SHORT TERM LOAD FORECASTING FOR AN INTERCONNECTED POWER SYSTEM

By

M. TAHIR ABBAS KHAN

Submitted to the Faculty of Computer Science,

National University of Sciences and Technology, Rawalpindi,

in partial fulfillment for the requirement of an M S Degree.

April, 2000

Dedicated to my beloved parents, my wife and my daughter

ABSTRACT

Accurate load forecasting is very beneficial for the efficient and economical operation of a power supply utility. The power system load forecast can be categorized into long-term forecast and short-term forecast. Long term forecast usually covers a time span of one year to ten years and is needed for decision making regarding capacity expansion and long term capital investment return studies. Short term forecast on the other hand provides information about system load characteristics from one hour to twenty four hours or upto a few days into the future. Short term forecast is necessary for the efficient and reliable operation of an electric utility. It is needed for economic dispatch, unit commitment, energy sale/purchase decision and load management.

Keeping in view the importance of accurately forecasting the system load an improved algorithm has been developed for short term load forecasting of an electric utility. In this perspective the approach adopted is based upon time series analysis. Emphasis is on modeling the interconnected power system in state space form. Stochastic Approximation algorithm and Kalman Predictor are applied for load forecasting upto one hour. A software package is developed with the proposed implementation method for electric supply utilities. The application of this software package to real load data obtained from WAPDA (Water and Power Development Authority) has shown good results. The salient feature of this software package is that it can provide a breakup of load demand at each grid station, which is not currently available with WAPDA. This breakup can result in reduced transmission losses with corresponding economic benefits.

ACKNOWLEDGMENTS

First of all I want to thank "Almighty Allah" for the blessings he has conferred on me and for giving me the wisdom, knowledge and understanding to perform this research. I am a humble human being and without His help I would not have been able to carry out this task.

I wish to express my deepest appreciation to my thesis advisor Dr. Muhammad Ali Maud for his invaluable guidance, inspiration, dedication, commitment, encouragement and affection during the course of the research and the completion of the dissertation. I am also thankful to Lt. Col. Waseem Suleman, Dr. Nauman Jaffery and Dr. Nauman Ahmad Khan for taking keen interest and providing guidance.

I am heartily indebted to the members of my family for their immeasurable faith, wishes, forbearance and confidence in me, which had always been a source of inspiration and encouragement. Specially I would like to thank my parents who agreed to move to Rawalpindi so that I could carry out my studies peacefully and my wife for providing the moral support.

My sincere gratitude goes to National University of Sciences and Technology, Military College of Signals and especially to Computer Science Department for the provision of facilities for this research. A great deal of help was provided by the College of Electrical and Mechanical Engineering for which I am grateful to its Commmandant.

In the end I would again thank Dr. Muhammad Ali Maud because without his support and keen interest the completion of this thesis was very difficult.

Table of Contents

Contents **Page Nos.**

Contents **Page Nos.**

List of figures

Figure Nos.

Page Nos.

Chapter 1

Introduction

Identifying the characteristics of a system and its modeling is essential for controlling that system. Control applications in many diverse fields like chemical processes, biomedical systems, transportation, hydrology, socio-economic systems, aeronautics and electrical power systems require identification and modeling of the system. In all these cases, a model consists of a set of mathematical equations that can be used for an understanding of system's behavior for prediction and control.

There are two basic types of modeling problems and both are studied by means of mathematical (differential or difference) equations. In the first type a number of measurable inputs (causes) can be associated to a number of measurable outputs (effects). The mathematical model is determined by relating the outputs and inputs through mathematical equations. Some typical examples are modeling of a stirred-tank chemical reactor, a multi

machine electrical power system and a nuclear power plant. In these examples one can straightaway identify certain input and output quantities and develop mathematical model relating them.

In the other type one can identify a quantity as a measurable output however the causes (inputs) corresponding to the changes in output are not well defined. The examples of this type include the daily variations in the share value of a certain company enlisted in the stock market, annual flow in a river and fluctuations in values of different currencies in the money market. In these cases a sequence of outputs is available but the inputs are numerous and often unobservable. The sequence of outputs is called a time series and the models formulated in such cases, because of their probabilistic nature, are called stochastic models.

The first of the above mentioned two types of modeling problems is referred to as the problem of system identification and the second one as problem of stochastic modeling. However, both of them are closely related and in both cases one must be able to select the optimum model from the available set of models, that is, the model which represents the system of interest in the best possible way.

Thus in system identification the system models are determined using records of system operation. The problem can be represented diagrammatically as follows:

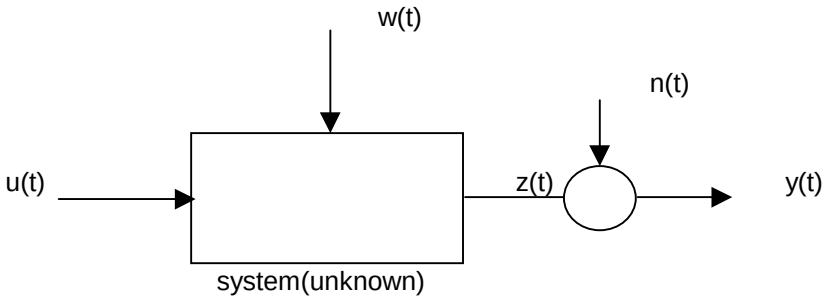


Fig 1.1 System identification problem

Where:

u(t) is the known input vector

z(t) is the output vector

w(t) is the input disturbance vector

n(t) is the observation noise vector

y(t) is the measured output vector

Thus the problem of system identification is the determination of the system model from records of u(t) and y(t).

What distinguishes a system from a model is that system basically is an ordered collection of objects which in some sense is goal oriented. What

constitutes a system depends upon the view point of the system designer. For example an amplifier may itself be considered a system for a certain application. This amplifier can form a part of the system forming a feed back control for a plant. Furthermore, this feedback control may be a part of a chemical process comprising of many such loops.

A model may be defined as the representation of important aspects of a system, which presents knowledge of that system in a usable manner. A model should not be so complicated that it cannot be understood and thus unsuitable for predicting the behavior of the system under a certain set of conditions. Here it would be pertinent to mention the famous "Butterfly effect" which states that when a butterfly in Tokyo moves its wings the disturbance caused by them affects the weather in Newyork. Thus an exact model for forecasting Newyork's weather should account for this effect but this is virtually impossible. On the other hand a model should not be so superficial that important aspects of the system are not taken care of and thereby generating results about the system behavior which are grossly inaccurate.

A basic problem in system identification is choice of nature of the model.

Practically most systems are nonlinear with distributed parameters, however, such systems are approximated by linear models because of simplicity. In most of the cases incremental linear models can very well explain behavior of the system they are modeling. While using such models the user must be aware of the limits within which the models yield valid results. Following are some of the problems in system identification:

a) Determining order of the linear model

b) Selection of a suitable criterion for determining accuracy of the model

c) Designing an input signal that will maximize the accuracy of estimates of model parameters.

Most of the systems are continuous-time type, however, the application of digital computers for identification requires use of discrete-time models. Since this research requires implementation on digital computer, discrete-time models will be used. The determination of parameters of discrete-time models is much easier. Provided that the sampling interval fulfils certain conditions, the determination of continuous-time models from discrete-time models is quite straightforward.

Large number of applications require an on-line identification of the system

rather than off-line identification. In case of off-line identification a large amount of input and output data about the system is collected. This data may be stored in a computer or recorded in some manner, which is then processed in a batch to estimate model parameters. The accuracy of estimates for off-line identification can be made fairly high because there is a greater flexibility in selecting computational methods without any restriction on computing time. Also, in off-line identification one may often select the type of input most suitable.

In a number of control applications, especially adaptive control, the system has to be identified in a fairly short time and on-line identification methods are used in such cases. For an identification scheme to qualify to be on-line type it should fulfill following conditions:

a) It does not require a special input

b) All the data need not be stored

c) A recursive algorithm is used for adjusting the estimates of the parameter after each sampling instant

d) The amount of computation required for model adjustment is a fraction of the sampling period.

Therefore it is understood that mostly on-line methods will not lead to as accurate models as possible with off-line models, which can use a much larger amount of data as compared to on-line models. However, in many practical situations one cannot afford to wait for the time required to collect all the data.

A large variety of system identification methods both off-line and on-line have been applied. The methods can be classified in many ways and one classification scheme is as follows:

1. Classical Methods
    a) Frequency Response Identification
    b) Impulse Response Identification
    c) Step Response Identification
    d) Identification from correlation functions
2. Equation-error Approach
    a) Least-squares
    b) Generalized least squares
    c) Maximum likelihood
    d) Minimum variance
3. Gradient Methods

a) Model Adjustment Techniques

b) Least-squares (recursive)

c) Generalized least squares (recursive)

d) Instrumental variables

e) Bootstrap

f) Maximum likelihood (recursive)

g) Correlation (recursive)

h) Stochastic approximation

Some of the above mentioned techniques will be used in this research to address the system identification problem for performing the load forecasting for dynamic economic dispatch of the WAPDA (Water And Power Development Authority) system.

Chapter 2

Literature Survey

Short term load forecasting (STFL) is necessary for the efficient and economical operation of a power supply utility. The one-hour to 168-hour load forecast is needed for dispatch, unit commitment, energy sale/purchase decision and load management. The less than one-hour forecasts are needed for the system stability and dynamic economic dispatch A variety of techniques exist for short term load forecasting and a survey of the literature available on this topic will be mentioned in this chapter. The volume of literature on the subject is so much that to achieve a comprehensive and complete survey is beyond the scope of this research. However, an attempt has been made to, at least briefly, give an introduction to the literature published on the subject.

K. Y. Lee et al [1] have developed a composite load model for 1-24 hours ahead prediction of hourly electric load. The load model is composed of

three components: the nominal load, the type load and the residual load. Nominal load is modeled such that the Kalman filter can be used and the parameters of the model are adopted by the exponentially weighted recursive least squares method. Type load component is extracted for weekend load prediction and updated by an exponential smoothing method. Residual load is predicted by an auto-regressive model and parameters of the model are estimated using the recursive least squares method.

Alex D. Papalexopoulos and Timothy C. Hesterberg [2] have described a linear regression based model for calculation of short term system load forecasts. The model's distinguishing characters are; innovative model building, including accurate holiday modeling by using binary variables, temperature modeling by using heating and cooling degree functions, robust parameter estimation and parameter estimation using heteroskedasticity by using weighted least squares linear regression techniques, the use of reverse errors-in-variables techniques to mitigate the effects on load forecasts of potential errors in the explanatory variables and distinction between time-independent daily peak load forecasts and the maximum of the hourly load forecasts in order to prevent peak load forecasts from being negatively

biased. The impact of these issues on the accuracy of a model's results has been established through testing of an existing load forecasting algorithm.

I. Komprej and P. Zunko [3] have applied the exponential smoothing method in conjunction with the Box-Jenkins approach to time series analysis methods for STLF. The application of both methods to load forecasting has been shown to produce good results for a public electric utility in Slovenia.

Takeshi Haida and Shoichi Muto [4] have presented a regression based daily peak load forecasting method with a transformation technique. In order to forecast the load precisely through an year, one should consider seasonal load change, annual load growth and the latest daily load change. To deal with these characteristics in the load forecasting, a transformation technique has been presented. This technique consists of a transformation function with translation and reflection methods. The transformation function is estimated with the previous year's data points, in order that the function converts the data points into a set of new data points with preservation of the shape of temperature-load relationships in the previous year. Then, function is slightly translated so that transformed data points will fit the shape of temperature-load relationships in the year. Finally, multivariate regression

analysis, with the latest daily loads and weather observations, estimates the forecasting model. Large forecasting errors caused by the weather-load nonlinear characteristic have been reduced using this technique.

I. Erkmen and T. Adanir [5] have also suggested an STLF algorithm for dynamic economic dispatch. The algorithm is based on using a finite auto-regressive (AR) time series process in conjunction with a Kalman filter. It predicts load value with five-minute intervals up to one hour into the future, with sufficient accuracy to be used for targeting dynamic economic dispatch.

J. D. McDonald and J.Y. Fan [6] have presented a practical real-time implementation of weather adaptive STLF for distribution power utilities. The implementation has been accomplished by utilizing a comprehensive load forecasting model consisting of time series, nonlinear load-weather functions and a residual load function represented by an auto-regressive moving average (ARMA) model. Model parameters have been estimated and updated online using the weighted recursive least squares (WRLS) algorithm. A variable-forgetting factor (VFF) technique has been incorporated in the WRLS algorithm for improved model tracking and numerical performance in real-time operation.

G.A.N. Mbamalu and M.E. El-Hawary [7] have proposed sub-optimal least squares or iteratively reweighted least-squares (IRWLS) procedures for estimating the parameters of a seasonal multiplicative auto-regressive (AR) model encountered during power system load forecasting. The method involves using an interactive computer environment to estimate the parameters of a seasonal multiplicative AR process. The method comprises five major computational steps. The first determines order of the seasonal multiplicative AR process, and the second uses least squares or IRWLS to estimate the optimal non-seasonal AR model parameters. In the third step one obtains the intermediate series by back forecast, which is followed by using least squares or IRWLS to estimate the optimal seasonal AR parameters. The final step uses estimated parameters to forecast future load.

