Prediction of Time Series Data with Linear Modeling

by

Mohammed Owahidur Rahman

A project submitted to the Department of Electrical and Electronic Engineering, BUET, in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL AND ELECTRONIC ENGINEERING

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
DHAKA

MAY, 2000

The project titled **Prediction of Time Series Data with Linear Modeling**
Submitted by **Mohammed Owahidur Rahman** Roll No.: **9506101 P**
Session **1994-95-96** has been accepted as satisfactory in partial fulfillment of the requirements for the degree of **Master of Engineering in Electrical and Electronic Engineering** on **May 17, 2000**.

BOARD OF EXAMINERS

1. Dr. Md. Aynal Haque — Chairman
   Assistant Professor
   Dept. of EEE , BUET , Dhaka

2. Dr. Md. Rezwan Khan — Member
   Professor
   Dept. of EEE , BUET , Dhaka

3. Dr. Md. Kamrul Hasan — Member
   Assistant Professor
   Dept. of EEE , BUET , Dhaka

CANDIDATE'S DECLARATION

It is hereby declared that this project or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Mohammed Owahidur Rahman

Countersigned by:

Dr. Md. Aynal Haque
Assistant Professor
Dept. of EEE, BUET, Dhaka

ACKNOWLEDGEMENTS

The author would like to express his indebtedness and gratitude to his supervisor Dr. Md. Aynal Haque for his endless patience, friendly supervision and invaluable assistance in making a difficult task a pleasant one.

The author wishes to express his thanks and regards to the Head of the Department of Electrical and Electronic Engineering, BUET, Dhaka, for his support during the work.

Sincerest thanks to friends and colleagues for their constant support and criticism of the project work.

CONTENTS

ABSTRACT

Prediction of a time series data plays an important role in determining the underlying physical mechanism generating the data. Linear parametric modeling techniques can be used in data prediction in many diversified fields. In this project linear autoregressive (AR) model is used to predict two types of data, namely, heart rate and electric power load. The heart rate is calculated from the measured ECG of 5 adult human subjects. 5 sets of power load data are collected from central load dispatch center of Bangladesh power development board.

The Burg method is applied to determine the AR model parameters. The model order is determined by the use of Akaike information criterion statistics. For heart rate time series data 1000 data points are used as input, while 500 data points are for predicting power load data. The prediction of data is performed starting from the immediate next data point after model order. For heart rate the data is predicted up to 2000 points and for power load it is up to 1000 points. Analysis of variance technique is used to determine the difference between original and predicted data. The results of prediction show that there is no statistically significant difference between original and predicted data in each data set.

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Historical Background

Linear prediction is an important topic in digital signal processing with many practical applications. Two of the most important applications of linear estimation are power spectral estimation and linear filtering. Many of the phenomena that occur in nature are best characterized statistically in terms of averages. Most signals in nature are characterized as random processes. For example, meteorological phenomena such as the fluctuations in air temperature and pressure are best characterized statistically as random processes. Thermal noise voltages generated in resistors and electronic devices are additional examples of physical signals that are well modeled as random proceses. Due to the random fluctuations in such signals, one must adopt a statistical viewpoint, which deals with the average characteristics of random signals. In particular, the autocorrelation function of a random process is the appropriate statistical average that is used for characterizing random signals in the time domain, and the Fourier transform of the autocorrelation function provides the transform from time domain to frequency domain.

Given a mathematical description of a physical system, one can analyze its overall behavior and predict the response of its output to different inputs. The difficulty, however, lies in determining such a mathematical description. Indeed, although one can derive a model for some simple systems, such as electrical machines, using Newton's laws, there are many systems whose outright complexity makes such a task impossible. In such cases, one has to resort to the numerical techniques of system identification [1]. In attempting to determine empirically the mathematical descriptions associated with linear time invariant systems, linear modeling techniques are used. Linear prediction theory of time series anlysis has a long and rich history of developmant over the last four decades. The classical theory of time series analysis has been well developed over the past two decades, and excellent accounts of this theory are available, for example in [2-6] and in many other books. An important assumption that is made in the classical theory is that the structure of the series can be described by a linear model such as an autoregressive (AR), moving average (MA) or mixed autoregressive moving average (ARMA) model.

The assumption of linearity is often a very dubious one. The theory of Volterra [7] and Wiener [8] on functional series representation has provided great stimulus to the development of non-linear models, but unfortunately Wiener's representation is too general and the statistical estimation of the Wiener kernels is not universal.

In spite of the limitations of linear techniques, AR and ARMA models have been used in the prediction of a time series in a wide variety of applications. AR model has a highly flexible structure that can be used to parameterize the dynamics of a system [9].

1.2 Linear Modeling Techniques

Two broad categories of methods are used in the modeling of time series data, namely, parametric and nonparametric methods. Nonparametric methods make no assumption about how the data are generated, and mainly use Fourier based approach and periodogram. Nonparametric methods are relatively simple, well understood and easy to compute using the FFT algorithm. However, these methods require the availability of long data records in order to obtain necessary frequency resolution required in many applications. Furthermore, these methods suffer from the leakage effects due to windowing that are inherent in finite-length data records. Often, the leakage masks weak signals that are present in the data.

On the other hand, parametric methods of data modeling eliminates the need of window functions. As a consequence, parametric or model based methods avoid the problem of leakage and provide better resolution than do the FFT based nonparametric methods. This is especially true in applications where short data records are available due to time-variant of transient phenomena.

AR and ARMA are the mostly used parametric methods for linear prediction. For a linear time invariant system, the general system function $H(z)$ is given by [10],

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{1}$$

where b_k and a_k are the system coefficients that determine the location of the zeros and poles of $H(z)$, respectively. The parameters p and q are the system orders. For the linear system with the rational system function $H(z)$ given by eqn. (1), the output $y(n)$ is related to the input $x(n)$ by the difference equation,

$$y(n) + \sum_{k=1}^{p} a_k y(n-k) = \sum_{k=0}^{q} b_k x(n-k) \tag{2}$$

The model described by eqn. (2) is known as autoregressive moving average (ARMA) model. This model has both finite number of poles and zeroes. Depending on the system parameters, there are two subclasses of ARMA model, namely, autoregressive (AR) model and moving average (MA) model.

If, in eqn. (1), $b_0 = 1$, $b_k = 0$, $k > 0$, the linear system has $H(z) = 1/A(z)$, and the system is an all-pole system. In this case, the differnce equation for the input-output relashionship is,

$$y(n) + \sum_{k=1}^{p} a_k y(n-k) = x(n) \tag{3}$$

This model is known as AR model.

On the other hand, if in eqn. (1), $a_k = 0$, $k \geq 1$, the linear system has $H(z) = B(z)$, and the system is an all-zero system. In this case, the differnce equation for the input-output relashionship is,

$$y(n) = \sum_{k=0}^{q} b_k x(n-k) \tag{4}$$

and the model is known as MA model.

1.3 Aim of the Project

The objective of this research work is to apply linear modeling techniques in the prediction of time series data. Out of the different techniques, AR method will be used. Two types of time series data namely heart rate and electrical power load will be predicted. The predicted data will be compared with the corresponding original one by statistical analysis and the validity of the model will be tested for each data set.

2.4 Organization of the Dissertation

Several different algorithms have been are available in the determination of AR parameters and model order. Every method has some advantages as well as limitations over the other. The methods are described in Chapter 2. The application of AR modeling technique and the results of prediction are provided in Chapter 3. Chapter 4 represents the conclusions of the findings.

CHAPTER 2

AUTOREGRESSIVE MODEL

2.1 Introduction

The most popular of the time series modeling approaches to prdiction is the autoregressive (AR) estimator. This is because accurate estimates of the AR parameters can be found by solving a set of linear equations. For accurate estimation of the parameters of moving average (MA) or autoregressive moving average (ARMA) process, we need to solve a set of highly nonlinear equations.When the AR modeling assumption is valid, estimators are obtained which are less biased and have a lower variability than conventional Fourier based estimators. Other names by which the AR estimator is known are the maximum entropy estimator and the linear prediction estimator. Although the theoritical foundations for the latter two estimators differ from those of the AR estimators, in practice all the approaches are identical. The difficulty with adopting either the maximum entropy or linear prediction philosophies is that the all-pole filter assumption implicit in both approaches is not highlighted. As a consequence, application to non-AR time series usually results in poor quality estimates with no clues provided as to the reasons why. The AR modeling is the vehicle used to describe this class of high resolution estimators.

In the prediction of time series data, we have a single function. Let us assume that we are provided with N data as $x[n], n = 0, 1, 2, ..., N-1$ and we need to predict $x[N]$. In that case the AR model can be written as,

$$x[N] = -\sum_{i=0}^{P} a_P[i]x[N-i] \tag{5}$$

where a current data is a linear function of P (known as model order) former data multiplied by the AR parameters $a_P[i]$. To design an effective AR model for prediction, we need to calculate model order P and AR parameters $a_P[i], i = 1, 2, .., P$. There are various techniques to determine the values of $a_P[i]$ and P. This chapter describes the estimation methods of AR parameters and model order which are used in the prediction of data for this study.

2.2 Determination of AR Parameters

The methods normally used to estimate AR parameters are the Yule-Walker method, the Burg method, unconstrained least-squares method and sequential estimation method. In the Yule-Walker and unconstrained least-squares methods, we simply calculate the autocorrelation from the data and use it to solve for the AR parameters [11]. Sequential estimation of AR parameters is used in situations where data are available on a continuous basis and the estimates can be updated as new data points become available [11]. In contrast to the other three methods which estimate the AR parameters directly, the Burg method [12] estimates the reflection coefficients from the data and uses Levinson recursion [13]. Due to some advantages, the Burg method has been used in this study for the determination of AR parameters. The major advantages are:

(a) It results in high frequency resolution.

(b) It yields a stable AR model.

(c) It is computionally efficient.

