VOLTA RIVER FLOWS

STOCHASTIC MODELLING AND FORECASTING

VOLTA RIVER FLOWS

STOCHASTIC MODELLING AND FORECASTING

by

C. K. O. ADDO, B.Sc.(HONS)

A Project Report
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Engineering
McMaster University

December 1987

ABSTRACT

The Volta River Authority (VRA) is responsible for the generation and transmission of power in Ghana. For this purpose, VRA owns and operates two hydroelectric generating stations (at Akosombo and Kpong) with a combined installed capacity of 1060 Mw. The Akosombo plant is served by the Lake Volta Reservoir. Prediction of inflows into the Volta Lake is one of the important functions of the reservoir management group.

For this project, some of the more recent methods of mathematical modelling are investigated with a view to building a simple stochastic model which adequately represents and forecasts the Volta river average monthly flow. The Box-Jenkins family of models are employed in this exercise. A parsimonious model in the form of a seasonal autoregressive integrated moving average (SARIMA) model is arrived at which adequately models and forecasts the available data.

The selected model is reasonably easy to set up, has few parameters to estimate and therefore making the updating of these parameters a relatively simple task.

ACKNOWLEDGEMENTS

I wish to extend my gratitude to Dr. A. A. Smith who supervised this project and enthusiastically assisted me in various ways. I also wish to thank Dr. G. Patry for the valuable discussions we had. I acknowledge with thanks the assistance of the Volta River Authority who financed and provided the data for this work. I also want to thank Dr K. W. Hipel and Dr. A. I. McLeod who kindly permitted the use of their time series package for this exercise. I wish also to acknowledge the assistance of Acres Consulting Services of Canada, who permitted me access to computer facilities. I also wish to thank Ontario Hydro for administering the program. A special thank you to the Engineering Department of VRA for allowing me the time to complete this project and finally to my wife for bearing with me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND SCOPE OF WORK

The Volta River Authority (VRA) is responsible for generation and transmission of power in Ghana. For this purpose, VRA owns and operates two hydroelectric generating stations with a combined installed capacity of 1060 Mw.

VRA supplies power directly to major consumers in Ghana and to Electricity Corporation of Ghana (ECG). In addition VRA exports power to Communauté Electrique du Bénin (CEB) which supplies power to Togo and Bénin and also VRA has an exchange agreement with Energie Electrique de la Côte d'Ivoire (EECI).
The two generating stations operated by VRA are the Akosombo and Kpong generating stations whose locations are as shown on Figure 1. It should be noted that though there exists some diesel generation in Ghana the total output capability does not amount to more than 30 Mw in total, accordingly hydro power plays an important role in the Energy sector of Ghana and it's neighbouring countries.

The Kpong generating station, a run-of-river plant situated 25 km downstream of the Akosombo generating station, has pondage of approximately 0.1% of the storage capacity of the Akosombo plant.

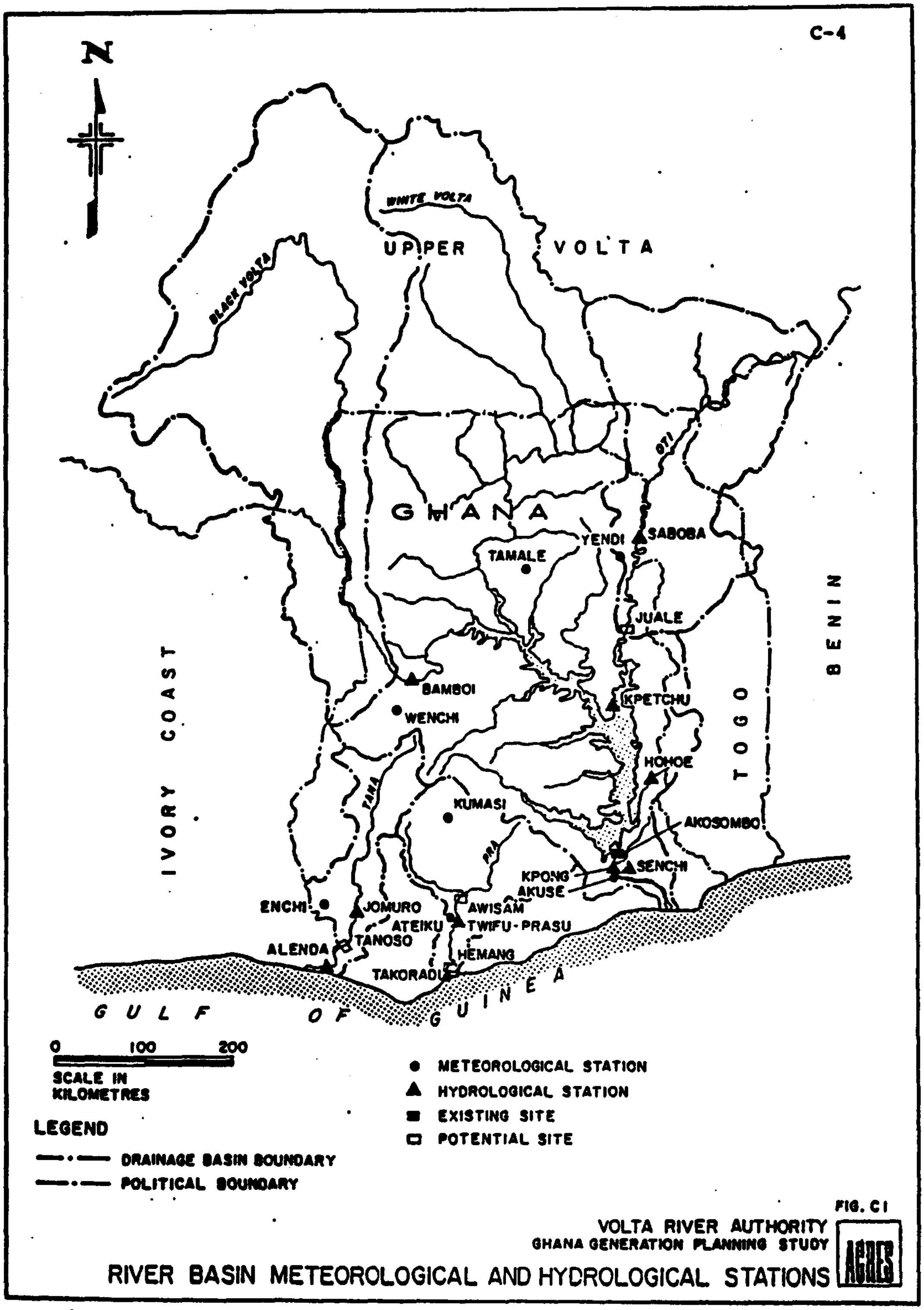


Figure 1.1 - River Basin
(Reproduced with permission of Acres International Ltd.)

Hence the effect of local inflows into the Kpong headpond is negligible in terms of power flows. The Akosombo plant however is served by the Lake Volta Reservoir which has a storage capacity of 149900 mcm covering an area of approximately 8400 km^2 at full supply level. The drainage area is estimated at 390 000 km^2.

The Akosombo reservoir (Lake Volta) is a large reservoir with multi-year storage capability. It is usually allowed to fluctuate between full supply level of 84.13 m and 75.59 m the minimum operating level for regulation of power flows. Inflow into this reservoir was estimated from measurements made at a gauging station near Senchi, a town midway between Akosombo and Kpong. Following the regulation of flow since the commissioning of Akosombo in 1964, natural monthly runoff has been synthesised based on operation records at Akosombo with adjustments for rainfall and evaporation. This reservoir has three major tributaries: White Volta, Black Volta and Oti river on which gauging stations are located. Gauging stations are also located at the periphery of the reservoir.

Computation of inflows into the Volta Lake is one of the critical functions of the reservoir management group. This function includes prediction of flows into the lake especially in months preceding attainment of peak levels at the dam site. This activity though conducted year round attains heightened importance during the 'wet season' of the year due to several reasons including flood prediction/protection, power and energy studies relating to optimum operations over the next hydrologic year as well providing the necessary data base for negotiations concerning contractual obligations to VRA's international customers.

Traditionally, prediction of inflows has been effected using various methods including graphical adaptive forecasts and regression equations relating rainfall and runoff at upstream gauges to measured flow at Akosombo. These methods involved plotting discharges on graph paper, then extrapolating curves drawn through the resulting points, and then updating the forecasts by modifying the extrapolations by eye. In the case of the regression method, equations used are based on data available as at 1975. At worst these equations will have to be updated in light of more current data and coded into a computer to facilitate use of such equations.

It would however be recommended that some of the more recent methods of forecasting should be employed by way of building a simple deterministic / stochastic model which ideally would incorporate as many physically-based parameters as possible. Some consideration can be given to utilising rainfall / runoff data available for areas including the catchments of the three main tributaries. By way of data availability, daily and monthly streamflow records are available for gauging stations located on

these tributaries. Monthly rainfall summaries are also available for various stations across the country. The length of available data varies from as short as 42 years to over a hundred.
The above method of forecasting is being suggested for various reasons including the ease of not only setting up the model in question but also ease of updating relevant parameters of the model(s). This model will, in conjunction with existing methods of forecasting in VRA, , provide convenient computational methods suitable for implementation on Personal Computers.

It is believed that the development of such model(s) could be attained using some of the well established algorithms from Box and Jenkins [1976] as well the extensive work done by personalities like Hipel [1977], McLeod [1977], Kottegoda [1980].

2. THE VOLTA RIVER SYSTEM

2.1. Catchment Basin

The Volta river basin is situated between 6°N - 14°N latitudes and 5°15'W - 2°10'E longitudes, and lies in six countries namely Ghana, Bourkina Faso, Togo, Bénin, Mali and Côte D'ivoire. Most of the upper catchment is in Bourkina Faso and the lower catchment in Ghana. A sizeable portion of one of the sub-catchments is in Togo.

The catchment area of the Volta River is 398 373 km² of which 163,382 km² are in Ghana. The principal tributaries to the Volta River proper are the White Volta, Black Volta and the Oti Rivers with individual catchment areas of 104 753, 148 498 and 72779 km² respectively. The sources of the Black Volta and the White Volta are in Bourkina Faso flow generally south through Ghana joining to form the Volta proper at about 483 km upstream from its mouth. The Oti river however rises in Bénin, runs through the northern part of Togo, joins the Volta proper about 250 km downstream of the confluence of the White and Black Volta rivers. The Volta river itself runs generally south-east across the southern part of Ghana joining the Gulf of Guinea about 89 km east of Accra, the capital. At the Akosombo dam site however, the area of the Volta basin is 396 503 km^2.

2.2. Climate

The climate in the Volta River basin is dictated by the movement of two main principal air masses, the southerly monsoons and the northerly trade winds. The boundary where these two air masses converge is known as the Inter-Tropical Boundary (ITB) or Inter-Tropical Convergence Zone (ITCZ). The monsoon air masses originate from the South Atlantic Ocean anti-cyclone and generally have a long history of sea track before reaching the West African coast. Since the coast is to the north of the South Atlantic cyclone, the air masses cross the coast from the south-east and turns to the south-west on meeting the north-east trade winds (Sahara anti-cyclone) to the north of the ITCZ. In general the south-easterly wind is heavily moisture laden and convectively unstable on its northward pass over Ghana. The Northerly trades, which originate from the Azores anti-cyclone extending over the Sahara desert, have a long history of desert track before reaching the Volta Basin area as a north-easterly wind known locally as the Harmattan.

2.2.1. Rainfall

The average annual rainfall for the Volta basin as a whole is about 1100 mm with a complex distribution governed by number of factors which have a wide range of movement. In the northern areas where the effect of the monsoons is least, mean annual rainfall is about 650 mm while it can be as high as 1500 mm in the south. The rain belt is closely associated with the seasonal movement of the ITCZ, following it in a northerly direction from April to mid-September and preceding it in a southerly direction from mid-September to December. As a result, the northern portion of the basin experiences only one rain season when the ITCZ reaches the apogee of its northward pass while the lower, southern belt of the basin experiences two rain seasons reaching maxima in June to mid-July and mid-September to October, the latter season being invariably drier than the former.

2.2.2. Temperature

The mean temperature over the basin shows only moderate variations, varying from about 26°C near the coast (latitude 5°N to 6°N) to about 29°C in the far north of the catchment. Variations in the mean temperature from year to year is also small. Seasonal change in average daily temperature is only 3 to 6°C. The maximum daily temperature occurs in March preceding the rain season and the minimum temperatures in August. The mean daily range of temperatures varies from about 7°C in the coastal areas to about 11°C in the north.

2.2.3. Evaporation

Evaporation values have been estimated using various methods including Penman's method. As shown on Figure 2.1, these vary from about 1300 mm (51.18 in.) on the coast to 2000 mm (78.74 in.) in the north. Over the Volta reservoir itself it has been estimated that annual evaporation loss from the reservoir surface is between 1400 mm (55 in.) and 1780 mm (70 in.).

2.2.4. Surface Wind

Surface winds are generally very low with average speeds varying from 8 km/h to 16 km/h on the coast to about 8 km/h inland. The maximum gust associated with thunderstorms and line squalls has been observed up to 110 km/h at Tamale and 105 km/h at Accra.

2.3. Hydrology

The hydrology of the Volta river is best characterised by the hydrology of its main tributaries, the White Volta, Black Volta,

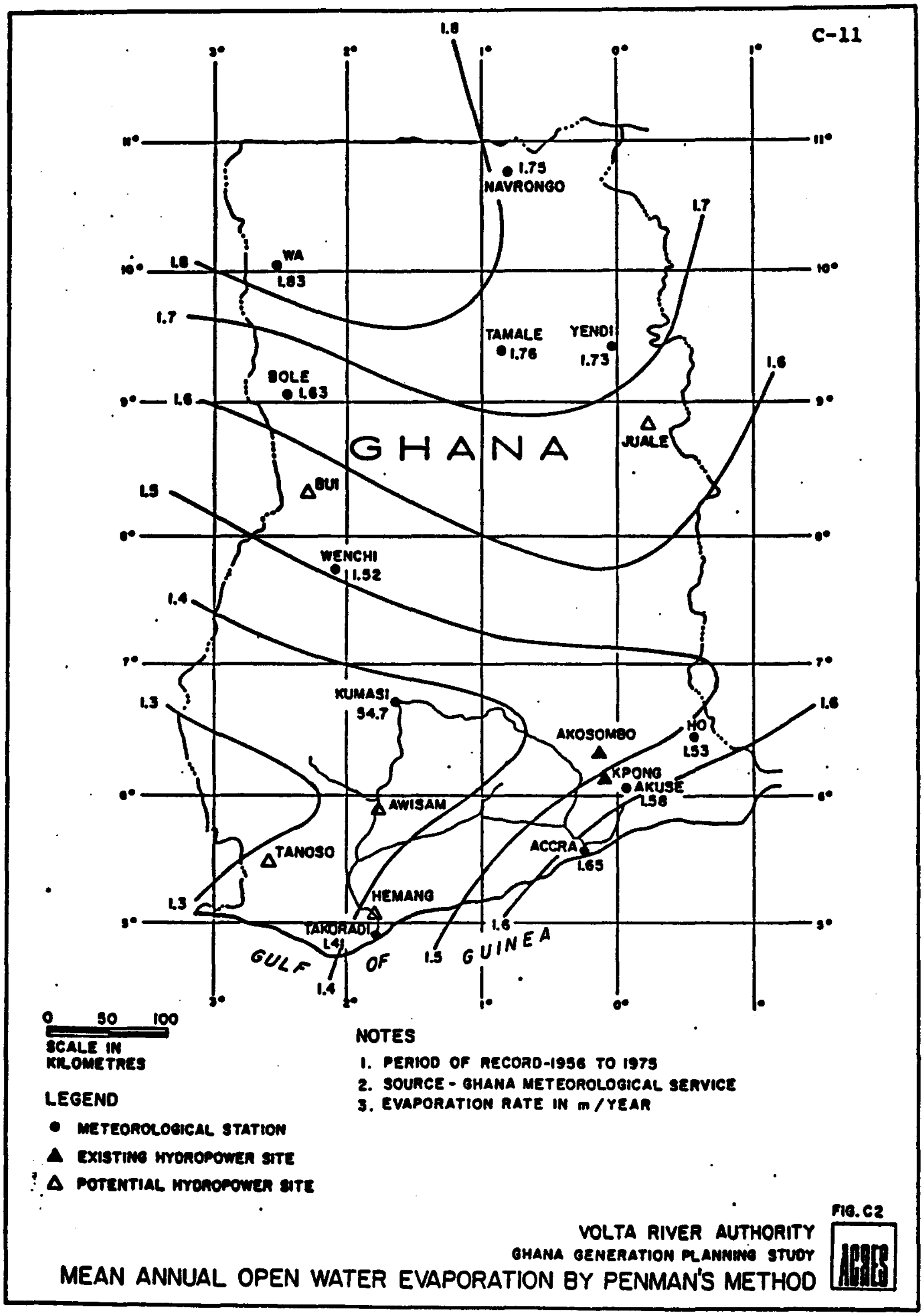


Figure 2.1 – Mean Annual Evaporation
(Reproduced with permission of Acres International Ltd.)

Oti river and the Afram river.

Mean annual flow for the Black Volta at Bui is about 200 m^3/s, White Volta at Yapei is about 300 m^3/s (estimate) and Oti at Saboba 272 m^3/s. This is a total of 772 m^3/s compared with the average annual flow of 1183 m^3/s measured at Akosombo.

3. EXISTING FORECASTING SYSTEM

3.1. Introduction

The existing flow forecasting system is essentially spawned from work done on a flood management system designed primarily to minimise flood risk levels at Kpong during construction of Kpong Hydroelectric Station. It was also meant to improve general effectiveness of a flood management system at Akosombo. It may be worthy of mention here that the largest peak flow of the Volta river in recorded history occurred at Akosombo during construction of the dam in 1963.

3.2. HISTORICAL AND REAL-TIME DATA COLLECTION

The above mentioned flood management program identified five rainfall stations located in the lower reaches of the drainage basin as well as two flow stations whose data was judged as being reliable and pertinent to prediction of flows at Akosombo. See Figure 1.

Rain Stations:

- **Kintampo**
- **Bui**
- **Kpando Prison**
- **Ho**
- **Kete Krachi**

Flow Stations:

- **Bui**
- **Saboba**

Data from the above stations are collected by various local and governmental agencies.

3.3. DATA TRANSMISSION

Transmission of data from the collection sites to the central offices is mostly done by telephone and hardcopy. Processed data is then sent on to head offices of organizations concerned which are mostly located in the capital Accra.

Personnel from VRA however also visit data collection stations regularly to obtain hardcopies of data pertinent to flow prediction at Akosombo. The frequency of collection is increased during the wet season when data is required at shorter time intervals.

Plant and reservoir data from hydrostations are transmitted by

telephone to the VRA hydrologist at the Engineering office which is located at the Kpong Hydroelectric station at Akuse.

3.4. DATA BASE MANAGEMENT

The Engineering department of the Volta River Authority has responsibility for, among other things, the management of pertinent data relating to system planning and operation.
The Civil Engineering section of the above department is entrusted with the hydrometeorological data base management which includes processing, organization and storing of incoming data. The data resides on files in the form of hardcopies as well as on magnetic media for computers. This facilitates access to data for applications including forecasting and power and energy studies.

3.5. FORECASTING PROCEDURE

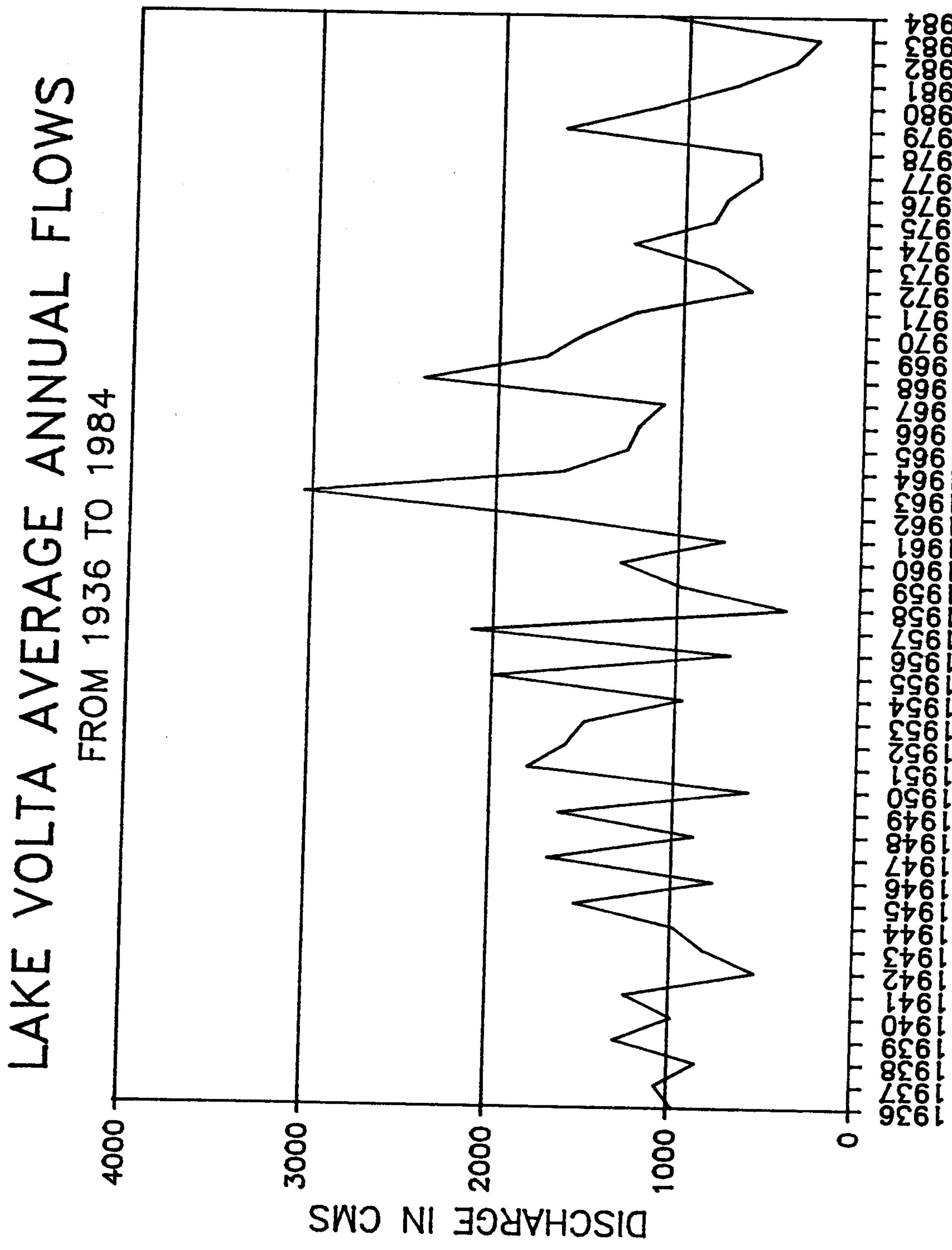
Prediction of inflows has been effected using various methods including graphical adaptive forecasts and regression equations relating rainfall and discharge at the previously mentioned seven stations to measured flow at Akosombo.
The previous method involves plotting discharges on graph paper, then extrapolating curves drawn through the resulting points. The forecast is updated by modifying the extrapolations by eye.

Regression equations were derived which related a combination of rainfall and runoff at selected upstream stations with discharge at Akosombo for various lag periods ranging from one to five months. The equations used are based on data available as at 1975. These equations should be updated in light of the extended recorded data currently available and coded into a computer for ready accessibility.

3.6. FORECAST DISSEMINATION SERVICES

When forecasts are made solely for purposes of flood forecasting, forecast information dissemination to various parties becomes very important since adequate time would be needed in order to pursue the necessary activities to mitigate the effects of floods. A cursory look at the plot of the average annual flow at Akosombo (FIG 3.1) however quickly reveals the fact that the current hydrological regime can hardly be associated with any expectation of floods.
Nevertheless it is still necessary to predict inflow into the Volta reservoir in order to enable power and energy studies to be made. This usually culminates in the operating strategy over subsequent months as well as a draw down policy for the next hydrological year. Fortunately flow predictions and energy studies are undertaken by the same department and therefore the information disseminated is usually in the form of expected

Figure 3.1


LAKE VOLTA AVERAGE ANNUAL FLOWS
FROM 1936 TO 1984
DISCHARGE IN CMS
0
1000
2000
3000
4000
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984

reservoir levels and available energy as well recommended plant operating strategy.

3.7. FORECAST EVALUATION AND UPDATING

Predicted values of inflow, reservoir elevations at Akosombo and expected energy generation are constantly updated as new information becomes available.

Graphical methods of prediction are updated by plotting observed values of flows and thereby correcting extrapolated curves drawn by eye.

Regression equations are more difficult to modify in light of individual deviates from predicted values. The tendency is not to modify regression equations unless they show persistent and consistent errors.

4. STOCHASTIC METHODS

4.1. Introduction

Modelling of climatic phenomena has engaged the attention of people for a long time in view of the importance climate plays in survival or decline of civilizations. In particular modelling of river flow is of considerable importance considering the fact that most civilizations were founded close to available sources of water [Kottegoda, 1980]. Kottegoda [1980] cites for example the waters of the Nile, the rivers of Mesopotamia, the irrigation system of the Indus and China's Yellow River. The use of water is of no less importance today than it has ever been. The many ways in which available water plays important roles in today's civilization include the source of cheap hydroelectric power. From the design stages through operation of a hydroelectric power system, it is necessary to characterise the nature and quantity of inflow into reservoirs so as to provide an efficient design of the reservoirs and also to facilitate efficient operation of the power plant. It is, however, impossible to exactly characterise or model the nature of river flow. This fact is easily seen on examining some of the various processes which interact to produce river flow. River flow is primarily a function of antecedent precipitation. The relationship between these two variables however should take cognisance of river catchment physiography, spacial and temporal distribution of effective rainfall. A host of other factors include antecedent soil moisture conditions, variable evapotranspiration, vegetation, aquifer conditions, land use (interventions) etc. Even if all these processes were fully understood, measurements of these parameters are bound to include errors which are random in nature. Therefore there is a strong argument for modelling river flows as random or probabilistic processes.

4.2. STOCHASTIC PROCESS

A set of observations that measure the variation in time of some aspect of some phenomenon, such as the rate of river flow, water level, dissolved oxygen level in a reservoir is termed a time series [Kottegoda 1980]. These observations can further be described as the various states of this phenomenon at corresponding stages. A time series can consist of either continuous or discrete measurements. However with the ease and speed of computation afforded with the advent of electronic computers which are generally digital machines, continuous measurements are invariably transformed into discrete series for their processing on computers. Events like river flow which can be quantified are termed random or stochastic events since their outcomes are uncertain. A collection or set of observations of actually measured values are termed a stochastic process if these outcomes are uncertain as in the case of river flow. If the set of outcomes is arranged chronologically then it forms a time series. It is important to characterise a set of historical data of measured river flow as one realization out of several equally likely sets of an underlying stochastic process.

4.3. STOCHASTIC MODELS

Stochastic modelling sets out to represent the generating mechanism of the time series including the various trends and periodicities inherent in the series. A family of models now commonly known as Box-Jenkins models are typically used in the formulation of these models.

These models are generally linear models which can be classified into:

1. Stationary models
2. Nonstationary models
3. Seasonal models

4.3.1. Stationary models

A process is said to be stationary if all moments including the mean, standard deviation and higher moments (e.g. μ, σ, g) are independent of the time of observation, in other words its statistical properties are constant over time. This is termed strict (or strong) stationarity.