Artificial neural networks (ANN) have recently received considerable attention regarding STLF and a number of publications concerning ANN-based STLF methods have been added to the literature available on the subject. D. Srinivasan, A. C. Liew and J. S. P. Chen [8] have demonstrated how an ANN can be used for this purpose. The network used is based on non-statistical neural paradigm back propagation, which is found to be effective for accurate load forecasting. The advantage of this technique is

that the network produces immediate decision with minimal computation for given input data. Performance of the proposed network has been compared to that by some traditional methods.

Hsu Yuan-Yih and Yang Chien-Chuen [9] have proposed a new approach using ANN for STLF. To forecast the hourly loads of a day, the hourly load pattern and the peak and valley loads of the day must be determined. First, a neural network based on self-organizing feature, which maps to identify those days with similar hourly load patterns, is developed. These days with similar load patterns are said to be of the same day type. The load pattern of the day under study is obtained by averaging the load patterns of several days in the past, which are of the same day type as the given day. The short term load forecasting of the Taiwan Power Company has demonstrated the effectiveness of the proposed neural network.

A. H. Noureddine, A. T. Alouani and A. Chanderasekaran [10], T.Matsumoto et al [11], A. D. Papalexopoulos, S. Hao and T. M. Peng [12], Y. Shimakura et al [13], K. Y. Lee et al [14], Y. Mizukami and T. Nishimori [15] and D. Park et al [16] have proposed various approaches for solving the problem of short term load forecasting using Artificial Neural Networks.

Time series approaches, as conventional algorithms in the short-term load forecasting of electric power systems yield considerably high accuracy under circumstances of no abrupt disturbances such as sharp changes of weather and special holiday, or, if the results of these algorithms are manually adjusted by experienced operators. Thus expert systems used in conjunction with conventional time series method can yield good results. Chen Dong, Chen Bingxin and Li Tonghao [17] have proposed an expert system for STLF. The time series approach is combined with heuristic inference and a composition method is developed to forecast the hourly load up to 48 hours in advance. The influence factors such as weather, holidays, season and dispatch control are considered. The structure of expert system, programmed in PROLOG language, has also been described. This expert system is tested by actual power system load data of 210 days.

A. U. Asar, J. R. McDonald and M. I. Khan [18] and W. Rattray, J. R. McDonald and A. U. Asar [19] have discussed the prospects for applying a combined solution using artificial neural networks and expert systems to the STLF problem. The proposed method called as the expert network may provide a better solution to the forecasting problem than either system alone can provide.

G. Shrestha and T. T. Lie [20] have presented a hybrid forecasting technique that utilizes the attractive features of both the statistical and expert system based methods. The priority vector based load forecasting technique uses pair wise comparisons to extract relationships from pre-sorted historical hourly load and weather records for up to two years. The pre-sorting is done to identify seasonal boundaries and to categorize the day types (weekdays, weekends, holidays, etc.). The technique is adaptive in the sense that it internally generates the coefficients for relative influence of relevant variables (i.e., weather parameters) on the load. As these relationships change over time, such coefficients are automatically updated. The technique is extended by investigating qualitative use of continuous variable (e.g. temperature) as a means to overcome some of the difficulties confronted during the implementation when record high or record low values of these variables are encountered. This technique has been applied to forecast the hourly loads for a week, in summer when record high temperatures were observed, using 168-hour lead-time. Results obtained by implementing the technique by using temperature as both continuous variable and qualitative variable, using the same set of historical data from a utility company are presented. Most forecast errors are below 5% and many

of the large errors are reduced by qualitative treatment of the variable temperature.

2.1 Conclusion

Load forecasting is being done using auto-regressive modeling, time series analysis and in some cases utilizing Kalman filter algorithm. In this perspective, the approach being adopted in this research is also based upon time series analysis. The emphasis is on modeling the interconnected power system in state space form, applying estimation theory algorithms such as Least Squares and Kalman filtering for short term forecasting.

Chapter 3

Off-Line System Identification Methods

3.1 Introduction

In real world the processes are not deterministic but, because of their probabilistic nature, are stochastic. This means that outcome of a specific instance of a process cannot be predicted because of lack of sufficient apriori knowledge of the process. However, statistics of a phenomenon can be determined and they do not change while the actual process may change. Thus probable outcome of a process can be predicted. This is done by identifying the system and then applying estimation theory algorithms to predict system output. The estimation theory algorithms may be off-line or on-line (as explained in chapter 1). In this chapter we'll mention some off-line methods of system identification.

One off-line system identification method is known as Least Squares which was proposed by Karl Friedrich Gauss at the end of eighteenth century. This

method to determine the orbits of planets and asteroids. Gauss stated that the unknown parameters of the system should be estimated in such a way that the sum of squares of the differences between the actually observed values and the computed values is a minimum [21].

Let there is a time function $y(t)$ such that

$$y(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \ldots \theta_n x_n(t) \quad (3.1)$$

If measurements are taken at discrete intervals t=1,2,3…m the following equations are obtained

$$y(1) = \theta_1 x_1(1) + \theta_2 x_2(1) + \ldots \theta_n x_n(1)$$

$$y(2) = \theta_1 x_1(2) + \theta_2 x_2(2) + \ldots \theta_n x_n(2)$$

.

.

.

$$y(m) = \theta_1 x_1(m) + \theta_2 x_2(m) + \ldots \theta_n x_n(m)$$

The above equations can be written in the matrix form as follows

$$\underline{Y} = X\underline{\theta}$$

Where

$$\underline{Y}=\begin{bmatrix} y_1 \\ y_2 \\ : \\ . \\ y_m \end{bmatrix}$$

$$X=\begin{bmatrix} x_1(1) & x_2(1) & \ldots & x_n(1) \\ x_1(2) & x_2(2) & \ldots & x_n(3) \\ : & : & : & : \\ . & . & . & . \\ x_1(m) & x_2(m) & \ldots & x_n(m) \end{bmatrix}$$

And

$$\underline{\theta}=\begin{bmatrix}\theta_1 & \theta_2 & \ldots & \theta_n\end{bmatrix}^T$$

If n = m

$$\underline{\theta}=X^{-1}\underline{Y} \qquad (3.2)$$

However, the equation can be applied only if 'X' is a non-singular, square matrix. Actually m > n and hence X is not a square matrix.

3.2 Ordinary least squares theory

In order to find the best solution for $\underline{\theta}$,in equation (3.2), if m > n least squares method is used which is explained as follows

Define $\underline{e} \triangleq [e(1)\ e(2)\ \ldots\ldots e(m)]^T$

where $\underline{e}$ is known as the equation error

Let $\underline{e} = \underline{Y} - X\underline{\theta}$ (3.3)

In least squares theory $\hat{\underline{\theta}}$ (estimate of $\underline{\theta}$) is chosen such that a criterion 'J' is minimized. J is the square of error and is given by

$$J = \sum_{i=0}^{m} \underline{e}^{2}(i)$$

$$= e^{T} e$$

$$= (\underline{Y} - X\underline{\theta})^{T} (\underline{Y} - X\underline{\theta})$$

$$= (\underline{Y}^{T} - \underline{\theta}^{T} X^{T}) (\underline{Y} - X\underline{\theta})$$

$$= \underline{Y}^{T}\underline{Y} - \underline{Y}^{T} X\underline{\theta} - \underline{\theta}^{T} X^{T}\underline{Y} + \underline{\theta}^{T} X^{T} X\underline{\theta} \quad (3.4)$$

J is minimized if

$$\partial J / \partial \theta = 0$$

Taking partial derivatives of both sides

$$-\partial J / \partial \theta = 0 - \underline{Y}^{T} X - X^{T}\underline{Y} + 2X^{T} X\underline{\theta} = 0$$

Thus following normal equation is obtained

$$X^{T}\underline{Y} = X^{T} X\underline{\theta}$$

$$\underline{\theta} = (X^{T} X)^{-1} X^{T}\underline{Y}$$

The term $(X^{T} X)^{-1} X^{T}$ is called pseudo inverse of the matrix X. Thus inverse of a rectangular matrix is taken by its pseudo inverse. This method of getting the estimated value of $\underline{\theta}$ is called Ordinary Least Squares (OLS). Results of an ordinary least squares algorithm applied on random data sequences having

different number of data points are shown in Fig 3.1 to Fig 3.4. It can be seen that as number of data points increases the forecasted value converges to actual value, thus least squares is a consistent estimator.

3.3 Generalized least squares

This method is more powerful than ordinary least squares in that it can provide more accurate estimates when random disturbances present in the system. Application of the ordinary least squares method in such conditions would yield biased estimates. This is because of the fact that a least squares estimator weighs all errors equally which corresponds to the assumption that all measurements have the same precision. Different weighting of the errors can be accounted for by using a weight matrix which is a diagonal matrix with weights in the diagonal.

In ordinary least squares method the loss function J was given by:

$J = \underline{e}^T \underline{e}$

In weighted least squares

$J_w = \underline{e}^T W \underline{e}$ where W is the weight matrix

$= (\underline{Y} - X\underline{\theta})^T W(\underline{Y} - X\underline{\theta})$

it can be shown that

$$\underline{\hat{\theta}}_W = (X^T WX)^{-1} X^T W\underline{Y} \qquad (3.5)$$

3.4 Statistical properties of generalized least squares

3.4.1 Unbiasedness

$\underline{\hat{\theta}}$ is said to be unbiased if

$$E\{\underline{\hat{\theta}}\} = \underline{\theta}$$

Where E stands for expectation.Thus estimate is said to be unbiased if the expectation of the estimated value is equal to the true value.

From equation (3.5)

$$\underline{\hat{\theta}}_W = (X^T WX)^{-1} X^T W\underline{Y}$$

Taking expectation of both sides

$$E\{\underline{\hat{\theta}}_W\} = E\{(X^T WX)^{-1} X^T W\underline{Y}\}$$

Assume $\underline{Y} = X\underline{\theta} + \underline{e}$ where $\underline{e}$ is the measurement noise and is assumed to be White Gaussian Noise (WGN).

$$E\{\underline{\hat{\theta}}_W\} = E\{(X^T WX)^{-1} X^T WX\underline{\theta} + (X^T WX)^{-1} X^T W\underline{e}\}$$

But $(X^T WX)^{-1} X^T WX = I$ where 'I' denotes the identity matrix, thus

$$E\{\underline{\hat{\theta}}_W\} = E\{\underline{\theta}\} + E\{(X^T WX)^{-1} X^T W\underline{e}\}$$

$$E\{\underline{\hat{\theta}}_W\} = \underline{\theta} + (X^T WX)^{-1} X^T W E\{\underline{e}\}$$

Thus least squares theory will give unbiased result if the noise is actually measurement noise and is not due to our improper modeling.

3.4.2 Variance

The variance of the estimator in effect measures the width of the probability density. A small value of variance suggests that the probability density is concentrated around its mean value, which, if the estimator is also unbiased, will be the true value of the parameter.

When $E\{\underline{\hat{\theta}}_W\} = \underline{\theta}$

Define the covariance ψ_w as follows:

$$\psi_w \underline{\underline{\Delta}} E\{(\underline{\hat{\theta}}_W - \underline{\theta})(\underline{\hat{\theta}}_W - \underline{\theta})^T$$

But $\underline{\hat{\theta}}_W = (X^T WX)^{-1} X^T W\underline{Y} = (X^T WX)^{-1} X^T W (X\underline{\theta} + \underline{e})$

$$= (X^T WX)^{-1} X^T WX\underline{\theta} + (X^T WX)^{-1} X^T W\,\underline{e}$$

$$= \underline{\theta} + (X^T WX)^{-1} X^T W\,\underline{e}$$

Thus

$$(\underline{\hat{\theta}}_W - \underline{\theta}) = (X^T WX)^{-1} X^T W\,\underline{e}$$

Because $\psi_w \underline{\underline{\Delta}} E\{(\underline{\hat{\theta}}_W - \underline{\theta})(\underline{\hat{\theta}}_W - \underline{\theta})^T\}$

$$\psi_w = E\{(X^T WX)^{-1} X^T W\,\underline{e}\,.\,\underline{e}^T WX (X^T WX)^{-1}\}$$

If 'X' is deterministic

$\psi_w = (X^T WX)^{-1} X^T W\ E\{ \underline{e} . \underline{e}^T \} WX (X^T WX)^{-1}$

Thus the covariance depends only on expected covariance of error.

Say $R = E\{ \underline{e} . \underline{e}^T \}$

If somehow $W=R^{-1}$

$\psi_w = (X^T R^{-1} X)^{-1} X^T R^{-1}\ E\{ \underline{e} . \underline{e}^T \} R^{-1} X (X^T R^{-1} X)^{-1}$

$= (X^T R^{-1} X)^{-1} = \psi_{MV}$

Then $\underline{\theta}_{MV}$ is the minimum variance estimator and is also called best linear unbiased estimator (BLUE).