The Burg method for estimating the AR parameters may be viewed as an order recursive least square lattice method based on the minimization of the forward and backward errors in linear predictors, with the constraint that the AR parameters satisfy the Levinson recursion. The reflection coefficients used in determining AR parameters are obtained by minimizing prediction error power for different order predictors in a recursive manner. Specially, based on the Levinson recursion algorithm, if estimates of the reflection coefficients $k_1, k_2, ..., k_P$ are available, the AR parameters may be obtained as follows:

The autocorrelation of N data is,

$$r_{xx}[0] = \frac{1}{N} \sum_{n=1}^{N-1} |x[n]|^2 \tag{6}$$

The first AR parameter is,

$$a[1] = k_1 \tag{7}$$

For $k = 2, 3, ..., P$,

$$a_k[i] = a_{k-1}[i] + k_k a_{k-1}[k-i], \text{for i=1, 2, ..., k-1} \tag{8}$$

and $a_k[i] = k_k$ for $i = k$. The AR filter parameters are $a_P[1], a_P[2], ..., a_P[P]$. It remains only to calculate the reflection coefficients. In deriving the kth reflection coefficient, Burg assumed that the $(k-1)$st order prediction error filter coefficients had already been calculated as $a_{k-1}[1], a_{k-1}[2], ..., a_{k-1}[k-1]$, having been obtained by minimizing $(k-1)$st order prediction error power. Burg proposed to calculate k_k by minimizing average of the estimates of the forward and backward prediction error powers. Defining forward and backward prediction error powers as ρ_k^f and ρ_k^b, respectively, we need to minimize the average error power ρ_k, where ρ_k is,

$$\rho_k = \frac{1}{2}(\rho_k^f + \rho_k^b) \tag{9}$$

Let us assume that we have an AR model with model order k and $\hat{x}[n]$ and $\hat{x}[n-k]$ be the forward and backward predicted values of the data $x[n]$ and $x[n-k]$, respectively. Then we may write,

$$\hat{x}[n] = -\sum_{i=1}^{k} a_k[i]x[n-i] \tag{10}$$

and,

$$\hat{x}[n-k] = -\sum_{i=1}^{k} a_k[i]x[n-k+i] \tag{11}$$

Defining forward and backward errors of prediction as e_k^f and e_k^b, respectively, we can write,

$$e_k^f[n] = x[n] - \hat{x}[n] = x[n] + \sum_{i=1}^{k} a_k[i]x[n-i] \tag{12}$$

$$e_k^b = x[n-k] - \hat{x}[n-k] = x[n-k] + \sum_{i=1}^{k} a_k[i]x[n-k+i] \tag{13}$$

The forward and backward prediction error powers ρ_k^f and ρ_k^b then become,

$$\rho_k^f = \frac{1}{N-k} \sum_{n=k}^{N-1} e_k^f[n]|^2 \qquad (14)$$

$$\rho_k^b = \frac{1}{N-k} \sum_{n=k}^{N-1} e_k^b[n]|^2 \qquad (15)$$

The lattice filter relations which describe the model order update of the forward and backward prediction error time series can be obtained from Levinson recursion algorithm [13]. These are,

$$e_k^f[n] = e_{k-1}^f[n] + k_k e_{k-1}^b[n-1] \qquad (16)$$

and,

$$e_k^b[n] = e_{k-1}^b[n-1] + k_k e_{k-1}^f[n] \qquad (17)$$

where, $e_0^f[n] = e_0^b[n] = x[n]$.

When the relations of eqns. (16) and (17) are substituted into eqns. (14) and (15) and then into (9), we obtain,

$$\rho_k = \frac{1}{2(N-k)} \sum_{n=k}^{N-1} [|e_{k-1}^f[n] + k_k e_{k-1}^b[n-1]|^2 + |e_{k-1}^b[n-1] + k_k e_{k-1}^f[n]|^2 \qquad (18)$$

Differentiating ρ_k with respect to k_k and solving for k_k, we obtain,

$$k_k = -2 \sum_{n=k}^{N-1} e_{k-1}^f[n] e_{k-1}^b[n-1] / (\sum_{n=k}^{N-1} |e_{k-1}^f[n]|^2 + |e_{k-1}^b[n-1]|^2) \qquad (19)$$

Equation (19) is used for calculating reflection coefficients k_k and the AR parameters $a_k[i]$ are determined by eqn. (8).

2.3 Determination of AR Model Order

One of the most important aspects of the use of the AR model is the selection of the order P. As a general rule, if we select a model order with too low an order, we obtain a highly smoothed data fit. On the other hand, if p is selected too high, we run the risk of introducing spurious low-level peaks in the data variation. One indication of the performance of the AR model is the mean square value of the residual error, which, in general, is different for each of the estimators described above. The characteristics of this residual error is that it decreases as the order of the AR model is increased. We can monitor the rate of decrease and decide to terminate the process when the rate of decrease becomes relatively slow. It is apparent, however, that this approach may be imprecise and ill-defined, and other methods should be investigatd. Much work has been done by various researchers on this problem and many experimental results have been given in the literature.

Two of the better known criterion for selecting the model order have been proposed by Akaike [14-15]. With the first, called the final prediction error (FPE) criterion [14], the order is selected to minimize the performance index,

$$FPE(P) = \sigma^2 (\frac{N+p-1}{N-p+1}) \tag{20}$$

where σ^2 is the estimated variance of the linear prediction error. This performance index is based in minimizing the mean-square error for a one-step predictor.

The second criterion proposed by Akaike [15] is called the Akaike Information Criterion (AIC) statistics which is based on selecting the order that minimizes,

$$AIC(P) = \ln\sigma^2 + \frac{2P}{N} \tag{21}$$

Note that the term σ^2 decreases and therefore $\ln\sigma^2$ also decreases as the order of the AR model is increased. However, $2P/N$ increases with an increase in P. Hence a minimum value is obtained for some P.

2.4 Discussion

The Burg has been extensively used in calculating AR parameters due to its high frequency resolution, stability and computational efficiency. However, it has several disadvantages as well. First, it exhibits line spilliting at high signal-to-noise ratios. By line splitting, we mean that the of $x(n)$ may have a single sharp peak, but the Burg method many result in two or more closely spaced peaks in specially the power spectrum of the signal. For high order models, the method also introduces spurious peaks. Furthermore, for sinusoidal signal in noise, The Burg method exhibits a sensitivity to the initial phase of a sinusoid, especially in short data records. This sensitivity is manifested as a frequency shift from the true frequency resulting in an frequency bias that is phase dependent frequency bias.

Several modifications have been proposed to overcome some of the more important limitations of the Burg method namely the line splitting, spurious peaks, and frequency bias. Basically, the modifications involve the introduction of a weighting (window) sequence of the squared forward and backward errors.

Despite all the limitations, Burg method is widely used in AR parameter determination. Despite its bias to the data length, AIC statistics is the mostly used method for model order selection. The applications of the Burg method in determining AR parameters and AIC statistics in selecting AR model order and hence to predict time series data are presented in the next chapter.

CHAPTER 3

PREDICTION OF DATA

3.1 Introduction

This chapter describes the application of AR model to predict two types of data, namely instantaneous heart rate (IHR) and power load (PL). Together with the electro-cardiogram (ECG), mean heart rate (MHR) is oftenly used to assess cardiac activities. The heart rate can vary widely even in a short period of time due to natural variability. MHR can not explain the issues of arrhythmias which are the major causes of sudden cardiac death. A parameter other than MHR may illuminate more light on the better understanding of the cardiac functions and other phenomena associated with it. The beat to beat analysis of cardiogenic signals, i.e., (IHR) is such a parameter. It has long been recognized that the heart rate (HR), arterial blood pressure, and other hemodynamic parameters fluctuate on a beat-to-beat basis [16]. The beat-to-beat fluctuations in hemodynamic parameters reflect the dynamic response of the cardio-vascular control systems to a host of naturally occurring physiological perturbations. The sympathetic and parasympathetic nervous systems are usually considered to be the principal systems involved in short-term cardiovascular control on the time scale of seconds to minutes. Heart rate fluctuations reflect the regulatory activity of the autonomic nervous system (ANS), which modulates the intrinsic sinus node firing rate. In fact, the IHR can be thought of as the output of a feedback control loop that is continuously regulated by the ANS. Lack of beat-to-beat heart rate variability is an important predictor of impending death.

In recent years, linear modeling techniques such as AR and ARMA models have been used to study HR dynamics [17]. AR model has the ability to correlate patho-physiological mechanisms to HR dynamics. Much works have been devoted to the modeling of HR with AR models.

The time to time variation of electrical power is necessary to predict future load in any power system. Load forcasting is an important branch in power engineering.

The following sections depict the data used in this study and the results found in predicting HR and PL.

3.2 Data Description

Two types of data were predicted, HR and PL. The calculation of IHR and PL is described below.

Heart Rate Calculation

ECG was measured from 5 healthy human subjects aged 22-30 years. The ECG was stored in a microcomputer through a 12-bit A/D converter. The digitized signals were temporarily saved on the hard-disk driver and then transferred to a file server. The frequency at which the signals are to be sampled may cause errors in the calculation of IHR [18]. Due to the limited capacity of the computer memory, the signals were sampled at a low frequency of 50 Hz. On the other hand, in order to calculate the IHR with an inaccuracy of less than one beat/min (bpm), the sampling frequency (f) which satisfies the following relation is needed,

$$f \geq 2(60 - \Delta t)/(\Delta t)^2 \tag{22}$$

where $\Delta t = 60/\mathrm{IHR}$. For example, for a maximum IHR value of 120 bpm, the minimum sampling frequency is 476 Hz. The signals sampled at 50 Hz were restored by the sinc function,

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin\left[2\pi f_M(t - nT)\right]}{2\pi f_M(t - nT)} \tag{23}$$

where T is sampling interval (i. e., 20 msec) and f_M is Nyquist's reflection frequency (25 Hz). In order to minimize time of calculation, the signal restoration was made only for the period of 40 msec when three points sampled were found to constitute a peak of ECG (R wave) corresponding to each heartbeat. After restoration of the R wave, the time interval between two successive peaks was determined and the IHR was then calculated from the time interval. The number of sampled points used for restoration by eqn.(23) was determined to be 401 (i.e., $n = 200$) in a preliminary experiment. In the preliminary experiment, the ECGs were sampled at a frequency of 500 Hz and the IHR was calculated from the time intervals of two successive peaks (referred to as HR_{500}). Then, single values of every 80 sampling points (which correspond to the signals sampled at 50 Hz) were extracted and the ECG waves were restored by eqn.(23) changing n-value from 100 with a step of 5. The IHR was calculated for each wave restored with different values of n. It was found that the

IHR so calculated was consistent with HR_{500} when n was more than 200 in all data sets. IHR of a person was stored for more than 1 hour and in this way 5 data sets were constructed. The calculated IHR of the 5 persons are depicted in Figs. 1(a) to 1(e) where, the HR is presented for 30 minutes in each case.