In reality strict stationarity is not realised and it therefore becomes necessary to assume a weaker form of stationarity. For practical purposes stationarity is often limited to the mean and standard deviation. When a process possesses a constant mean it is said to be first-order stationary. When the variance is also stationary it is said to possess second order stationarity. An example of a data series exhibiting stationary behaviour is depicted in Figure (4.1).

4.3.2. Nonstationary models

Many time series occur which do not appear to have a fixed mean or level. These series often constitute data which does not fluctuate around any apparent state or level and thus the statistical "mean" will have little practical significance. Some of these series nevertheless exhibit some homogeneity in that different parts of the series behave in similar fashion though at different levels and may also show some trends. Others posses no

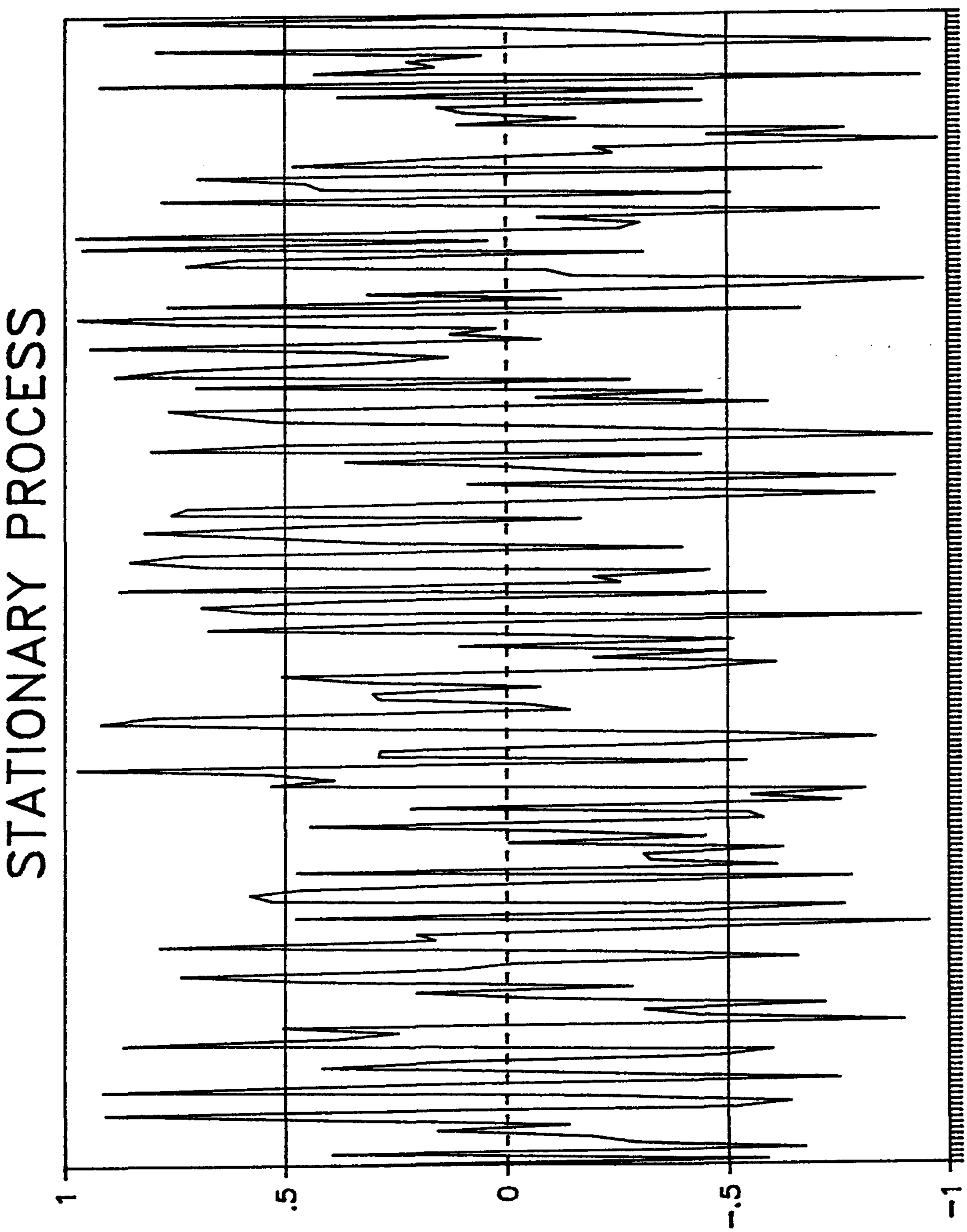


Figure 4.1

apparent trends and have different levels at different time periods. An example of a data series exhibiting nonstationary behaviour is depicted in Figure (4.2).

4.3.3. Seasonal models

Many geophysical phenomena like monthly river flow exhibit periodic behaviour. Periodic effects in hydrological time series are however deterministic in nature [Kottegoda 1980] with regard to their frequency of occurrence since they are caused by cyclic phenomena with fixed periods. The main periodic component is caused by the earth revolving around the sun in the elliptical orbit while itself rotating on an axis which is inclined to the orbital plane. The consequence is the seasonal effect shown in most closely spaced observations like monthly river flow measurements. An example of a data series exhibiting seasonal behaviour is depicted in Figure (4.3). From the plots, one can also observe the changing variance of the data.

4.4. FORMULATION OF STOCHASTIC MODELS

Stationary time series are generally represented by the class of linear models known as the Box-Jenkins family of models. This same formulation can be utilised in modelling non-stationary time

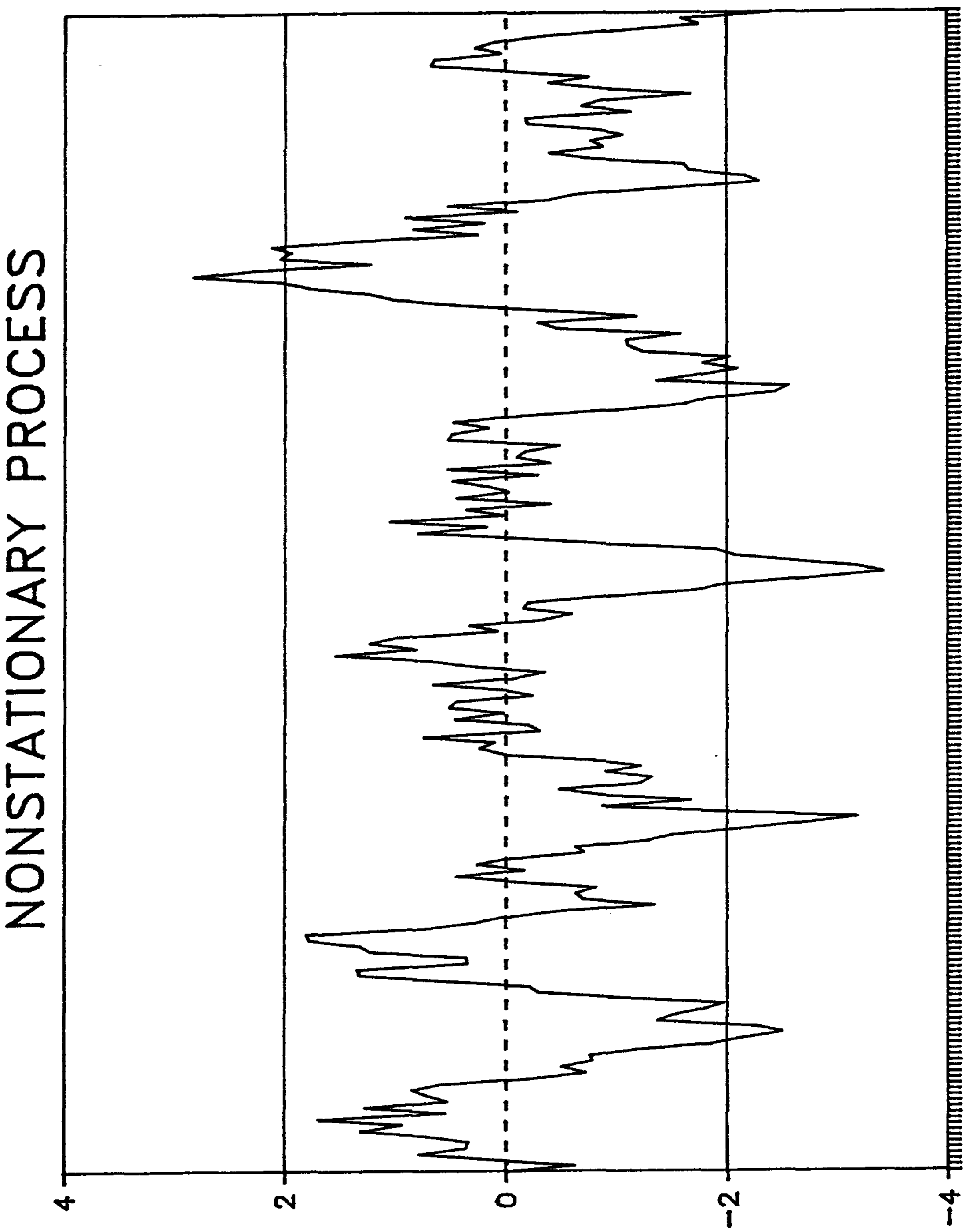


Figure 4.2

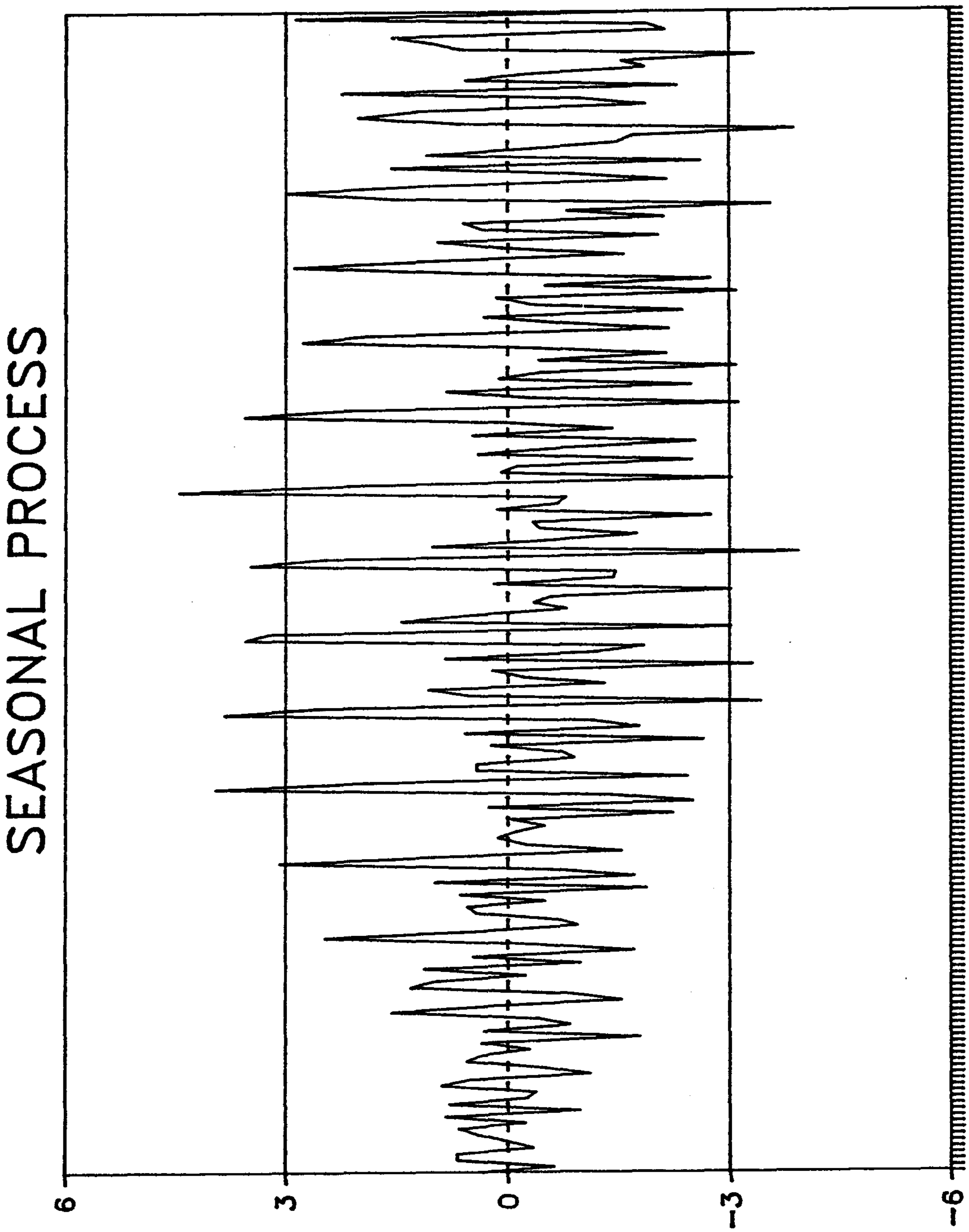


Figure 4.3

series. When dealing with data with nonstationarities such as trends, seasonality, time varying variance (heteroscedasticity), stationarity can be induced by invoking a suitable transformation such as applying a power transformation like the Box-Cox transformation and / or differencing the original data.

Consider a highly dependent time series (ie a series which is serially correlated at one or more lags) represented by the set of observations z_t, z_{t-1}, z_{t-2},... measured at time t, t-1, t-2, This series can be thought of as being generated from a series of independent "shocks" a_t [Box and Jenkins, 1976]. These shocks are assumed to be randomly drawn from a fixed distribution usually assumed to be Normal with zero mean and variance σ_a^2. This sequence of random variables a_t, a_{t-1}, a_{t-2},... is known as a white noise process.

The white noise process a_t is supposed to be transformed into the z_t series by a suitable linear filter of the form which takes a weighted sum of previous values of z_t as shown below

$$z_t = \mu + a_t + \varepsilon_1 a_{t-1} + \varepsilon_2 a_{t-2} + \ldots + \qquad (4.1)$$

$$= \mu + \varepsilon(B) a_t \qquad (4.2)$$

where B is the backward shift operator ie $Bz_t = z_{t-1}$, $B^k z_t = z_{t-k}$ and

$$\varepsilon(B) = 1 + \varepsilon B_1 + \varepsilon B_2^2 + \ldots + \qquad (4.3)$$

In this case μ can be thought of as a parameter that defines the level of the process and $\varepsilon(B)$ is known as the transfer function of the linear filter that transforms a_t into z_t.

4.4.1. AUTOREGRESSIVE MODELS (AR)

In this formulation, the current value of z_t is expressed as a linear aggregate of previous values of the series z_{t-1}, z_{t-2}, etc If z_t, z_{t-1}, z_{t-2}, ..., represent the observations at equispaced times t, t-1, t-2, ... of a process at equilibrium about μ, and denote the deviations of z_t from μ by $\bar{z}$ i.e. $\bar{z}_t = z_t - \mu$, then

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \phi_3 \bar{z}_{t-3} + \dots + \phi_p \bar{z}_{t-p} + a_t \quad (4.4)$$

is called the autoregressive process of order p. The equation

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (4.5)$$

called the autoregressive operator enables us to write the above AR(p) model compactly as

$$\phi(B)\bar{z}_t = a_t \quad (4.6)$$

The above model contains p + 2 parameters (μ, $\phi_1,\dots,\phi_p$, σ_a) which are estimated from the available data. It can be shown that

$$\phi(B)\bar{z}_t = a_t \quad (4.6)$$

is equivalent to

$$\bar{z}_t = \varepsilon(B) a_t \quad (4.7)$$

with

$$\varepsilon(B)^{-1} = \phi(B) \quad (4.8)$$

The parameters ϕ_1, ϕ_2, ..., ϕ_p, of the equation

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \phi_3 \bar{z}_{t-3} + \dots + \phi_p \bar{z}_{t-p} + a_t$$

i.e.

$$(1 - \phi_1 B - \dots \phi_p B^p)\bar{z}_t = \phi(B)\bar{z}_t = a_t \quad (4.9)$$

must satisfy certain conditions in order to be stationary. If we denote the following

$$\phi(B) = 0 \tag{4.10}$$

as the characteristic equation for the autoregressive process, then the stationarity condition for the autoregressive process can be expressed by saying that the roots of the characteristic equation must lie outside the unit circle. Equivalently, the zeroes of the polynomial $\phi(B)$ must lie outside the unit circle. It must be noted that the series

$$\Pi(B) = \phi(B) = 1 - \phi(1)B - \phi(2)B^2 - \ldots - \phi(q)B^p \tag{4.11}$$

is finite so there are no constraints on the parameters of the autoregressive process to ensure invertibility [Box and Jenkins].

4.4.2. MOVING AVERAGE MODELS (MA)

The autoregressive model describes a stationary series as a linear aggregate of p previous values of the series $z_{t-1}, z_{t-2}, \ldots$ plus a random shock a_t. Equivalently it expresses z_t as an infinite weighted sum of a's.

The moving average model expresses z_t as a finite linear function of q previous values of a's. Therefore

$$\bar{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{4.12}$$

is called the moving average process of order q. The equation

$$\theta(B) = 1 - \theta_1 B - \theta_2^2 B - \ldots - \theta_q^q B \tag{4.13}$$

called the moving average operator, enables us to write the above MA(q) model compactly as

$$\bar{z} = \theta(B) a_t \tag{4.14}$$

The above model contains q + 2 parameters (μ, $\theta_1,\ldots,\theta_p$, σ_a) which are estimated from the available data.

The parameters θ_1, θ_2, ..., θ_q, of the equation

$$\bar{z} = a_t - \theta_1\bar{z}_{t-1} - \theta_2\bar{z}_{t-2} - \ldots - \theta_q\bar{z}_{t-q} \tag{4.12}$$

must satisfy invertibility conditions. In order for the moving average representation to be invertible, the roots of

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q = 0 \tag{4.15}$$

the characteristic equation for the moving average process, must lie outside the unit circle. Equivalently, the zeros of the polynomial $\theta(B)$ must lie outside the unit circle.

It is worthy of note that the series

$$\varepsilon(B) = \theta(B) = 1 - \theta(1)B - \theta(2)B^2 - \ldots - \theta(q)B^q \tag{4.16}$$

is finite so there are no constraints on the parameters of the moving average process to ensure stationarity.

4.4.3. AUTOREGRESSIVE-MOVING AVERAGE MODELS (ARMA)

The autoregressive and moving average models can usefully be combined to achieve greater flexibility in the fitting of stationary stochastic models [Box and Jenkins 1976]. These models are of the form

$$\bar{z}_t = \phi_1\bar{z}_{t-1} + \phi_2\bar{z}_{t-2} + \ldots + \phi_p\bar{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{4.17}$$

or

$$\phi(B)\bar{z}_t = \theta(B)a_t \tag{4.18}$$

The above involves (p+q+2) unknown parameters μ, ϕ_1, ..., ϕ_p, $\theta_1, \ldots, \theta_q$, σ_a^2, which are estimated from the actual time series. Box and Jenkins [1976] note that in practice p and q are hardly greater than 2 for AR(p) or MA(q) or ARMA(p,q) models which are used in representing actual time series.

4.4.4. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODELS (ARIMA)

The linear stochastic models shown [AR(p), MA(q) and ARMA(p,q)] are adequate in representing data which are stationary. Many series encountered in practice have some form of nonstationarity or the other. Most series however exhibit homogeneity of some kind. These can conveniently be represented by introducing a generalised autoregressive operator $\varphi(B)$, in which one or more of the zeros of the polynomial $\varphi(B) = 0$ is unity [Box and Jenkins, 1976]. The operator $\varphi(B)$ can then be expressed as

$$\varphi(B) = \phi(B)(1 - B)^d \qquad (4.19)$$

where $\phi(B)$ is a stationary operator. Therefore a general model which will adequately represent homogeneous nonstationary behaviour is of the form

$$\varphi(B)z(t) = \phi(B)(1 - B)^d z(t) = \theta(B)a(t) \qquad (4.20)$$

or

$$\phi(B)w(t) = \theta(B)a(t) \qquad (4.21)$$

where

$$w(t) = \nabla^d z(t) \qquad (4.22)$$

and ∇ = backward difference operator ie $\nabla z(t) = z(t) - z(t-1)$. In practice, therefore, an actual series is differenced d times until stationarity is attained after which a stationary model is fitted to the resulting data. This general model is the autoregressive integrated moving average model ARIMA(p,d,q) of order (p,d,q). For example Figure (4.4) shows a plot of Figure (4.2) after differencing the data of Figure (4.2) once. Comparing the plots Figure (4.2) and Figure (4.4) one can immediately discern that stationarity has been induced in the data both in terms of the level and variance.

4.5. STOCHASTIC MODELLING

It is generally recommended to follow an identification, estimation and diagnostic checking procedure as developed by Box and Jenkins [1976].

4.5.1 Identification

The purpose of identification is to determine the differencing required to produce stationarity and also to ascertain the order of both the seasonal and nonseasonal AR and MA parameters for the time series under consideration [Hipel, McLeod & Lennox 1977]. At this stage it is determined whether the original data series will

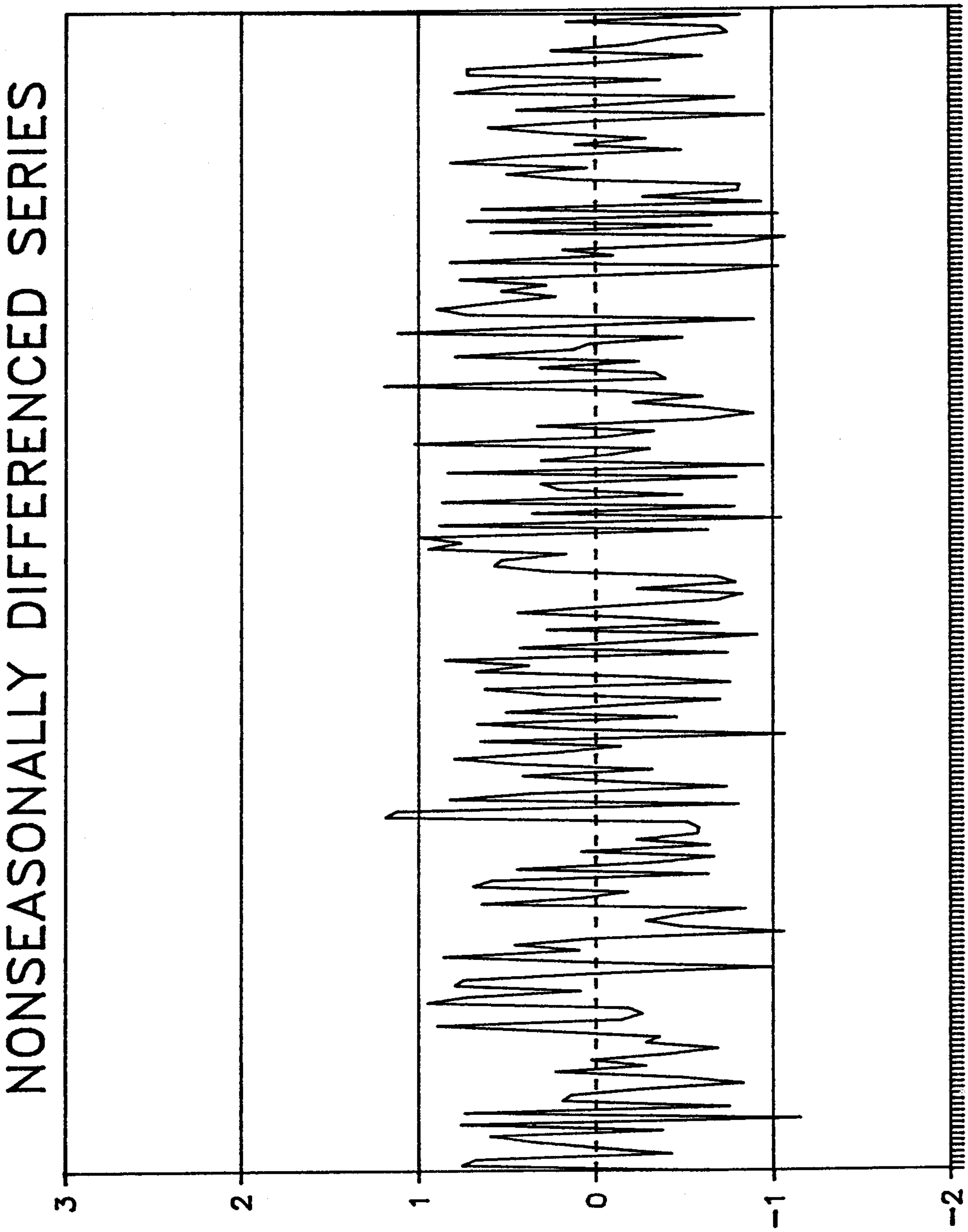


Figure 4.4

need some form of transformation and / or deseasonalization in order to induce stationarity and also normality in the residuals, in order to satisfy the basic computational requirements underlying the theoretical assumptions of the linear stochastic models. In certain cases a transformation (such as the Box-Cox transformation) may change the form of the model to be fitted to the data. This will be evidenced when the diagnostic checks are performed and therefore a proper model could be fitted. It is also recommended by Box and Jenkins [1976] to use at least 50 data points in order to obtain reasonably accurate maximum likelihood estimates of the parameters.

4.5.1.1. Plot of the original series

The plot of the original time series reveals some obvious characteristics of the series. Seasonality, trends in either the mean level or variance, persistence, long term cycles and extreme values can be observed in a visual inspection of this plot. From the plot of the available data (Figure 6.1), the inherent within-year periodic nature of the flows are revealed. Also an examination of the annual flows plot (Figure 3.1) reveals some form of oscillatory behaviour about the mean. This can also be an indication of some cyclical characteristic of the climatic patterns prevailing in the Sahelian zone. In fact Lavender and Anderson [1984] postulate a 30 to 35 year periodicity.

4.5.1.2. Plot of autocorrelation function

The autocorrelation function is a measure of the amount of linear dependence between observations in a time series at fixed time periods apart. The covariance between a value z(t) and another value z(t+k) separated by k time periods apart is defined in terms of the theoretical autocovariance at lag k which is given by

$$\tau_k = \mathrm{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \tag{4.23}$$

From this definition when k=0, the autocovariance becomes the variance of the series ie $\tau_0 = \sigma_z^2$.

Similarly the autocorrelation at lag k is defined as

$$d_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{\{ E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2] \}}} \tag{4.24}$$

$$= \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} \tag{4.25}$$

since for a stationary process, the variance $\sigma_z^2 = \tau_0$ is the same at time t+k as at time t.

Therefore the theoretical autocorrelation at lag k is given by

$$d_k = \frac{\tau_k}{\tau_0} \tag{4.26}$$

from which $d_0 = 1$

The autocorrelation coefficient is therefore dimensionless and also independent of the scale of measurement. The possible range of values of d_k is from -1 to 1 and the value at lag zero is 1.

The plot of autocorrelation versus lag is known as the theoretical autocorrelation function (TACF). The ACF is symmetric about lag zero and therefore in practice it is only necessary to plot the positive lags from lag one onwards against the autocorrelation values.

In practice, the autocorrelation function has to be estimated from the available sample data of the time series. From various methods studied by Jenkins and Watts (1968) it is concluded that the most satisfactory estimate of the lag k autocorrelation d_k is given by

$$r_k = \frac{c_k}{c_0} \tag{4.27}$$

where

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z}), \ k = 0,1,2,3,\ldots,K$$

is the estimate of the autocovariance τ_k, and $\bar{z}$ is the mean of the time series. In practice it is necessary to use at least 50 observations in the estimation of the sample autocorrelation function and then plot values of $r(k)$, $k = 0,1,\ldots,K$ where K is not greater than about N/4 or 5s where s is the seasonality of a periodic series and $5s < N/4$. Procedures for the computation of the standard errors of the ACF are given by Box and Jenkins [1976].