3.4.3 Consistency

Estimator is said to be consistent if the estimated value converges to the true value as the number of observations increases towards infinity.

Let $R = \sigma^2 \mathrm{I}$

$\Psi = (X^T \frac{1}{\sigma^2 \mathrm{I}} X)^{-1}$

$\Psi = \sigma^2 (X^T X)^{-1}$

Dividing and multiplying by 'm' where 'm' is the number of rows in matrix 'X'

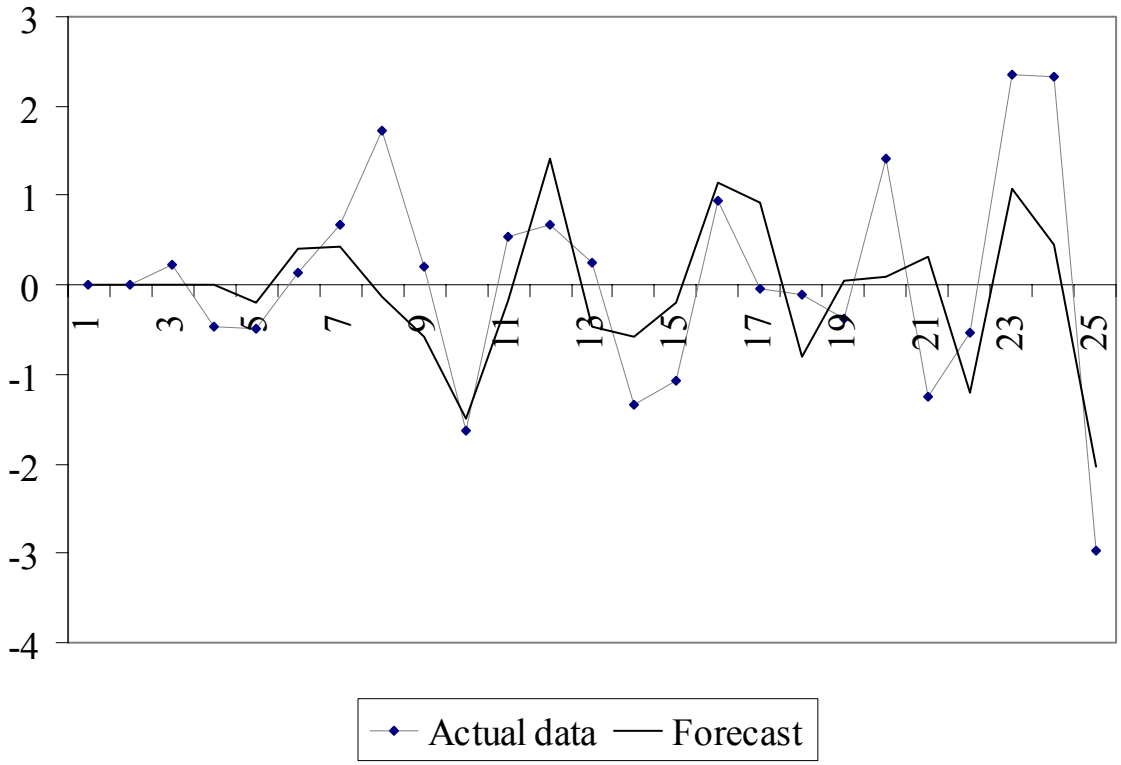
$$\Psi = \frac{\sigma^2}{m} \left(\frac{1}{m} X^T X \right)^{-1}$$

As $m \to \infty$, $\frac{1}{m} X^T X = \Gamma$

Or $\Psi = \frac{\sigma^2}{m} (\Gamma)^{-1} \to \underline{\theta}$

$\underset{m \to \infty}{\text{Limit}} \underline{\hat{\theta}} = \underline{\theta}$

Or estimator is no more a random variable but a constant. Thus when there are infinite data points the variable acquires a constant value and is known as a consistent estimator.

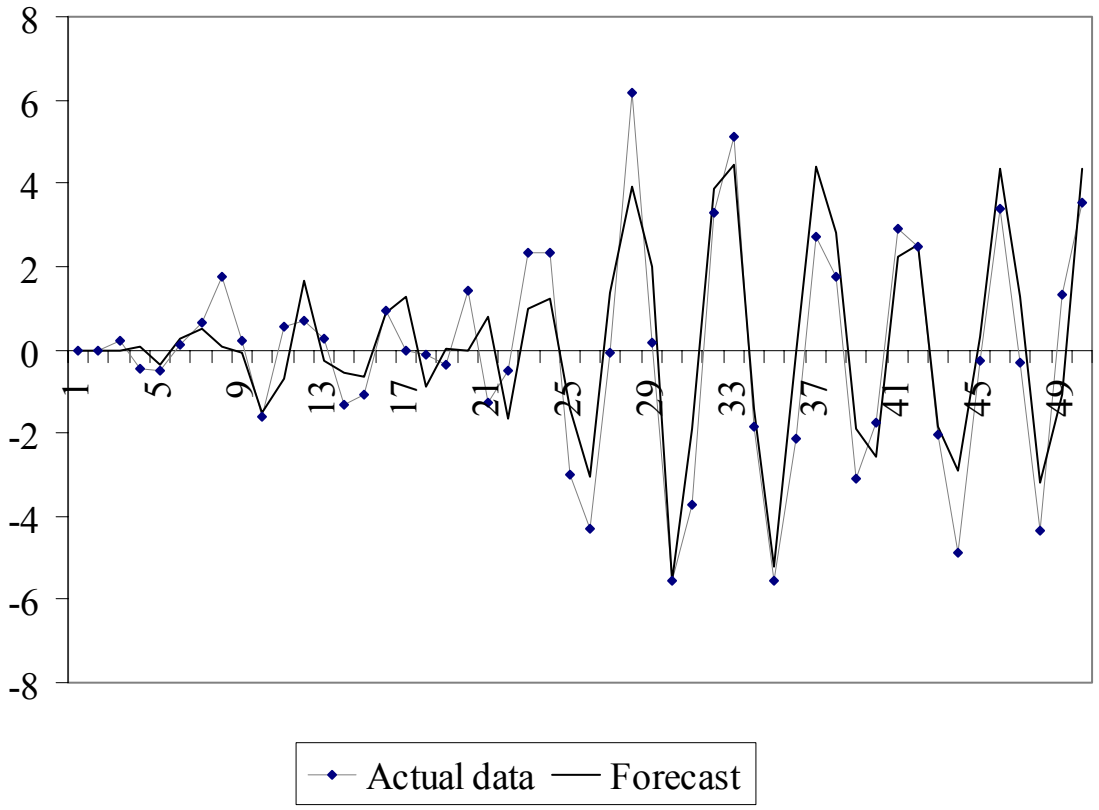


Standard Deviation = 1

Mean = 0

Number Of Data points = 25

Fig 3.1 Graph of actual sequence having 25 data points and its forecast

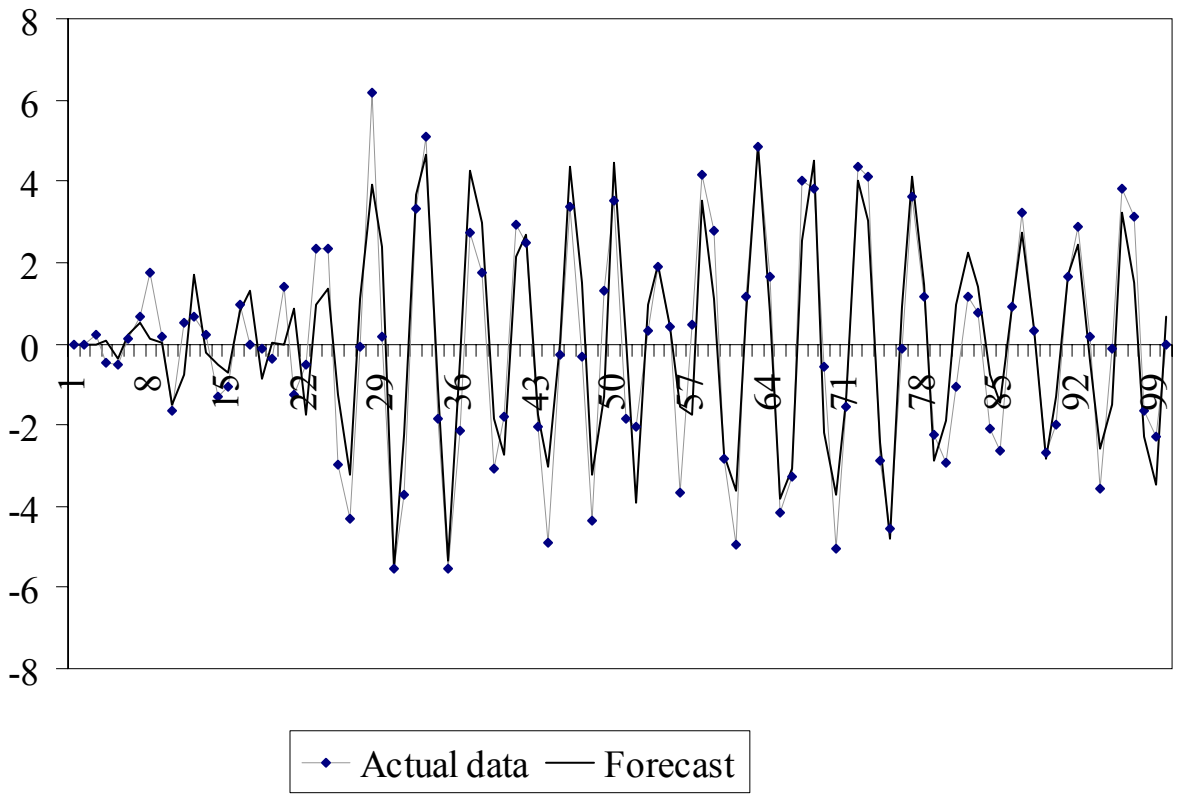


Standard Deviation = 1

Mean = 0

Number Of Data points = 50

Fig 3.2 Graph of actual sequence having 50 data points and its forecast

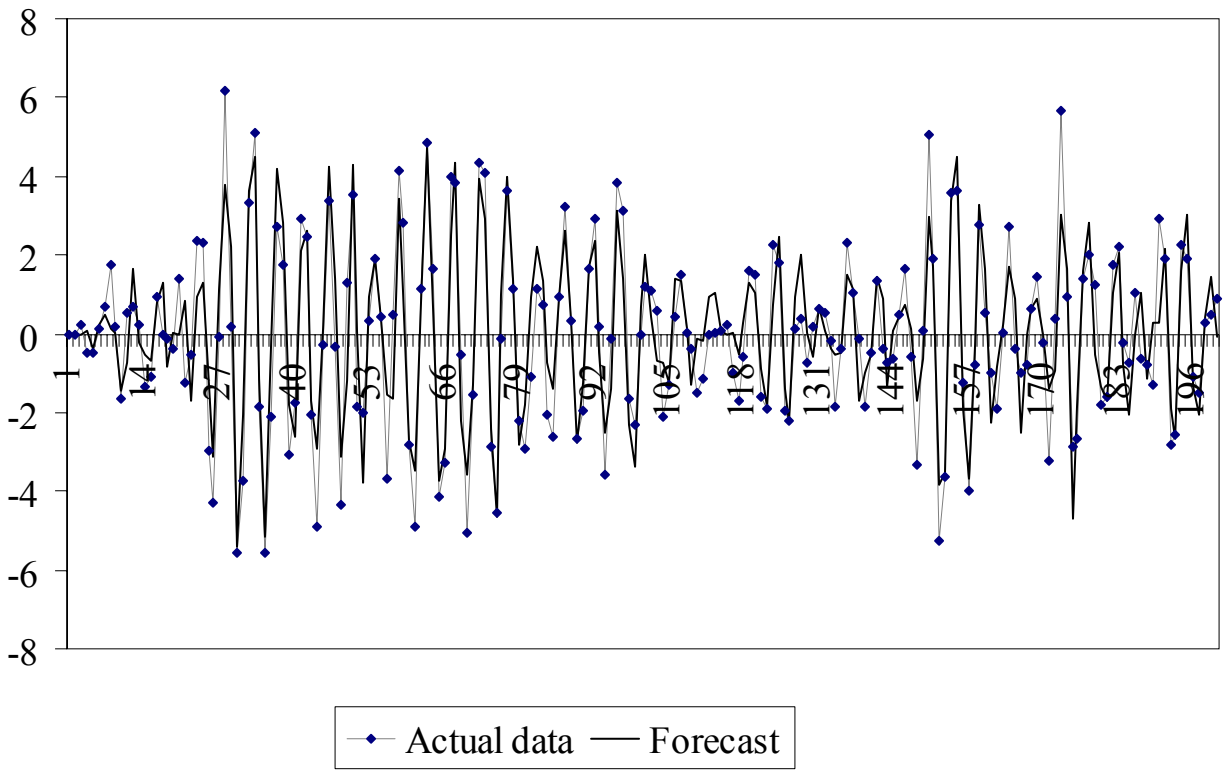


Standard Deviation = 1

Mean = 0

Number Of Data points = 100

Fig 3.3 Graph of actual sequence having 100 data points and its forecast



Standard Deviation = 1

Mean = 0

Number Of Data points = 200

Fig 3.4 Graph of actual sequence having 200 data points and its forecast

Chapter 4

On-Line System Identification Methods

The least squares method of system identification discussed in chapter 3 is an off-line method and requires a lot of computational effort because it involves matrix inversion. The parameter estimation in this case requires storage of all the previous data. Although the results so generated are quite accurate, but its off-line nature renders this method unsuitable for real time applications. In this chapter derivation of a recursive algorithm for least squares estimation, that requires no matrix inversion, has been carried out. This method is known as Recursive least squares or On-line least squares. Another name used for this method is sequential least squares. Later in the chapter, another on-line system identification method known as Stochastic approximation has been mentioned.