Power Load Data

Hourly power supply by the Bangladesh Power Development Board (BPDB) was collected from the Central Load Dispatch Center (CLDC), Siddhirganj. The data collected is of 3 years. The data were grouped into 5 categories. The daily load at 1 PM is the group 1 data, that of 2 PM is the group 2 data, and similary group 3, group 4 and group 5 data belong to 3 PM, 4 PM and 6 PM, respectively. In this way 5 data sets were constructed. The 5 sets of data are presented in Figs. 2(a) to 2(e).

3.3 Results

The AR model as described in chapter 2 was applied to predict the IHR and PL. The AR parameters were determined by Burg method and the model order by AIC statistics, the detail descriptions of both are provided in chapter 2.

In case of IHR, for each data set, first 1000 IHR was provided as the input of the AR model. All the parameters of the model were then determined and from the model order the next data were predicted. If the model order is P, the predicted data will be of the data point from $p + 1$ to 2000. The AR parameters for IHR data set 1 are provided in Table I. The original data and the error of prediction (differnce between original and corresponding predicted data) are presented in Figs. 3(a) to 3(e), where in each figure, the upper plot shows the original data and the lower one depicts the error of prediction.

In the prediction of PL, 500 data was provided as the input of the AR model and upto 1000 data starting from model order was predicted in a similar manner as in IHR prediction. The original data (upper graph) and the error of prediction (lower graph) are presented in Figs. 4(a) to 4(e).

The mean and standard deviation (SD) (mean± SD) of original as well as predicted IHR with the model order (MO) are presented in Table 2, while those for PL in Table 3. The F-values provided in both the tables were calculated applying one-way analysis of variance technique [19] between the predicted and original data in each case. The statistically significant difference is taken when $F \geq 3.84$. It is seen that statistically there is no significant difference between original and predited data in all data sets of IHR as well as of PL.

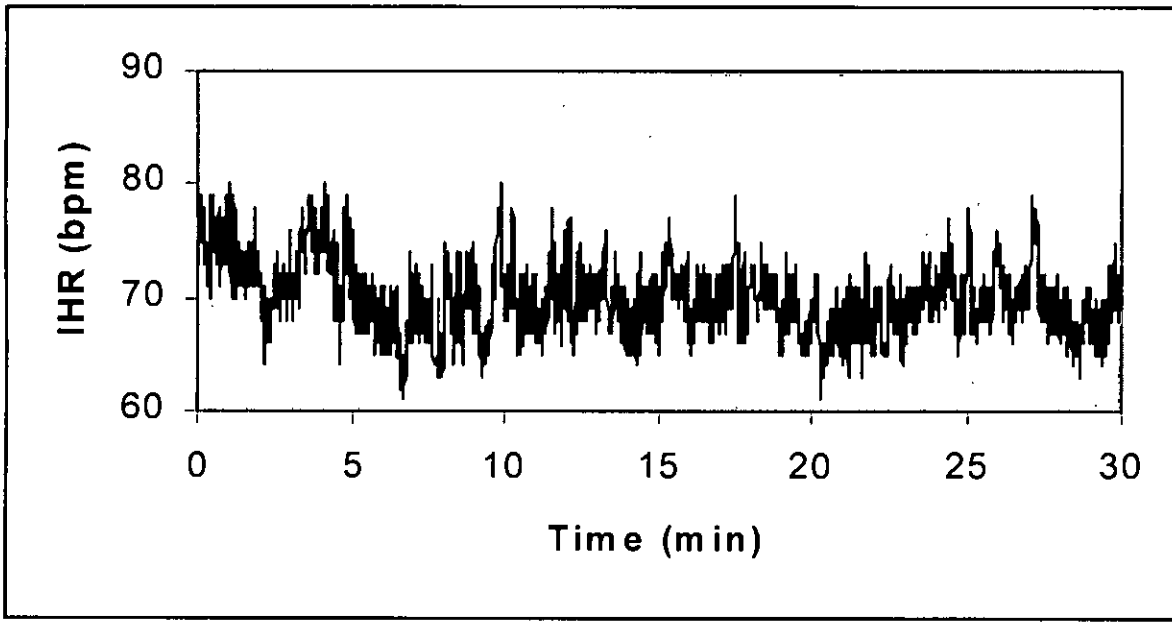

Fig. 1(a) Instantaneous heart rate for data set 1.

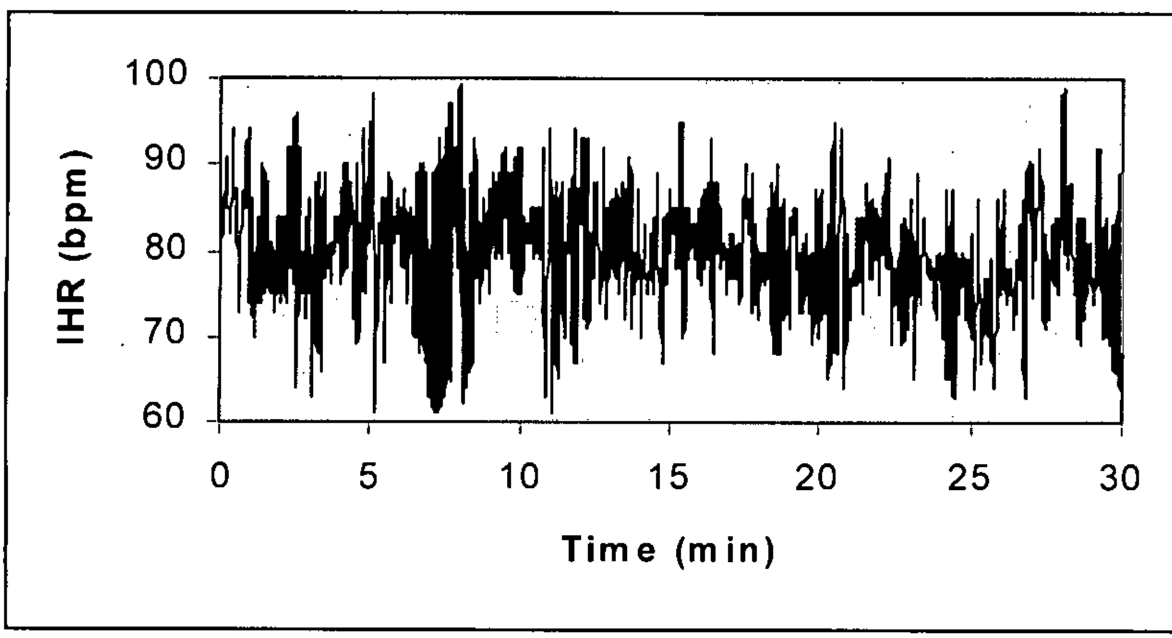

Fig. 1(b) Instantaneous heart rate for data set 2.

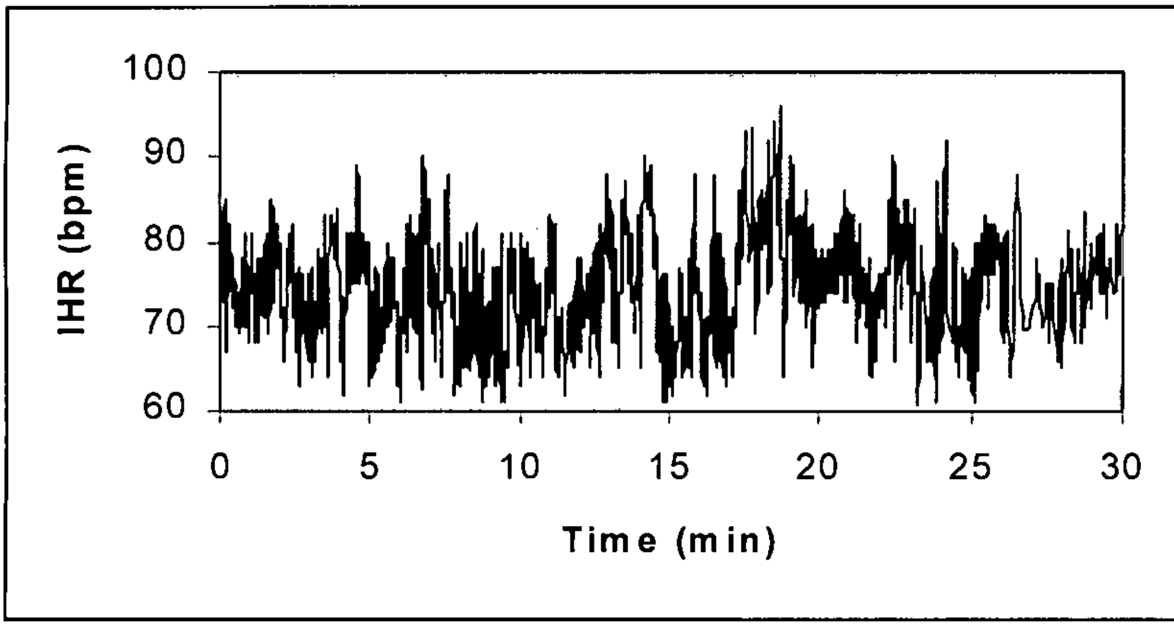

Fig. 1(c) Instantaneous heart rate for data set 3.