The first step in the use of the above is to examine the plot of the sample autocorrelation function to determine the presence or absence of nonstationarity in the time series. For nonseasonal data, if the sample autocorrelation function fails to damp out quickly this indicates nonseasonal nonstationarity. This therefore indicates a need to difference the time series. For the case of seasonally correlated data with seasonality s, the sample autocorrelation function will follow a wave pattern with peaks at a period s. As in the case of the nonseasonal data, if the sample autocorrelation function fails to damp out at integral multiples of s, then seasonal differencing is required to induce seasonal stationarity. Also if the sample autocorrelation function fails to damp out at values in-between integral multiples of s, then additional nonseasonal differencing may be required to induce stationarity in the data.

After the original data has been differenced accordingly as required, the sample autocorrelation function of the resultant differenced series can be examined to determine the order of the autoregressive (AR) and moving average (MA) parameters needed in order to correctly specify the model.

In some cases no differencing of the data is required. It may also be possible to obtain a series which is white noise in which case the sample autocorrelation function is normally independently distributed with mean zero and variance 1/n i.e. NID(0,1/n). This result enables one to test whether a series is white noise by plotting approximate confidence limits on the sample autocorrelation function to see if a number of the plotted autocorrelation values fall outside the limits. This test will show if a significant number of the values are significantly different from zero. If the series is shown not to be white noise then the following can be used as a guide to identify the type and order of model required.

For nonseasonal models, r(k) truncates after lag q for a moving average process MA(0,d,q). If the r(k) values tail off and do not truncate, this indicates that autoregressive parameters are needed in the model.

In the case of seasonal models, r(k) truncates and is not significantly different from zero after lag q + sQ for a seasonal moving average process MA(0,d,q)x(0,D,Q). If r(k) attenuate at lags that are integral multiples of s then seasonal AR parameters are required in the model. If also the r(k) values attenuate for lags in between the seasonal peaks then nonseasonal AR parameters are required.

4.5.1.3. Plot of partial autocorrelation function

Another useful tool in the identification stage of a stochastic process is the partial autocorrelation function [MacGregor 1986]. The partial autocorrelation function at lag k represents the residual autocorrelation between points separated by k time periods (ie z[t], z[t+k]) after the correlation effects at lags 1,2,..,k-1 have been taken into account. This can also be shown to be the last coefficient ø(kk) in an AR(k) approximation to the stochastic process. A plot of ø(kk) against lag k is known as the partial autocorrelation function.

The partial autocorrelation function (PACF) behaves differently and opposite to the sample autocorrelation function. This is advantageous considering the fact that for an autoregressive process, the sample autocorrelation function tails off infinitely. In contrast, the PACF for a pure nonseasonal autoregressive process AR(p) cuts off after lag p. This means the partial autocorrelation function is nonzero before and at lag p and is significantly not different from zero after lag p. After lag p, the $\phi(kk)$ is approximately NID(0,1/n).

If the partial autocorrelation function fails to truncate after lag p, moving average parameters may be needed in the model. In fact it can be shown that the partial autocorrelation function for a pure moving average process consists of a mixture of sinusoids and exponentials [MacGregor 1986].

In the case of data which exhibits seasonal behaviour, for a seasonal autoregressive process AR(p,d,0)x(P,D,0) the partial autocorrelation function cuts off and is not significantly zero after lag p + sP. After lag p + sP, the $\phi(kk)$ is approximately NID(0,1/n).

If the partial autocorrelation function $\{\phi(kk)\}$ exhibits a pattern which damps at lags that are integral multiples of s then seasonal MA parameters are required in the model. If also the $\phi(kk)$ values fail to truncate at lags in-between the seasonal lags then nonseasonal MA parameters are required.

4.5.2. Estimation

After the identification stage a tentative formulation of the model is arrived at. The estimation process is then to provide efficient estimates of the parameters after which diagnostic and goodness of fit tests are performed to verify the chosen model. It is important that efficient estimates are arrived at else the diagnostic checks may invalidate a model though the model form is appropriate [Box and Jenkins 1976]. Box and Jenkins suggest that the approximate maximum likelihood estimate (MLE) for the autoregressive integrated moving average model (ARIMA) parameters be obtained by using the unconditional sum of squares method. With this method the unconditional sum of squares function is minimised to obtain the least squares estimates of the parameters.

Various optimization techniques exist for the minimization of the unconditional sum of squares. During the estimation stage, values are computed for the AR and MA parameters unless the exact values are known beforehand. This is usually the case for the estimation of the mean of the series [Hipel et al. 1977]. If the data has been differenced either seasonally or nonseasonally then the mean value is assumed to be zero unless a trend component is to be incorporated into the model.

4.5.3. Diagnostic checks

After the identification and parameter estimation stages have been completed, one arrives at a tentative model to represent the original or transformed time series. This tentative model is then subjected to diagnostic checks to determine the "goodness of fit" and adequacy. It is important to carry out this step of the model building in order to find out whether the major assumptions of the model appear to be valid [Hipel et al.]. The assumption of independent identically distributed (white noise) residuals or innovations is very important if the model is to be used for simulation and/or forecasting. Simultaneously the estimated parameters could be inefficient when the above conditions are not fulfilled. The model order is verified through a technique called overfitting but the other assumptions are mainly done through tests on the estimated residuals. If found to be inappropriate changes are made to the model accordingly.

4.5.3.1. Overfitting

Overfitting is the term given in Box and Jenkins [1976] to the situation where after arriving at a model believed to be the correct one, a more elaborate model is fitted to the data. This is done especially when the direction in which the model is likely to be inadequate is known or suspected. Caution needs to be exercised when adding parameters to either side of the ARMA equation. In particular, care must be taken to avoid parameter redundancy by adding parameters to both sides of the equation simultaneously. This procedure can be employed if the autocorrelation function appears to have significant values at certain lags. If the maximum likelihood estimate of the extra parameter fitted has a magnitude three or four times its standard error this could indicate that a more elaborate model is needed to adequately describe the process.

4.5.3.2.. Independence of residuals

The residuals or innovations are assumed to be normally independently distributed with zero mean and variance $\sigma(a)$. Tests done on the residuals to validate this assumption constitute one of the main verification procedures for the identified model. A visual inspection of the residuals is recommended since it could immediately reveal any discrepancies that may exist.

An important verification procedure is to plot the residual autocorrelation function (RACF) to check the independence criterion. If the residual series is white noise the residual autocorrelation function would be expected not have any values significantly different from zero i.e. not to be autocorrelated. Further the residual autocorrelation function would be normally independently distributed with zero mean and variance $1/N$. The approximate standard error can then be computed from the square root of the variance. A plot of the autocorrelation function with approximate significance limits would reveal if a significant number of the autocorrelations lie outside the chosen confidence limits. Box and Jenkins [1976] refer to the work of Box and Pierce [1970] where it is pointed out that at low lags the residuals tend to be highly correlated and the standard errors tend to be much less than 1/N. There could be an underestimation therefore of the statistical significance of apparent departures from zero of the autocorrelation at low lags. They explain that this method can however be used for higher lags.

Another technique which is suggested to address the above difficulties is the Box-Pierce portmanteau lack-of-fit test. This method seeks to take into account the joint effect of say the first 10 to 25 values of the serial correlations r(l) of the sequence of estimated residuals. It can be shown that if the model is appropriate,

$$Q = n\sum_{l=1}^{K} r_l^2 \tag{4.28}$$

is approximately distributed as a Chi-squared distribution with (K-p-P-q-Q) degrees of freedom and n = N - d is the number of data points used in fitting the model. Since this method has low power for small samples n should be 15 to 25 for nonseasonal data and 4s for seasonal data. In practice the chi squared value computed from the data is compared to the actual value read from a table for a chosen significance level. If the computed value is

higher than the tabulated value the identified model is rejected and appropriate changes made to it.
An alternative method commonly used especially in the case of seasonal data is the cumulative periodogram of the residuals. As pointed out by Hipel et al. [1977] however, this test is known to be inefficient in the case of residuals and that the cumulative periodogram often fails to indicate model inadequacy due to dependence of the residuals unless the model is a very poor fit to the available data.

4.5.3.3. Homoscedasticity of residuals

McLeod (1974) has proposed a test for checking whether the constant variance (homoscedasticity) assumption for the residuals is valid. As mentioned before, the residuals are assumed to follow a normal independent distribution with constant variance. If the residuals are found to be heteroscedastic then an appropriate procedure such as the Box-Cox transformation can be applied to the data and in most cases this is sufficient to rectify the problem.

4.5.3.4. Normality of residuals

The normality assumption for the residuals can be confirmed through several methods. The most direct fashion is to compute the skewness and kurtosis of the residuals. The skewness g_1 is computed from the residuals a(t) as follows

$$g_1 = \left(\frac{1}{n}\sum_{t=1}^{n} \hat{a}_t^3\right) / \left(\frac{1}{n}\sum_{t=1}^{n} \hat{a}_t^2\right)^{3/2} \qquad (4.29)$$

where g_1 is approximately N (0, 6/n).

The kurtosis g_2 is computed from the residuals a(t) as follows

$$g_2 = \left(\frac{1}{n}\sum_{t=1}^{n} \hat{a}_t^4\right) / \left(\frac{1}{n}\sum_{t=1}^{n} \hat{a}_t^2\right)^{2} - 3 \qquad (4.30)$$

where g_2 is approximately N (0, 24/n).

If the residuals are normally distributed they would posses no significant skew or kurtosis. If these statistics are found to be significant then a suitable Box-Cox transformation would usually rectify the situation.

4.5.4.. Box-Cox Transformation

As can be inferred from the above a suitable power transformation is sometimes required to aid in the identification of a time series. In Box-Jenkins modelling, as the above linear stochastic models are often called, the residuals are assumed to be independent, homoscedastic and usually normally distributed. The independence assumption being the most important for the correct specification of the identified model. The constant variance and independence assumptions are required for the computation of efficient parameter estimates.

For practical purposes it is advantageous to satisfy the normality assumption in order to be able to compute confidence intervals for forecasted data. This is achieved by transforming the original data series by a suitable power transformation as the Box-Cox shown below as follows

$$z_t^{\delta} = \delta^{-1} [(z_t + cons)^{\delta} - 1] \quad \delta \neq 0 \qquad (4.31)$$

$$z_t^{\delta} = \ln (z_t + cons) \quad \delta = 0 \qquad (4.32)$$

where cons = constant.

The transformation described above can be specified in various ways. Sometimes it is known in advance the appropriate value of δ is for a particular type of series. In such cases the value of δ is specified before the estimation stage. Some data series are normalised when they undergo square root transformation ($\delta = 0.5$) or natural logarithm ($\delta = 0.0$) as in the case of average monthly river flow. In this case the value of the constant is set to zero if there are no zero values present in the series. If some of the data have zero value then the constant is set to a small positive value in order to take the natural logarithm.

There are cases however where the standard transformations of square root and natural logarithm do not result in the necessary normal and homoscedastic distributions and therefore require that the maximum likelihood estimate of δ be computed. This is usually done at the estimation stage of the model building. In such cases, the value of the constant is set to a value such that all the data in the series assume positive values and the maximum

likelihood value of the value of δ is estimated.
The computation of the maximum likelihood value of the value of δ involves a significant increase in computer time compared to when the best value of δ is not computed. It is therefore not recommended to do this computation unless the diagnostic checks reveal inadequacies in the model with respect to the normality and homoscedasticity assumptions. To further economise on computer time or in situations where the facility for estimating the maximum likelihood value of the value of δ is not available then an approach can be adopted where different values of δ are used and then the most appropriate one is selected.

4.5.5. Akaike information criterion (AIC)

Box and Jenkins [1976] emphasise the need to use as few parameters as possible i.e. the model should be parsimonious. Certain situations also arise where there are competing models for the same set of data. In such cases one needs to discriminate between the models to arrive at the one which performs best under the diagnostic checks. One mathematical formulation of the parsimony requirement is the Akaike information criterion (AIC) which is given by [Hipel, 1981] as

$$AIC = -2\ln ML + 2k \qquad (4.33)$$

where ML denotes maximum likelihood and k the number of independently estimated parameters within the chosen model. It will be noted that this formulation includes a term for good statistical fit (first term on right hand side) and also an inclusion for the parsimony requirement (second term on right hand side). This formula allows for the selection of a "best" model from alternatives by calculating the AIC for the various models and then choosing the one with the minimum AIC. This is referred to as the minimum AIC estimation (MAICE). Although sometimes a "best" model can be obtained from the estimation and diagnostics stage, sometimes it becomes necessary to entertain two or three models. The final one can then be selected using the MAICE formulation.

5. FORECASTING SEASONAL TIME SERIES

5.1. Introduction

Many geophysical phenomena exhibit cyclic behaviour of one form or another. Average monthly river flow measurements constitute an example of a time series which shows a periodicity of twelve, corresponding to the number of months in a year. The methods described for fitting linear stochastic models to nonstationary data can be extended to the case of seasonal data. Various methods exist and are described in water resources literature. Hipel [1981] and Noakes, McLeod & Hipel [1985] describe various methods applicable to modelling seasonal time series. These methods usually involve fitting a separate autoregressive model to each season in the period, removing the seasonal component of the time series and thereafter fitting an appropriate form of Box-Jenkins type of model or differencing the data series seasonally to induce stationarity.

5.2. MODELLING MONTHLY RIVER FLOW SERIES

The original use of seasonal models applicable to monthly time series is credited to Thomas and Fiering [1962]. Later developments in research has led to a number of seasonal models. The three main seasonal models are:

1. Periodic Autoregressive model (PAR)
2. Deseasonalized model
3. Seasonal Autoregressive Integrated Moving Average (SARIMA)

5.2.1. PERIODIC AUTOREGRESSIVE MODEL (PAR)

The periodic autoregressive model (PAR) which is also known as monthly autoregressive model Hipel [1981], is basically a system of taking cognisance of the fact that the autocorrelation structure between different seasons within a year may be different from season to season. It also affords a way of avoiding differencing the data in order to induce stationarity.
The periodic autoregressive model (PAR) model is simply formulated by fitting a separate autoregressive (AR) model to each month of the year Noakes, McLeod & Hipel [1985]. Therefore one arrives at essentially s different independent models each of which describe a different season within the year where s is the periodicity of the available data. It has been noted that though

the above technique could theoretically be extended into a periodic ARMA (PARMA) formulation Noakes, McLeod & Hipel [1985], introduction of moving average parameters into the model also introduces difficulties in obtaining maximum likelihood estimates of model parameters.

5.2.2. DESEASONALIZED MODELS

This approach is commonly used to model seasonally varying geographic time series. The methodology involves removing the seasonality inherent in the data by introducing a deterministic component into the model and then modelling the resulting nonseasonal series by means of a nonseasonal autoregressive moving average model (ARMA).

The data may be transformed using an appropriate power transformation like the Box-Cox transformation to remove nonnormality and correct any heteroscedasticity present in the residuals of the ARMA model which is fit to the deseasonalized data. To deseasonalize the data series the following standard equations have been used

$$w_{i,j} = Z_{i,j}^{(\delta)} - \mu_j \tag{5.1}$$

and

$$w_{i,j} = (Z_{i,j}^{(\delta)} - \mu_j) / \sigma_j \tag{5.2}$$

where $Z_{i,j}^{(\delta)}$ = the transformed observation for the i^{th} year and j^{th}

month, μ_j is the fitted mean for month j, σ_j is the fitted standard deviation for month j, and δ is the exponent of an appropriate Box-Cox transformation applied to the data. This process of transformation is often referred to as prewhitening. The first transformation (5.1) is applicable in situations where monthly means tend to be different from month to month but the monthly variances are approximately the same across the months within a year. In situations where both means and variances of the z series change from month to month, the second transformation (5.2) is then more appropriate. Hipel [1981] gives the AIC formula for the deseasonalized model.

5.2.3. SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODELS

This model formulation is also known as the multiplicative seasonal model and is described in detail by Box and Jenkins [1976]. The seasonal autoregressive integrated moving average (SARIMA) model follows directly from ideas introduced in the formulation of nonseasonal linear stochastic models. For a time series of periodicity s, it is easily seen that data sets s time periods apart tend to behave in similar patterns. It therefore becomes necessary to introduce operators which act on data s time periods apart. The multiplicative model takes seasonality into account by including operations on data separated by one time period as well as s time periods apart. Consider a seasonal time series z(t), z(t-1), ... occurring at time periods t, t-1, ... where data points s time periods are similar in pattern. The data s time periods apart can be related by a model of the form given by

$$\Phi(B^{s})\nabla_{s}^{D} z_{t} = \Omega(B^{s})\alpha_{t} \tag{5.3}$$

where s = 12 for monthly river flows, $\nabla_s = 1 - B^s$, $B^s z_t = z_{t-s}$ $\Phi(B^s)$,$\Omega(B^s)$ are polynomials in B^s of degrees P and Q respectively and also satisfy conditions of stationarity and invertibility as explained in Box and Jenkins [1976]. If similar models are specified for each of the s periods in the year or cycle, it would seem reasonable to assume that values of parameters which form the polynomials Φ and Ω would be similar for different months. Considering equation (5.3) it is easily seen that the error terms (α's) will be correlated with each other since equation (5.3) takes only seasonal correlation into account. Therefore to account for residual correlation between the α's, a second model of the form below is postulated

$$\phi(B)\nabla^{d} \alpha_{t} = \theta(B)a_{t} \tag{5.4}$$

where a_t is now a normally independently distributed (white noise process), and $\phi(B)$,$\theta(B)$ are polynomials in B of degrees p and q respectively and also satisfy conditions of stationarity and invertibility respectively. The ∇ in (5.4) is also $\nabla = 1 - B$. Substituting for α in equation (5.4) from (5.3) gives

$$\phi_{p}(B)\Phi_{P}(B^{s})\nabla^{d}\nabla_{s}^{D} z_{t} = \theta_{q}(B)\Omega_{Q}(B^{s})a_{t} \tag{5.5}$$

which is the so called general multiplicative seasonal model of order $(p,d,q) \times (P,D,Q)_s$.

Box and Jenkins [1976] postulate that seasonalities of multiplicity greater than two can be modelled along similar lines by extending the arguments outlined above.

In practical applications of the seasonal multiplicative model the available data is differenced seasonally enough times to induce seasonal stationarity and then again differenced nonseasonally to induce stationarity. The combined operator for seasonal and nonseasonal differencing is

$$w_t = (1 - B)^d (1 - B^s)^D z_t^{(\delta)} \qquad (5.6)$$

where δ is the exponent of a suitable Box-Cox transformation which has been applied to the raw data. For the case of monthly river flow it has been observed that a lognormal transformation (corresponding to $\delta = 0$) is usually sufficient to correct any problems due to nonnormality and / or heteroscedasticity inherent in the data. To illustrate the effect of the seasonal differencing operation the seasonal data plotted in FIG(4.3) is seasonally differenced once and the resultant series plotted in FIG(5.1). From this plot it can be seen that seasonal variation of the variance is more or less confined to constant limits.

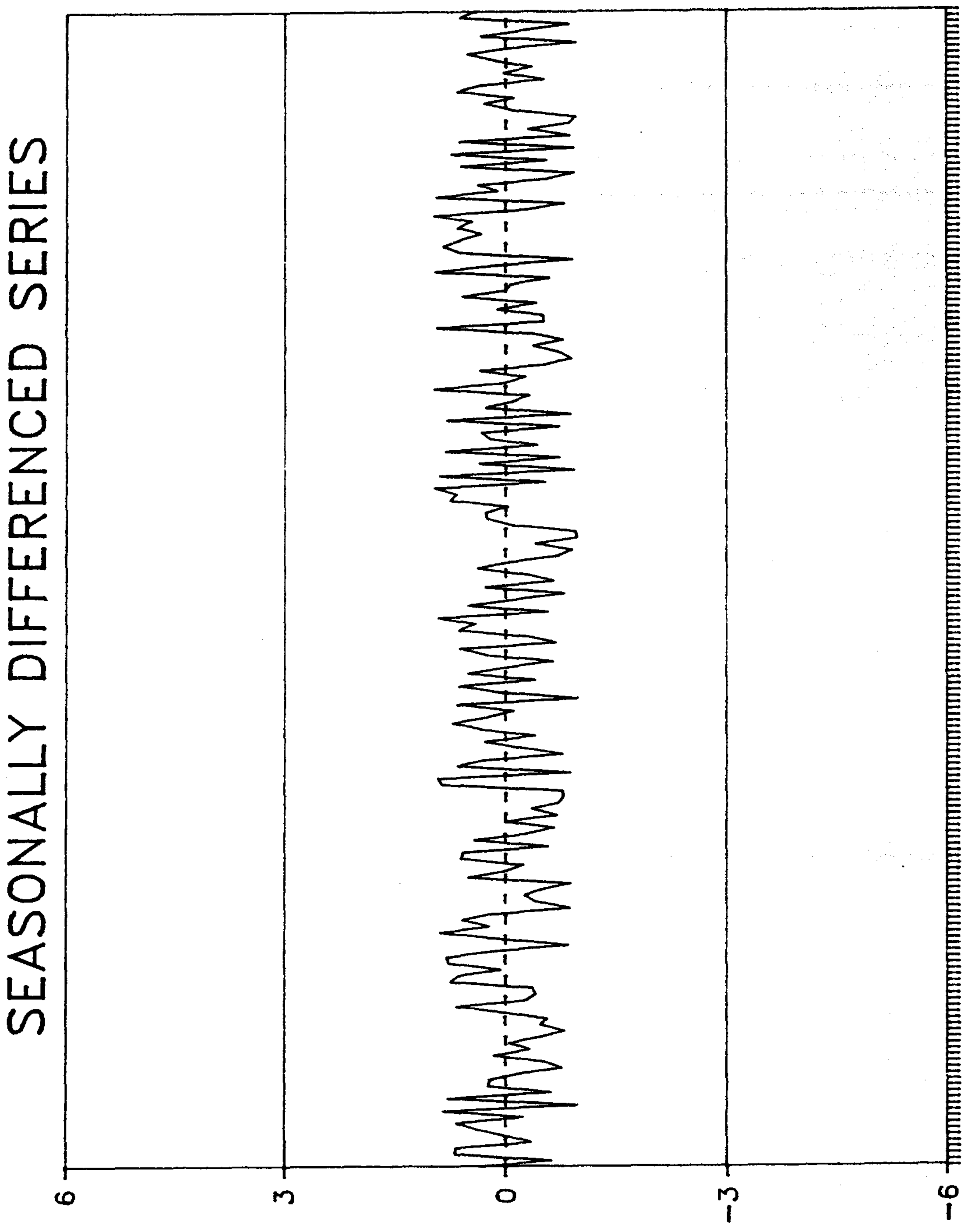


Figure 5.1

5.3. FORECASTING MONTHLY RIVER FLOW SERIES

5.3.1. Introduction

After identifying a suitable model to adequately represent a set of real data, diagnostic checks are performed to test that the model satisfies the basic assumptions and criteria underlying the formulation of the model namely, stationarity, normality of residuals and constant variance. Estimates are computed for the parameters included in the model to a sufficient degree of accuracy to ensure that a good fit is obtained.

The model thus identified can be used to generate synthetic sets of data which bear a statistical resemblance to the original data. These sets of data can be used to test the design and planning of water resource systems by simulating different possible inputs into the various alternative designs.

The identified model can also be used to forecast future inflows into a hydroelectric system with a view to finding the optimal policy which maximizes the hydrological output subject to physical, economical, and political constraints [Hipel 1985].

Box and Jenkins [1976] show how forecasts can be directly computed from a linear stochastic model once a suitable form of model has been identified and its parameters estimated.

5.3.2. THREE FORMS OF THE ARIMA MODEL

To show how minimum mean square error forecasts are obtained it is useful to consider the three explicit forms in which the general autoregressive integrated moving average model can be expressed. The current value z(t) of a series described by the ARIMA(p,d,q) model of the form (4.20)

$$\varphi(B)z_t = \theta(B)a_t \tag{5.7}$$

where $\varphi(B) = \phi(B)\nabla^d$, can be expressed in the following three explicit forms

a. in terms of previous values of z's and current and previous values of a's, by the use of the difference equation.

b. in terms of current and previous shocks a(t-j) only

c. in terms of a weighted sum of previous values z(t-j) of the process z and the current shock a(t).

5.3.2.1. DIFFERENCE EQUATION FORM OF ARIMA MODEL

Recalling equation (4.17)

$$\bar{z}_t = \phi_1\bar{z}_{t-1} + \phi_2\bar{z}_{t-2} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \tag{5.8}$$

and also equation (4.20)

$$\varphi(B)z(t) = \phi(B)(1 - B)^d z(t) = \theta(B)a(t) \tag{5.9}$$

a general difference equation for z_t can then be written as

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \dots + \phi_{p+d} z_{t-p-d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \tag{5.10}$$

This form of the model therefore allows for computation of the current value of z(t) using previous values of z as well as the current value of the random shock a(t) and previous values of a's. Box and Jenkins [1976] emphasise that the above formulation of the ARIMA(p,d,q) model is useful for many purposes especially in the computation of minimum mean square error forecasts.

5.3.2.2. RANDOM SHOCK FORM OF ARIMA MODEL

This form of the model computes z(t) as a weighted sum of the current random shock a(t) and previous values of a's. From (4.1) the current value of z(t) can be expressed as

$$z_t = \mu + a_t + \varepsilon_1 a_{t-1} + \varepsilon_2 a_{t-2} + \dots + \tag{5.11}$$

or

$$(z_t - \mu) = a_t + \varepsilon_1 a_{t-1} + \varepsilon_2 a_{t-2} + \dots + \tag{5.12}$$

$$= a_t + \sum_{j=1}^{\infty} \varepsilon_j a_{t-j}$$

$$= \varepsilon(B) a_t \tag{5.13}$$

However since we are considering nonstationary processes whose models include a differencing parameter $d \geq 1$, μ can be taken to be effectively zero. The z values can therefore be used in place of deviations from the mean μ which in this case is equal to zero. The formulations arrived at are however valid for cases where d = 0 if deviations of z from the mean are substituted for values of z in the equations arrived at. Equation (5.12) can therefore be written as

$$z_t = \varepsilon(B)a_t \tag{5.14}$$

which is a form expressing the process z(t) as an output of a linear filter whose input is white noise or a series of uncorrelated shocks a(t). This form of the model is also useful in several ways but its main function with respect to forecasting is that it allows for computation of the variance of generated forecasts.

It is also useful to express the random shock form of the ARIMA process as a sum of a weighted finite sum of t-k current and previous shocks occurring after some reference point k, and a complementary function $C_k(t-k)$ as:

$$z_t = C_k(t-k) + \sum_{j=k+1}^{t} \varepsilon_{t-j}a_j \tag{5.15}$$

from which the complementary function is seen to be the truncated infinite sum

$$C_k(t-k) = \sum_{j=-\infty}^{k} \varepsilon_{t-j}a_j \tag{5.16}$$

The complementary function of (5.16) is later shown to be the minimum mean square error forecast of z(t) made at time origin k.