4.1 Recursive least squares

Denoting $\underline{Y}_m$ and X_m for a vector and matrix involving 'm' equations. Thus

least square estimator $\underline{\theta}$ is given by

$$\underline{\hat{\theta}}(m) = \left[X_m^T X\right]^{-1} X_m^T \underline{Y_m}$$

Which is derived from the following equation

$$\underline{Y}_m = X_m \ \underline{\theta}$$

Suppose the present state is obtained by another equation

$$y\ (m+1) = \theta_1 x_1(m+1) + \theta_2 x_2(m+1) + \ldots \theta_n x_n(m+1) \text{ -}$$

Or

$$y\ (m+1) = \underline{x}^T(m+1)\ \ \underline{\theta}$$

Where

$$\underline{x}^T(m+1) = \left[x_1(m+1) \quad x_2(m+1) \ldots x_n(m+1)\right]$$

and

$$\underline{\theta}^T = \left[\theta_1 \quad \theta_2 \quad \ldots \quad \theta_n\right]$$

Thus

$$\underline{\hat{\theta}}(m+1) = \left[X_{m+1}^T X_{m+1}\right]^{-1} X_{m+1}^T \underline{Y}_{m+1}$$

Where

$$X_{m+1} = \begin{bmatrix} X_m \\ \ldots\ldots\ldots\ldots \\ \underline{x}^T(m+1) \end{bmatrix} \text{ and } \underline{Y}_{m+1} = \begin{bmatrix} \underline{Y}_m \\ \ldots\ldots\ldots\ldots\ldots \\ y(m+1) \end{bmatrix}$$

Hence one new line has been added to 'X' matrix and '$\underline{Y}$' vector because now one more equation has been taken at point m + 1.

Denote $P(m) = \left[X_m^T X_m\right]^{-1}$

And $P(m+1) = \left[X_{m+1}^T X_{m+1}\right]^{-1}$

$$\Rightarrow P(m+1) = \left[\begin{Bmatrix} X_m^T & \vdots & \underline{x}(m+1) \end{Bmatrix} \begin{Bmatrix} X_m \\ \cdots \\ \underline{x}^T(m+1) \end{Bmatrix}\right]^{-1}$$

$$P(m+1) = \left[X_m^T X_m + \underline{x}(m+1)\ \underline{x}^T(m+1)\right]^{-1} \tag{4.1}$$

If A,C and $[A+BCD]$ are non singular, square matrices then applying the matrix inversion lemma it can be proved that the following identity holds

$$\left[A + BCD\right]^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1} DA^{-1}$$

Putting

$$A = X_m^T X_m$$

$$B = \underline{x}(m+1)$$

$C = I \rightarrow$(i.e. identity matrix)

and $D = \underline{x}^T(m+1)$ in equation (4.1) we get LHS of the equation for matrix inversion lemma. Therefore

$$P(m+1) = \left[X_m^T X_m\right]^{-1} - \left[X_m^T X_m\right]^{-1} \underline{x}(m+1)\left[I + \underline{x}^T(m+1)\left(X_m^T X_m\right)^{-1} \underline{x}(m+1)\right]^{-1} \underline{x}^T(m+1)\left[X_m^T X_m\right]^{-1}$$

$$p(m+1) = P(m) - P(m)\underline{x}(m+1)\left[I + \underline{x}^T(m+1)P(m)\underline{x}(m+1)\right]^{-1} \underline{x}^T(m+1)P(m)$$

Since $\underline{x}^T(m+1)P(m)\underline{x}(m+1)$ is a scalar quantity we replace 'I' by 1. Also we

know that $A^{-1} = \frac{1}{A}$

$$P(m+1) = P(m) - \frac{P(m)\underline{x}(m+1)\underline{x}^T(m+1)P(m)}{\left[1 + \underline{x}^T(m+1)P(m)\underline{x}(m+1)\right]} \tag{4.2}$$

Now

$$\underline{\hat{\theta}}(m+1) = \left[X_{m+1}^T X_{m+1}\right]^{-1} X_{m+1}^T \underline{Y}_{m+1}$$

$$\underline{\hat{\theta}}(m+1) = P(m+1)\, X_{m+1}^T \underline{Y}_{m+1}$$

$$\underline{\hat{\theta}}(m+1) = P(m+1) \left\{ \left[X_m^T \quad \vdots \quad \underline{x}(m+1) \right] \begin{bmatrix} \underline{Y}_m \\ \cdots \\ y(m+1) \end{bmatrix} \right\}$$

$$\underline{\hat{\theta}}(m+1) = P(m+1) \left\{ X_m^T \underline{Y}_m + \underline{x}(m+1)\, y(m+1) \right\}$$

$$\underline{\hat{\theta}}(m+1) = P(m+1)\, X_m^T \underline{Y}_m + P(m+1)\, \underline{x}(m+1)\, y(m+1)$$

Substituting the value of $P(m+1)$ from equation (4.2)

$$\underline{\hat{\theta}}(m+1) = P(m)\, X_m^T \underline{Y}_m - \frac{P(m)\, \underline{x}(m+1)\underline{x}^T(m+1)\, P(m) X_m^T \underline{Y}_m}{1 + \underline{x}^T(m+1)P(m)\, \underline{x}(m+1)} +$$

$$P(m)\underline{x}(m+1)y(m+1) - \frac{P(m)\, \underline{x}(m+1)\underline{x}^T(m+1)\, P(m)\underline{x}(m+1)y(m+1)}{1 + \underline{x}^T(m+1)P(m)\, \underline{x}(m+1)}$$

But $\underline{\hat{\theta}}(m) = P(m)\, X_m^T \underline{Y}_m$

So

$$\underline{\hat{\theta}}(m+1) = \underline{\hat{\theta}}(m) - \frac{P(m)\, \underline{x}(m+1)\underline{x}^T(m+1)\, \underline{\hat{\theta}}(m)}{1 + \underline{x}^T(m+1)P(m)\, \underline{x}(m+1)} +$$

$$P(m)\underline{x}(m+1)y(m+1)\left[1 - \frac{\underline{x}^T(m+1)\, P(m)\underline{x}(m+1)}{1 + \underline{x}^T(m+1)P(m)\, \underline{x}(m+1)}\right]$$

Order of $y(m+1)$ can be changed because it is a scalar.

$$\underline{\hat{\theta}}(m+1) = \underline{\hat{\theta}}(m) - \frac{P(m)\ \underline{x}(m+1)\underline{x}^T(m+1)\ \underline{\hat{\theta}}(m)}{1+\underline{x}^T(m+1)P(m)\ \underline{x}(m+1)} +$$

$$\frac{P(m)\underline{x}(m+1)y(m+1)}{1+\underline{x}^T(m+1)P(m)\ \underline{x}(m+1)}\left[1+\underline{x}^T(m+1)P(m)\ \underline{x}(m+1) - \underline{x}^T(m+1)P(m)\ \underline{x}(m+1)\right]$$

$$\Rightarrow \underline{\hat{\theta}}(m+1) = \underline{\hat{\theta}}(m) + \frac{P(m)\underline{x}(m+1)}{1+\underline{x}^T(m+1)P(m)\ \underline{x}(m+1)}\left[y(m+1) - \underline{x}^T(m+1)\ \underline{\hat{\theta}}(m)\right] \qquad (4.3)$$

Thus, the new estimate of the parameter vector is given by its previous estimate plus an error term. Equation (4.3) has a strong intuitive appeal. The new estimate $\underline{\hat{\theta}}(m+1)$ is obtained by adding a correction term to the previous estimate. The correction term is proportional to $y(m+1) - \underline{x}^T(m+1)\ \underline{\hat{\theta}}(m)$, where the term $\underline{x}^T(m+1)\ \underline{\hat{\theta}}(m)$ can be interpreted as the value of '*y*' at time $m+1$ predicted by the model of equation (4.3). The correction term is therefore, proportional to the difference between the measured value of the output $y(m+1)$ and the prediction of $y(m+1)$ based on the previous parameter estimate.

Equations (4.2) and (4.3) together form the recursive least squares method of parameter estimation. Thus, matrix inversion requirement of the ordinary least squares method has been eliminated. Although, one matrix inversion is

necessary to compute *P(m)* at start. However, for practical purposes this inversion can be avoided by putting *P(m)* equal to the identity matrix multiplied by a large constant.

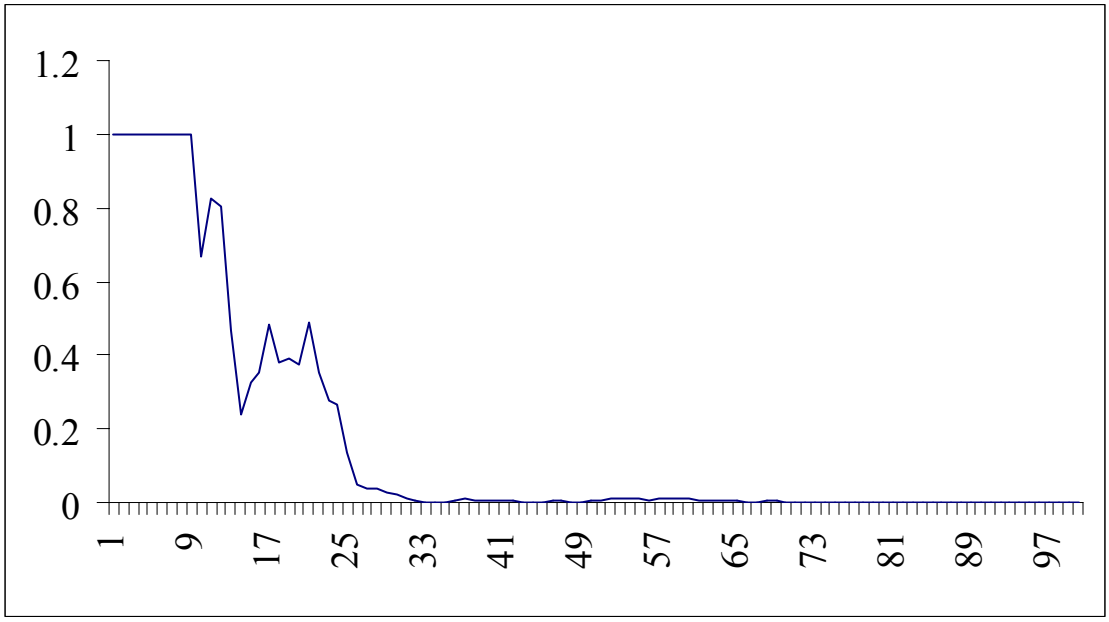
To summarize

$$P(m) = \left[X_m^T X_m\right]^{-1}$$

$$P(m+1) = P(m) - \frac{P(m)\underline{x}(m+1)\underline{x}^T(m+1)P(m)}{\left[1+\underline{x}^T(m+1)P(m)\underline{x}(m+1)\right]}$$

$$\hat{\underline{\theta}}(m+1) = \hat{\underline{\theta}}(m) + \frac{P(m)\ \underline{x}(m+1)}{1+\underline{x}^T(m+1)P(m)\ \underline{x}(m+1)}\left[y(m+1) - \underline{x}^T(m+1)\ \hat{\underline{\theta}}(m)\right]$$

Results of a recursive least squares algorithm applied on random data sequences are shown in Fig 4.2 to Fig 4.4. These figures show the normalized value of forecasting error against each data point for different values of mean and standard deviation of the sequence and also the mean and standard deviation of the noise. Different values of *P(m)* have been taken to see the effect.