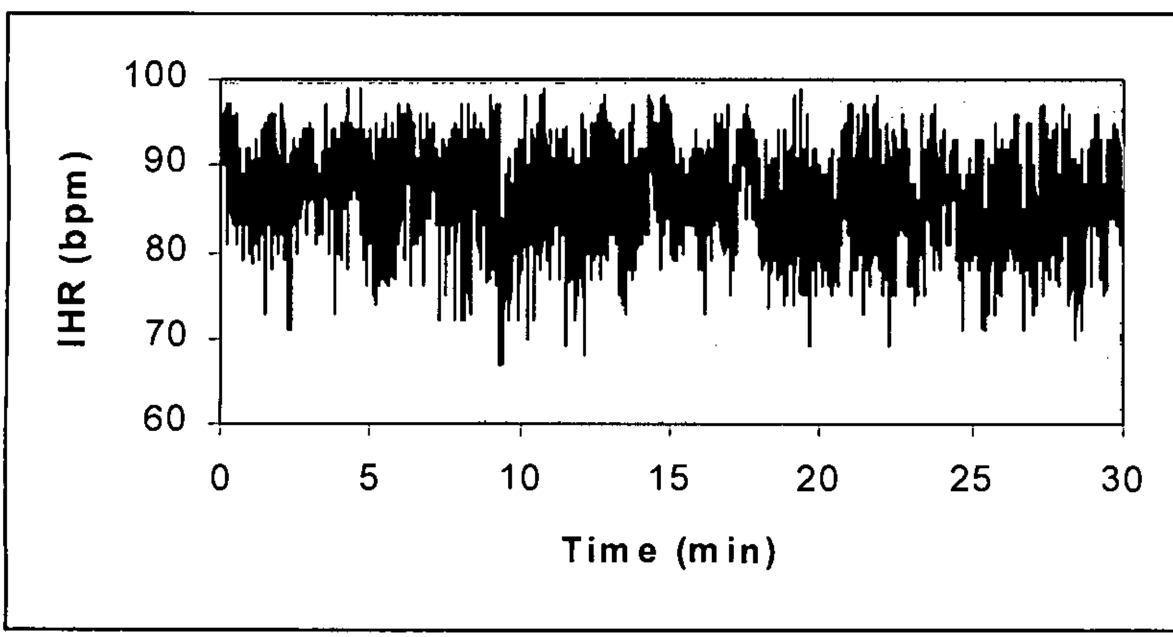

Fig. 1(d) Instantaneous heart rate for data set 4.

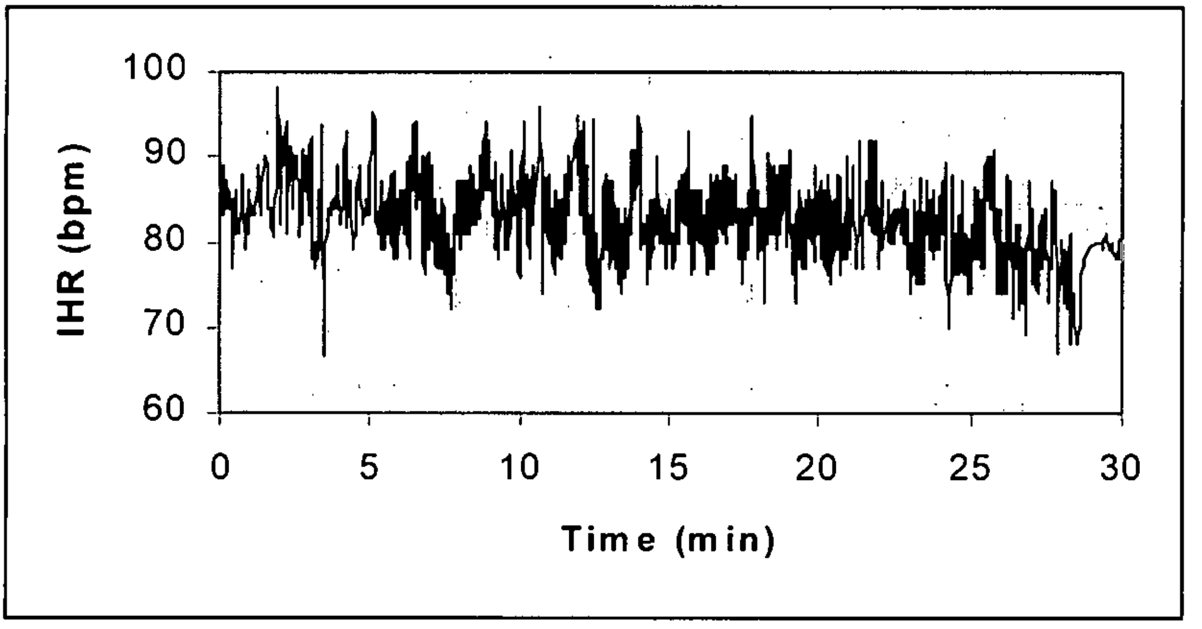

Fig. 1(e) Instantaneous heart rate for data set 5.

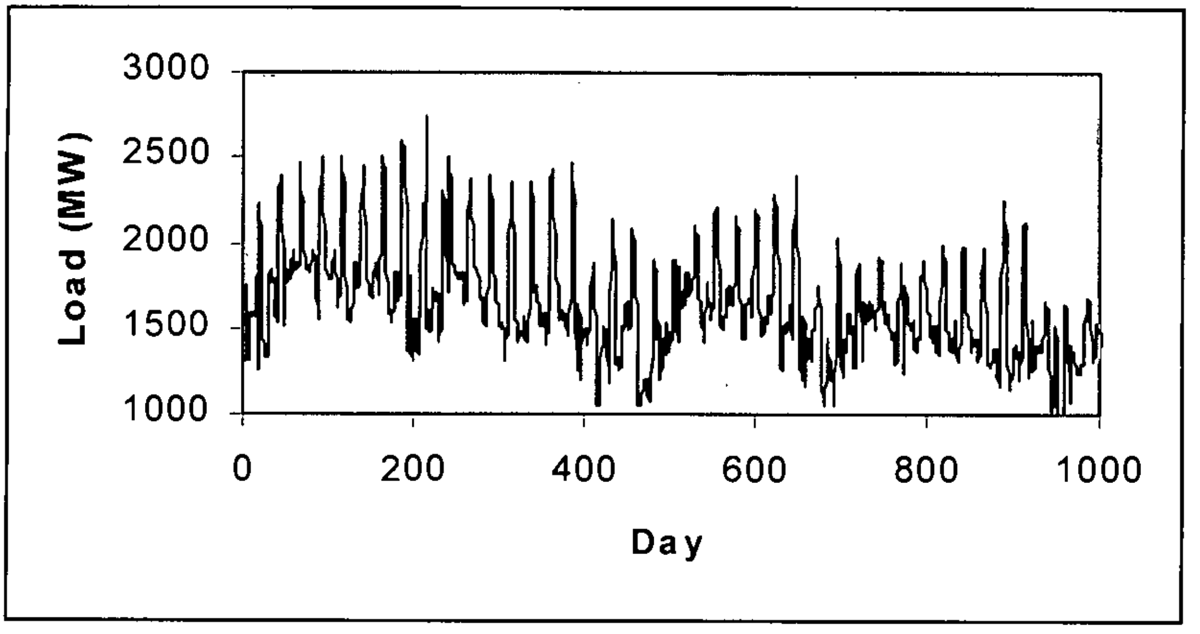

Fig. 2(a) Power load data at 1 PM.

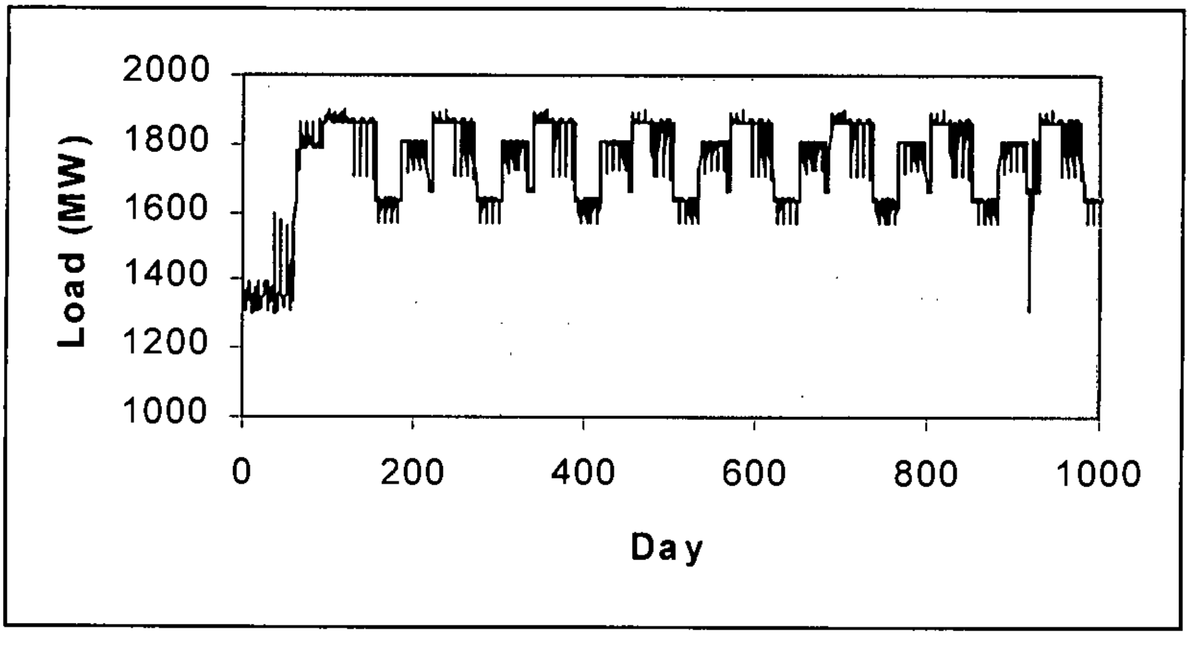

Fig. 2(b) Power load data at 2 PM.