5.3.2.3. INVERTED FORM OF ARIMA MODEL

Since $\varepsilon(B)$ can be treated as an algebraic operator (5.14) can therefore be expressed as

$$\varepsilon^{-1}(B)z_t = a_t \qquad (5.15)$$

or

$$\pi(B)z_t = (1 - \sum_{t=1}^{\infty}\pi_j B^j)z_t = a_t \qquad (5.16)$$

z_t can then be expressed as

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-2} + \ldots + a_t \qquad (5.17)$$

which is an infinite weighted sum of previous values of z, plus a random shock a(t). The $\pi(B)$ expression must however satisfy invertibility conditions ie $\pi(B)$ must converge on or within the unit circle.

Box and Jenkins [1976] show that for $d \geq 1$, the π's sum up to unity. This provides an interesting interpretation of (5.17), viz., the current value z(t) of a nonstationary process can be computed from a weighted average of an infinite number of previous values of z, plus a random shock a(t). In practice however it is known that the convergent π series die out rapidly and therefore implying that though theoretically z(t) is supposed to be dependent on an infinite number of π weights, it is actually dependent only on recent past values.

5.3.3. MINIMUM MEAN SQUARE ERROR FORECASTS

The minimum mean square error forecast can be obtained from the three formulations of the general autoregressive integrated moving average model. To forecast a value z_{t+l}, where $l \geq 1$ from a reference point t, it is said that the forecast is made at origin t for lead time l. The three explicit forms of the model can be rewritten replacing t by t+l. This gives the following

5.3.3.1. Difference equation form

Replacing t by t+l in (5.10) gives

$$z_{t+l} = \varphi_1 z_{t+l-1} + \dots + \varphi_{p+d} z_{t+l-p-d} + a_{t+l} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q} \qquad (5.18)$$

5.3.3.2. Random shock form

Replacing t by t+l in (5.14) gives

$$z_{t+l} = \sum_{j=-\infty}^{t+l} \psi_{t+l-j}\, a_j = \sum_{j=0}^{\infty} \psi_j\, a_{t+l-j} \qquad (5.19)$$

where $\psi_0 = 1$. The ψ's can be obtained as follows. Operate on both sides of (5.14) with the generalised autoregressive parameter $\varphi(B)$ to obtain

$$\varphi(B) z_t = \varphi(B) \psi(B) a_t$$

However from (5.9)

$$\varphi(B)z(t) = \theta(B)a(t)$$

therefore

$$\varphi(B)\varepsilon(B)a_t = \theta(B) \tag{5.20}$$

Expanding (5.19) gives

$$(1 - \varphi_1 B - \dots - \varphi_{p+d}B^{p+d})(1 + \varepsilon_1 B + \varepsilon_2 B^2 + \dots$$
$$= (1 - \theta_1 B - \dots - \theta_q B^q \tag{5.21}$$

The ε weights are obtained from (5.20) by equating coefficients of B, B^2, B^3, ..., and computing recursively from the following

$$\varepsilon(1) = \phi(1) - \theta(1)$$
$$\varepsilon(2) = \phi(1)\varepsilon(1) + \phi(2) - \theta(2)$$
$$\varepsilon(3) = \phi(1)\varepsilon(2) + \phi(2)\varepsilon(1) + \phi(1) - \theta(3)$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$
$$\varepsilon(j) = \phi(1)\varepsilon(j-1) + \dots + \phi(p+d)\varepsilon(j-p-d) - \theta(j) \tag{5.22}$$

Box and Jenkins give the minimum mean square error forecast as

$$\hat{z}_t(l) = \varepsilon_l a_t + \varepsilon_{l+1}a_{t-1} + \dots = \underset{t}{E}\,[z_{t+l}] \tag{5.23}$$

where $\underset{t}{E}\,[z_{t+l}]$ is the conditional expectation of z_{t+l} given all knowledge of z's up to time t. Expanding (5.19) gives

$$z_{t+l} = (a_{t+l} + \varepsilon_1 a_{t+l-1} + \dots + \varepsilon_{l-1}a_{t+1}) + (\varepsilon_l a_t + \varepsilon_{l+1}a_{t-1} + \dots$$
$$= e_t(l) + \hat{z}_t(l) \tag{5.24}$$

from which it is seen that $e_t(l)$ is the error of forecast $\hat{z}_t(l)$ at lead time l from origin t.

From (5.23) the computed ε weights may be used directly to compute the forecasts $\hat{z}(l)$ and when (5.23) is regarded as a function of l, the equation is known as the forecast function for the origin t. It is understood that t is fixed.

The variance of the forecast error can also be shown to be

$$V(l) = \mathrm{var}[e_t(l)] = (1 + \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_{l-1}^2)\sigma_a^2 \qquad (5.25)$$

where σ_a^2 is the variance of the white noise process a_t.

Substituting $l = 1$ into (5.24) gives the one step ahead forecast as

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1} \qquad (5.26)$$

This result implies that the random shocks which generate the process are in reality one step ahead forecast errors. Though one step ahead forecast errors must be uncorrelated, errors for longer lead times tend to be correlated. In fact forecast errors for different lead times made from the same origin tend to be correlated. This fact leads to the forecast function lying either wholly above or below the actual series when these become known.

5.3.3.3. Inverted form

Recalling (5.17) and substituting t+l for t gives

$$z_{t+l} = \pi_1 z_{t+l-1} + \pi_2 z_{t+l-2} + \dots + a_{t+l} \qquad (5.27)$$

$$= \sum_{j=1}^{\infty} \pi_j z_{t+l-j} + a_{t+l}$$

It will be noted that though theoretically an infinite number of π's are required to compute a forecast from (5.27), in practice due to the invertibility constraints, only a few π's are really needed to compute forecasts to a reasonable degree of accuracy.

5.3.4. METHODOLOGY

In summary all three formulations can be used to compute minimum mean square error forecasts from the ARIMA model. However the simplest and most elegant means of deriving the forecast in practice is to employ the difference equation (5.18) directly [Box and Jenkins, 1976].

$$z_{t+l} = \varphi_1 z_{t+l-1} + \dots + \varphi_{p+d} z_{t+l-p-d} + a_{t+l} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q}$$

After expressing the forecast $\hat{z}(l)$ in terms of the difference equation (or any of the other two explicit forms) one then proceeds according to the following rules (from Box and Jenkins, 1976)

1. the z(t)'s (j=0,1,2,...) which have already occurred up to time t are left unchanged.
2. the z(t+j)'s (j=1,2,...) which are yet to occur are replaced by their forecasts $\hat{z}(l)$ at origin t.
3. the a(t-j)'s (j=0,1,2,...) which have occurred are available from $z_{t-j} - \hat{z}_{t-j-1}(1)$.
4. the a(t+j)'s (j=1,2,...) which are yet to occur are replaced by zero.

The techniques outlined here for forecasting nonseasonal nonstationary models can easily be extended to the case of seasonal models. As mentioned before the best method for direct computation of forecasts is by means of the difference equation.

Equation (5.5) for example could be rearranged into the difference form of representation from which the forecasts are readily obtained. As in the case of nonseasonal models, the minimum mean square error forecast at lead time l is given by the conditional expectation of z(t+l) taken at origin t. The application of this formulation is basically possible, because the invertible models which are fitted to actual data, usually have forecasts which depend mostly on recent values of the series. It has also been demonstrated in Box and Jenkins [1976] that forecasts are insensitive to small changes in parameter values such as those introduced by estimation errors.

6. MODEL SELECTION METHODOLOGY

6.1. Introduction

The previous chapter demonstrates how data exhibiting seasonality can be modelled. It is also shown how minimum mean square error forecasts can be obtained directly from the identified models. There are however alternative methodologies for representing seasonally varying time series. In this chapter, we attempt to find which class of seasonal model best describes the average monthly flow measured at Senchi.

6.2. DATA AVAILABILITY

Various climatological, meteorological and hydrological data are collected and documented by various agencies and organizations in the Volta basin area.

The Hydrological Services Department of the Ministry of Works and Housing in Ghana collects and maintains a data base which includes rainfall, evaporation, temperature, gauge and discharge data for the Volta river basin as well other major rivers in

Ghana. Data is also collected by the Meteorological Services Department in Ghana.

In view of the role which the Volta river system plays in the region, daily flow data are collected at various sites for the Black Volta, White Volta and the Oti rivers (Figure 1.1). Average monthly flow data are then obtained by processing the daily flow data.

Daily rainfall data are also collected at various locations around the country. These are also summarised into average monthly values. With respect to reservoir water balance computation, three rainfall stations which are very often used are situated at Tamale, Akuse and Ho all located around the periphery of the reservoir area.

In the case of evaporation data, numerous studies have shown that annual evaporation is almost constant at 1524 mm (60 in.) in the region of the Volta Reservoir. The monthly distribution of evaporation is then obtained by utilising measured pan evaporation data around the reservoir area.

In addition to the above, average monthly natural flow data exist for the Volta river at Senchi (Figure 1.1) for the period 1936-1963 after which impoundment of the Volta Lake commenced. From 1964 todate, natural flow data has been synthesised using the reservoir characteristics (area, elevation, volume curves) various plant operating records including power flows, reservoir elevation as well as long term average evaporation and monthly rainfall values for Akuse and Tamale (Figure 1.1).

For the purpose of developing a stochastic model with the objective of forecasting natural river flow at Senchi, data from six rainfall stations as well as average monthly flow from four flow gauge sites on the Black Volta, White Volta, the Oti River and the Volta river itself were examined. The data availability is as follows :

6.2.1. Rain gauge stations

- Wa : Rainfall data (1939 to 1982)
- Wenchi : Rainfall data (1939 to 1982)
- Yendi : Rainfall data (1939 to 1982)
- Ho : Rainfall data (1945 to 1982)
- Tamale : Rainfall data (1945 to 1982)
- Kumasi : Rainfall data (1945 to 1982)

6.2.2. Flow Stations:

- Nawuni : White Volta River flow (1954 to 1976)
- Bui : Black Volta River flow (1944 to 1984)
- Saboba : Oti River flow (1944 to 1984)
- Senchi : Volta River flow (1936 to 1984)

6.3. ANALYSIS

In order to develop suitable models adequately preserving statistical properties of measured natural flow at Senchi, all the above data was examined with a view to finding out which set or data set combination would be most useful in the formulation of a suitable Box-Jenkins linear stochastic model.

A plot of average monthly natural flow at Senchi (Figure 6.1) immediately reveals the seasonal nature of the data. It is therefore logical to consider seasonal types of models namely,

1. The Periodic Autoregressive Model
2. The Deseasonalized Model
3. The SARIMA (p,d,q,) X (P,D,Q)

6.3.1. The Periodic Autoregressive Model

As mentioned previously in section 5.2.1., the periodic autoregressive model consists of s independent autoregressive models representing data for each separate season of the year. To build the necessary models, 12 different AR(p) moldels will have to be identified for each of the twelve months of the year Noakes, McLeod & Hipel [1985]. This will therefore result in a collective model with at least 12 parameters. For the purpose of

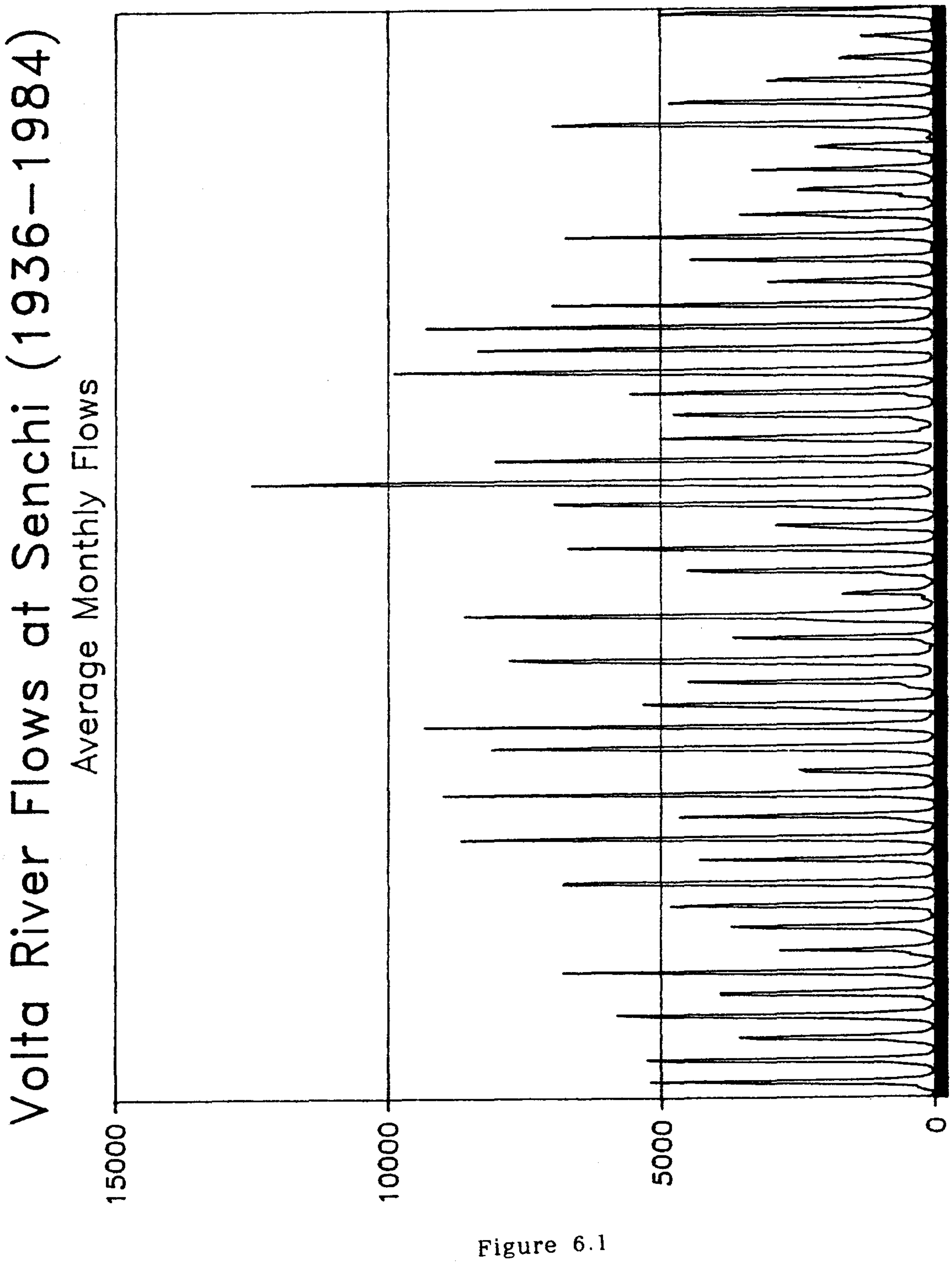

Volta River Flows at Senchi (1936–1984)
Average Monthly Flows
15000
10000
5000
0

Figure 6.1

this exercise, it was felt that a model of this nature would defeat the objective of finding a very simple and parsimonious representation of the average monthly flows at Senchi. The periodic autoregressive model was therefore rejected as being unsuitable model. It was therefore decided to investigate other alternatives as regards seasonal models namely deseasonalized models and seasonal multiplicative models.

6.3.2. The Deseasonalized Model

This approach to modelling seasonal data basically employs the introduction of a deterministic component into the formulated model to account for seasonality. The resulting nonseasonal series is then modelled as a nonseasonal autoregressive moving average model (ARMA). Sometimes, the data will need to be transformed using an appropriate power transformation to induce normality in the residuals. This method is applied to measured average monthly flow at Senchi with a view to fitting a suitable deseasonalized model. The approach outlined in Chapter 5 is followed in this case.

6.3.2.1. Plot of the original data

The plot of original series is shown in Figure 6.1. It is difficult to discern any apparent trends though seasonality is evident in the plot. The sample autocorrelation function is therefore plotted.

6.3.2.2. Plot of sample autocorrelation function

The sample autocorrelation function plot is shown in Figure 6.2. This clearly exposes the seasonal character of the data. The peaks and troughs of the SACF coincide with integral multiples of s and s/2 respectively (where s in this case is 12). The period s indicated in the SACF would imply data 12 time periods apart are similar (either high or low). The troughs at lags s/2 which implies negative correlation, would imply that data 6 time periods apart are opposite in behaviour (ie high and low and vice versa). The sample autocorrelation function is also plotted for lags which are integral multiples of 12 and shown in Figure 6.3. The values of the SACF are seen to tail off as opposed to truncating at a particular lag. This would indicate the necessity to include seasonal AR parameters to account for the seasonal dependence in the data. However, for this method of approach, an attempt is made to account for seasonality by employing the standardization equation (5.2)

$$w_{i,j} = (Z_{i,j}^{(\delta)} - \mu_j) / \sigma_j \qquad (6.1)$$

where $Z_{i,j}^{(\delta)}$ = the transformed observation for the i^{th} year and j^{th}

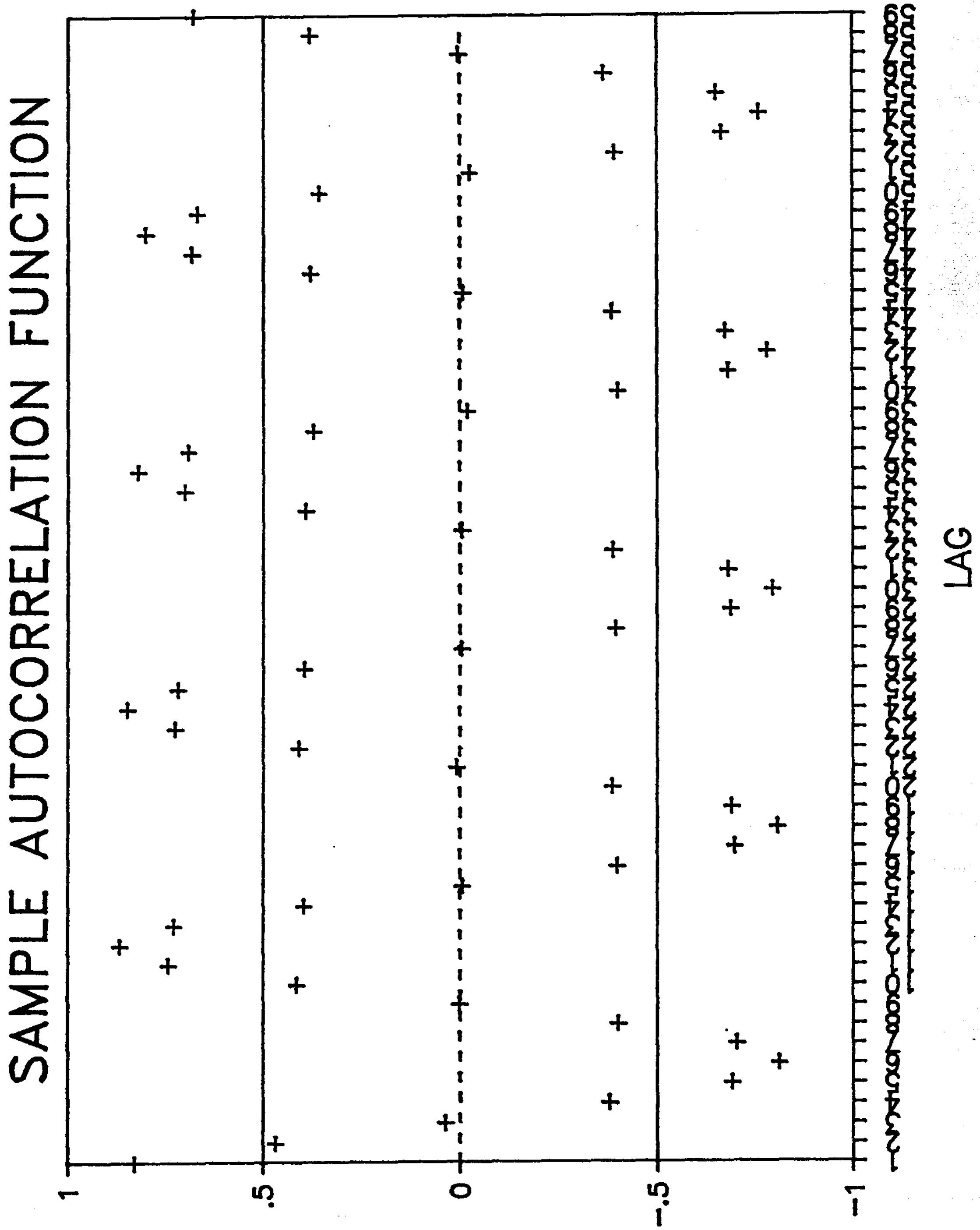


Figure 6.2

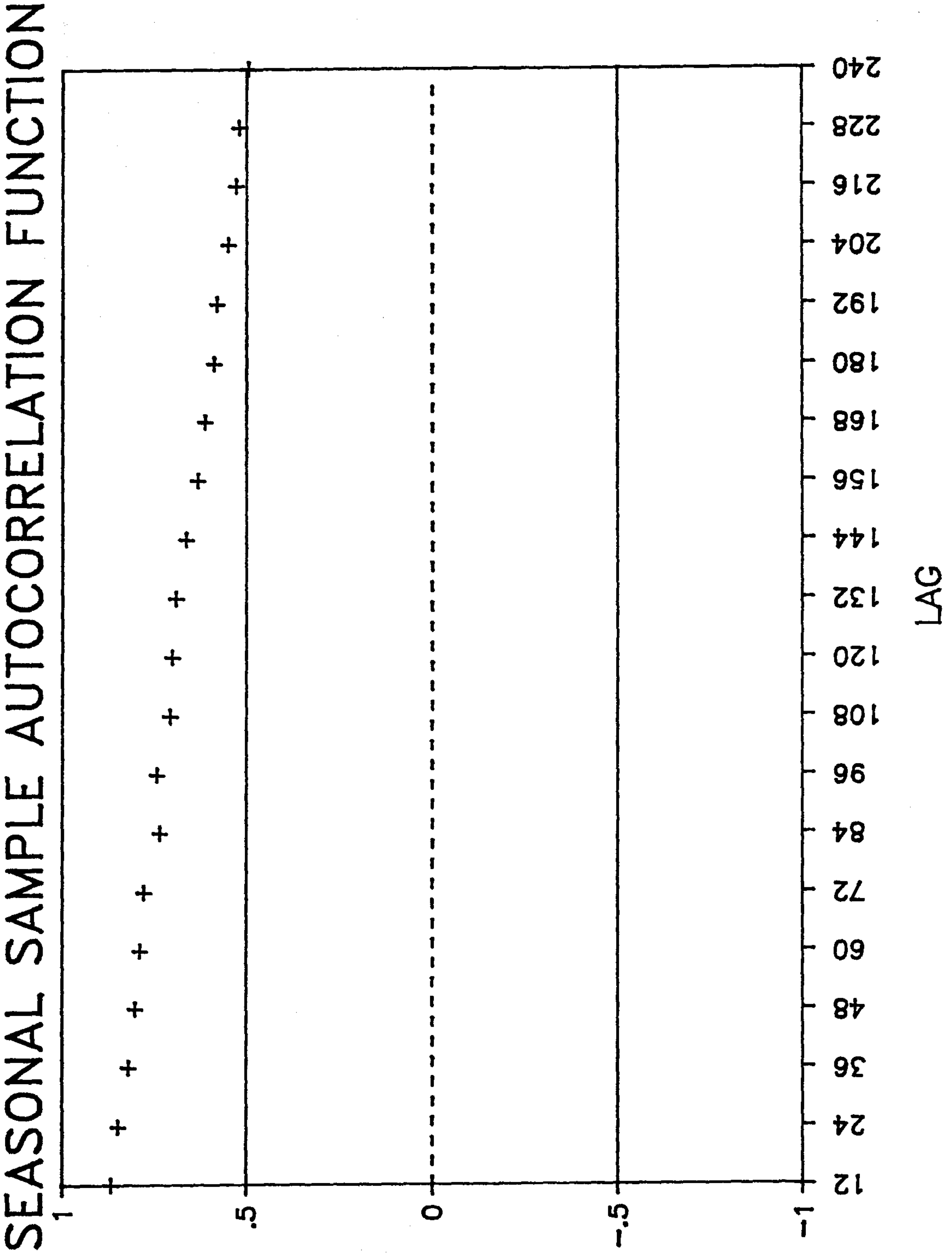

SEASONAL SAMPLE AUTOCORRELATION FUNCTION
1
.5
0
-.5
-1
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
192
204
216
228
240
LAG

Figure 6.3

month, μ_j is the fitted mean for month j, σ_j is the fitted standard deviation for month j, and δ is the exponent of an appropriate Box-Cox transformation applied to the data.

The plot showing the autocorrelation function of the data after deseasonalization is shown in Figure 6.4. This shows a remarkable reduction in magnitude of the values of the autocorrelation function at seasonal lags (ie integral multiples of s). This fact is shown more clearly in Figure 6.5 where the autocorrelation function of the deseasonalized data is plotted for seasonal lags. This deseasonalization procedure is supposed to effectively remove seasonal trends in the data.

It was however noticed that the coefficient of skew was significant and therefore the log transform of the original data was performed to induce normality in the data. Figures 6.6 and 6.7 show replotted values of the sample autocorrelation function for serial and seasonal lags. It is noticed that the log transform does not aid in removal of seasonality. It only induces normality in the data. The coefficient of skew, used in determining normality of the data was computed as follows

$$g_1 = \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^3\right) / \left(\frac{1}{n}\sum_{t=1}^{n}\hat{a}_t^2\right)^{3/2} \qquad (6.2)$$

The value g_1 was reduced from 1.2097 to -0.4382 in this case. After the seasonal component had been removed and normality accounted for, an ARMA model was fitted to the data.