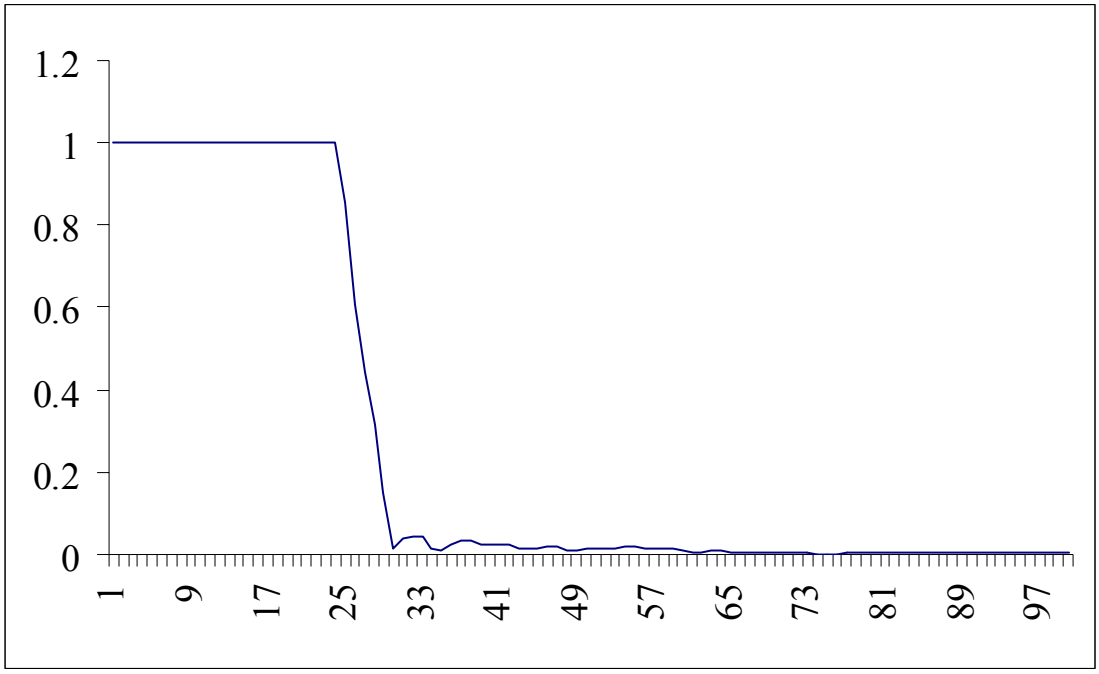
Standard Deviation = 1

Mean = 0

Number of data points = 100

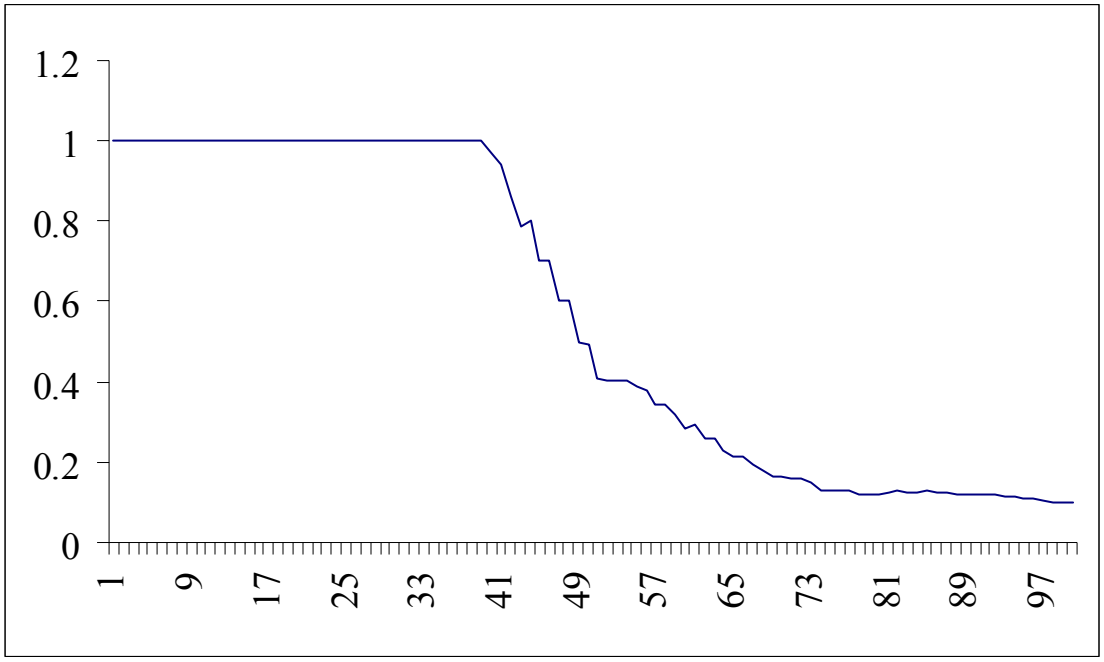
Constant multiplying identity matrix to form *P(m)* = 50

Fig 4.1 Normalized error curve for multiplying constant equal to 50



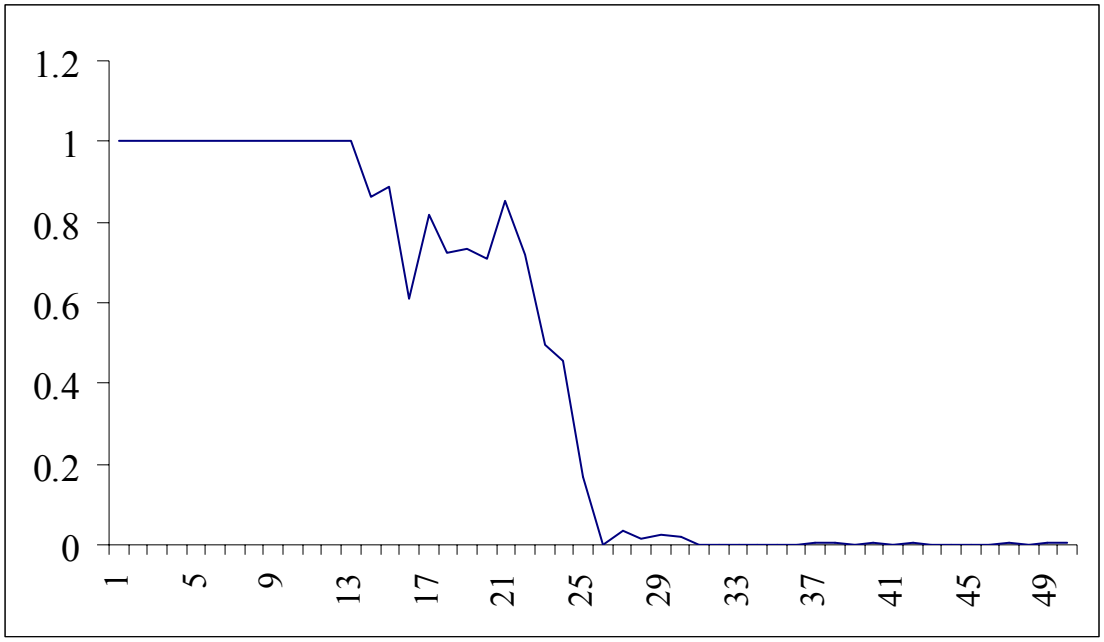
Standard Deviation	= 1
Mean	= 0
Number of data points	= 100
Constant multiplying identity matrix to form *P(m)*	= 500

Fig 4.2 Normalized error curve for multiplying constant equal to 500



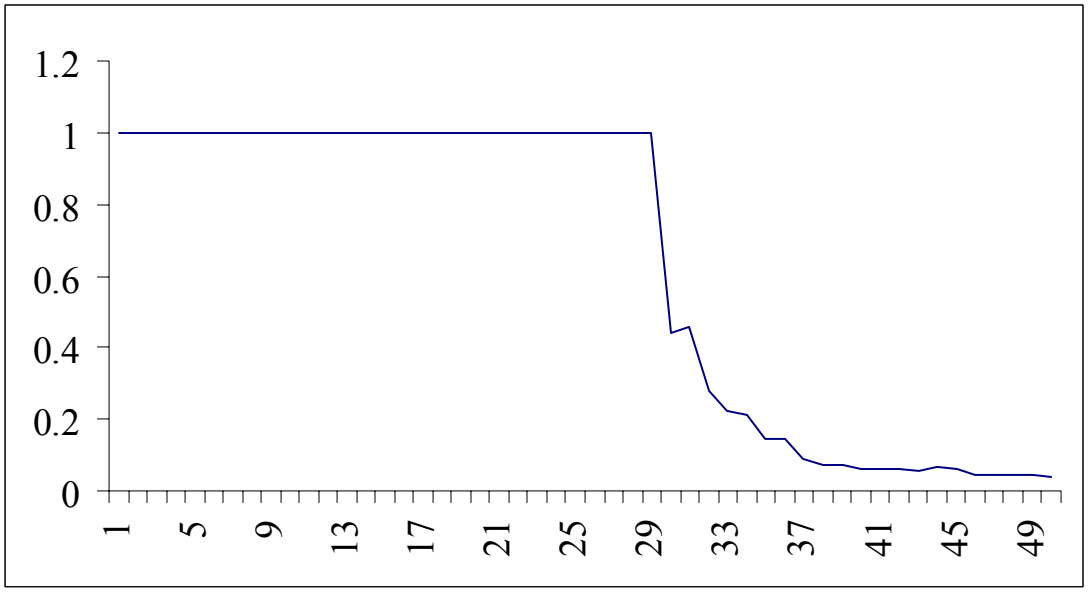
Standard Deviation	= 1
Mean	= 0
Number of data points	= 100
Constant multiplying identity matrix to form *P(m)*	= 5000

Fig 4.3 Normalized error curve for multiplying constant equal to 5000



Standard Deviation	= 1
Mean	= 0
Mean of noise	= 0.01
Standard deviation of noise	= 1
Number of data points	= 50
Constant multiplying identity matrix to form *P(m)*	= 50

Fig 4.4 Normalized error curve for data having noise with mean =0.01



Standard Deviation	= 1
Mean	= 0
Mean of noise	= 1
Standard deviation of noise	= 1
Number of data points	= 50
Constant multiplying identity matrix to form *P(m)*	= 50

Fig 4.5 Normalized error curve for data having noise with mean=1

4.1.1 Observations

Fig 4.1, Fig 4.2 and Fig 4.3 show the normalized error curve for a data sequence having 100 data points, mean equal to zero and standard deviation equal to 1. However, for the first figure *P(m)* was formed by multiplying the identity matrix by 50, the second figure by 500 and the third one by 5000. It can be seen that for the first case, the normalized error (usually called as error norm) converged to zero after about thirtieth data point. In the second case although the curve touched zero at about thirty sixth data point, complete convergence was achieved beyond sixtieth data point. For the third case, it can be seen that the curve will converge even after the hundredth data point. We have already mentioned that instead of taking the inverse of matrix *P(m),* we can put it equal to identity matrix multiplied by a large constant but Fig 4.1 to Fig 4.3 show that for a very large constant the rate of convergence will be slowed down.

Fig 4.4 shows that convergence can also be achieved using recursive least squares when the data is contaminated with noise. However from Fig 4.5 it can be seen that the normalized error has converged to zero even after the fiftieth data point when the noise has a high value of mean.

4.2 Stochastic Approximation

This method of parameter estimation is an on-line one and can be applied to any problem which can be formulated as some form of regression in which repeated observations are made. Its main characteristic is the enormous simplicity of its implementation. It may be defined as a scheme for successive approximation of a sought quantity when observations involve random errors due to stochastic nature of the problem.

Robbins and Monro [22], Kiefer and Wolfowitz [23] and Dvoretzky [24] have made important contributions to the area of stochastic approximations. Robins and Monro first introduced the method. They developed an algorithm which is the statistical equivalent of gradient method for finding unique roots of equation:

$h(x) = 0$

which is

$$x_{i+1} = x_i - K_i\, h(x)$$

where K_i is a sequence of random numbers satisfying certain conditions to ensure convergence and the conditions are:

1) $\lim_{i \to \infty} K_i = 0$

2) $\sum_{i=1}^{\infty} K_i = \infty$

3) $\sum_{i=1}^{\infty} K_i^2 < \infty$

Kiefer and Wolfowitz extended the method proposed by Robbins and Monro to include finding the extremum of an unknown unimodel regression function θ(u). Dvoretzky presented a generalized algorithm which is as follows:

$$x_{n+1} = x_n + \gamma_{n+1}\{ f(r_1, r_2, \ldots\ldots\ldots, r_{n+1}) - x_n \}$$

Where

x_n = n th estimate of x

γ_{n+1} = gain sequence

r_n = n th observation

$f(r_1, r_2, \ldots\ldots\ldots, r_{n+1})$ = scalar functional of observations $r_1, r_2, \ldots\ldots\ldots, r_{n+1}$

Another algorithm proposed by Kwatny is as follows:

$$\underline{\hat{\theta}}_{k+1} = \underline{\hat{\theta}}_k - \frac{\gamma}{k+1} \frac{\left(h_{k+1}^T \underline{\hat{\theta}}_k - y_{k+1} \right) \underline{h}_k^T}{\left\| \underline{\hat{h}}_k \right\|^2}$$

Where

$\underline{\hat{\theta}}_k$ = kth estimate of $\underline{\theta}$

γ = a positive constant

h =information vector

y =observed output

The algorithm used by Kwatny has been used in this research for parameter estimation.