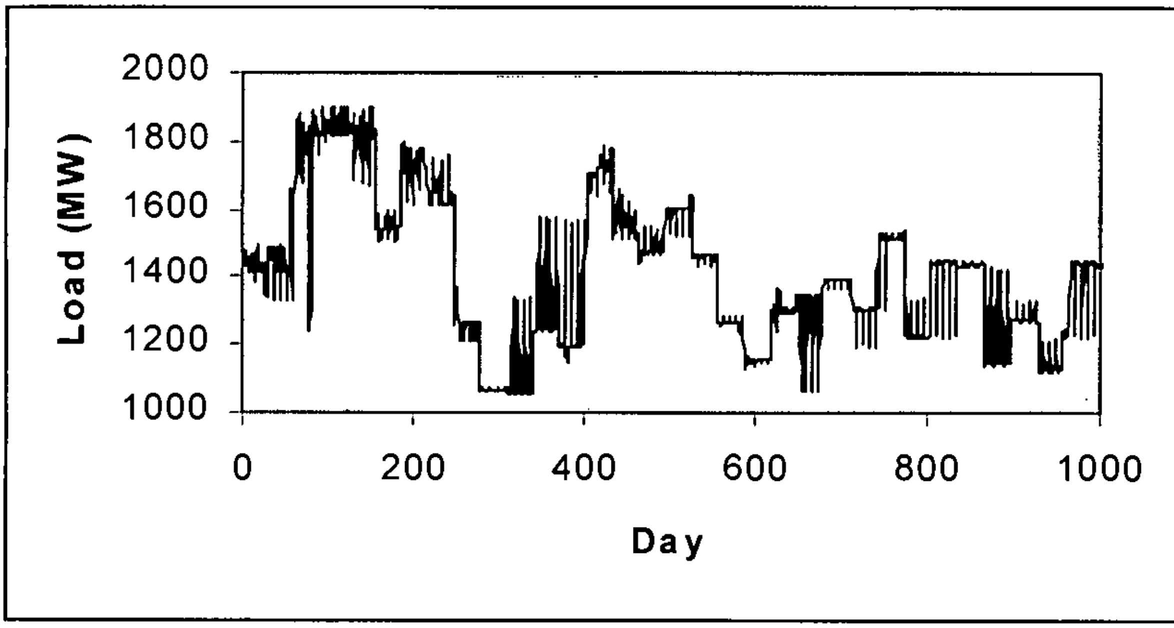

Fig. 2(c) Power load data at 3 PM.

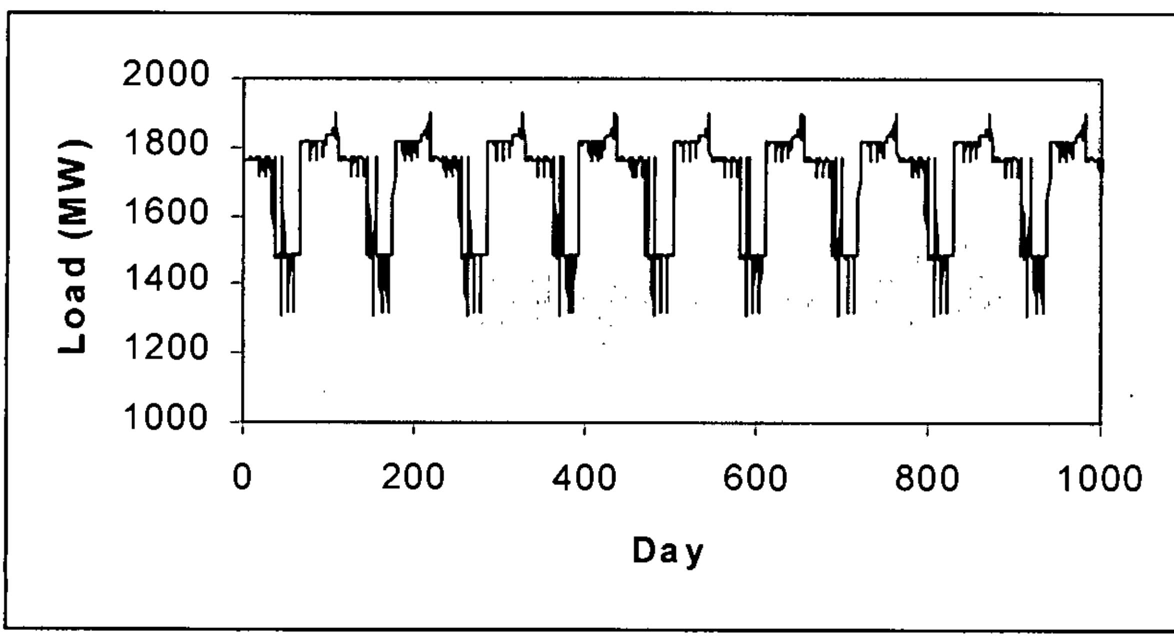

Fig. 2(d) Power load data at 4 PM.

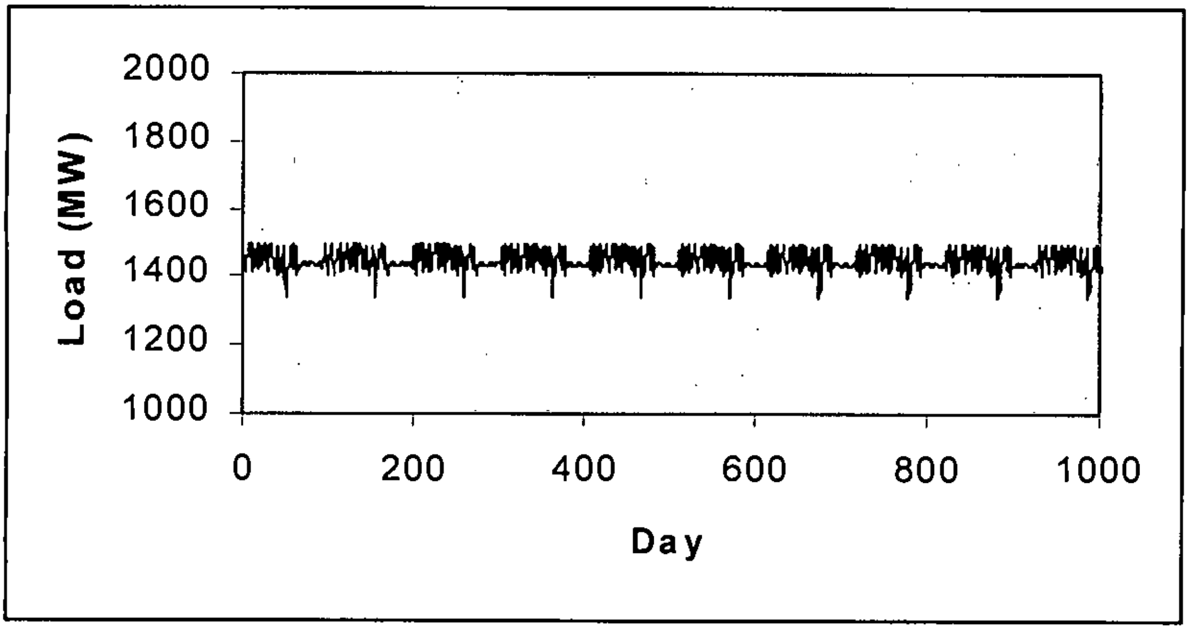

Fig. 2(e) Power load data at 6 PM.

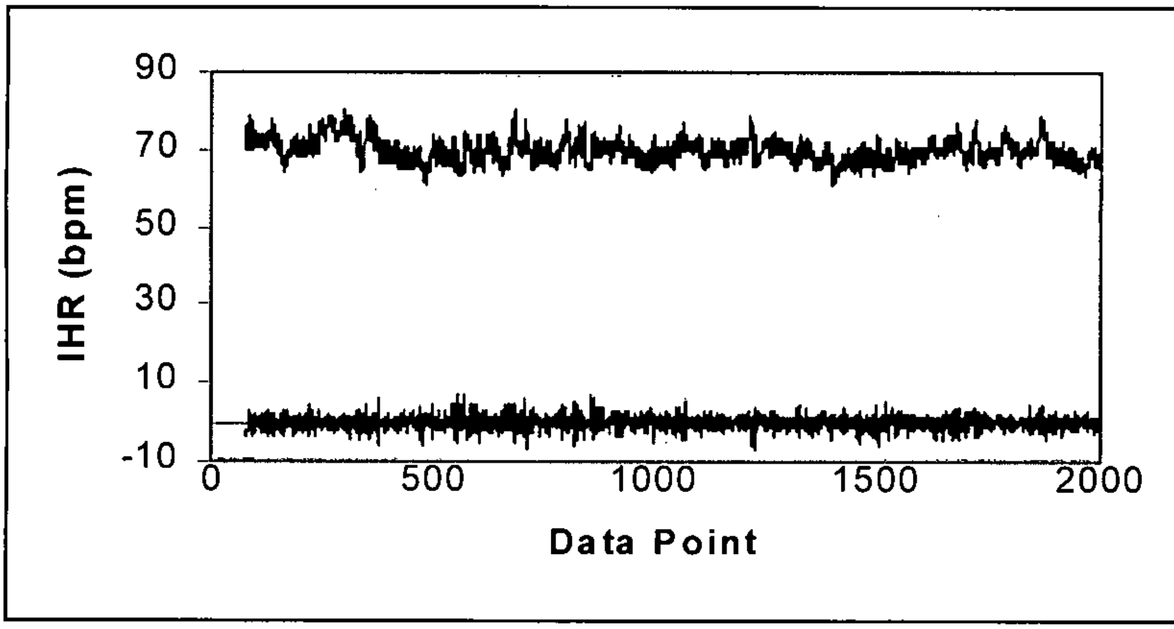

Fig. 3(a) Original IHR and error of prediction for data set 1.