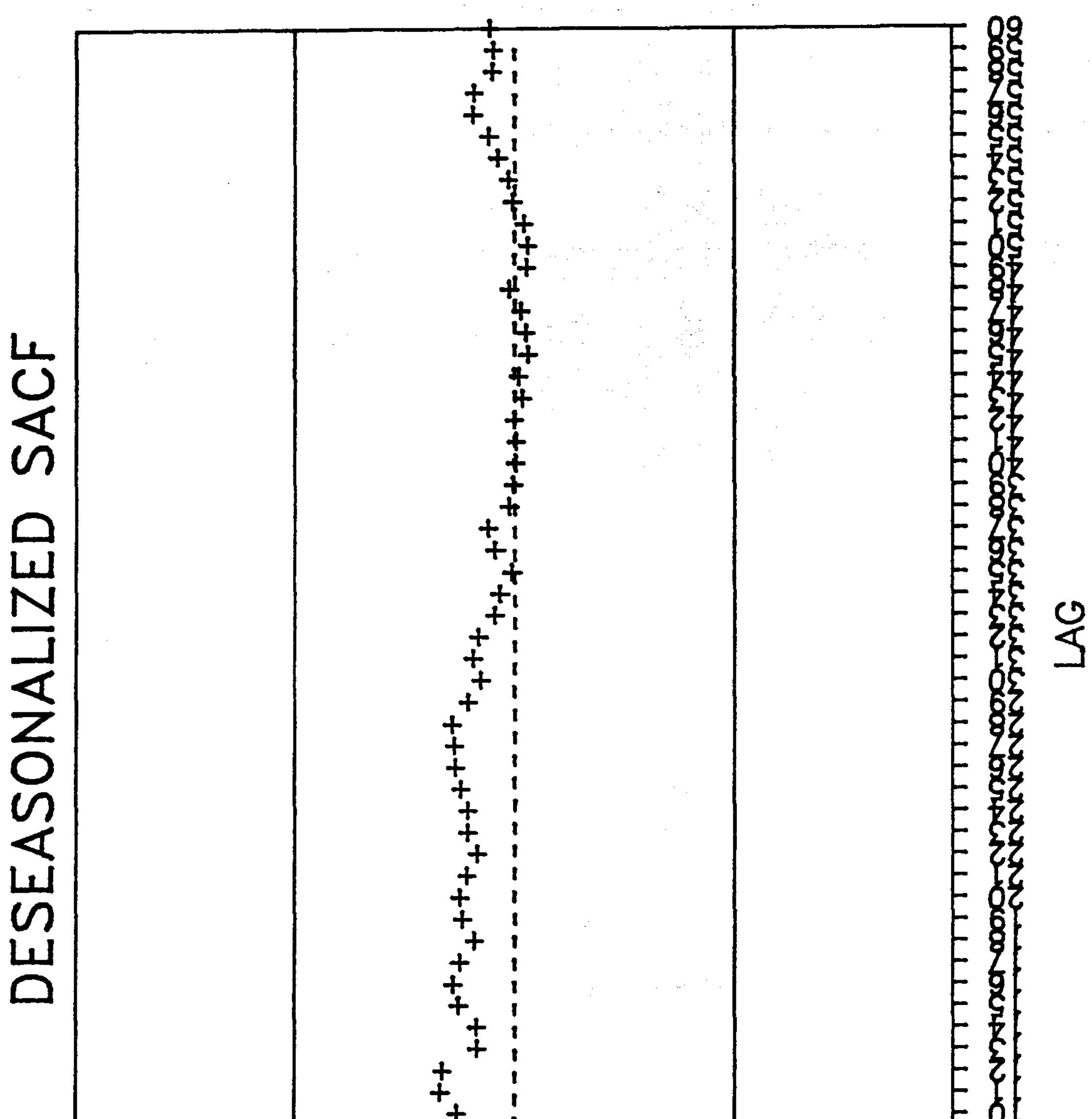


Figure 6.4

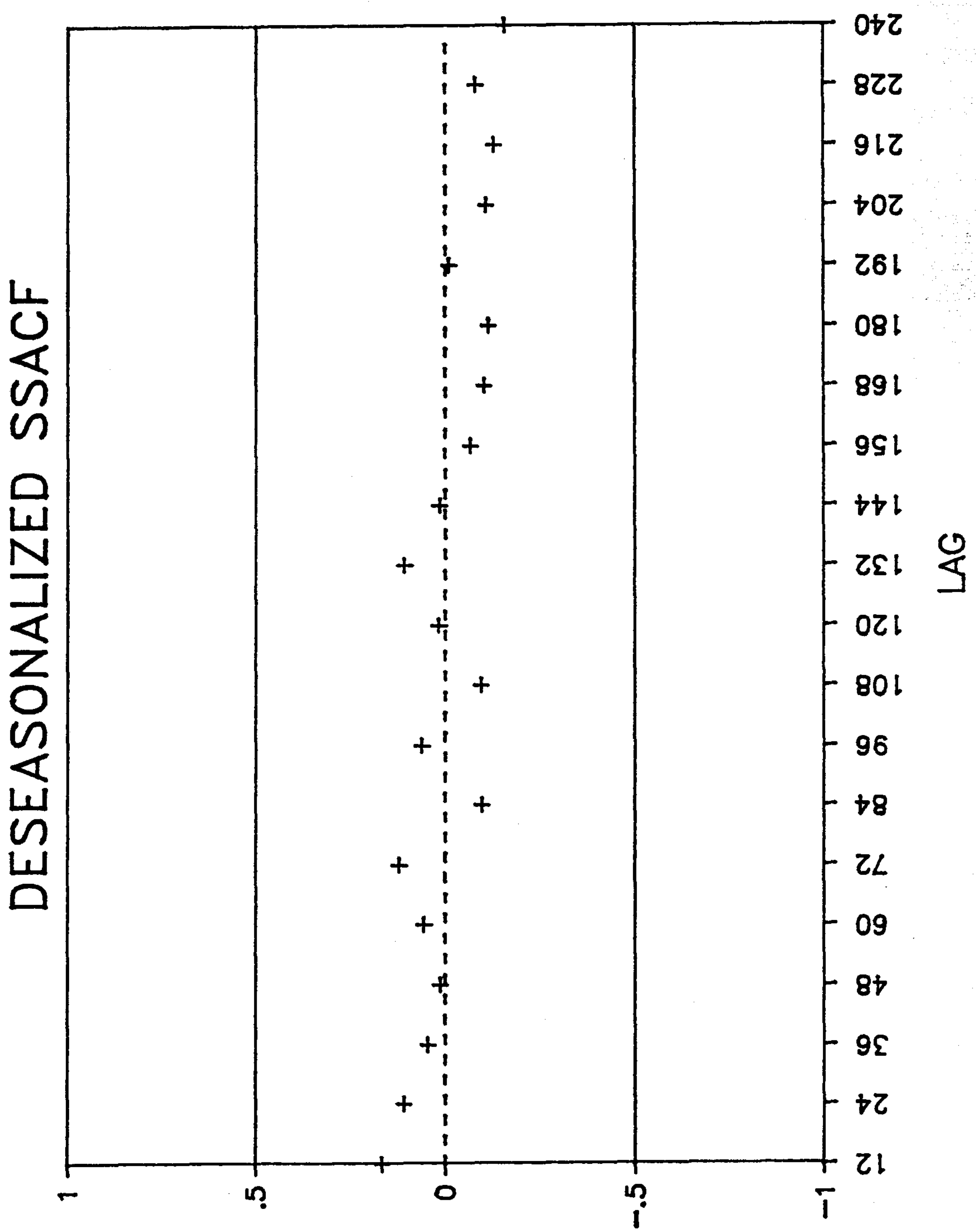

DESEASONALIZED SSACF
LAG
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
192
204
216
228
240
1
.5
0
-.5
-1

Figure 6.5

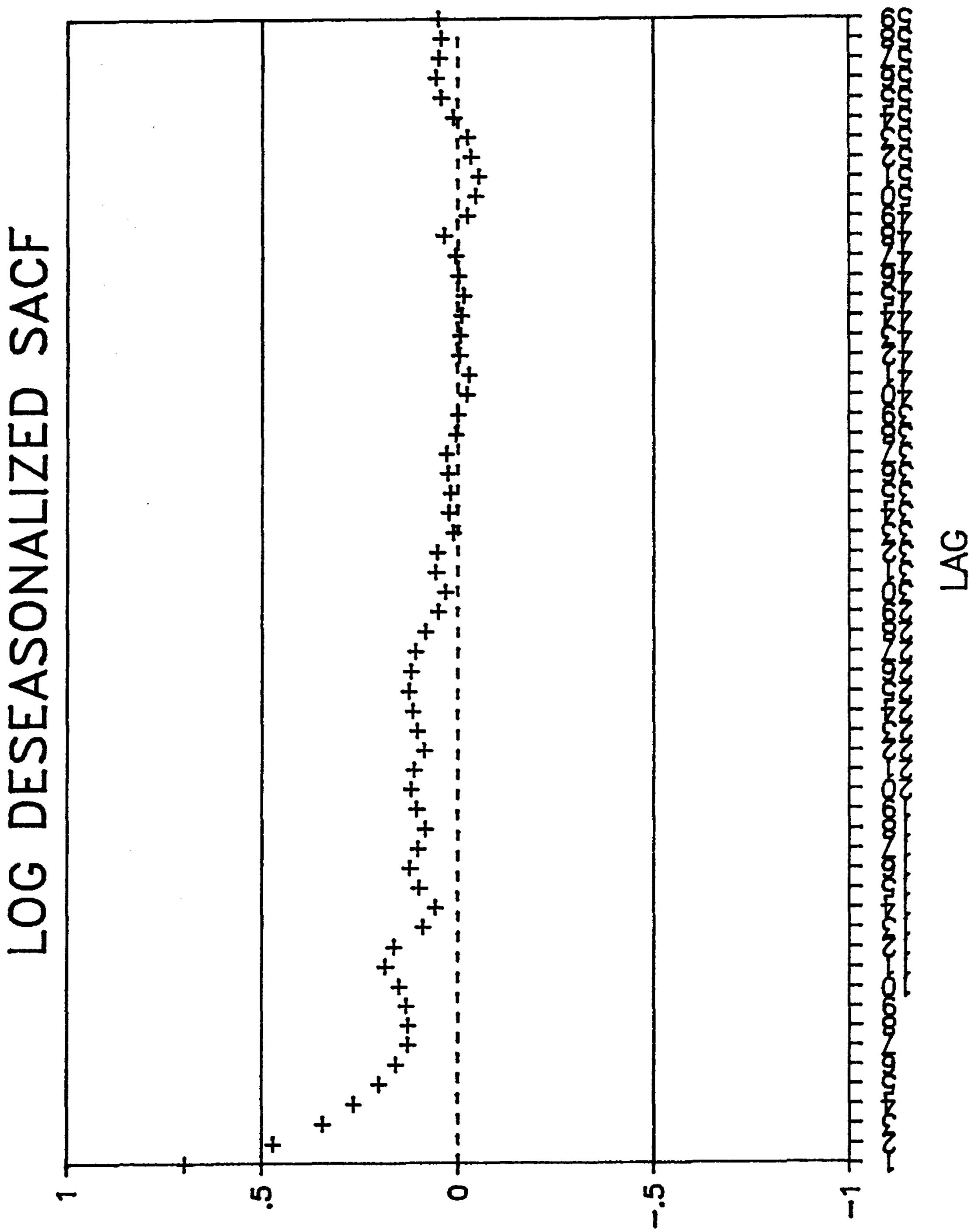


Figure 6.6

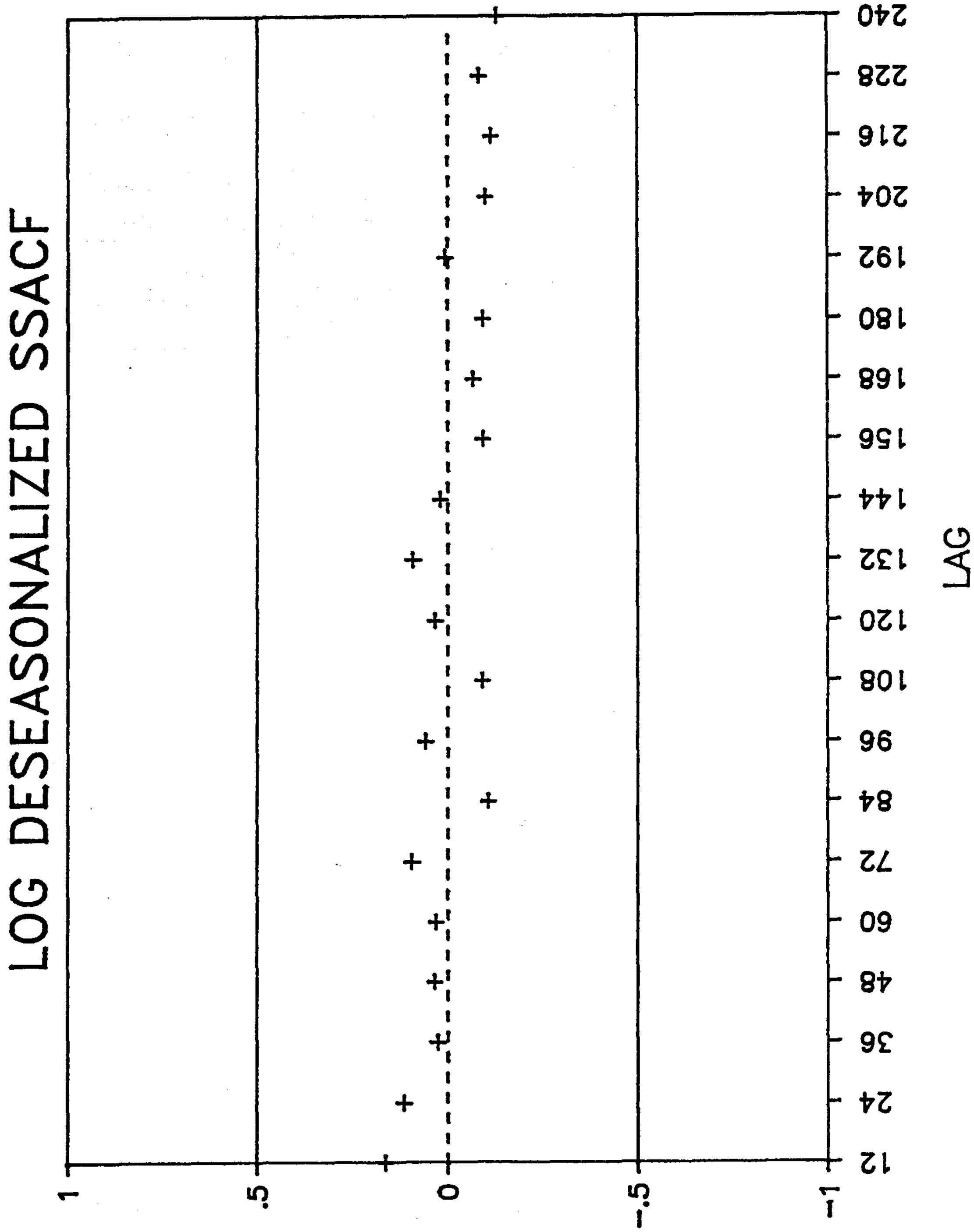

LOG DESEASONALIZED SSACF
LAG
12 24 36 48 60 72 84 96 108 120 132 144 156 168 180 192 204 216 228 240
1 .5 0 -.5 -1

Figure 6.7

6.3.2.3. ARMA model identification

To aid in identifying the appropriate model to fit to the deseasonalized log transformed data, the plot of the corresponding sample autocorrelation function (Figure 6.6) was reexamined. It is noticeable that the first few lags exhibit high values yet the whole function does not truncate at any particular lag. It is therefore reasonable to include AR parameters in the model. An AR(1) model was fitted to the data and the autocorrelations of residuals plotted. The residual autocorrelation function (RACF) is plotted for both serial and seasonal lags and shown in Figures 6.8 and 6.9 respectively. Figure 6.8 shows that at most lags, the RACF is contained within the 95% confidence limits. This fact is mathematically confirmed using the portmanteau lack-of-fit test. The value of Q (from 4.28) computed for the residuals is 69.47 on 59 degrees of freedom, which compares with 79.1 from the table of the chi squared distribution, shows that the model could be considered adequate. However two facts militate against acceptance of this model especially with a view to using it as a forecasting tool.

1. The Q statistic computed for residuals at the seasonal lags does not pass the portmanteau lack-of-fit test. The Q statistic value computed on 20 degrees of freedom is 49.43 which compares with chi squared value of 31.4 from the tables. This implies some form of seasonal

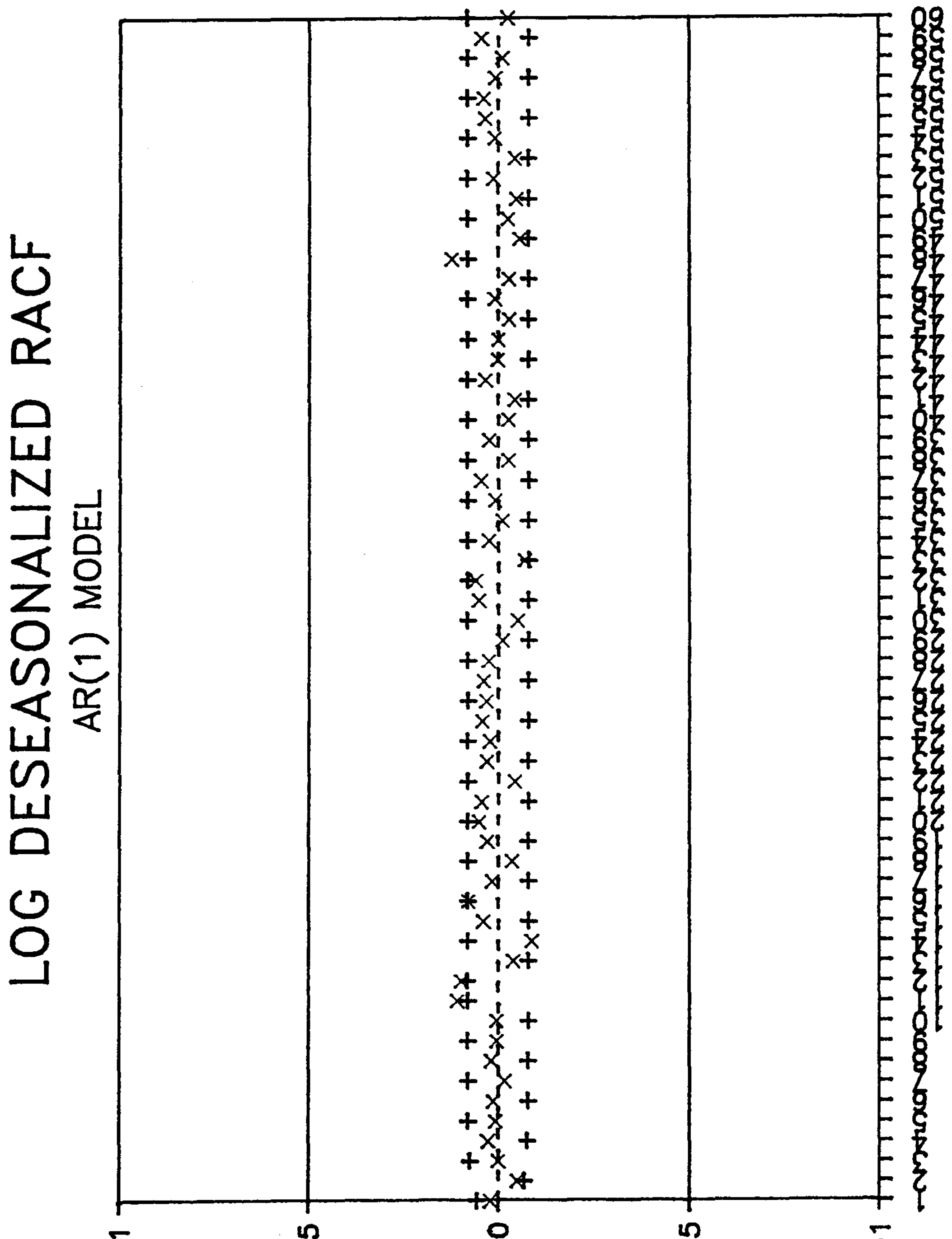
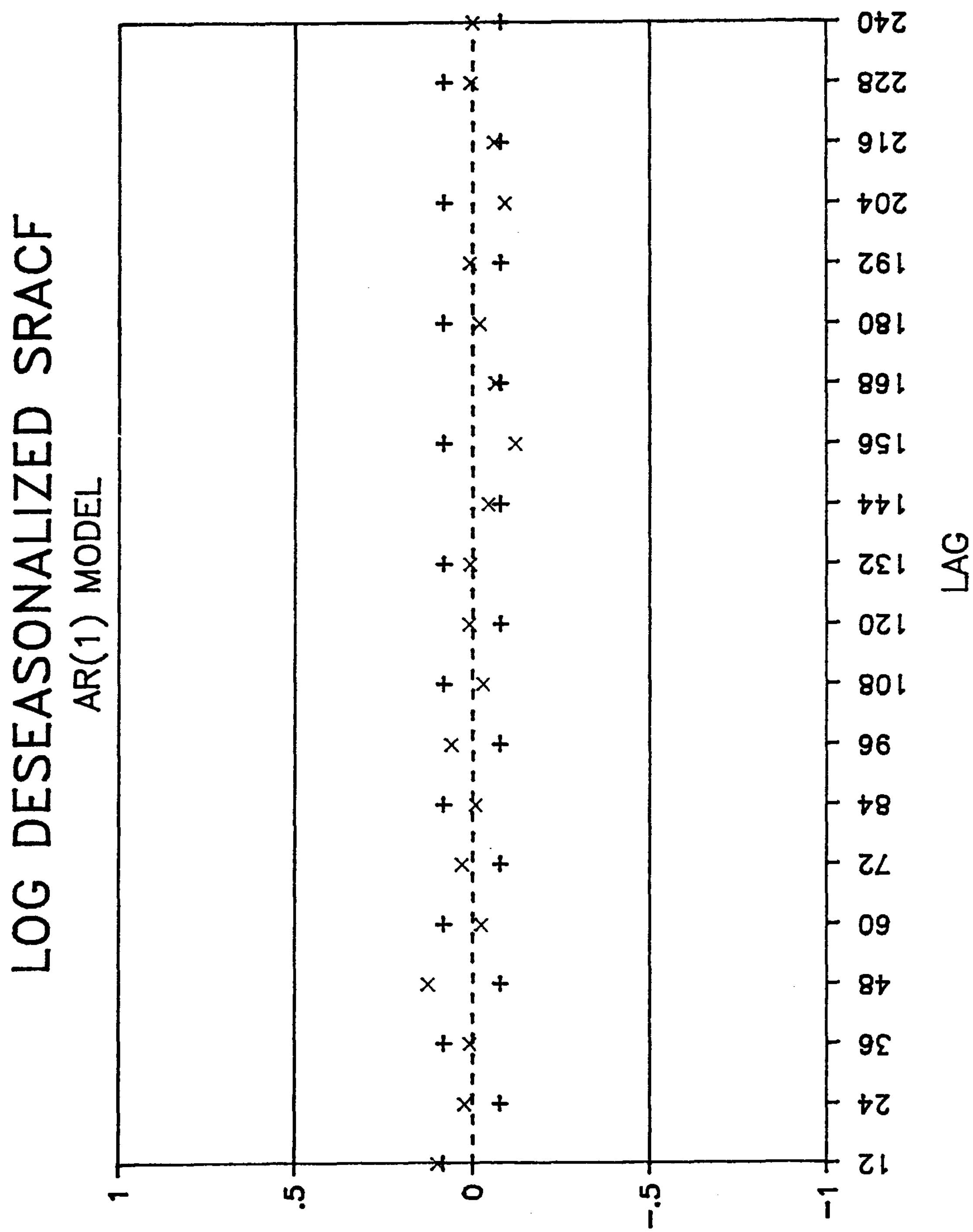


Figure 6.8



LOG DESEASONALIZED SRACF
AR(1) MODEL
LAG
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
192
204
216
228
240
1
.5
0
-.5
-1

Figure 6.9

dependence in the residuals. This can be observed from the plot of seasonal residual autocorrelation function Figure 6.9.

2. The value obtained for the coefficient of determination (R squared) is 48.43%. The R squared value indicates degree of, or as in this case, lack of forecastibility of the fitted model. Kottegoda [1980] indicates that the R squared value, which is also the square of the multiple correlation coefficient, gives a measure of predictability of the given series.

6.4. SUMMARY

Due to the above, it is deemed inappropriate to accept the deseasonalized AR(1) model as a suitable model for average monthly flows at Senchi. It is therefore necessary to investigate the alternate approach of Box and Jenkins [1976] viz the general multiplicative model or seasonal autoregressive integrated moving average model (SARIMA).

The SARIMA model building is described in the next chapter.

7. SELECTED MODEL FOR THE VOLTA RIVER

7.1. Introduction

The previous chapter describes attempts at identifying a suitable model adequately preserving statistical properties of the Volta river in the Periodic Autoregressive and Deseasonalized model forms. These attempts have not yielded any acceptable models and therefore an attempt is made to fit a general multiplicative seasonal model.

7.2. SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODELS

This form of dynamic representation of a data series which exhibits stochasticity or randomness, is usually applied to data which is seasonal in nature. Since this model has previously been described, [Sec. 5.2.3.] we proceed directly to the model building.

7.2.1. Identification

The identification stage of the model building process will indicate if seasonal and / or nonseasonal differencing is required to produce stationarity in the data, what transformation is needed to normalise the data and also to ascertain the order of both seasonal and nonseasonal AR and MA parameters for the time series under consideration. The data available consists of average monthly flow measurements of the Volta river at Senchi from 1936 to 1984. It will be noted therefore, that, there are 588 data points available for use in model parameter estimation. It will therefore be expected that efficient maximum likelihood estimates will be found for these parameters. The identification procedure followed is as outlined in Chapter 4.

7.2.1.1. Plot of original data series

The plot of the original series is shown in Figure 6.1 and its pertinent characteristics are discussed under deseasonalized models. To gain further insight into the characteristics of the data, the sample autocorrelation function plot is examined.

7.2.1.2. Plot of sample autocorrelation function

The sample autocorrelation function plot is as shown in Figure 6.2. As discussed previously under deseasonalized model formulation, the seasonal nature of the data is clearly exposed. The peaks and troughs of the SACF coincide with integral multiples of s and s/2 respectively (where s in this case is 12), indicating the period of seasonality. The seasonal sample autocorrelation function is also shown in Figure 6.3. Values of the seasonal sample autocorrelation (SSACF) are seen to tail off indefinitely. This indicates some form of seasonal differencing may be required to induce seasonal stationarity in the data. It is also known before hand that a log transform (corresponding to $\delta=0$ in the Box-Cox transformation) is appropriate for monthly river flow series to induce normality. The log of original data is therefore taken. Following the observation of the seasonal SACF, the data is differenced once seasonally according to the procedure

$$\nabla^s z_t = z_t - z_{t-s} \tag{7.1}$$

The sample autocorrelation function for the differenced data is shown in Figures 7.1 and 7.2 for both serial and seasonal lags. The SACF for serial lags shows values which though significant at lag 1, decay rapidly to zero. This would suggest the need to incorporate an autoregressive parameter in the nonseasonal component of the model.

The seasonal SACF also shows a relatively significant value at lag 1 after which the next few values appear to be close to zero. This behaviour is symptomatic of a pure moving average process where values of the autocorrelation function are truncated after lag q. This suggests the inclusion of a moving average parameter in the seasonal component of the multiplicative model.