Chapter 5

Kalman Filter

5.1 Introduction

In previous chapters we have been talking about single input, single output systems. Actually we may have systems having multiple inputs and outputs. Such systems are known as multivariable systems. The models used for multivariable systems are different from those used for single variable systems. Following are the models commonly used for multivariable systems:

1) The State-Space model
2) The Transfer Function Matrix representation
3) The Impulse Response Matrix representation
4) The Input Output Difference Equation model

Out of the above mentioned, most commonly used form is the State-Space form in which a dynamical system is described by a set of variables called

states. The state contains all necessary information about behavior of the system such that, given the present and future values of input, we can find future state and output of the system. Kalman Filter is one of the methods that can be used to find estimates of states given the system in state-space form. The application of Kalman Filter theory provides a set of difference equations, the solution of which can be found recursively. The updated estimate is computed from previous estimate and new input data, as such, only the previous estimate is needed to be stored instead of entire past data. Thus Kalman Filter is ideally suited for implementation on a digital computer.

Suppose a multivariable system is modeled by the following stochastic state-space equations:

$$x(n+1) = Ax(n) + Bu(n) + v_1(n) \tag{5.1}$$

$$y(n) = Cx(n) + v_2(n) \tag{5.2}$$

Where

$v_1(n) \rightarrow$ is the process noise

$v_2(n) \rightarrow$ is the observation noise

$x(n) \rightarrow$ is the n dimensional state vector

A, B and C $\rightarrow$ are constant matrices

$v_1(n)$ and $v_2(n)$ are assumed to be zero mean, stationary, white noise processes. These processes have changed the above system from deterministic to stochastic. Covariance is given by:

$$E\left\{\begin{bmatrix} v_1(n) \\ v_2(n) \end{bmatrix}\begin{bmatrix} v_1(k)^T & v_2(k)^T \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\delta(n-k)$$

$\delta(n-k)$ is known as Kronecker Delta and it exists when $n = k$ and is zero otherwise i. e.

$\delta(n-k) = 1$ when $n = k$

$\delta(n-k) = 0$ when $n \neq k$

5.2 Kalman Filter Theorem

Consider the stochastic signal model given by equations (5.1) and (5.2) and assume that the initial noise and state sequences are jointly Gaussian.

Let $\hat{x}(n+1/n)$ or simply written as $\hat{x}(n+1)$ denote best estimate of $x(n)$, based on $\{y(0), y(1), ..., y(n)\}$ observations, called as conditional mean of $x(n+1)$. Then $\hat{x}(n+1)$ satisfies the following recursion (the Kalman Filter):

$$\hat{x}(n+1) = A\hat{x}(n) + K(n)\left[y(n) - C\hat{x}(n)\right] + Bu(n)$$

Where $K(n)$ is the filter gain given by

$$K(n) = \left[A\,\Sigma(n)\,C^T\right]\left[C\,\Sigma(n)\,C^T + R\right]^{-1}$$

$\Sigma(n) \rightarrow$ is the state error covariance vector given by

$$\Sigma(n) \underline{\underline{\Delta}} E\left\{\left[\hat{x}(n)-x(n)\right]\left[\hat{x}(n)-x(n)\right]^T\right\}$$

$\Sigma(n)$ satisfies the following Riccati Difference Equation(RDE):

$$\Sigma(n+1) = A\Sigma(n)A^T + Q - K(n)\left[C\Sigma(n)C^T + R\right]K(n)^T$$

where $\Sigma(0) = \Sigma_0$

In case of Kalman Filter it is assumed that C, A, S and R are known.

5.3 Properties of Kalman Filter

Following are the main properties of Kalman Filter:

1) In the case of Gaussian noise, $\hat{x}(n)$ is the conditional mean of $x(n)$,that is,

$$\hat{x}(n) = E\left[x(n)/\mathrm{y}(n-1)\right]$$

Where

$$\mathrm{y}(n-1) = \left[y(n-1), y(n-2),.....y(n_0)\right]$$

2) A consequence of property (1) is that

$$E\left[\hat{x}(n)-x(n)/\ \mathrm{y}(n-1)\right] = 0 \quad \text{and}$$

$$\Sigma(n) \underline{\underline{\Delta}} E\left\{\left[\hat{x}(n)-x(n)\right]\left[\hat{x}(n)-x(n)\right]^T\right\} \le \Sigma_F$$

Where Σ_F is the error covariance given by any other filter.

Therefore, for Gaussian noise, Kalman Filter is the best estimator.

3) It can be seen that the Kalman gain $K(n)$ and the conditional error

Covariance $\Sigma(n)$ are independent of $y(n-1)$ provided that A, C, R, Q and S are all independent of $y(n-1)$. In this case $K(n)$ and $\Sigma(n)$ are pre-computable $\Sigma(n)$ is also the unconditional error covariance.

4) If the Gaussian assumption is removed, Kalman Filter becomes the Linear Minimum Variance estimator of $x(n)$.

In this research it has been assumed that both the process noise and observation noise are Gaussian. Therefore, Kalman Filter forms the best estimator in this case and that's why it has been used for state estimation.

Chapter 6

Load Forecasting

6.1 Introduction

Load forecasting is a necessary part of a power system utility. Long term load forecasts of five years to twenty years into the future are needed for construction of new power generation plants, framing of regulatory policies and decision-making regarding power purchase agreements. Intermediate term forecasts from a few months to five years are used for maintenance scheduling and negotiations for setting up power purchase prices. Short term load forecasts of 24 hours to 168 hours are required for economic generation plans, unit commitment decisions and short term maintenance scheduling. Very short term load forecasts of up to one hour into the future are important for the real time control and reliable operation of the power system and for dynamic economic dispatch.

Various methods that have been used in the past for short term prediction of

load. The methods include multiple regression analysis, exponential smoothing, spectral decomposition, time series analysis using Box-Jenkins method and the combined parameter identification and state estimation. The combined parameter and state estimation method has received a lot of attention in the past [25-26]. Initially, the problem was solved using the extended Kalman filter. However the extended Kalman filter requires excessive computations and is prone to divergence. Hence, several approaches have been adopted in the past to avoid these problems [25-26].

El-Sherief proposed a computationally simple algorithm to solve the problem of combined state estimation and parameter identification of linear multivariable stochastic systems [25]. The algorithm starts by transforming the state space equations of the system into a canonical innovations representation. A pseudo parameter measurement equation is derived which contains the inputs, outputs and states of the system. Then assuming an initial estimate of the states a normalized stochastic approximation algorithm is used to find estimates of the parameters of pseudo parameter measurement equation. These parameters are estimated in ‘m’ separate steps from ‘m’ pseudo parameter measurement equations where ‘m’ is the number of outputs. Afterwards the system matrices and the Kalman gain matrix are

determined directly and these are then used for state estimation of the system by a linear Kalman filter. The procedure is continued in a bootstrap manner. This algorithm uses a normalized stochastic approximation criterion and hence works well in noisy environments and avoids any complicated operations.

6.2 Combined parameter and state estimation [27]

Assume there is a linear, time-invariant, discrete time, multivariable stochastic system given by the following state-space equations:

$$x^*(k+1) = A^* x^*(k) + B^* u(k) + D^* w(k)$$

$$y(k) = C^* x^*(k) + v(k)$$

Where

$x^*(k) \rightarrow n$ dimensional state vector

$u(k) \rightarrow p$ dimensional input vector

$w(k) \rightarrow r$ dimensional state noise vector

$v(k) \rightarrow m$ dimensional output noise vector

$y(k) \rightarrow m$ dimensional measured output

Applying innovations theory and assuming the system is completely observable and controllable it can be transformed to the following system:

$$x(k+1) = A\,x(k) + B\,u(k) + K\,e(k)$$

$$y(k) = C\,x(k) + e(k) \tag{6.1}$$

where the matrices A and C are in the following canonical form

$$A_{ii} = \begin{bmatrix} 0 & & & \\ : & I(n_i - 1) & & \\ . & & & \\ a_{ii}(1) & a_{ii}(2) & \ldots & a_{ii}(n_{ij}) \end{bmatrix}$$

$$A_{ij} = \begin{bmatrix} 0 & & 0 & \\ : & & & \\ . & & & \\ a_{ij}(1) & a_{ij}(2) & \ldots & a_{ij}(n_{ij}) \end{bmatrix}$$

And

$$C = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ : & \ldots & 1 & 0 \\ 0 & \ldots & 0 & 1 \end{bmatrix}$$

n_{ij}'s are the structural parameters of the system.

Defining

$$B = \begin{bmatrix} b^1 \\ b^2 \\ : \\ b^n \end{bmatrix}$$

$$K = \begin{bmatrix} k^1 \\ k^2 \\ : \\ k^n \end{bmatrix}$$

And

$$A = \begin{bmatrix} a^1 \\ a^2 \\ : \\ a^n \end{bmatrix}$$

Where b^j, k^j and a^j represent the j th row of the system matrices B, K and A. Hence the j th subsystem of system (6.1) can be represented by the following equations:

$$y_j(k+n_j) = \underline{h}_j(k+n_j-1)\underline{\theta}_j + e_j(k+n_j) \qquad j = 1, 2, ..., m \tag{6.2}$$

$$\underline{h}_j(k+n_j-1) = [\hat{x}(k)\, u(k) ... u(k+n_j-1)\, e(k)\, e(k+1) ..... e(k+n_j-1)] \tag{6.3}$$

$$\underline{\theta}_j = [a^j\, b^{n_1+.....+n_j} ... b^{n_1+....+n_{j-1}+1}\, k^{n_1+....+n_j} .... k^{n_1+....+n_{j-1}+1}] \tag{6.4}$$

The following stochastic approximation algorithm can be used for recursively estimating parameters of equation (6.2):

$$\hat{\underline{\theta}}_j(k+1) = \hat{\underline{\theta}}_j(k) + \frac{\nu(k) o(k) \underline{\underline{h}}_j(k+n_j-1)}{\left\| \underline{\underline{h}}_j(k+n_j-1) \right\|^2} \tag{6.5}$$

Where

$$o(k) = y_j(k+n_j-1) - \underline{\underline{h}}_j(k+n_j-1)\hat{\underline{\theta}}_j(k) \quad j = 1,2,....., m$$

and $\nu(k)$ is a sequence satisfying Dvoretzsky's [24] conditions and the innovation will be estimated by :

$$\hat{e}_j(k+n_j) = y_j(k+n_j) - \hat{\underline{h}}(k+n_j-1)\hat{\underline{\theta}}_j(k+1)$$

First we assume an estimate of the state, then the system parameters are

estimated using the stochastic approximation algorithm (6.5). Hence an estimate of the system matrices $\hat{A}(k), \hat{B}(k)$ and $\hat{K}(k)$ can be obtained. After having this estimate the expression for $\hat{x}(k)$ in equation (6.3) can be replaced by $\hat{x}(k/k)$ to get a new estimate of the parameter vector $\underline{\theta}_j$. The procedure is repeated between the two stages in a bootstrap manner.

6.3 Load forecasting [27]

For load forecasting we consider an interconnected power system having 'n'loading nodes. Let

$$x_1(k), x_2(k), ..., x_n(k) \qquad k = 0,1,2,...$$

denote loads at various nodes at instant k. We assume that the load at any loading node at the time instant $k+1$ depends upon loads at all other loading nodes at the instant k. Thus the mathematical model for this system will be an AUTO-REGRESSIVE one of the following form:

$$x_i(k+1) = \sum_{j=1}^{n} a_{ij}(k+1/k)\, x_j(k)$$

Then the entire interconnected power system can be represented by the following stochastic state- space model:

$$\underline{x}(k+1) = A(k+1/k)\, \underline{x}(k) + w(k)$$
$$\underline{y}(k+1) = C\underline{x}(k) + v(k)$$

Where

$\underline{x}(k)$ $\rightarrow n$ dimensional state vector representing loads at individual stations

$A(k+1/k) \rightarrow$ model matrix

$\underline{y}(k)$ $\rightarrow m$ dimensional vector representing data from a few major loading nodes

$w(k)$ $\rightarrow n$ dimensional model noise vector

$v(k)$ $\rightarrow n$ dimensional measurement noise vector

Both noise sequences are assumed Gaussian, zero mean and uncorrelated with each other. We know that when two Gaussian, zero mean, uncorrelated sequences are added they result in another Gaussian sequence with zero mean. From this and applying the innovation approach the above state-space model with two noise sequences can be represented by equivalent single white noise source $e(k)$ as shown below:

$$\underline{\hat{x}}(k+1/k) = A(k+1/k)\,\underline{\hat{x}}(k/k-1) + Ke(k)$$

$$\underline{y}(k) = C\underline{\hat{x}}(k/k-1) + e(k) \qquad (6.6)$$

Where

$\underline{\hat{x}}(k/k-1) \rightarrow$ best estimate of $\underline{x}(k)$ based on $\{y(0), y(1), ..., y(k-1)\}$ observations

$K \rightarrow$ optimum steady state filter gain of the system

$e(k) \rightarrow$ innovations sequence of system

From equation (6.6)

$$\underline{y}(k+1) = CA(k+1/k)\,\underline{\hat{x}}(k/k-1) + CKe(k) + e(k+1) \tag{6.7}$$

If measurements from all nodes are available then $m = n$ and $c^i = u^i$ where c^i denotes ith row of matrix C and u^i denotes ith unit row vector. We can write the ith row of equation (6.7) as follows:

$$\underline{y}_i(k+1) = \underline{a}^i(k+1/k)\,\underline{\hat{x}}(k/k-1) + k^i e(k) + e_i(k+1) \tag{6.8}$$

Where a^i and k^i are ith rows of matrices A and K respectively. Augmenting the two vectors $\underline{a}^i(k+1/k)$ and $\underline{k}^i$ into a new vector $\underline{\theta}_i(k+1)$ equation (6.8) can be written as:

$$\underline{y}_i(k+1) = \underline{h}_i^T(k)\,\underline{\theta}_i(k) + e_i(k+1)$$

Where

$$\underline{h}_i^T(k) \underline{\underline{\Delta}} \left[\underline{\hat{x}}^T(k/k-1) \quad e^T(k)\right] \tag{6.9}$$

and $\underline{\theta}_i^T \underline{\underline{\Delta}} \left[\underline{a}^i(k+1/k) \quad k^i\right]$

Define $\underline{\theta}_i^T$ as the estimate of unknown parameter vector of equation (6.5). It is shown in [28] that the parameter vector can be estimated by normalized stochastic approximation algorithm of equation (6.5).