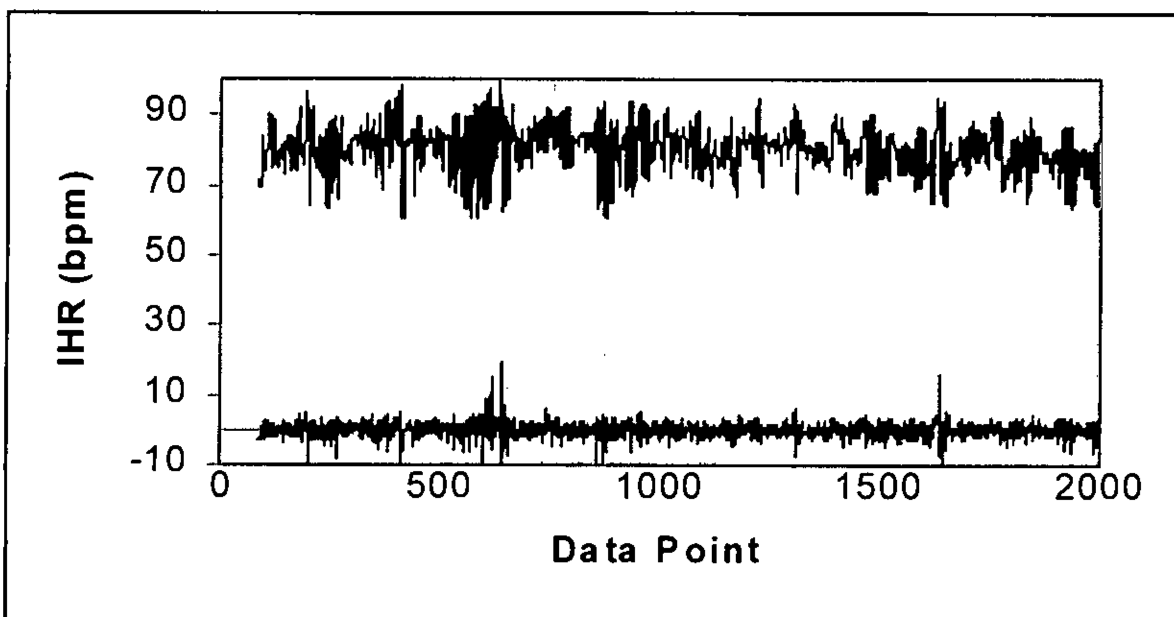

Fig. 3(b) Original IHR and error of prediction for data set 2.

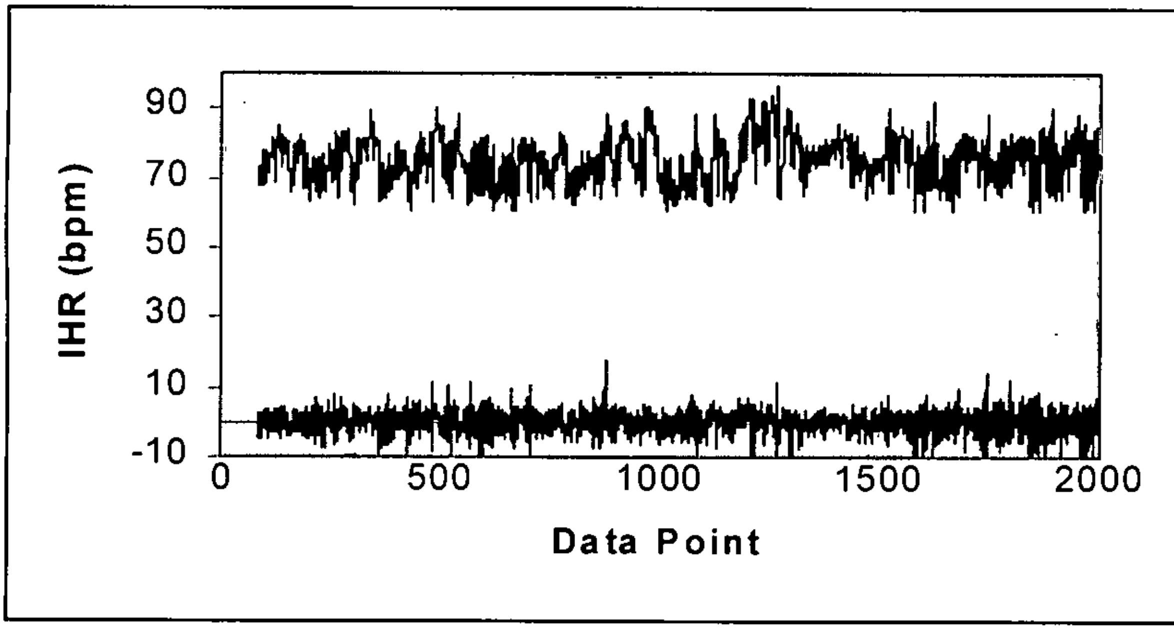

Fig. 3(c) Original IHR and error of prediction for data set 3.

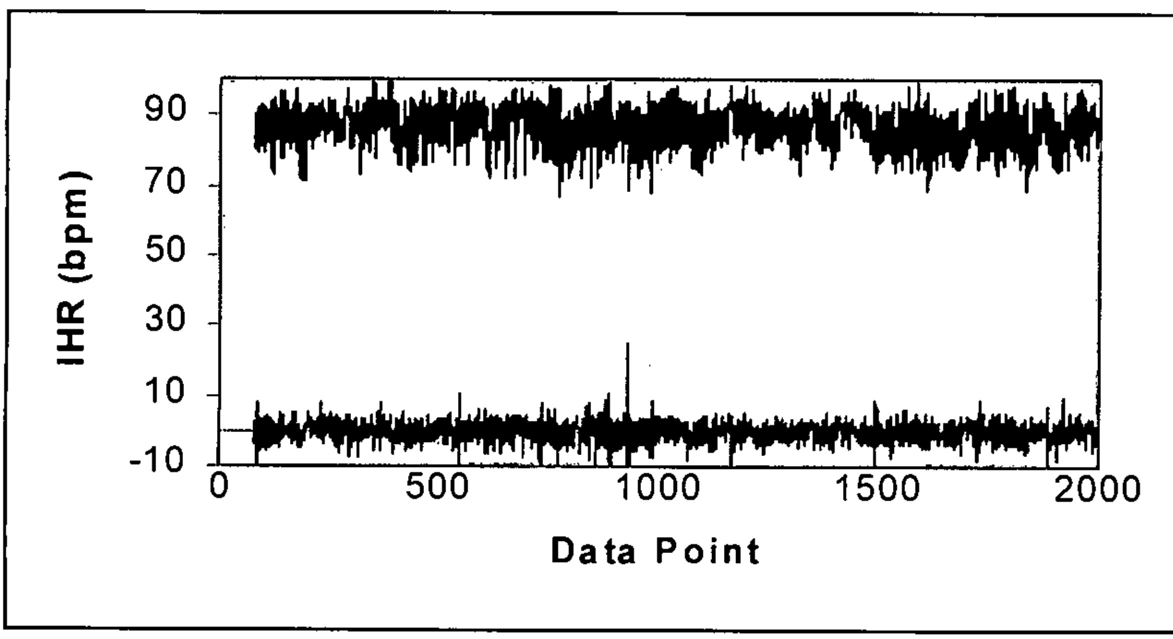

Fig. 3(d) Original IHR and error of prediction for data set 4.

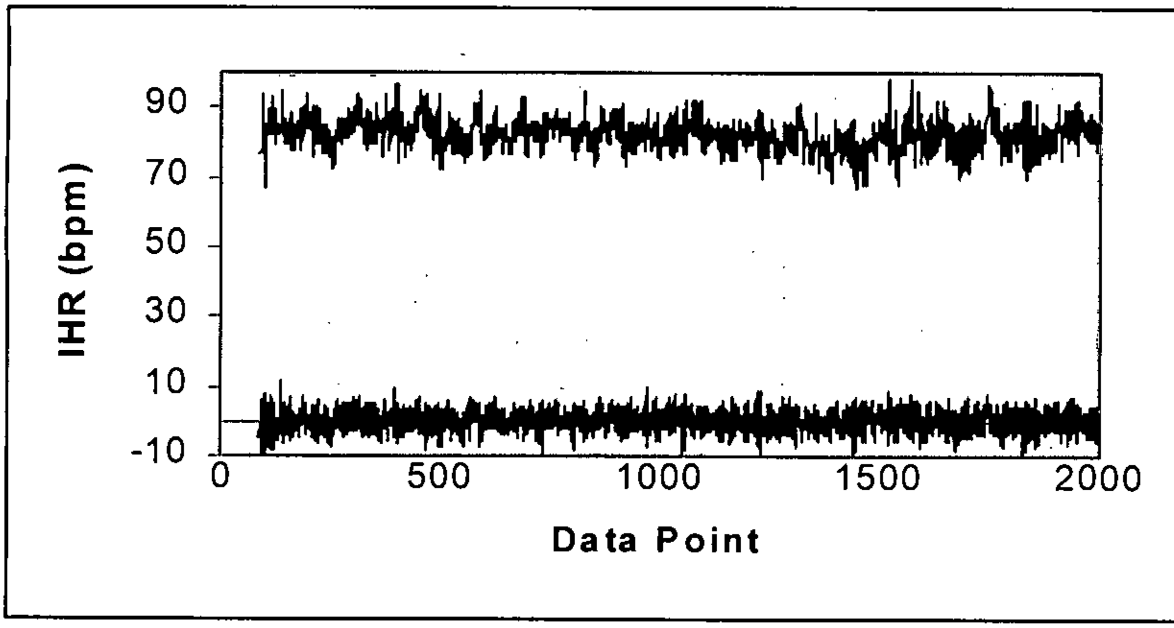

Fig. 3(e) Original IHR and error of prediction for data set 5.