7.2.2. Estimation

Following deductions from plots of the sample autocorrelation functions, a SARIMA(1,0,0) X (0,1,1) model was fitted to the data and the resulting residuals examined. The model parameters estimated were as follows

$$\phi(1) = 0.6955 \ , \ \Omega(1) = 0.9146$$

7.2.2.1. The Mathematical Model

To arrive at the final mathematical representation of the identified model, we consider the general multiplicative seasonal process model which is expressed as

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D \, z_t = \theta_q(B)\Omega_Q(B^s)a_t \qquad (7.2)$$

with order (p,d,q) X (P,D,Q). For a SARIMA (1,0,0) X (0,1,1)

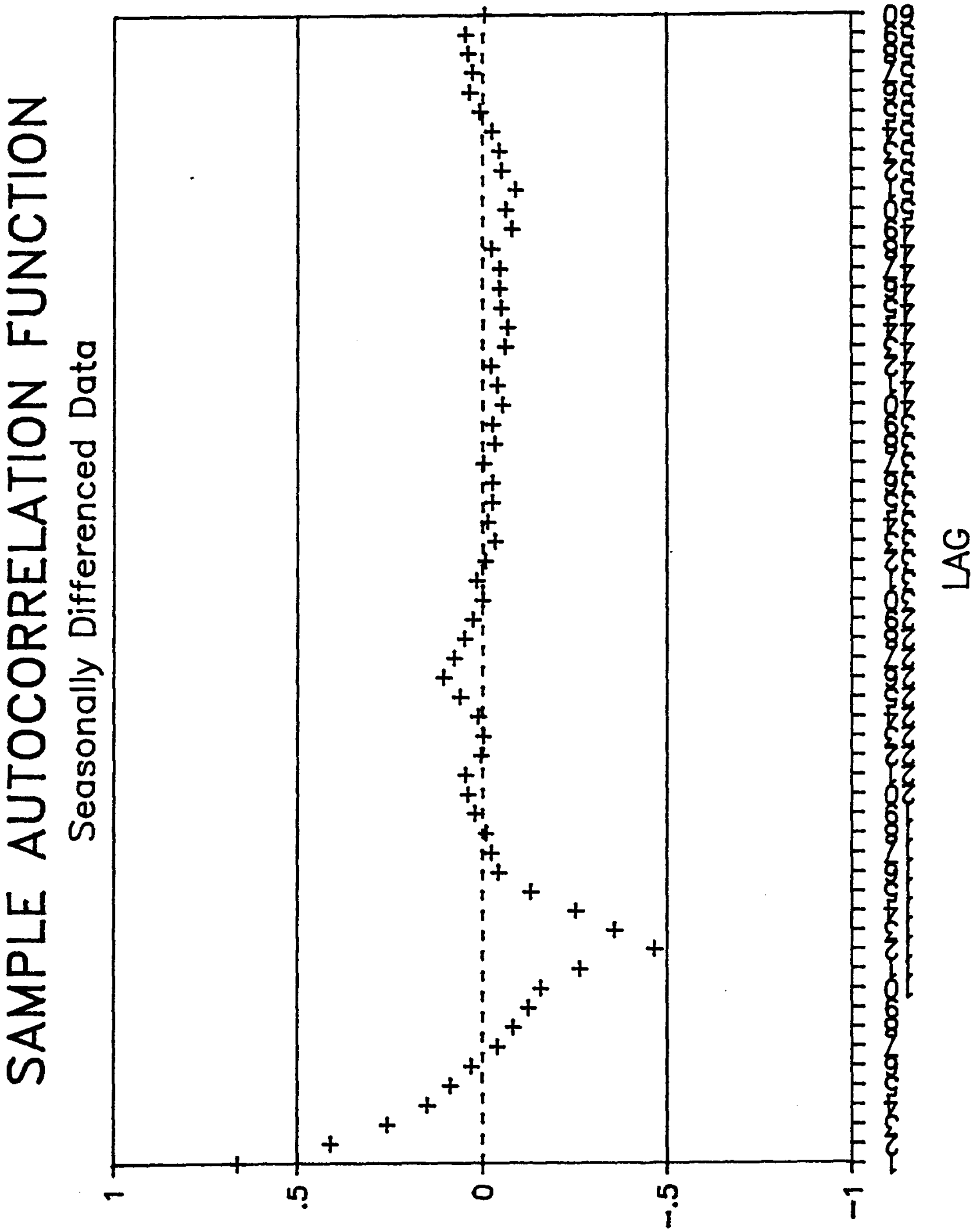

SAMPLE AUTOCORRELATION FUNCTION
Seasonally Differenced Data
LAG
1
.5
0
-.5
-1

Figure 7.1

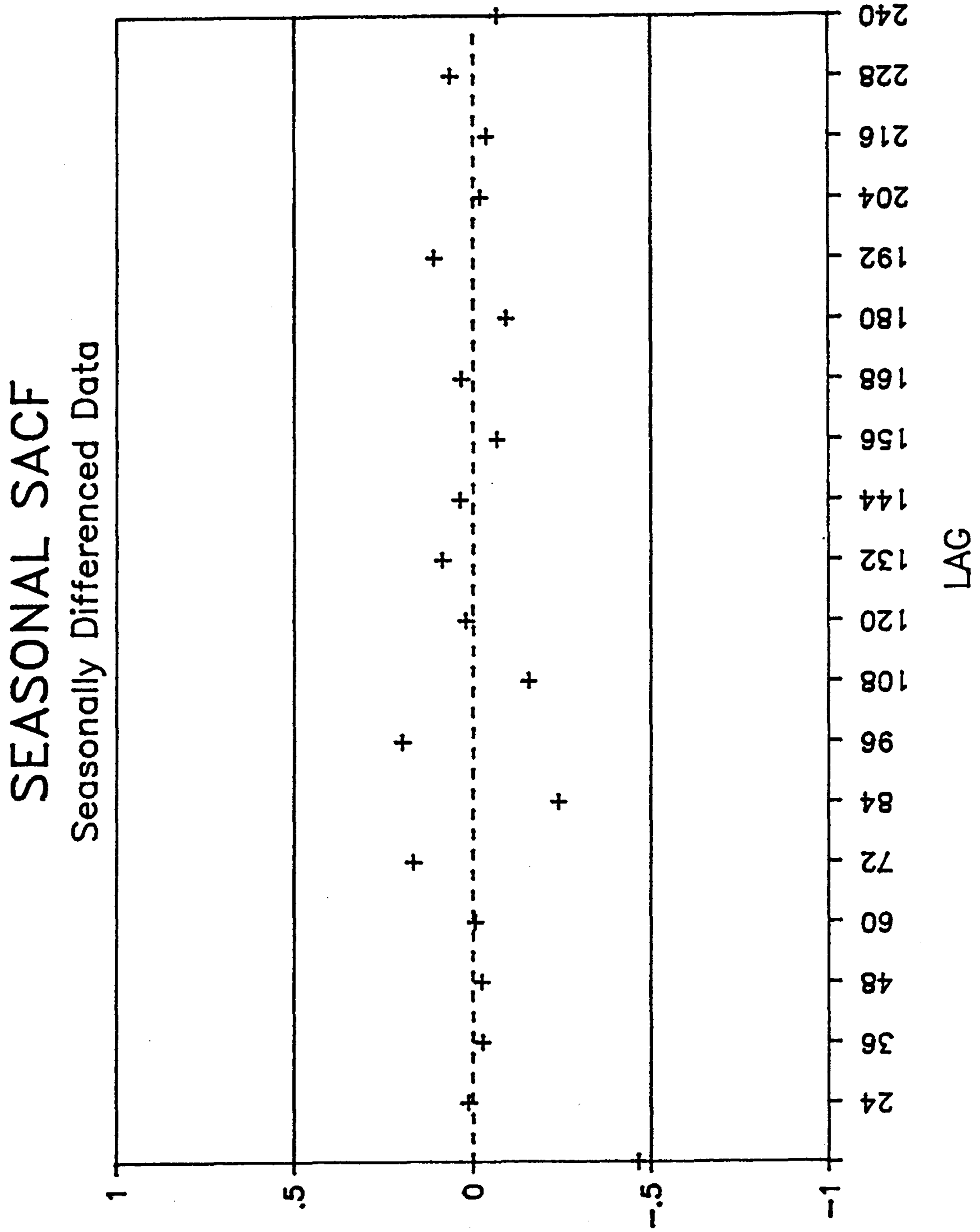

SEASONAL SACF
Seasonally Differenced Data
LAG
24
36
48
60
72
84
96
108
120
132
144
156
168
180
192
204
216
228
240
1
.5
0
-.5
-1

Figure 7.2

model, d = q = P = 0 . Deleting these values in (7.2) gives the more parsimonious

$$(1 - \phi_1 B)\ \nabla_{12}\ z_t = (1 - \Omega_1 B^{12})\ a_t \qquad (7.3)$$

Substituting for estimated parameters in equation (7.3) obtains

$$(1 - 0.6955B)\ \nabla_{12}\ z_t = (1 - 0.9146B^{12})\ a_t \qquad (7.4)$$

which can be written explicitly as

$$(1 - 0.6955B)\ (z_t - z_{t-12}) = a_t - 0.9146a_{t-12} \qquad (7.5)$$

or

$$z_t - z_{t-12} - 0.6955z_{t-1} + 0.6955z_{t-13} = a_t - 0.9146a_{t-12} \qquad (7.6)$$

z(t) can then be expressed as

$$z_t = a_t + z_{t-12} + 0.6955z_{t-1} - 0.6955z_{t-13} - 0.9146a_{t-12} \qquad (7.7)$$

It is observed therefore that twelve previous values of innovations and thirteen previous values of z's values will be required as starting values for the computation of forecasts.

7.2.3. Diagnostic checking of fitted model

The SARIMA (1,0,0) X (0,1,1) model obtained from the identification stage of the model building process is a very parsimonious representation of the original data. It must however be verified through a series of tests designed to expose any deficiencies which might be inherent in the fitted model. These tests are performed mainly on the residuals to ensure that the fitted model satisfies the basic requirements of independence and normality in the residuals. The order of the fitted model is also checked through the process of overfitting. These so called diagnostic checks are outlined in Chapter 4.

7.2.3.1. Overfitting

This mode of checking a model involves fitting a more sophisticated model than that which has been identified to see whether it brings any improvement to the identified model. It presupposes that one already has an indication of which direction the deficiency in the identified model might take. To decide in which areas the identified SARIMA (1,0,0) X (0,1,1) model might be deficient, the sample autocorrelation coefficient is examined. The plot of the SACF Figure 7.1 indicates that values of the autocorrelation function at the first few lags are significant. This could indicate the necessity of including nonseasonal autoregressive parameters. The fitted model however has only one nonseasonal AR parameter. One could therefore suppose that an increase in the nonseasonal AR parameters might improve the model.

As a diagnostic check a SARIMA (2,0,0) X (0,1,1) model is fitted to the data and compared with the SARIMA (1,0,0) X (0,1,1) model. It was found that the SARIMA (2,0,0) X (0,1,1) model resulted in a marginal increase in the value of the coefficient of determination R squared (from 92.95% to 92.97%). However the value of the second AR parameter computed was not significant.

SECOND AR PARAMETER

VALUE	STANDARD ERROR
-0.0496	0.0416

Additionally, the computed value for the Akaike Information Criteria (equation 4.33) increased. The values of AIC obtained for the two models are as follows

AIC FOR SARIMA (1,0,0) X (0,1,1) = 5788.29421

AIC FOR SARIMA (2,0,0) X (0,1,1) = 5788.84699

The extra autoregressive parameter was therefore deemed not necessary.

Considering the fact that the maximum likelihood estimate of the Box-Cox parameter δ can be computed, it was judged prudent to test the assumption that the appropriate value of δ was actually zero (log transform). This test was therefore carried out by fitting the SARIMA (1,0,0) X (0,1,1) model form with a simultaneous computation of the maximum likelihood estimate (MLE) of δ. The value of δ computed was 0.0369 with a standard error of 0.1274. This shows that the computed MLE of δ is not significantly different from zero. The model form SARIMA (1,0,0) X (0,1,1) is therefore fitted to the data leaving the Box-Cox transformation parameter at $\delta=0$. It is also believed that it is easier to explain a log transform than a transform (equations

4.31 and 4.32) with $\delta \neq 0.0$.

7.2.3.2. Independence of residuals

The residuals obtained after fitting an identified model are assumed to be normally, independently distributed with constant variance. The independence criterion is the most critical test which is performed. According to Hipel, McLeod and Lennox [1977] if some values of the residual autocorrelation function (RACF) are significantly different from zero, this indicates an inadequate model and therefore the model will have to be changed.

A visual inspection of the plot of the residual autocorrelation function Figure 7.3 shows that the function is mostly contained in the 95% confidence boundaries. This fact is also confirmed by the Box-Pierce portmanteau lack-of-fit test which is computed according to the following equation (4.28)

$$Q = n \sum_{l=1}^{K} r_l^2 \qquad (7.8)$$

The computed value of Q is compared with the value from the chi squared distribution. The test was done for both serial and seasonal autocorrelations. For both tests, no inadequacy was indicated in the identified model.

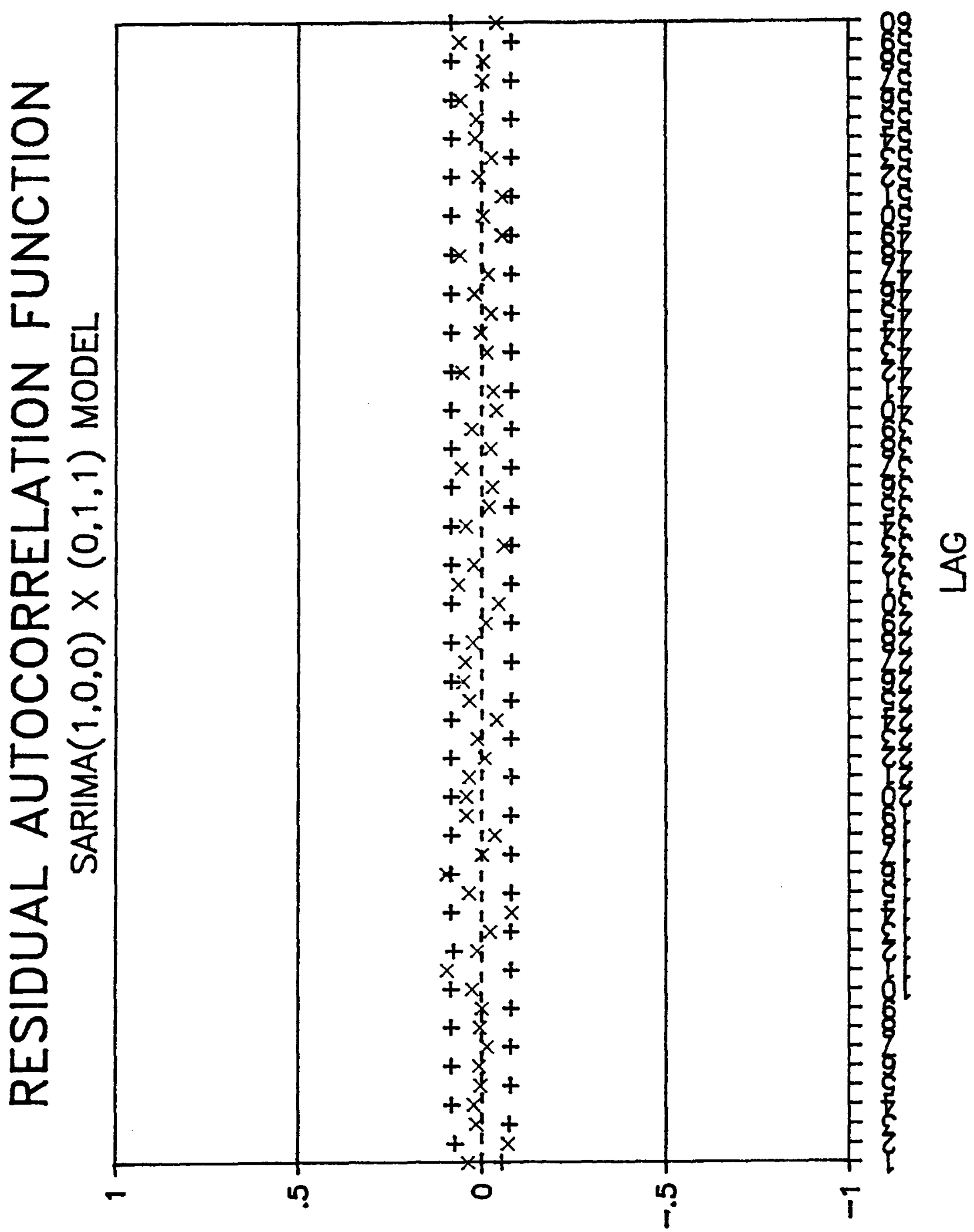

RESIDUAL AUTOCORRELATION FUNCTION
SARIMA(1,0,0) X (0,1,1) MODEL
LAG
1
.5
0
-.5
-1

Figure 7.3

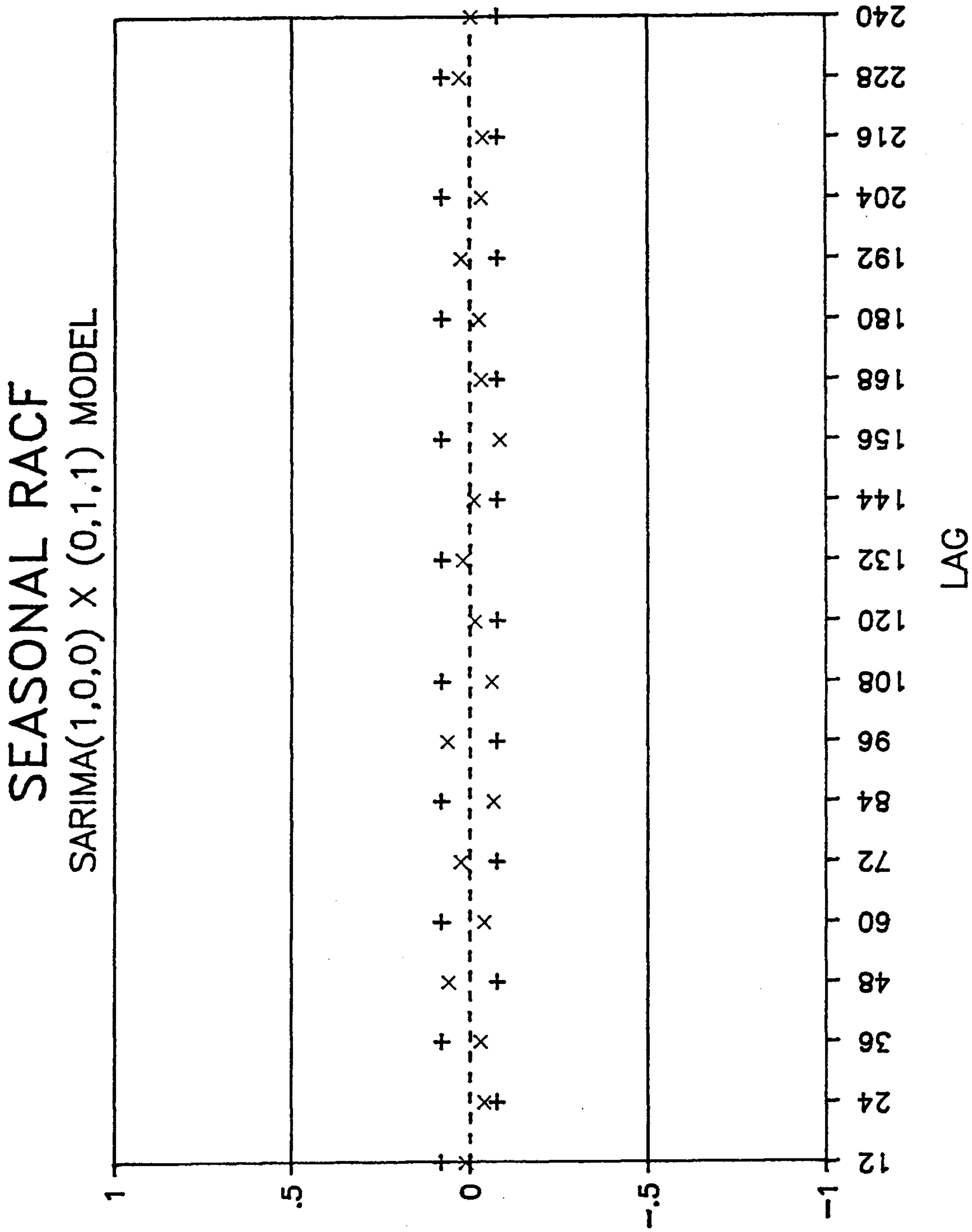

SEASONAL RACF
SARIMA(1,0,0) X (0,1,1) MODEL
LAG
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
192
204
216
228
240
1
.5
0
-.5
-1

Figure 7.4

7.2.3.3. Homoscedasticity of residuals

The less important assumption of constant variance of the residuals is also corroborated in the diagnostic checking procedure. It is generally known that an appropriate Box-Cox transformation is usually sufficient to induce homoscedasticity in the residuals. The output listing (see appendix) shows that the log transform induces constant variance in the residuals.

7.2.3.4. Normality of residuals

The normality assumption for the residuals, though of secondary importance with respect to parameter estimation, is critical in forecasting applications. This is because the computation of confidence limits of the forecasts is dependent on the normality assumption of the residuals [Hipel, McLeod and Lennox 1977]. To verify that the residuals are normal, the skewness coefficient is computed according to the following (4.29)

$$g_1 = \left(\frac{1}{n} \sum_{t=1}^{n} \hat{a}_t^3\right) / \left(\frac{1}{n} \sum_{t=1}^{n} \hat{a}_t^2\right)^{3/2} \qquad (7.9)$$

As seen in the output listing, the value of g computed is 0.0948 with a significance level of 0.343515. The residuals are therefore seen to be Normally distributed.

7.3. COMMENTS

From the foregoing discussions, it is seen that a parsimonious model [SARIMA (1,0,0) X (0,1,1)] is arrived at which adequately represents average monthly flow measured at Senchi. The model identification is greatly aided by sample and residual autocorrelation functions which are plotted to assist in the visual inspection of various characteristics exhibited by the data. The square of the multiple correlation coefficient obtained for this model is 92.95% which compares with 48.43% obtained for the Deseasonalized model fitted previously.

It is also noted that the model building procedure was accomplished with reasonable ease compared to other seasonal model forms. This fact is important because model parameters will need to be continually estimated since the data base is constantly updated.

For the purpose of this exercise, an adequate model has been selected through the process of identification, parameter estimation and subsequent diagnostic checks. It will however be prudent to validate the model form in addition to verification procedures (diagnostic checks) which have already been performed. This verification procedure involves taking subsets of the available data and comparing forecasts generated by the model against measured data.

8. FORECASTING THE VOLTA RIVER FLOWS

8.1. Introduction

In previous chapters, it is shown how a seasonal autoregressive integrated moving average model of the form denoted by SARIMA (1,0,0) X (0,1,1) is arrived at as a suitable and parsimonious representation of the average monthly flows of the Volta river. Diagnostic checks were performed on the identified model to verify underlying assumptions of normal independent residuals with constant variance. Adequacy of the number of estimated parameters was also tested by the overfitting process.

After its form is verified by means of diagnostic checks, it remains to show the model's adequacy for practical use. The model form is thus validated by demonstrating its ability to forecast the modelled monthly flows. This is accomplished by retaining the final four years of measured flow and employing the previous flow data to forecast these four years. This process is repeated for varying lengths of data selected from different periods in the historical data set.

To facilitate the computational effort required in processing the data, two programs (USED and USFO) [developed by Dr. Keith W. Hipel, Department of Systems Design Engineering, University of Waterloo; and Dr. A. Ian McLeod, Department of Statistical and Actuarial Sciences, University of Western Ontario] were obtained and modified into one program DRUSMAIN. A brief description of these computer packages are provided below.

8.2. COMPUTER SOFTWARE

The computer software used for this project was obtained by kind permission of Dr. Hipel and Dr. McLeod, from the Niagara Falls office of Acres Consulting Services Limited.
This program was then used to compute relevant parameters needed to help identify and build the correct model for the data and also compute both one step ahead (5.26) and minimum mean square error forecasts (sect. 5.3.3.).

8.2.1. Program "USED"

This is a univariate stochastic estimation and diagnostics program which forms part of the so called MCLEOD-HIPEL TIME SERIES PACKAGE. It is mainly a parameter estimation program which requires for input among others, model form and raw data. Its output includes values of residual autocorrelations and other parameters pertinent to residual analysis. This therefore makes it suitable for identification and diagnostic checking as well as parameter estimation.

8.2.1.1. Input data required

The input data required for correct running of program "USED" depends on which application it is being used for. In general however, the following constitute the main parameters which are required as input for this program:

IP = order of nonseasonal AR component, p

IQ = order of nonseasonal MA component, q

IPS = order of seasonal AR component, P

IQS = order of seasonal MA component, Q

ISEA = seasonal period, s

ITYPE1 = 0, normal setting

= 1, constrain some β parameters

= 2, constrain mean component μ

= 3, both 1 and 2

ITYPE2 = 0, normal setting

= 1, Box-Cox transformation

= 2, deseasonalize

= 3, both 1 and 2

IEST = 0, modified sum of squares estimator

= 1, conditional estimator

= 2, exact maximum likelihood estimator

IOUT = 0, create output file .USE

= 1, output to terminal

= 10, create output files .USE and .RSD

= 11, output to terminal and create file .RSD

= 20, create output files .USE, .RSD and .MOD

= 21, output to terminal and create output files .RSD and .MOD

ID = order of nonseasonal differencing, d

IDS = order of seasonal differencing, D

ALAMDA = Box-Cox exponent, δ

CONS = constant in Box-Cox transformation

Z(I) = raw data to be analyzed

8.2.1.2. Output data listed

The output furnished by this routine includes intermediate files (if requested) which are utilised as input into accompanying simulation (.RSD) and / or forecasting (.MOD) programs. It also produces a standard list of output (in file .USE) which includes the following:

a. model form for which parameters are estimated for example : SARIMA(1, 0, 0)(0, 0, 1)12

b. length of input time series

c. residual variance

d. coefficient of determination : R-SQ

e. Akaike information criteria : AIC

f. Box-Cox transformation parameters : δ, CONS

g. fitted seasonal means and standard deviations

(for deseasonalized model application)

h. estimated beta parameters and their standard errors

g. RESIDUAL ANALYSIS

1. Coefficients of SKEW and KURTOSIS
2. tests for heteroscedasticity
3. test for trends in the variance over time
4. residual and squared-residual autocorrelation functions
5. Box Pierce portmanteau statistic, Q for above
6. residual and squared-residual seasonal autocorrelations
7. Box Pierce portmanteau statistic, Q for above

It will be noted that the above listed data input and output though comprehensive, are just a subset of the total number of parameters involved.

8.2.2. Program "USFO"

This program is basically a univariate stochastic forecasting routine which accepts input from the estimation and diagnostics program through an intermediate file (.MOD). Required input includes the identified model form as well as the estimated parameters. It also requires as input "initial" values to be used in recursive computation of forecasts.

8.2.2.1. Input data required

FILE2 = name of .MOD file to be used in forecasting

The following data are transferred from "USED" through .MOD file

N = number of data used in parameter estimation

Z(I) = data used in parameter estimation

TITLE = Title for data series

MODEL = order of identified model is p, d, q, P, D, Q, s

NBETA2 = number of parameters estimated

BETA = estimated parameters

ALAMDA = exponent of Box-Cox transformation, δ

CONS = constant in Box-Cox transformation

ZMSEA = fitted seasonal means (for deseasonalized model)

ZSSEA = fitted seasonal standard deviations (for deseasonalized model)

A(I) = estimated residuals

The following are external data inputs:

UPDATE = name of file (if any) containing measured data to be used in one step ahead forecasts.

IORIG = origin (time period) after which forecasts are made

LEAD = lead time for which forecasts are required

8.2.2.2. Output data listed

Output from this routine includes model parameters, computed forecasts both in transformed and original data states as well some measures of forecastibility. The output includes the following:

a. title of data series being forecasted

b. model parameters

c. Box-Cox transformation parameters, δ and CONS

d. 1-step ahead forecasts in transformed state

e. average mean square error for 1-step ahead forecasts

f. coefficient of forecastibility for above

g. Box-Tiao test for comparison of forecast with actuality

h. forecasts in original data domain

i. average mean square error for MMSE forecasts

j. coefficient of forecastibility for above

k. generalized autoregressive coefficients

l. moving average term

m. coefficients in moving average expansion ie ε's

n. forecasting origin

o. initial observations used in forecasting

p. initial noise terms used for forecasting

The above output list is not exhaustive. Additional output is furnished depending on which application the program is used for.

8.2.3. Program "DRUSMAIN"

The two programs described above were obtained as source Fortran files suitable for compilation on a DEC VAX mini computer. To start with, these programs were adapted to be run on the IBM PC. Also in order to facilitate the use of these programs, especially by "non experts", an attempt is made to simplify data input and execution of both "USED" and "USFO". To this end, the two programs were modified and compiled together. Additionally, cursor control on the screen was handled by writing additional subroutines in order to provide an interactive "user friendly" environment for the running of programm "DRUSMAIN". As a result:

1. data input is simplified
2. on-line help is available for some input variables when required.