It is clear from equations (6.8) and (6.9) that the residual error is not correlated with the forcing function $\underline{h}_i(k),$ and it can be proved that the estimate $\underline{\hat{\theta}}_i(k)$ is unbiased and will converge to the true value in the mean square sense with probability one.

The unknown states of vector $\underline{h}_i(k)$ equation (6.9) are estimated in a bootstrap manner. Initial states are assumed and the parameter vector estimated. The estimated parameter vector $\underline{\hat{\theta}}_i(k)$ gives estimated model matrix $\hat{A}(k)$ and the estimated optimum filter gain matrix $\hat{K}(k).$

Using Kalman filtering and prediction theory we get the following equations:

$$\underline{\hat{y}}(k+1/k) = C\,\underline{\hat{x}}(k+1/k)$$

$$\underline{\hat{x}}(k+1/k) = \hat{A}(k)\,\underline{\hat{x}}(k/k)$$

Where $\underline{\hat{y}}(k+1/k)$ is the predicted load at the instant $k+1$ based on the observations $y(0), y(1), y(2),...,y(k)$

the vector $\underline{\hat{x}}(k/k)$ is a minimum variance filtered estimate of $\underline{x}(k)$ which can be obtained as follows:

$$\hat{x}(k/k) = \underline{\hat{x}}(k/k-1) + \hat{K}(k)[\text{y(k)} - \text{C}\underline{\hat{\text{x}}}(\text{k/k} - 1)]$$

Having obtained the estimate of state, next estimate of parameter vector $\underline{\theta}_i$ is estimated and this continues in a bootstrap manner.

6.4 Application and evaluation of the proposed algorithm for a real problem

The short term load forecasting algorithm explained in section 6.3 was applied to real data of hourly electricity generation, during the first week of May 2000, at WAPDA's power houses of Tarbella, Mangla and Warsak. The algorithm was applied with initial condition 0.1 for the parameter vector and initial states were assumed equal to the initial data. Results are shown in Fig (6.1) to Fig (6.4).

Fig 6.1 shows the actual and forecasted demand for the Tarbella power station. It can be seen that after vibrating to and fro about the actual demand, forecast began to track the demand after about thirty-fifth data point. From then onwards the forecast provided by the proposed algorithm is satisfactory.

However, the forecast curve seems to lag behind the actual curve at peak hours.

Fig 6.2 shows the actual demand and forecast for Mangla power station. Again it is clear from the figure that the forecast curve began to track the actual data curve successfully after about thirty-fifth data point. The forecast is fairly accurate except around data point 111 where the actual data curve has dipped below the forecast curve. However, this is not a forecasting error because corresponding to the same data point the actual data curve in figure for Tarbella power station has risen above forecast curve. Therefore the cause of this difference may be shut down of one or more units at Mangla power station, may be due to some fault or some kind of urgent maintenance needed.

Fig 6.3 shows the actual data curve and the forecast curve for Warsak Power station. The forecast curve does not seem to track the actual data curve very well. At some points it gives values higher than the actual data and at others the values are less than the actual data. This is because of the fact that Warsak is a very old hydel power station. The water storage reservoir has been completely filled with silt and as such, it is just run of the river plant

having no storage capacity for water. Therefore, a little flexibility is available in varying the amount of generation. This fact is also clear from actual data curve which is a smooth almost straight curve with very less fluctuations.

Finally Fig. 6.4 provides a comparison of the total actual data of the three power houses and the overall forecast. The forecast has been able to track the actual demand very well. However, at peak loads, the forecast seems to lag a little behind the actual data curve. This may be attributed to the following factors:

1) All the three power stations, whose data was used for load forecasting, are hydel power plants. Although because of silting Warsak is practically a run of the river plant, the other two power stations are basically used for providing irrigation water and electricity is only the by-product. The water reservoirs of these power plants are governed by IRSA (Indus River System Authority). It informs WAPDA, on daily basis, the amount of water that can be released. Accordingly WAPDA prepares its generation plan. In water shortage months, like May, IRSA allows only minimal water discharges. Therefore maximum amount of water is released at peak hours and at other times plants are normally run at

minimum load. Therefore, our forecast curve has lagged behind the actual curve at peak loads.

2) In a power system, electricity is generated at power houses which is then step up using step up transformers. On reaching the load centers this electricity is then step down in grid stations with the help of step down transformers and distributed. Therefore, the actual demand is given by the data of grid stations, which was not available. In Pakistan electric power is supplied by WAPDA through the National Grid (a network of transmission lines) and all power houses and grid stations are inter connected through this. It is clear that the load of power houses is a reflection of the total loads at grid stations. Therefore, data of power houses was used which may be the cause of forecast being slightly different than actual demand.