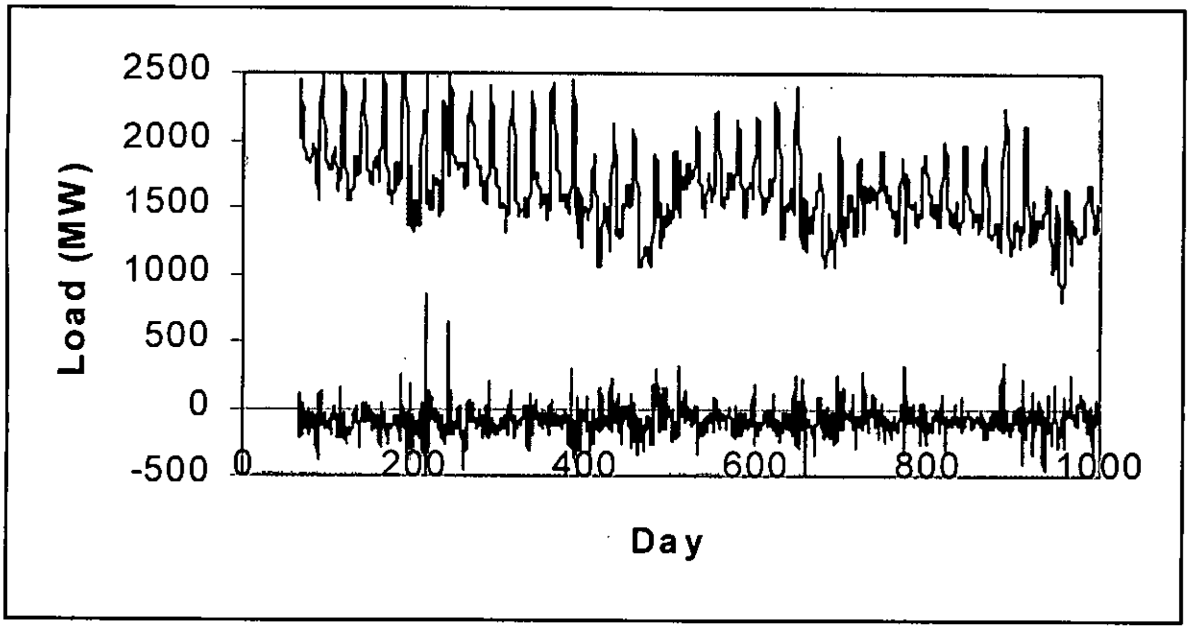

Fig. 4(a) Original power load and error of prediction at 1 PM.

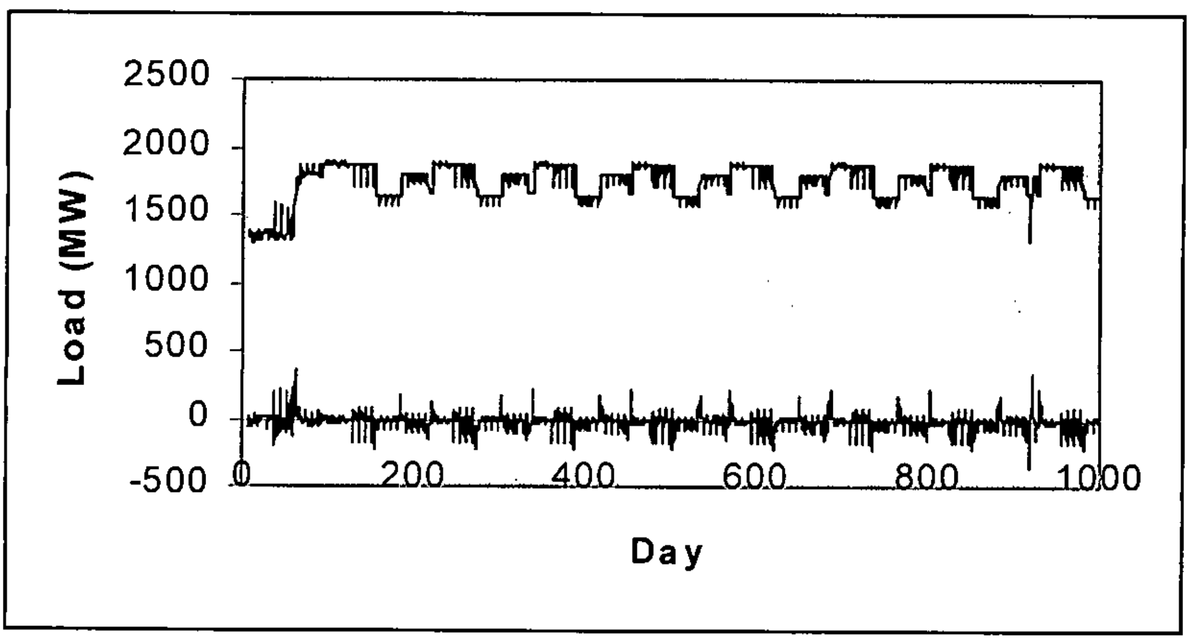

Fig. 4(b) Original power load and error of prediction at 2 PM.

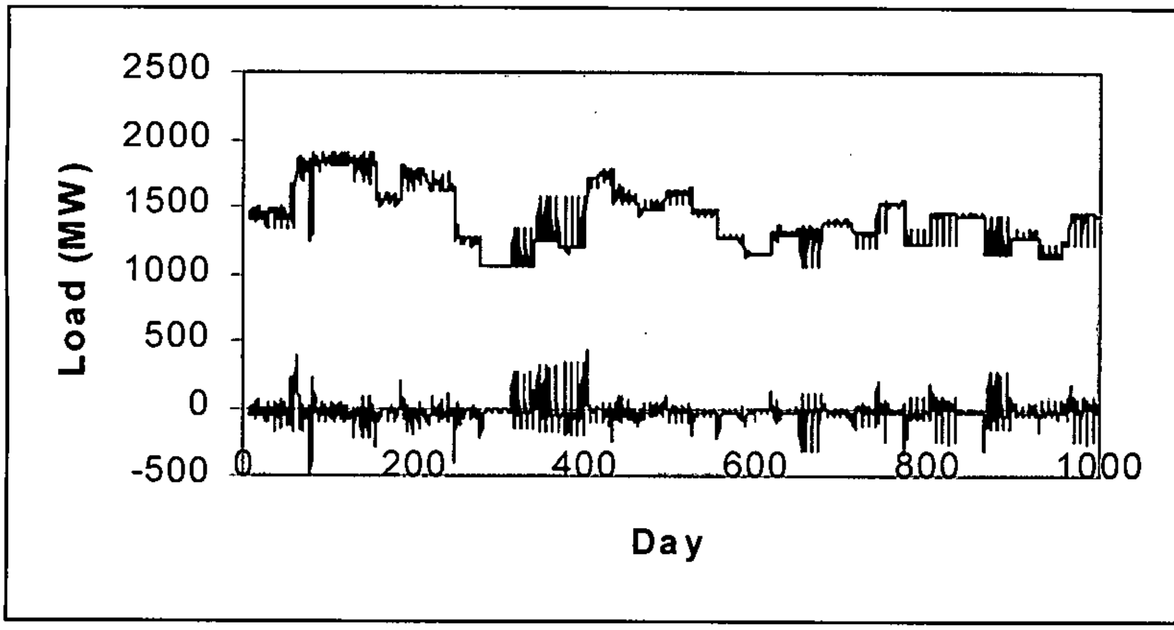

Fig. 4(c) Original power load and error of prediction at 3 PM.

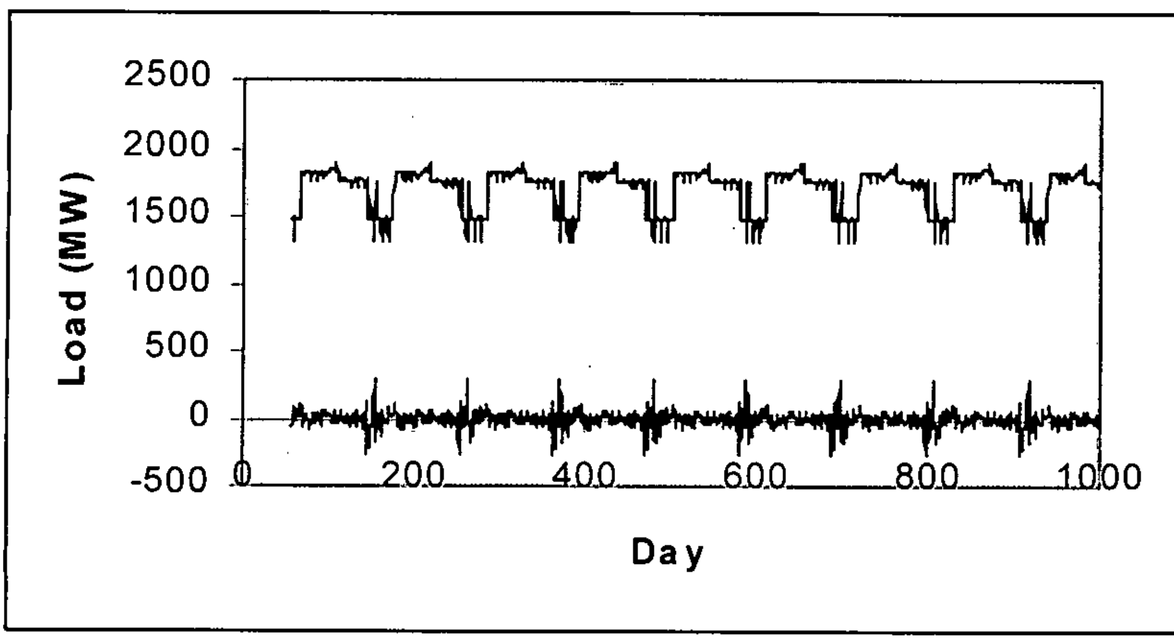

Fig. 4(d) Original power load and error of prediction at 4 PM.