3. input data correction can be made without quitting the program.
4. both estimation and forecasting can be done without exiting.
5. several sets of data can be processed and compared in one session.
6. pertinent results can be displayed after estimation portion to aid decision to rerun same set of data.
7. following 6. above, different model forms can be examined without the need to quit the program. This greatly helps at the identification stage of the process.
8. one selects exactly when to quit program.

After the software modification was done, the resulting program was used to perform the computations described in this report. This includes a model validation process which is next described.

8.3. MODEL VALIDATION

In order to test the practical performance of the identified model, it is used to forecast four years of data based on different portions of the available data. Since this is designed to test the model form only, that is SARIMA(1,0,0) X (0,1,1)12, the same formulation is used for all data sets. However, for each data set, parameters in the model are estimated.

8.3.1. Data formulation

The data arrangement for the model validation is as shown in Table 8.1.

data period	# years	Forecasted period
1936 to 1955	20	1956 to 1959
1936 to 1965	30	1966 to 1969
1936 to 1980	45	1981 to 1984
1951 to 1980	30	1981 to 1984
1955 to 1974	20	1975 to 1978
1961 to 1980	20	1981 to 1984
1971 to 1980	10	1981 to 1984
1976 to 1980	5	1981 to 1984
1977 to 1980	4	1981 to 1984

8.3.2. RESULTS

8.3.2.1. One Step Ahead Forecasts

The program output for parameter estimation as well as both one step ahead and minimum mean square error forecasts are presented in Appendix A. The forecastibility of individual data series at various stages of the analysis is shown in Table 8.2 below:

R-SQUARED VALUES

PERIOD	SAMPLE	TRANS.	STEP
1936 to 1955	93.79	87.84	91.10
1936 to 1965	93.30	94.28	76.40
1936 to 1980	92.62	95.28	86.86
1951 to 1980	91.60	95.25	87.15
1955 to 1974	92.63	83.76	28.56
1961 to 1980	91.53	95.03	86.72
1971 to 1980	89.74	93.99	83.57
1976 to 1980	82.87	92.71	70.55
1977 to 1980	86.33	90.27	57.71

Table 8.2

The results summarised in 8.2 show the coefficient of determination (R-SQ = square of multiple correlation coefficient) computed at the three stages of identification, estimation and forecasting.

Values under heading "SAMPLE" are computed values of R squared after model parameters are estimated. These values therefore reflect how well the estimated model fits the original data.

The values under "TRANS." is R squared computed for forecasted data in the transformed domain (ie forecasts obtained for the log and seasonal differenced data). This therefore represents the coefficient of forecastibility for transformed data. It must be noted that these are one step ahead forecasts. Observed data for the forecasted period are available in this case for one step ahead forecast computation.

The values shown under "STEP" are R squared values computed when forecasted data is in the original (ie untransformed) domain. The forecasts in this case are also one step ahead forecasts.

From column 2 of Table 8.2, it is observed that the model form performs reasonably well for shorter data spans even though identification was done on a longer data set. The values however fall below 90% for data sets of only 10 years or less. This would be expected since maximum likelihood estimates of model parameters will not be efficient with fewer data points.

Looking at column 3 of Table, one cannot discern any apparent patterns except the fact that data comprising of longer time spans (≥ 30 years) exhibit consistent high values of R squared. The highest R squared value is obtained with the largest amount of data but the reverse is not true. Inconsistencies in values of R squared for data of shorter time spans (≤ 20 years) create some doubt around their high R squared values.

Column 4 of Table 8.2 exhibits the reverse of what is observed in column 3. In this case, the only inference that can be made from the computed values is that data of shorter spans have the worst forecastibility when actual data is being examined. R squared values for longer data sets exhibit no particular patterns except they are higher than those obtained for shorter data sets. In this case the value of R squared corresponding to the period 1936 to 1955 appears to be anomalous.

8.3.2.2. Minimum Mean Square Error Forecasts

Results obtained for minimum mean square error forecasts are graphically compared with observed values as well as one step ahead forecasts in FIGs 8.1 to 8.9. It is obvious from these graphs that one step ahead forecasts are superior to those made for lead times up to four years. This is to be expected since one step ahead forecasts are continually updated at each time step.

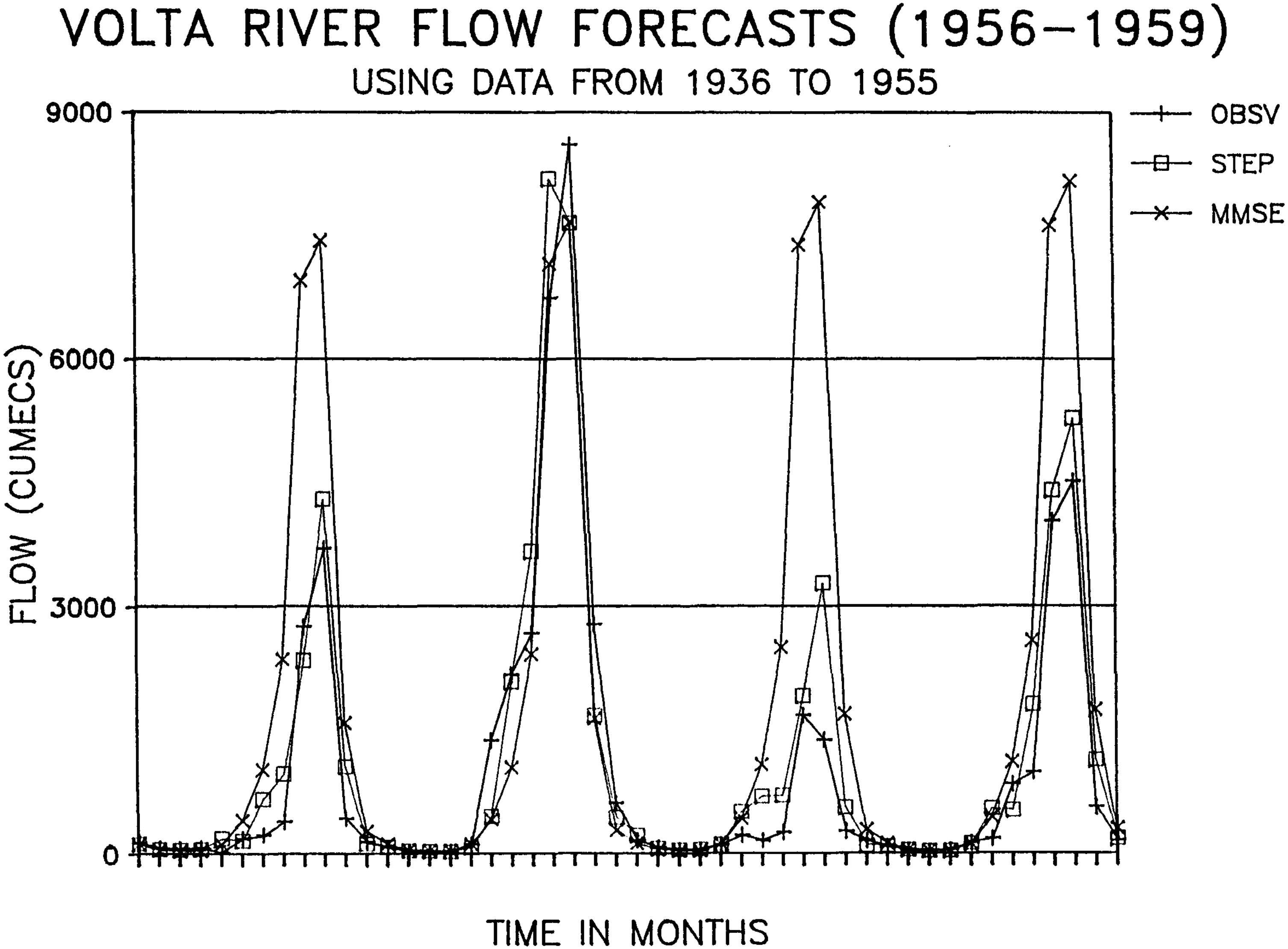

VOLTA RIVER FLOW FORECASTS (1956-1959)
USING DATA FROM 1936 TO 1955
9000
6000
3000
0
FLOW (CUMECS)
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.1

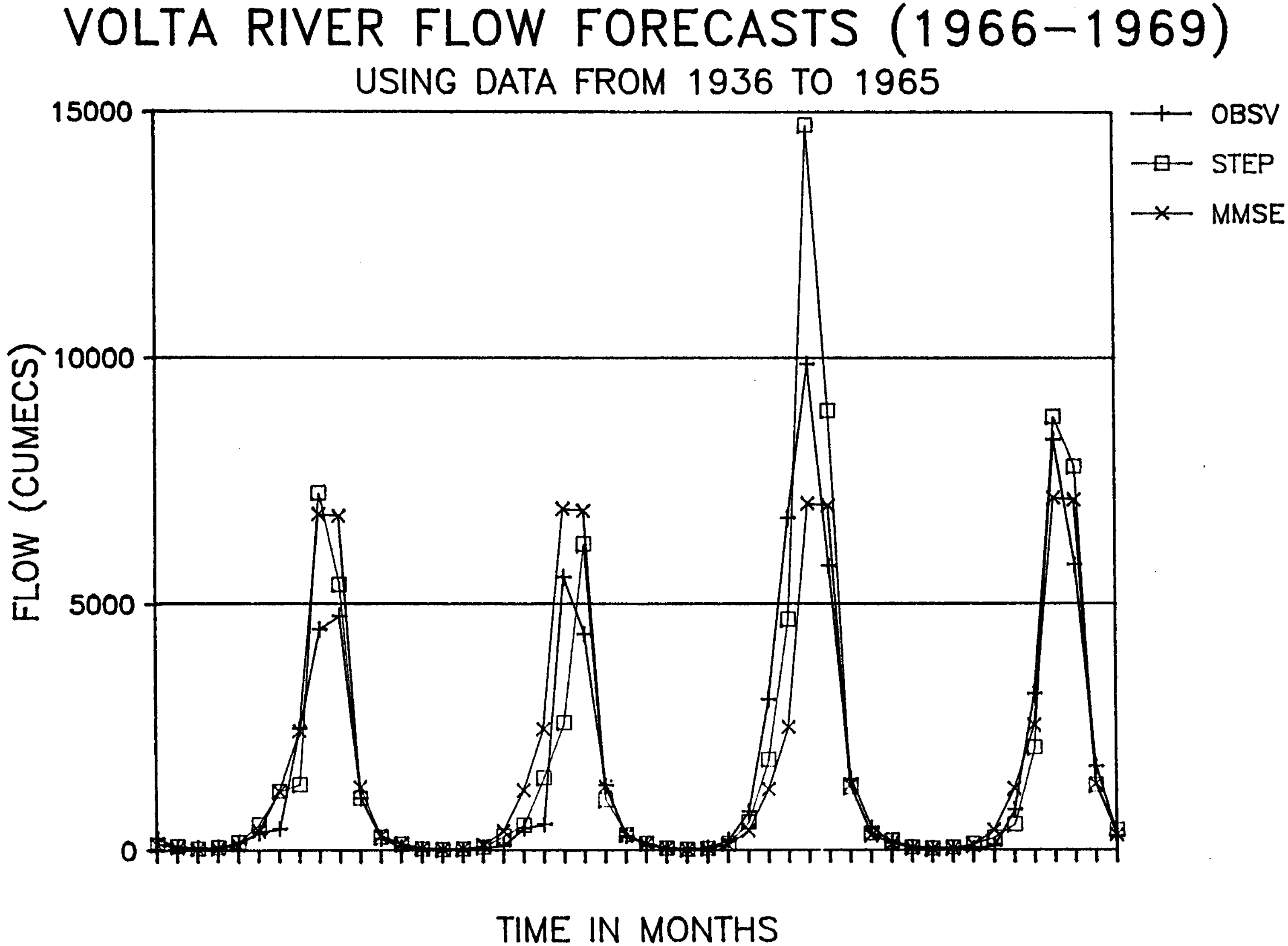

VOLTA RIVER FLOW FORECASTS (1966-1969)
USING DATA FROM 1936 TO 1965
15000
10000
5000
0
FLOW (CUMECS)
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.2

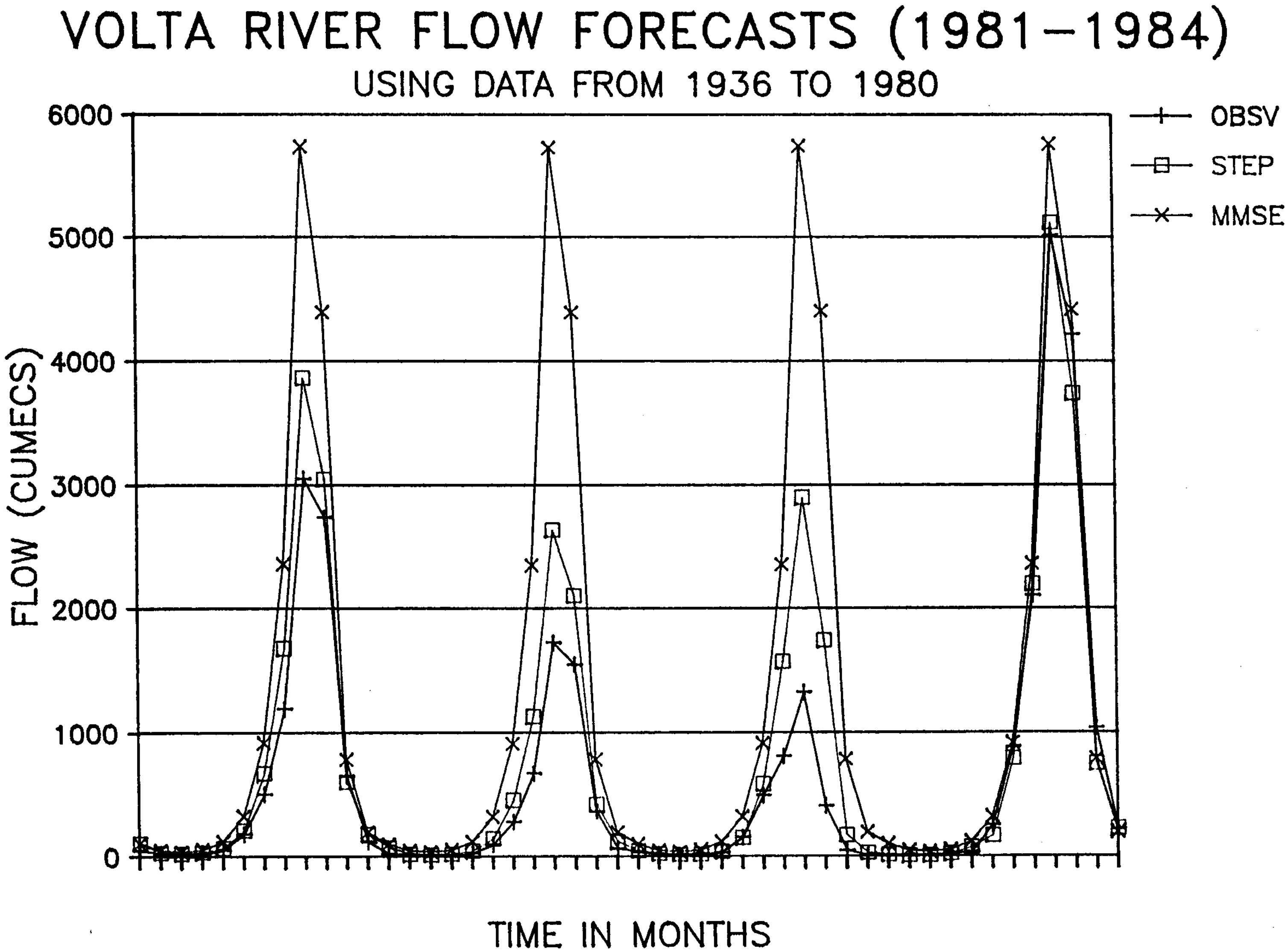

VOLTA RIVER FLOW FORECASTS (1981–1984)
USING DATA FROM 1936 TO 1980
FLOW (CUMECS)
6000
5000
4000
3000
2000
1000
0
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.3

VOLTA RIVER FLOW FORECASTS (1981–1984)

USING DATA FROM 1951 TO 1980

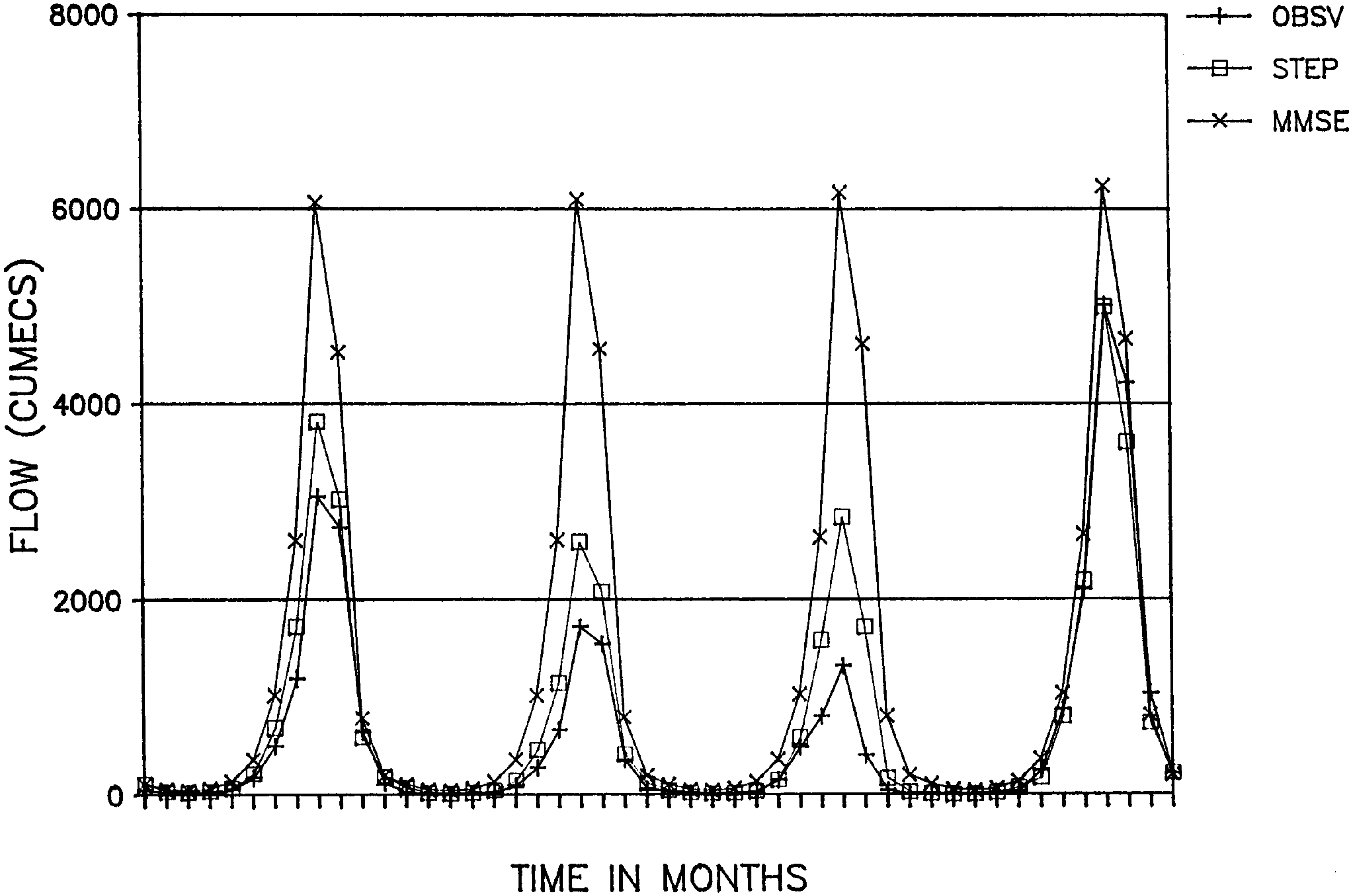


Figure 8.4

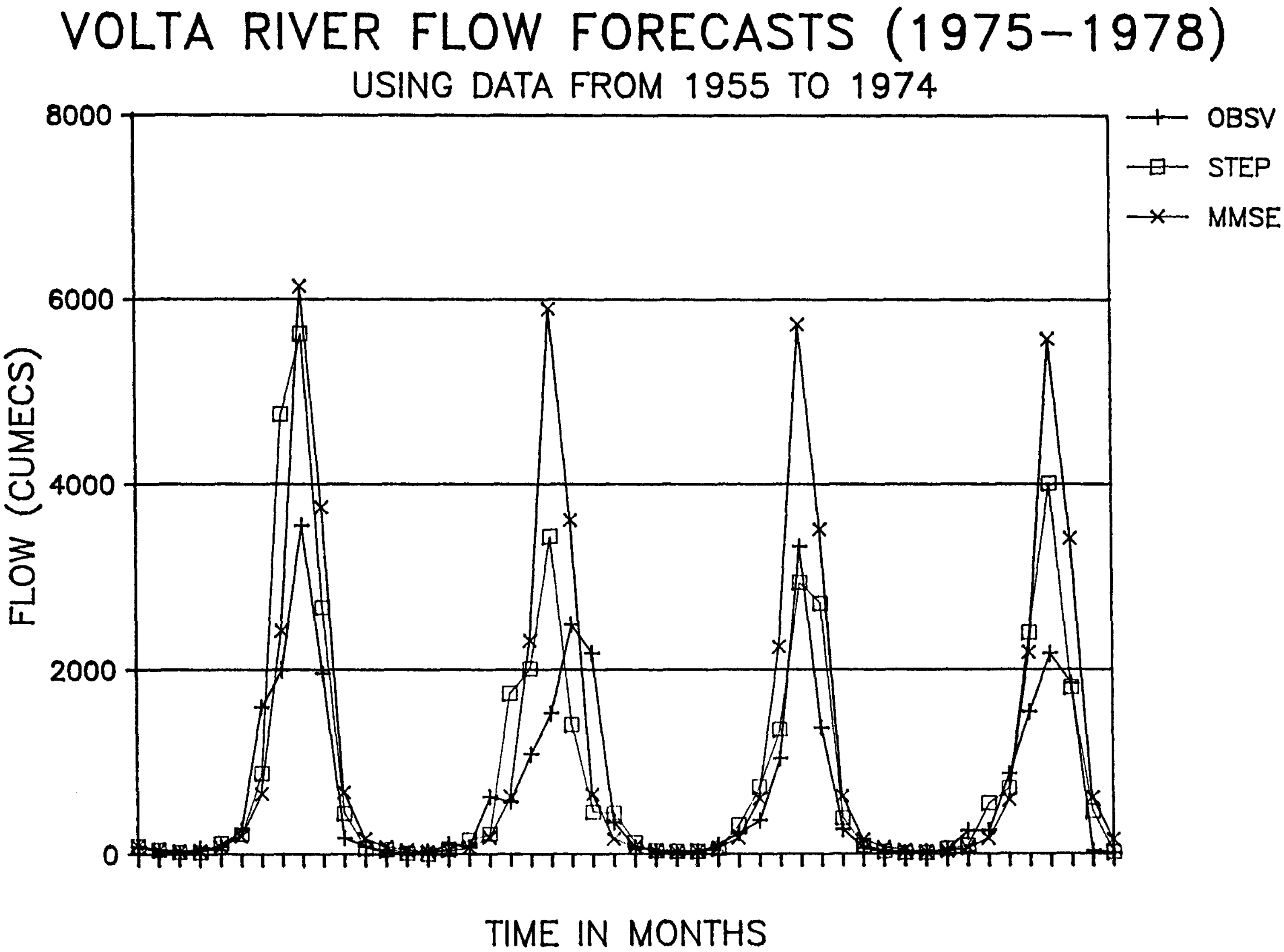

VOLTA RIVER FLOW FORECASTS (1975−1978)
USING DATA FROM 1955 TO 1974
8000
6000
4000
2000
0
FLOW (CUMECS)
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.5

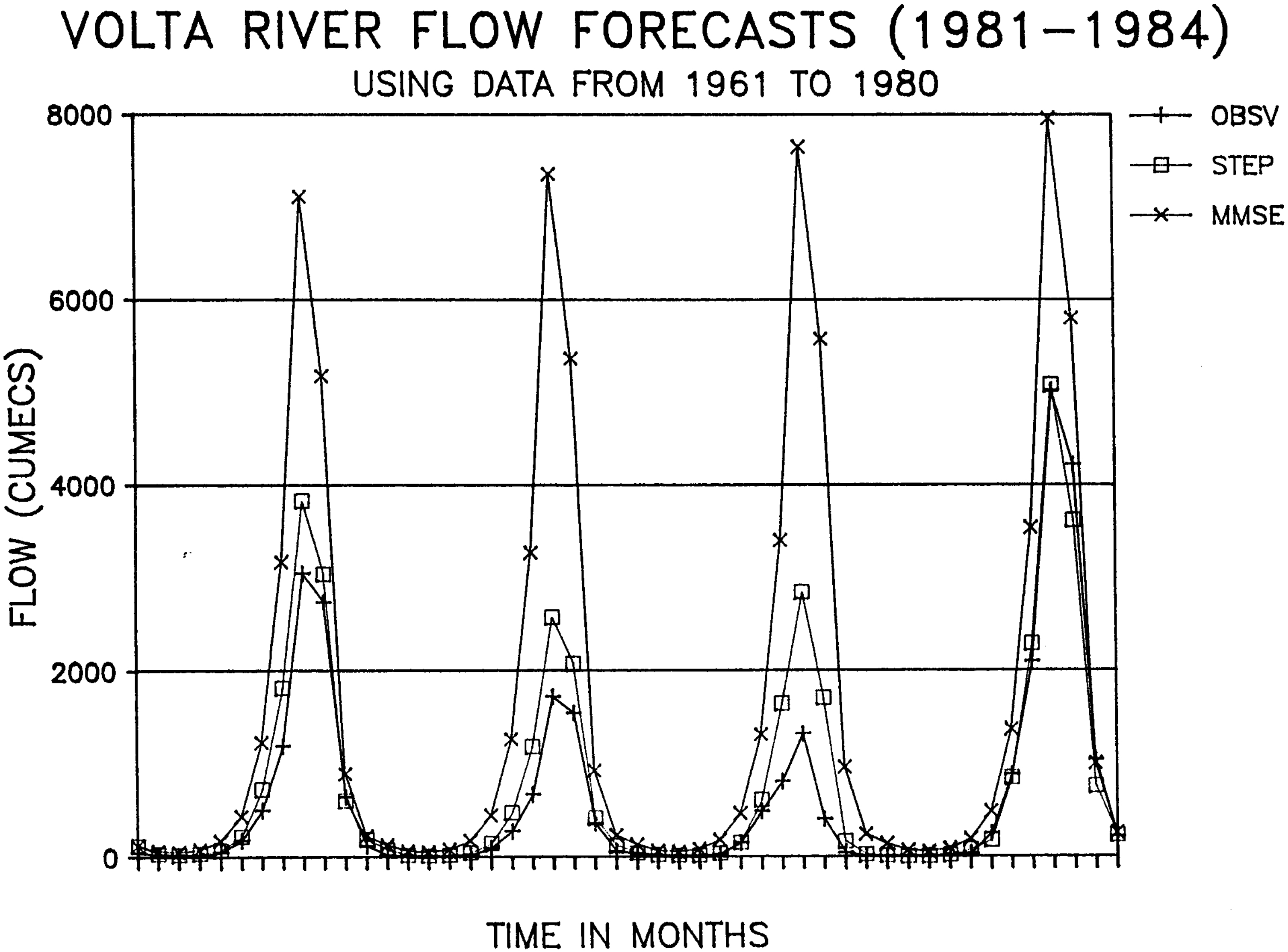

VOLTA RIVER FLOW FORECASTS (1981–1984)
USING DATA FROM 1961 TO 1980
FLOW (CUMECS)
8000
6000
4000
2000
0
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.6

