithm and particularly suitable for parallel processing where time is a factor.

(c) The DFTs can be most efficiently obtained by the fast Fourier transforms (FFT)<sup>8</sup> which, however, require a data set of  $2^n$  number of values of the variable (data), where  $n$  is a positive integer. Since, in hourly load data there are only 24 data points in a day, it is proposed to use padding technique<sup>9</sup> of augmenting the data set by eight points so that the new set becomes transformable now. In the forecasts made, the last eight values are rejected to get 24-hour forecast only.

#### 4. Mathematical formulations

Let  $y_{i,j}^m$  be the  $j$ th hourly load data of the  $m$ th daygroup of the  $i$ th week,

$$\begin{aligned} m &= 1, 2, 3, \dots, 7 \text{ (Sunday to Saturday)} \\ j &= 1, 2, 3, \dots, 24 \text{ (1st to 24th hour)} \\ i &= 1, 2, 3, \dots, \text{NW (1st - NW weeks)} \end{aligned}$$

Then,  $y_{i,j}^m$ ;  $j = 1, 24$  represents the daily load curve data.

Let  $y_{i,j}^m = Z_i^m(j)$ ;  $j = 1, 24$ .

Let the augmented set for the  $m$ th group, employing padding be considered as,

$$x_i^m(K) = Z_i^m(1), Z_i^m(2), \dots, Z_i^m(24), Z_i^m(25), Z_i^m(26), \dots, Z_i^m(32).$$

Now, DFT of this data sequence  $x_i^m(K)$  is

$$X_i^m(n) = \sum_{K=0}^{N-1} x_i^m(k) W_N^{nK}; n = 0, 1, \dots, N-1 \quad (1)$$

where  $W_N = \exp(-j2\pi/N)$ .

From a DFT series of, say the first NW, ( $x_{NW+1}^m(n)$ ), an ARMA ( $p, q$ ) model is extracted and  $\hat{X}_{NW+1}^m(n)$ 's (estimates of  $X_{NW+1}^m(n)$ ) are forecasted for the  $(NW+1)$ th week and the 24-hour forecasts are made and then, using IDFT,

$$\hat{x}_{NW+1}^m(k) = \frac{1}{N} \sum_{n=0}^{N-1} \hat{X}_{NW+1}^m(n) W_N^{-nk}; K = 0, 1, \dots, N-1. \quad (2)$$

By taking successive values of  $i$ , beyond NW (newly arrived data), DFTs are calculated and following recursive least square (RLS) algorithm, the parameter values are continuously updated; and forecasts  $\hat{x}_{NW+1}^m(K)$  are made for 32 points. It is to be noted that the weekly periodicity is made implicit in the formulation. The last eight values of the forecasts  $\hat{X}_{NW+1}^m(k)$  for  $k = 24, \dots, 31$  are ignored, since they correspond to the appended data.

Here we use only an AR ( $p$ ) model to forecast the different series employing RLS technique. Whilst an ARMA ( $p, q$ ) may be parsimonious, the computational time saved by FFT approach will be offset by the excessive number of non-linear equations involved in determining the parameters of an ARMA ( $p, q$ ) model.

### 1. Case study

In the present work the load demand data of one of the important sub-stations of the Tata Electric Company, Bombay, has been taken for the year, 1984 and has been analysed for three typical weekdays, viz. Monday (a normal working day), Saturday (a partly holiday) and Sunday (a weekly holiday). First 44 weeks (NW) data were used for modelling and the next few weeks (up to 52nd week) were used for model verification. Thus on the basis of 44-week data, we make the prediction of 24-hourly load for the 45th Monday, 45th Tuesday and so on, and thus get one complete week's forecast retaining the special features of each week day. It was found that for the AR( $p$ ) models for the DFTs, a maximum value of ' $p$ ' = 8 was sufficient. The residuals have been tested for whiteness in each case and the parameters are found to be 'significantly different from zero' by the Quenouilles criterion<sup>6</sup>. The correlograms of all the DFT models are found to be within the twice standard error limits (fig. 2 shows a typical one). The Chi-square test also shows that the 'hypothesis' is true for all the models.