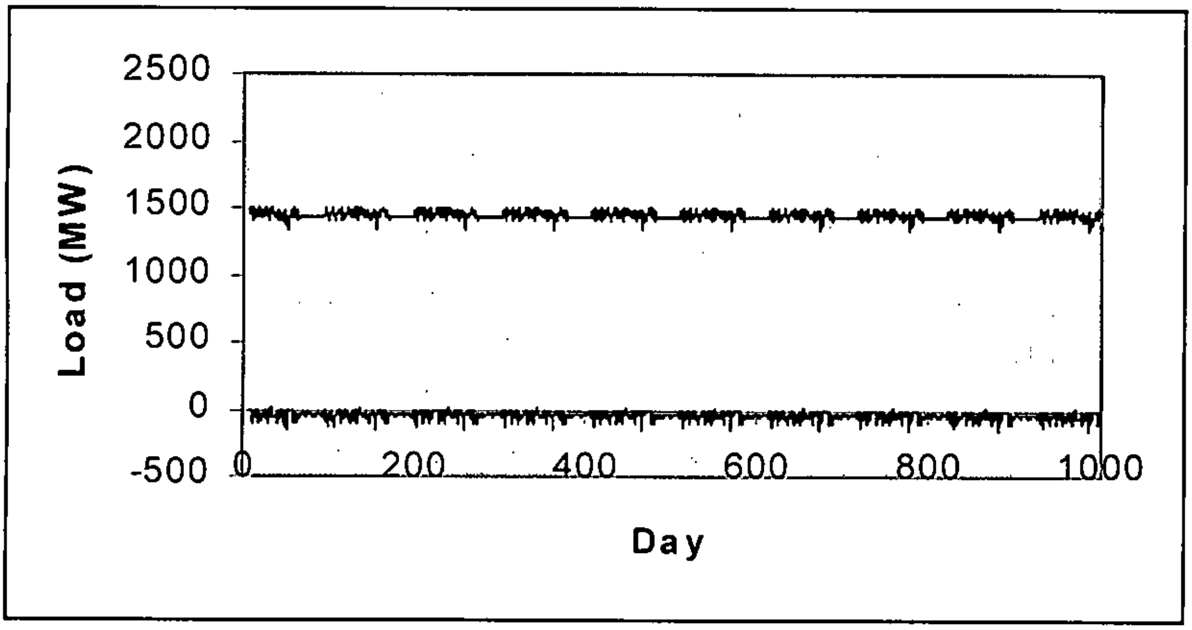

Fig. 4(e) Original power load and error of prediction at 6 PM.

Table 1: **AR Parameters for data set 1 of IHR**
(Model order = 78)

0.046232	-0.308710	-0.075185	-0.022001
0.007029	0.003834	-0.044492	0.004149
-0.080111	-0.644061	-0.001883	-0.054954
0.035720	0.019131	0.030095	0.066723
-0.028360	0.565168	-0.054121	-0.012262
-0.054510	0.000240	-0.035268	-0.084471
0.086283	-0.171771	-0.015284	0.033969
-0.024768	-0.002702	-0.091644	0.037702
-0.127906	-0.069023	0.022131	0.066812
-0.017354	-0.052009	0.098620	-0.060623
-0.043691	-0.051829	0.015746	-0.016378
0.071111	0.004062	-0.079379	0.014479
-0.021949	-0.076321	-0.011260	-0.039168
0.010978	-0.002375	-0.079185	-0.054031
0.164843	-0.048528	-0.018229	-0.023459
-0.006480	0.064510	-0.000734	0.027080
-0.115573	-0.054405	0.034729	0.029313
-0.001782	-0.029753	0.008955	0.065278
-0.549071	-0.033647	0.019079	
-0.000681	-0.007762	0.001375	

Table 2 Comparison of AR predicted IHR

Data set	Model order	Original IHR (bpm)	Predicted IHR (bpm)	F 0.01
# 1	78	69.8± 3.1	69.8±3.1	0.00
# 2	88	80.4± 6.4	80.4±6.0	0.00
# 3	86	75.2± 6.3	75.2±5.6	0.00
# 4	77	86.6± 6.1	86.6±5.3	0.00
# 5	89	82.2± 4.6	82.2±3.4	0.00

Table 3 Comparison of AR predicted power load

Data set	Model order	Original PL (MW)	Predicted PL (MW)	F
# 1	65	1635.95± 313.43	1726.97±305.69	0.000
# 2	5	1748.55± 136.50	1764.51±132.39	0.000
# 3	6	1417.30± 219.67	1447.16±216.01	0.000
# 4	56	1712.67± 146.37	1715.00±141.68	0.000
# 5	5	1447.27± 27.62	1486.45±20.73	0.001

3.4 Discussion

The present study was performed with an interest to know the applicability of AR model in predicting HR and PL. AR model is widely used for linear prediction of time series data. It is also a very useful tool in the spectral analysis of time series data. The optimal selection of model order is necessary to avoid the over-flatening and spurious peaks in the power spectra of the data and hence to have an optimum model fit. Although, various techniques can be used for model order selection depending upon the data length, AIC performs similarly or better irrespective of data length [11]. A large value of model order has been obtained in most of the data sets. The results of prediction show that AR model can precisely predict time series data of the kind provided here. The two types of data analyzed in this study have different characteristics. HR is a biological phenomenon while PL is man made. The prediction results show that AR model can be applied to a variety of fields for data prediction.

CHAPTER 4

CONCLUSIONS

4.1 Conclusions

This work describes the prediction of time series data with a linear technique. Out of the many linear methods of prediction, AR model is the most widely used one. Although ARMA model can provide superior prediction in comparison to AR model, the computational complexity of finding ARMA parameters makes it time consuming and less suitable for on-line prediction of data.

There many methods to determine the AR parameters. Out of them the Burg method is widely used due to its computational simplicity, since it is based on the recurrence of autocorrelation function. The Burg method uses the minimization technique of average error due to forward and backward prediction, and involves the solution of a set of linear difference equations. The major advantages of the Burg Method for estimating the parameters of the AR model are:

(i) It results in high frequency resolution.

(ii) It yields a stable AR model.

(iii) It is computionally efficient.

Out of the many methods of AR model order determination, AIC statistics is widely used due to its superior performance over other ones. All the other methods have the drawback that they are very much sensitive to data length. On the other hand, AIC statistics performs better or at least similarly irrespective of data length.

Two different types of time series data were predicted. Heart rate is a bilogical phenomenon. Although linear models are widely used for the prediction of HR, presently there is indication that heart rate may be nonlinear or even chaotic. But still, linear models provide a comprehensive view of the mechanism generating and controlling ECG and subsequently HR. On the other hand, the variation of electrical power can also be forcasted by linear models. Since all the input parameters for generating power such as intake of fuels, etc., are known, power generation can be predicted in a straight way by a deterministic procedure. But, the consumption of electricity is based on time, weather condition, human nature of the locality, etc. All these parameters make the power consumption forcasting a not so easy task as do in the case of generation. The power load collected from the Load Dispatch Centre is the hourly load consumed across Bangladesh.

The results in this work show that AR model can precisely predict HR as well as power load. Although the number of data sets are very few in both cases, and only normal conditions specially in case of HR have been taken into consideration, the results suggest that AR model can be applied in such cases.

4.2 Future Perspectives

The HR can very widely even for a short period of time, and even for the case of healthy subjects. The applicability of AR model can be tested for the HR with *certain* cardiac abnormality. The power load can be modeled based on the factors influencing the power consumtion, that is, the influences can be taken as input and power consumption as output and model the load to see how any of the influences alter the pattern of power load.

BIBLIOGRAPHY

[1]. L. Ljung, *System Identification: Theory for the user*, Englewood Cliffs, NJ: Prentice Hall, 1987.

[2]. G. E. P. Box and G. M. Jenkins, *Time series, Forcasting and Control* San Francisco: Holden-Day, 1970.

[3]. T. W. Anderson, *The Statistical Analysis of Time Series*, New York and London: Wiley, 1971.

[4]. E. J. Hannan, *Time Series Analysis*, London: Methuen Monographs, 1962.

[5]. E. J. Hannan, *Multiple Time Series*, New York and London: Wiley, 1970.

[6]. T. S. Rao and M. M. Gabr, A test for linearity of stationary time series, *J. Time Series Analysis*, Vol. 1, No. 2, 1981.

[7]. V. Volterra, *Theory of Functionals and an Integral and Integrodifferential Equations*, London: Dover, 1930.

[8]. N. Wiener, *Nonlinear Problems in Random Theory*, London: Wiley, 1958.

[9]. A. Papoulis, *Probability, Random Variables and Stochastic Processes*, New York, NY: McGraw-Hill, 1984.

[10]. J. G. Proakis and D. G. Malonakis, *Digital Signal Processing: Principles, algorithms and applications*, Prentice-Hall (India), 1997.

[11]. S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs: NJ, 1988.

[12]. J. P. Burg, A new analysis technique for time series data, NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, pp. 17-23, August, 1968. Reprinted in *Modern Spectrum Analysis*, D. G. Childers, ed., IEEE Press, New York.

[13]. J. Durbin, Efficient estimations of parameters in moving average models, *Biometrica*, Vol. 46, pp. 306-316, 1959.

[14]. H. Akaike, Power spectrum estimation through autoregressive model fitting, *Ann. Inst. Stat. Math.*, Vol. 21, pp. 407-409, 1969.

[15]. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automatic Control*, Vol. AC-19, pp. 716-734, 1974.

[16]. S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Barger and R. J. Cohen, Power spectrum analysis of heart rate fluctuation: a quantitative probe to beat-to-beat cardiovascular control, *Science*, Vol. 213, pp. 220-222, 1981.

[17]. D. J. Christini, M. Bennett, K. R. Lutchen, H. M. Ahmed, J. M. Hausdorff and N. Oriol, Application of linear and nonlinear time series modeling to heart rate dynamics analysis, *IEEE Trans. Biomed. Eng.*, Vol. 42, pp. 411-415, 1995.

[18]. M. Merri, D. C. Farden, J. G. Mottley and E. L. Titlebaum, Sampling frequency of electrocardiogram for spectral analysis of the heart rate variability, *IEEE Trans. Biomed. Eng.*, Vol. 37, pp. 99-106, 1990.

[19]. R. E. Walpole, *Introduction to Statistics*, Collier Macmillan, 2nd Ed.