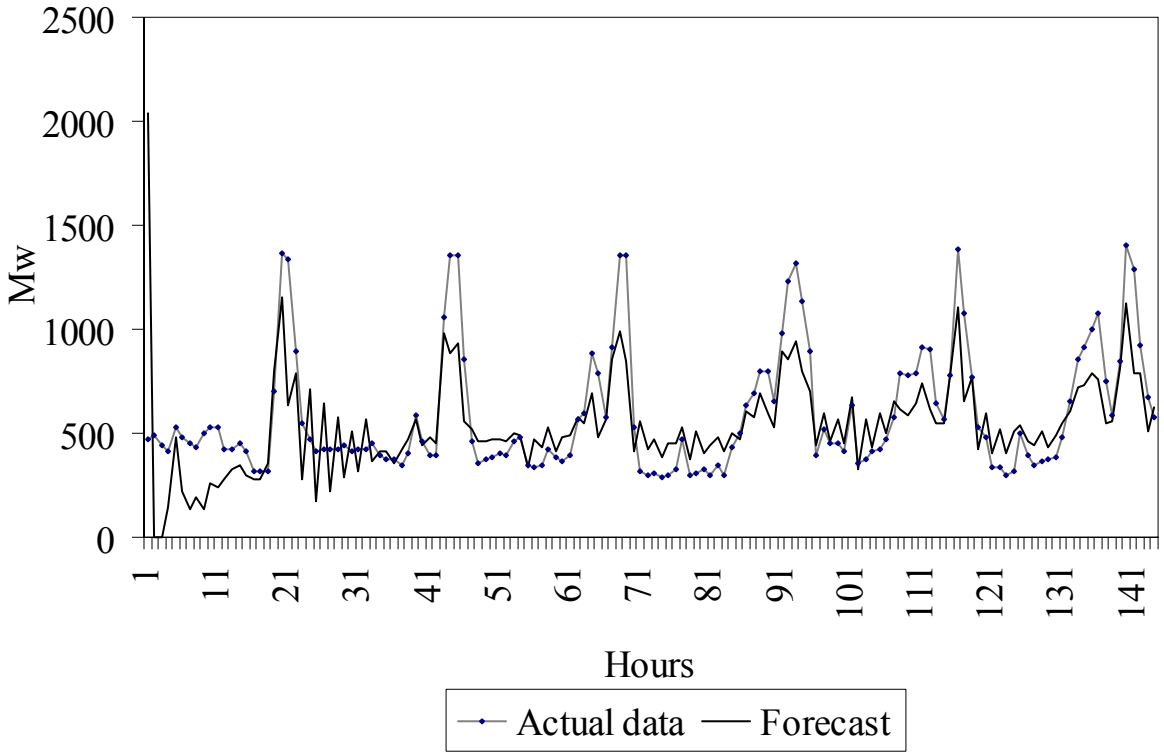


Fig 6.1 Actual and forecasted load of Tarbella power station

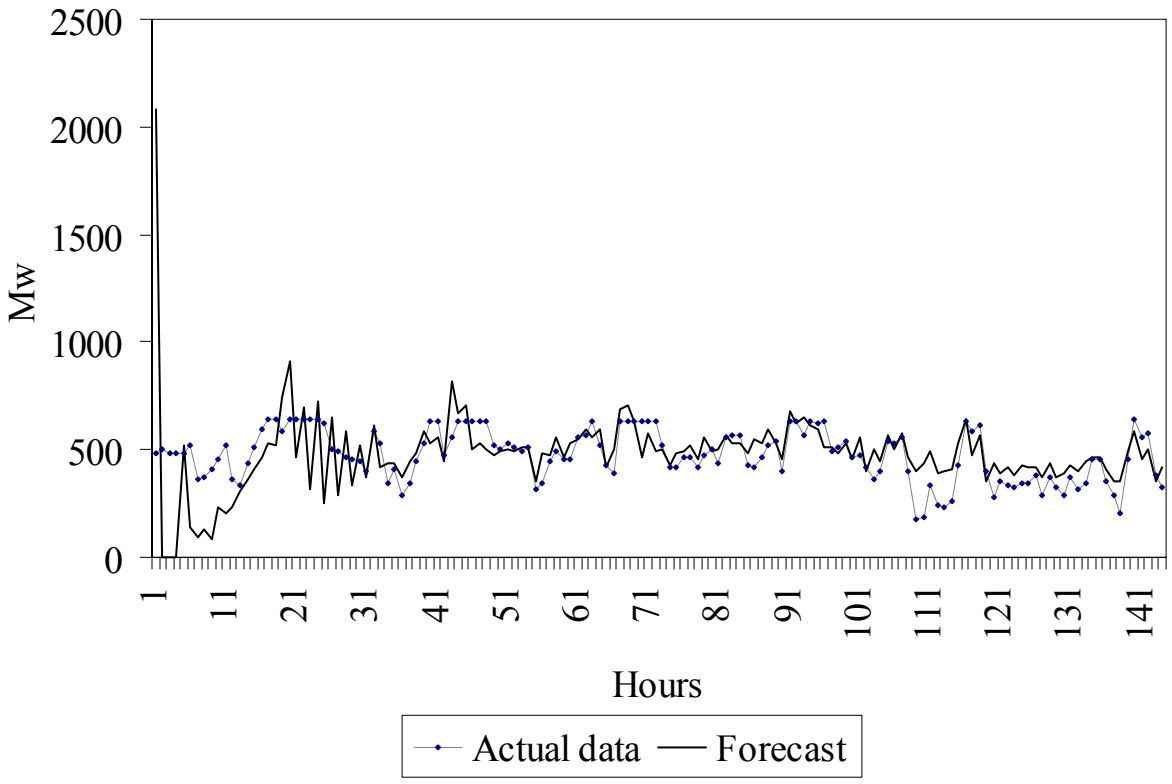


Fig 6.2 Actual and forecasted load of Mangla power station

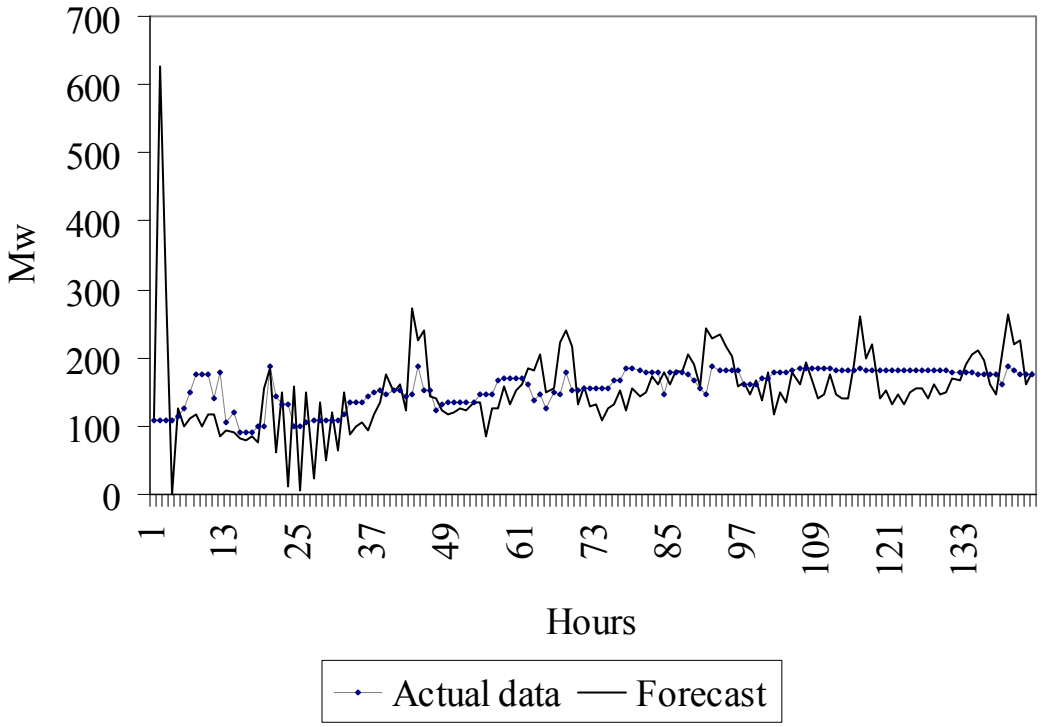


Fig 6.3 Actual and forecasted load of Warsak power station

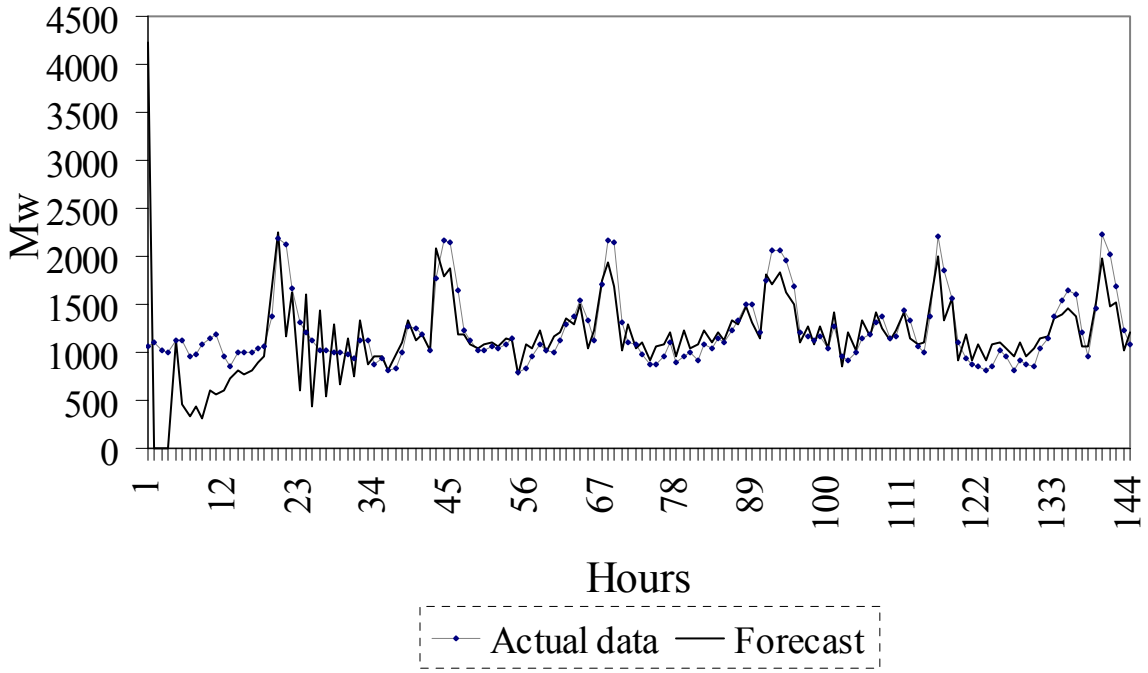


Fig 6.4 Sum of actual and forecasted loads of Tarbella, Mangla and Warsak

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

The existing load forecasting system being followed in WAPDA is not very sophisticated because of unavailability of sufficient data. Hourly load forecast of the next twenty four hours is prepared by taking a weighted average of three quantities i.e. hourly load of last day, last week same day and last year same day. This forecast is then given as input to a computer program, which provides the economic generation plan. However, the persons responsible for controlling the entire plan at the NPCC (National Power Control Center) WAPDA rely mostly upon their own intuition and experience instead of the power plan provided by the computer program, because it is not very accurate. The other problem is that people controlling the power system at NPCC only know total load of the whole system. They do not have any grid station wise break up of this total load which can be very useful for reducing transmission losses. For example, let the load

requirement is at Peshawar and surplus power is available both at Tarbella and Mangla power stations. If the person in charge at NPCC knows that power is required at Peshawar he would prefer to increase the generation at Tarbella, provided incremental cost of additional generation is same at both the power houses. This will reduce transmission losses substantially.

Presently, WAPDA's transmission losses are of the order of about 10%(actually they are reported to be around 8%-9% but for simplicity we have taken them to be 10%). The algorithm developed has been tested using real data and it has generated good results. If applied to hourly load data of grid stations, accurate forecast of individual grid stations' load after an hour can be generated. On the basis of this forecast, it will be possible to formulate most economic load dispatch plan for the next hour which can reduce transmission losses resulting in substantial savings.

Suppose average (daily) load demand = 7000 Mw

Then total generation for twenty four hours = 7000*24 Mwh

= 168,000 Mwh

If we add 10% transmission losses = 186,666 Mwh

So the average transmission losses per day are = 18,666 Mwh

If by using the proposed algorithm we can

decrease losses to 9% then savings = 1866 Mwh per day

= 1,860,000 Kwh per day.

If the average cost of one unit = 2 Rs

Savings in term of rupees = 2*1,860,000 Rs

= 3,720,000 Rs per day

Even if the losses can be restricted to 9.9%

savings are = 372,000 Rs per day.

If the proposed system is adopted, it will require the installation of telemetry equipment at main grid stations of WAPDA. The same equipment can be used for providing not only the present load, but the data regarding weather such as, ambient temperature, as well (which is the most important information required for longer term load forecasts). The hourly data provided by this equipment will currently be used for reducing transmission losses. However, this data can be stored, using a database management system. Such that after an year or two WAPDA will have complete data available with it, on the basis of which, accurate load forecasts of 24 hours to 168 hours into the future can be made. These forecasts can be used for

reducing unit commitment costs, that WAPDA has to pay to the Independent Power Producers (IPP's) like HUBCO, KAPCO etc.

Another benefit that WAPDA will have by installation of telemetry equipment is a lot of savings in labour man-hours. Presently, every hour's load of the grid station is written manually in a log sheet. It will no longer be needed. This will not only save man-hours but will also result in reducing stationary and storage costs plus the data will be available in a more useful form.

7.2 Recommendations for future research

The proposed algorithm may be improved to provide longer term load forecasts that is, 24 hours to 168 hours, which can be useful for reducing unit commitment costs and short term maintenance scheduling of power plants.

References:

[1] K. Y. Lee, J. H. Park, Y. M. Park, "Composite modeling for adaptive short-term load forecasting," *IEEE Transactions on Power Systems,* vol. 6, iss. 2, p. 450-7, May 1991

[2] A. D. Papalexopoulos and T. C. Hesterberg, " A regression-based approach to short-term system load forecasting," *IEEE Transactions on Power Systems*, vol. PWRS-5, 4, p. 1535-1547, 1990

[3] I. Komprej and P. Zunko, " Short term load forecasting," *proceedings of 6th Mediterranean Electrotechnical Conference*, vol.2, p. 1470-3

[4] T. Haida and S. Muto, " Regression based peak load forecasting using a transformation technique," *IEEE Transactions on Power Systems*, vol. 9, iss. 4, p. 1788-94, Nov. 1994

[5] I. Erkmen, T. Adanir, "A short-term load prediction algorithm for dynamic economic dispatch," *Proceedings of 7th Mediterranean Electrotechnical Conference,* vol.3, p. 885-8, April 1994

[6] J. Y. Fan, J. D. McDonald, "A real-time implementation of short-term load forecasting for distribution power systems," *IEEE Transactions on Power Systems,* vol. 9, iss 2, p. 988-94, May 1994

[7] G.A.N. Mbamalu and M.E. El-Hawary, " Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation," IEEE *Transactions on Power Systems,* vol.8, iss.1, p. 343-8, Feb. 1993

[8] D. Srinivasan, A. C. Liew, J. S. P. Chen, ” A novel approach to electrical load forecasting based on a neural network,” presented at 1991 IEEE International Joint Conference on Neural Networks, vol.2, p. 1172-7, 1991

[9] Hsu Yuan-Yih, Yang Chien-Chuen, ” Design of artificial neural networks for short-term load forecasting. I. Self-organizing feature maps for day type identification,” *Proceedings of IEEE conference on Generation, Transmission and Distribution,* vol. 138, iss. 5, p. 407-13, Sept. 1991

[10] A.H. Noureddine, A.T. Alouani, A. Chandrasekaran, ” A new technique for short-term residential electric load forecasting including weather and lifestyle influences,” *Proceedings of the 35th Midwest Symposium on Circuits and Systems,* vol.2, p. 1419-27, Aug. 1992

[11] T. Matsumoto, S. Kitamura, Y. Ueki, T. Matsui, ” Short-term load forecasting by artificial neural networks using individual and collective data of preceding years,” ANNPS 93. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, p. 245-50, April 1993

[12] A. D. Papalexopoulos, S. Hao, T. M. Peng, “Short-term system load forecasting using an artificial neural network,” ANNPS 93. P*roceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, p. 239-44, April 1993

[13] Y. Shimakura, Y. Fujisawa, Y. Maeda, R. Makino, Y. Kishi, M. Ono, J. Y. Fann, N. Fukusima, ” Short-term load forecasting using an artificial neural network,” ANNPS 93. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, p. 233-8, April 1993

[14] K. Y. Lee, T. I. Choi, C. C. Ku, J. H. Park, "Short-term load forecasting using diagonal recurrent neural network," ANNPS 93. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, p. 227-32, April 1993

[15] Y. Mizukami, T. Nishimori, "Maximum electric power demand prediction by neural network," ANNPS 93. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems,* p. 296-301, April 1993

[16] D. Park, O. Mohammed, A. Azeem, R. Merchant, T. Dinh, C. Tong, J. Farah, C. Drake, " Load curve shaping using neural networks," ANNPS 93. *Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, p. 290-5, April 1993

[17] Chen Dong, Chen Bingxin and Li Tonghao, "An expert system for short-term load forecasting," presented at APSCOM-91. 1991 International Conference on Advances in Power System Control, Operation and Management, vol.1, p. 330-4, Nov. 1991

[18] A. U. Asar, J. R. McDonald, M. I. Khan, " A specification of neural network applications in the load forecasting problem," *First IEEE Conference on Control Applications,* vol.1, p. 577-82, Sept. 1992

[19] W. Rattary, J. R. McDonald, A. U. Asar, "Experience with artificial neural network models for short term load forecasting in electrical power systems: a proposed application of expert networks," *Proceedings of Third International Conference on Artificial Neural Networks,* p. 123-7, 1993

[20] G. Shrestha and T.T. Lie, ” Qualitative use of forecast variables in hybrid load forecasting techniques,” presented at APSCOM - 93. 2nd International Conference on Advances in Power System Control, Operation and Management, vol.1, pp. 215-20, 1993

[21] K. J. Astrom and B. Wittenmark, “ Adaptive control,” Addison-Wesley Publishing Company, second edition, 1995

[22] H. Robbins and S. Monro, “A stochastic approximation method,” Annals of Math. Stat., vol. 22, 1951

[23] J. Kiefer and J. Wolfowitz, “Stochastic estimation of the maximum of regression problems,” Annals of Math. Stat., vol.23, 1952

[24] A. Dvoretzky, “On stochastic approximation,” *Proceedings of 3rd Berkeley Symposium on Mathematical Statistics and Probability*, vol.1, University of California Press, 1956

[25] H. El-Sherief and N. K. Sinha, ” Bootstrap estimation of parameters and states of linear multivariable systems,” *IEEE Transactions on Automatic Control*, vol. Ac-24, pp. 340-343, 1979

[26] H. El-Sherief, ” Combined state estimation and parameter identification of linear multivariable stochastic systems, ” *IEEE International Symposium On Circuits and Systems*, pp. 961-965, 1981

[27] M. A. Maud, ” Microprocessor based identification of stochastic multi-variable systems,” M.S. Thesis, EED, UPM, Saudi Arabia, 1981

[28] H. El-Sherief and Y. L. Abdel-Magid, ” An efficient on-line load modeling algorithm for short term load forecasting of interconnected power systems,” Submitted to IEEE PES Winter Meeting, 1982