The forecasts have been shown in fig. 3(a-c) for the three day-groups (Sunday, Saturday and Monday) from 45th to 47th week and it is observed that the forecasts are very close to the actual values. Of the different types of padding techniques available, circular padding technique is found to give good results.

## 6. Conclusions

For medium range forecasting of load of any particular day, the time-series analysis of DFTs is very helpful. It has the added advantage over normal time-series analysis in the sense that all the distinct periodicities and the divergent tolerance limits of multistep ahead forecasting are avoided. The well established technique of circular padding used in digital signal analysis has been gainfully employed for using the FFT algorithm. Another advantage of the method is that it is an automatic process and suitable to get the forecasts of many inter-connected sub-stations of different regions.

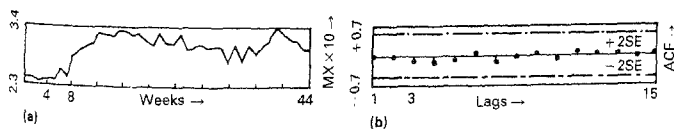


FIG. 2. (a) First DFT component; (b) ACF of model residue.

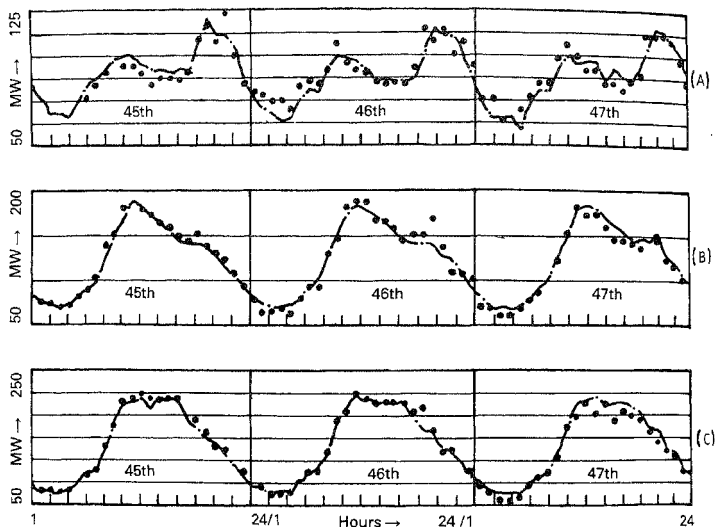


FIG. 3. Forecasts — (A) Sundays, (B) Saturdays, (C) Mondays. Actual . . . . Forecast - - - -

### Acknowledgement

The authors wish to express their appreciation to the Tata Electric Power Company, Bombay, for extending all help for collection of the data. They also express gratefulness to the Director, I.I.T., Kharagpur for providing the necessary facilities and the Government of Assam for sponsoring the first author under the Quality Improvement Program.

### References

1. SCHWEPPE, F.C. AND WILDES, J. Power system static state estimation, Parts I; II and III, *IEEE Trans.*, 1970, **PAS-89**, 120–135.
2. MAHALANABIS, A.K., HANMANDLU, M. AND RANA, R.S. Load modelling of an inter-connected power system for short term prediction, *Int. J. Systems Sci.*, 1980, **2**, 445–453.
3. BUNN, D.W. AND FARMER, E.D. (eds) *Comparative models for Electric load forecasts*. John Wiley, U.K. 1985.

4. KAIHO, S.                    *A System for time series analysis, time series*, Int. Conf. held at Nottingham Univ., March, 1979, pp. 295-299.
5. BOX, G.E.P. AND            *Time series Analysis Forecasting and control*, Helden-Dey, London, 1970.  
JENKINS, G.M.
6. O'DONOVAN, P. AND        *short term forecasting - An introduction to Box-Jenkins approach*  
THOMAS, M.                    John Wiley, N. Y., 1983.
7. JENKINS, G.M. AND        *Spectral analysis and its applications* Holden-Day, London, 1968.  
WATTS, D.G.
8. BRIGHAM, E. O.            *The fast fourier transform*, Prentice Hall, Inc., NY, 1974.
9. OTNIS, R.K. AND         *Digital time series analysis*, John Wiley, 1972.  
ENCINSON, L.