VOLTA RIVER FLOW FORECASTS (1981-1984)

USING DATA FROM 1971 TO 1980

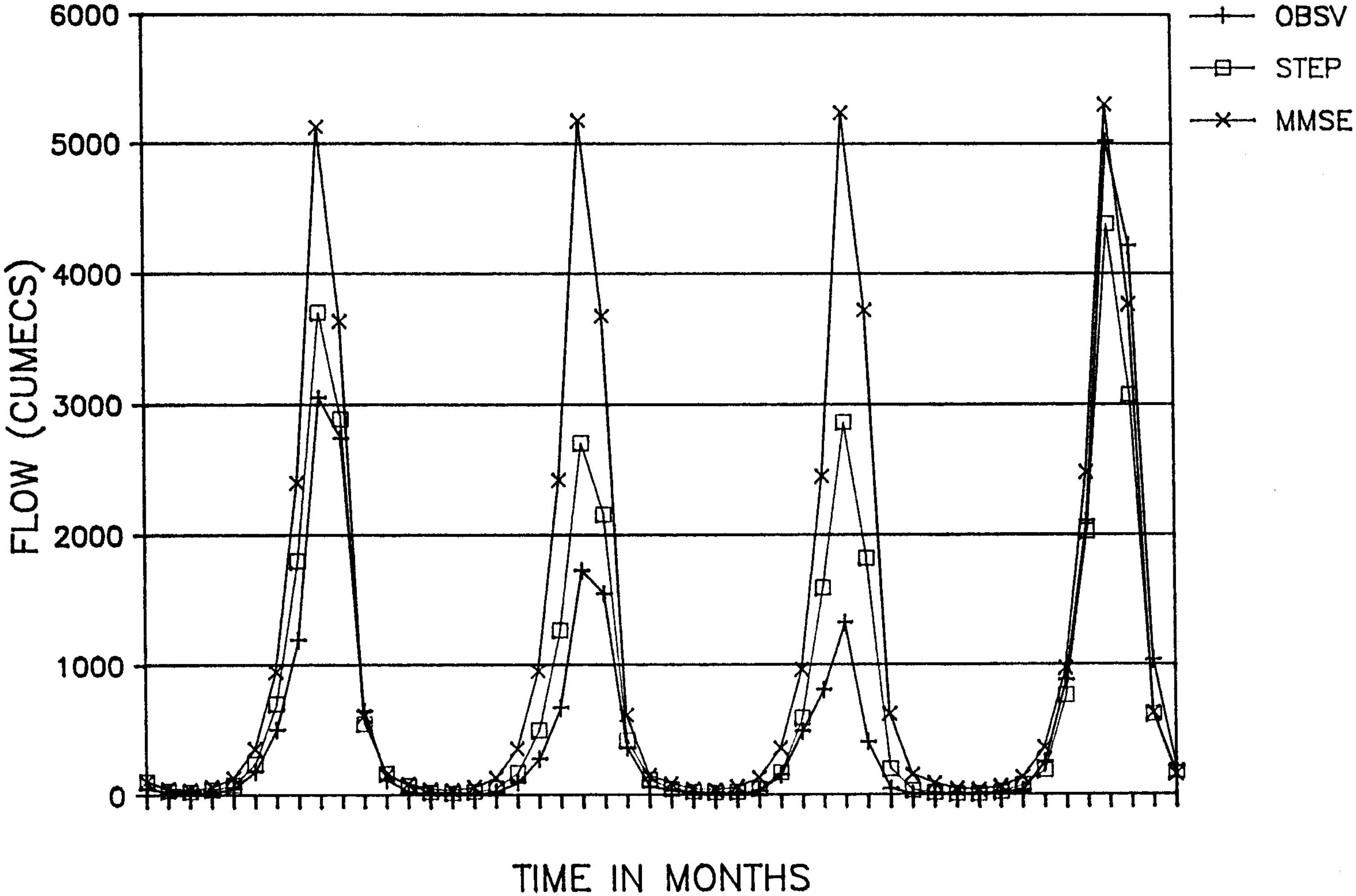


Figure 8.7

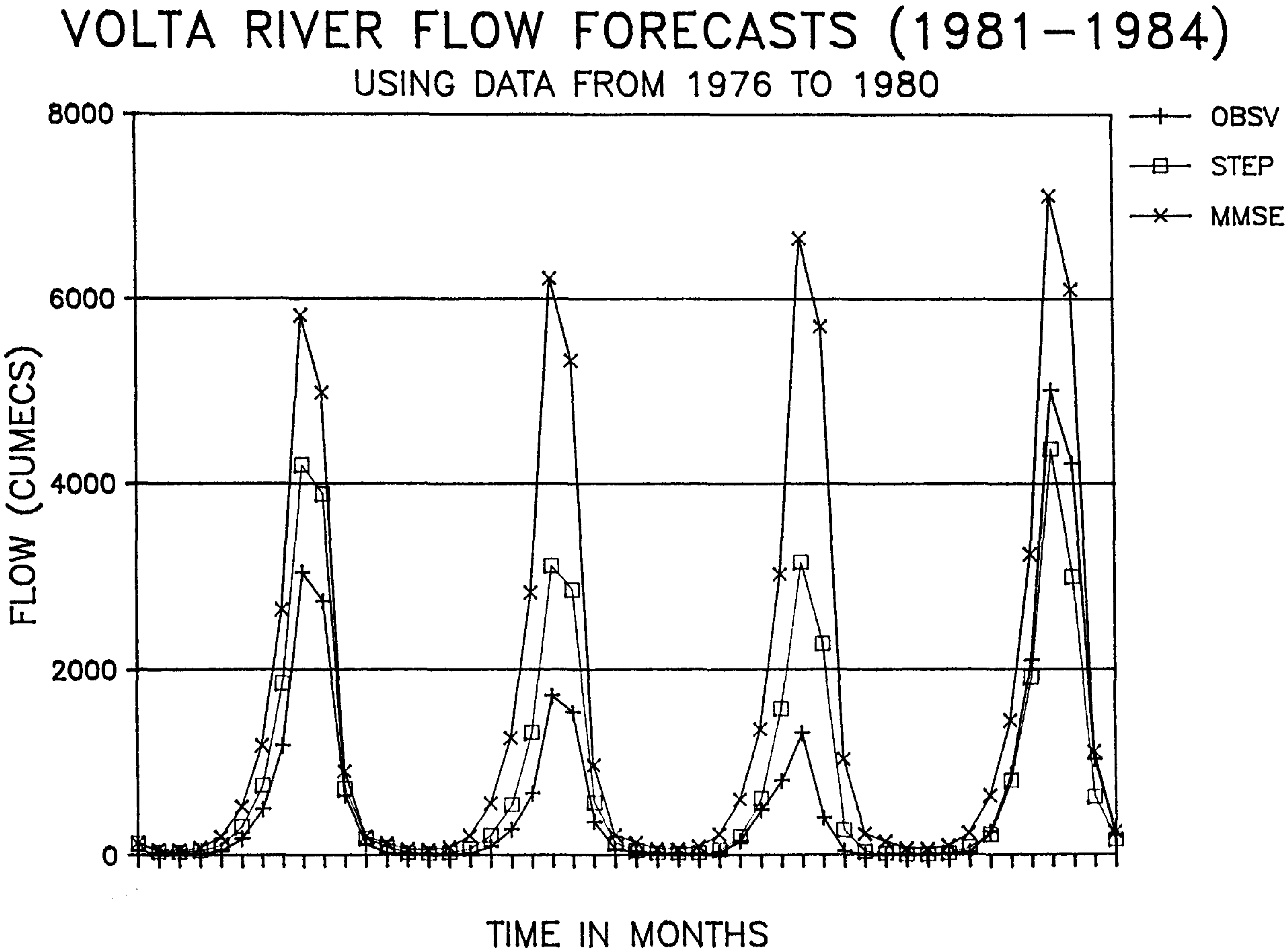

VOLTA RIVER FLOW FORECASTS (1981–1984)
USING DATA FROM 1976 TO 1980
OBSV
STEP
MMSE
FLOW (CUMECS)
8000
6000
4000
2000
0
TIME IN MONTHS

Figure 8.8

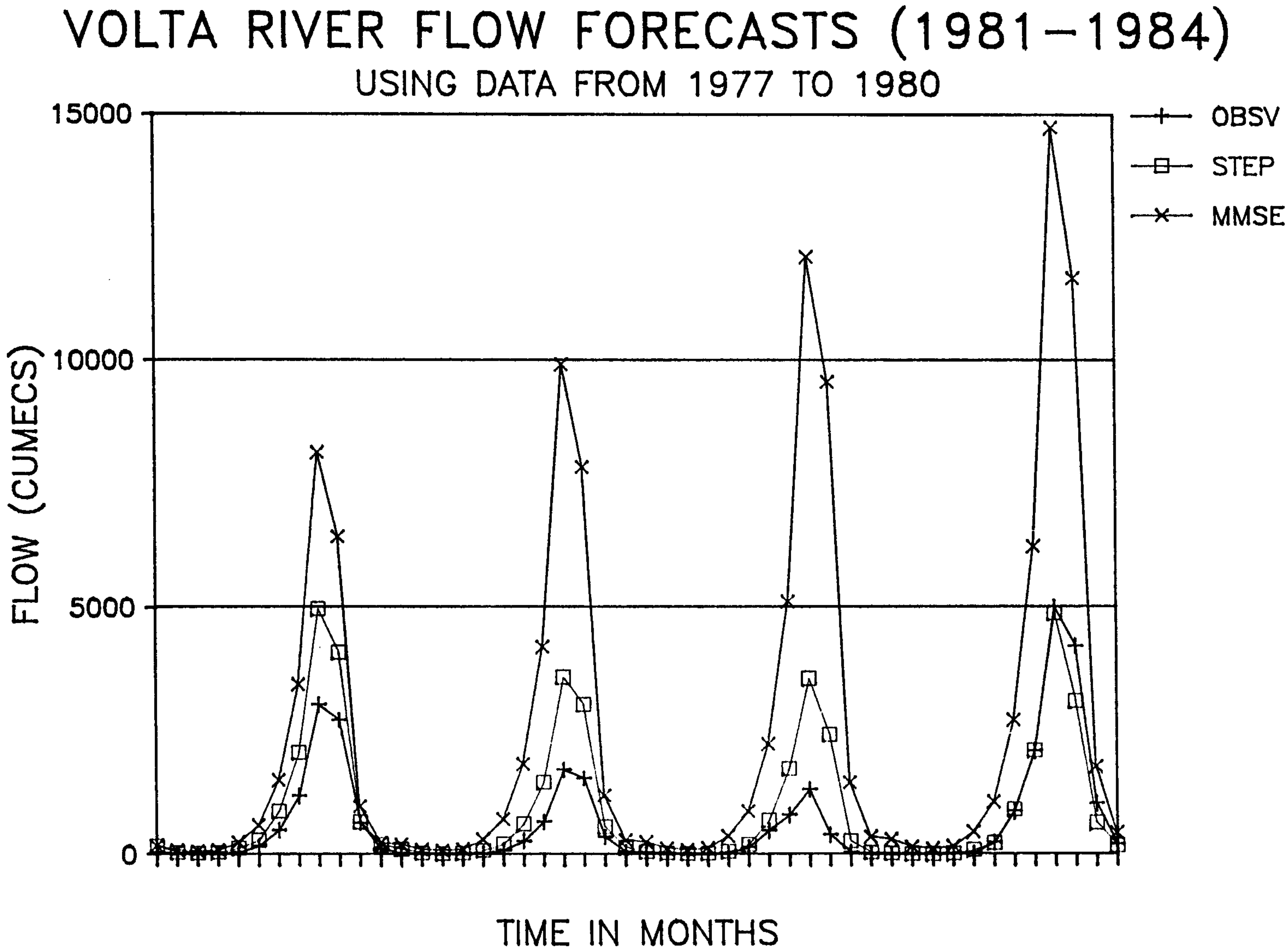

VOLTA RIVER FLOW FORECASTS (1981-1984)
USING DATA FROM 1977 TO 1980
15000
10000
5000
0
FLOW (CUMECS)
TIME IN MONTHS
OBSV
STEP
MMSE

Figure 8.9

8.4. SUMMARY

From the above observations, it is confirmed that shorter lengths of data do not provide models which forecast as well as those identified and estimated from longer data sets. In general however, the identified model form appears to be reasonably suitable even for shorter data spans taken from different parts of the historical data set. It is important however that model parameters and subsequently forecasts should be updated as soon as new observations are available.

9. UTILISING THE FORECASTED MONTHLY FLOWS

9.1. Introduction

The previous chapter shows forecasts obtained by employing the identified model SARIMA (1,0,0,) X (0,1,1) with different data sets. These forecasts were generated to test the validity of the model form which had been selected to represent the average monthly flows of the Volta river measured at Senchi.

In practice however, the forecasts generated by this model could be used in a number of ways especially as input into three Power and Energy Simulation programs available to the Volta River Authority. These computer programs HYDRO170, HYDRO824 and GFEPM developed at Acres International of Canada, are capable of short, medium and long term simulation respectively. Together, therefore, they can be employed for short term operational policy decisions as well as long term planning studies. A brief description of these programs is provided below.

9.2. HYDRO170

This is a general purpose single reservoir power and energy program. After the program is set up (ie after constructing the basic input data) different flow scenarios may be investigated with very little input data change.

9.2.1. REQUIRED DATA INPUT

This program is the least demanding of the three in terms of input data requirement. The following are standard data input:

1. name of input data file
2. name of output data file
3. title for particular scenario
4. choice of units, ie metric or imperial
5. number of years of flow data available
6. number of years being simulated
7. starting month of simulation
8. twenty five optional outputs to be printed
9. general hydro plant characteristics including,
    a. starting reservoir volume
    b. turbine head loss
    c. efficiency
    d. installed capacity

e. rated head

10. reservoir characteristics, (elevation, area, volume curves)
11. tail water elevation rating curve
12. net evaporation for 12 months
13. power rule curve
14. minimum rule curve
15. spill rule curve
16. irrigation rule curve
17. monthly average irrigation demand
18. monthly average energy demand
19. monthly average flow

It also has inputs which determine power releases during periods of rationing (load shedding).

The program is a "straight" simulation program which assesses the state of the reservoir at every stage and proceeds to allocate available energy (water) to power demand and irrigation after accounting for evaporation. This is done subject to the constraint of water level. All secondary energy is generated if there is enough power demand else the water is spilled after irrigation is taken into account. All energy is supplied subject to the constraint of the minimum rule curve. Any excess load is curtailed at the minimum rule curve. Below the irrigation rule curve, only power demand is supplied ie irrigation demand is curtailed below the irrigation rule curve. This program is handy for preliminary examination of reservoir behaviour on a short term in view of the output furnished.

9.2.2. OUTPUT LISTING

The main output from this program includes the following which can optionally be selected through "switches" included in the input data file.

1. monthly energy generated in gwh
2. secondary energy generation in gwh
3. primary energy generation in gwh
4. non-firm energy generation in gwh
5. power flows
6. month end reservoir volumes
7. month-end reservoir water elevations
8. monthly and annual energy generation coefficients
9. month-end reservoir spill volumes

The output also includes a water balance check on the water allocations if required. This program is ideal for a quick evaluation of reservoir performance given inflow and power demand conditions.

Traditionally, it has been used to compute the firm energy capability of a hydro plant by computing the expected energies produced using the whole of the historical inflow.

It is suggested here that this practice can be extended to include the use of forecasted inflow data to predict reservoir levels and energy availability over a short span of time. Specifically, it may be used to determine power and energy scheduling over the next hydrological year and also give an indication of expected reservoir elevations and power flows.

9.3. HYDRO824

This is a multi-purpose multi-reservoir hydro and thermal power and energy simulation model. It requires substantially more detailed data input than the single reservoir power and energy model described previously. It also affords the user the capability of defining thermal generating plants and when they come on line. Installation and availability of individual generating units at the hydro plants are required as input. The power demand is input as detailed monthly loads.

9.3.1. REQUIRED DATA INPUT

The main data input required are the following

1. the title for the particular scenario
2. name of hydro plants included in simulation

3. numerical identification of hydro plants and their priority scheduling.
4. reservoir characteristics at various plants
    a. maximum volume
    b. minimum volume (dead storage)
    c. irrigation rule curve
5. minimum flow for salinity control
6. reservoir operating data at various plants
    a. upper rule curve
    b. intermediate rule curve
    c. lower rule curve
    d. spill rule curve
7. availability of hydro plant units
8. detail scheduling of hydro and thermal plants
9. load duration curves for the entire period under simulation
10. volume elevation curves for reservoirs
11. area elevation curves for reservoirs
12. capacity head curves
13. tail water rating curves
14. capacity factors (ie percentage of capacity / head curve)
15. net evaporation data for various reservoirs (evaporation minus rainfall)
16. demand data
    a. peak monthly power demand
    b. average monthly demand
    c. irrigation demand

17. average monthly flow data

9.3.2. OUTPUT LISTING

The output from this program includes an echo of all the input data as well as the following:

1. monthly thermal peak output (expressed in megawatts)
2. monthly thermal energy output (expressed in gigawatt hours)
3. plant stacking in selected months (listed base to peak)
4. monthly irrigation diversion requirement
5. Load - Duration curve
6. monthly hydro peak output (expressed in megawatts)
7. monthly hydro energy output (expressed in gigawatt hours)
8. monthly final flow into the atlantic ocean (MCM)
9. monthly system load and hydro / thermal breakdown
10. monthly system energy and hydro / thermal breakdown
11. summary of energy generation for the run

The input data outlined above is set up and is hardly modified. It makes investigating different scenarios relatively easy once the data is set up. However, it demands more detailed knowledge about hydro plant characteristics and operation.As in the case of the single reservoir program, historical data is used in simulating "future" conditions and expectations taken over the whole set of available data. It is suggested that with the use of forecasted flows, more realistic conditions may be simulated.

The program allows the user to specify the total installed capacity at the inception of simulation. During the simulation period, additional generation may be selected from predefined hydro or thermal sets.

Some optimization is done with respect to reservoir draw down scheduling between the various reservoirs. When hydro plants have more than enough potential energy to meet demand, reservoir levels are raised in reverse priority sequence to minimize spill.

The program is currently configured to simulate the operation of three hydro plants in series corresponding to two existing plants and one planned addition. In addition, one can specify the addition of a number of fossil fueled, nuclear or mini hydro plants during the simulation period. This therefore makes it suitable for looking at the effects on the existing system (which is all hydro), of thermal complements.

It is believed that with the availability of forecasted flows, one may be able to compute to within certain confidence limits, the system characteristics during a specified period with thermal installation.

9.4. GFEPM

This program is the most elaborate of the three and is called the

Ghana Financial and Economic Planning Model [Ghana Generation Planning 1986 Update / Training Seminar]. It is basically an extension of the capabilities of the above two programs in that it has financial and economic routines included to deal with the financial consequences of current and additional installed generation during the simulation period. It computes the present net worth benefits by recognising benefits (energy sales) and costs (operating, maintenance and capital cost of constructing new facilities, depreciation and loan servicing). The program also computes detailed financial statements for the whole Volta River Authority taking cognizance of old loans as well as committed loans. As a result of the depth of detail, data input is exacting and demands multi disciplinary effort. This program is thus best run by both civil engineers and finance and economic personnel.

9.4.1. REQUIRED DATA INPUT

The data input required for this computer model includes financial, economic as well as reservoir and hydro plant parameters. In addition data is required for the planned generation expansion of the VRA system. A summary of the data input is as follows:

1. load forecast

2. general plant data
3. reservoir rule curves
4. tariff structure
5. processed average monthly flows
6. old loans
7. construction program
8. fixed assets
9. income sheet
10. balance sheet
11. rainfall data
12. reservoir data

9.4.2. OUTPUT LISTING

The output from this program could be voluminous depending on which outputs are selected for optional printing. A brief summary of the output is listed below:

1. an echo check of the input data
2. detailed reservoir simulation data
3. financial schedules
4. detailed financial statements.

It also prints out statistical summaries of the output data. Depreciation of fixed assets is taken into account while future costs and benefits are discounted to present worth.

It must be noted that the storage hydro portion of this program has been configured to take into account only the two existing hydro plants. It can however simulate the addition of different predefined thermal complements. The addition of thermal sets enable the program to be run to maximize annual average hydro energy in contrast to the practice before of maximizing firm hydro energy.

The economic and financial portion of this program is fairly detailed. It takes into account all the existing financial operations of the Authority. Additionally, it computes the financial and economic consequences of added generation during the simulation period.

This program therefore is the obvious choice for any study involving financial computations. It is especially suited to investigating the financial effect of any draw down policy scenario being studied. It also depicts the financial consequence of added generation during the period of simulation.

9.5. COMMENTS

To increase the authenticity of inflow assumptions made in any operational study, it may be more beneficial to employ predicted flows as input into the computer model in question. This could be a more realistic assumption than assuming a repeat of the historical flow or else taking the expected results from using all the flows on record.

To investigate the performance of a reservoir / hydro plant, it may be more useful to employ an identified stochastic model to generate different equally likely sequences of flow which preserve the statistical properties of the observed series of flow. This should provide a robust investigation of any study.

With reference to water resources studies, Kottegoda [1980] describes the restriction of using only historical flows for input data, as being a serious draw back in view of the fact that a particular sequence of flows will never repeat itself.

After all, it is highly unlikely that nature will run out of "random variables" and therefore decide to produce a "rerun" of the average monthly flow sequences in exactly the same order again and further more, decide to conveniently start the "reruns" exactly from the date of available record of flows.

10. DISCUSSION, CONCLUSION AND SUGGESTED FURTHER STUDY

10.1 DISCUSSION

The use of stochastic methods in the field of water resources especially hydrology has gained wide acceptance in recent years. This is due to a number of factors. Previously, purely deterministic methods were used to model and subsequently forecast hydrological phenomena which exhibited random behaviour. These models were however difficult to set up, involved too many parameters therefore making their updating very tedious [Fay, Watts and Watts, 1986]. This has led to the addition of probabilistic components in mathematical models which attempt to represent hydrologic phenomena.

Box and Jenkins [1976] greatly enhanced the wide acceptance of this methodology, which dates back to the beginning of this century, by introducing a systematic approach to the application of these methods. The success of these stochastic models which are sometimes referred to as Box-Jenkins models, can be attributed to the small number of parameters (parsimony) involved which therefore leads to simple calibration as well as easy and fast updating [Fay, Watts and Watts, 1986],[Olason and Watt, 1986]. Hipel [1985] mentions the use of stochastic modelling of river flow in the selection of the optimal design from alternative designs. This is accomplished by testing the physical and economic performance of the alternative designs by using different sets of simulated inputs into the system. After the design stage, stochastic modelling can assist in operations by forecasting input flows and demand on the hydroelectric system with a view to obtaining the optimal operating policy which maximizes production subject to physical, environmental, economic and political constraints [Hipel 1985].

For the purpose of this exercise, data consisting of average monthly flows of the Volta river at Senchi, were analyzed with a view to finding a stochastic model which adequately represents and preserves all the statistics of these flow series.
The process (recommended by Box and Jenkins [1976]), of identification, estimation and diagnostic checks was followed in building a suitable model from the available data. Three types of seasonal models were entertained namely periodic autoregressive models (PAR), deseasonalized models and finally the Box-Jenkins general multiplicative model or seasonal autoregressive integrated moving average model (SARIMA).

The first two seasonal types of models were rejected as being unsuitable candidates for modelling the available data. The periodic autoregressive model form has 12 AR(p) models which are fitted to each month of the year. This therefore results in many parameters to be estimated and demands a lot of computer time for the estimation of the parameters. Also for purpose of updating parameters, there would be too many parameters to be considered. Consequently, this method was rejected.

For the case of the deseasonalized model, the R squared value obtained was 48.43%. Since this indicated a low predictability of the given series, this method was also rejected.

The identified model form of SARIMA (1,0,0) X (0,1,1) was found to yield a model which passed the diagnostic checks done on the residuals. It was however necessary to invoke the Box Cox transformation (with $\delta=0$) corresponding to taking natural logs of the original data, in order to remove non-normality and heteroscedasticity in the variance. The process of overfitting was also used to confirm the order of the identified model. The parameters of this model were estimated as:

$$\phi(1) = 0.6955 \ , \ \Omega(1) = 0.9146 \ .$$

Substituting the values of these parameters into the general multiplicative model equation (7.2) and rearranging yields:

$$z_t = a_t + z_{t-12} + 0.6955z_{t-1} - 0.6955z_{t-13} - 0.9146a_{t-12} \qquad (10.1)$$

The conditional expectation of z[t+1] given all knowledge of the z's and a's up to time t gives the minimum mean square error forecast at lead time 1, from origin t. This simple expression is the basis of computations performed within the program "DRUSMAIN" to obtain minimum mean square error forecasts.

The model form is further validated by generating forecasts (both one step ahead forecasts and MMSE) for lead time of four years with nine sets of data. In each of these cases, the same model forms are utilised but the parameters are estimated to reflect the particular data series being used. The resulting forecasts are compared to the observed data in Chapter 8. It is perhaps worth noting that the period 1981 to 1983 marked the worst inflows on record and forms part of the critical period.

10.2 CONCLUSION

It is seen that, the SARIMA (1,0,0) X (0,1,1) model form, fitted to the average monthly flow data of the Volta river from 1936 to 1984, yielded a parsimonious representation of this set of data. Furthermore, the same model form was adequate in modelling subsets of the available data. The results presented in Chapter 8 indicate the inability of models identified using short flow records ($\leq$ 20 years) to forecast as well as those identified from longer records.

It has been demonstrated that it is reasonably simple and easy to identify this model type. The number of parameters involved are also found to be two. This therefore makes it easy to update the estimates of these parameters if additional data becomes available.

It is also noteworthy that this model is a univariate stochastic model. In effect the only data requirement is the historical flow at Senchi itself. This eliminates the problem of having to collect data for other stations (upstream) and the general handling of multiple inputs data. It therefore avoids the problem of verifying any other data sets except that from Senchi itself.

It may however be helpful if additional work is done with respect to finding the most accurate way of forecasting the average monthly flows of the Volta river.

10.3 SUGGESTED FURTHER WORK

The model identified for the purpose of this exercise was obtained after viewing the residual autocorrelations. It may be worthwhile to investigate the possibility of deriving a better fit to the data by using the partial autocorrelation function (PACF). Additionally, some of the more recent advances in model identification proposed by Hipel, McLeod and Lennox [1977] could be used to see whether any improvements in the model are obtained. These developments include the inverse autocorrelation function (IACF), inverse partial autocorrelation function (IPACF) and the cumulative periodogram white noise test.

From Chapter 6, data from six rain gauge stations and four flow stations are available and could be of help in identifying a multiple input single output (MISO) model which may yield some improvements over the univariate model just identified. In particular, a transfer function noise (TFN) type model could be identified to check for the effects, if any, of the construction of the Akosombo dam. Hipel, McLeod and McBean [1977] note that the construction of a reservoir could result in a significant reduction in the mean of the average monthly river flow downstream of the dam. The use of the TFN for intervention analysis is noted in the literature [Hipel, 1985], [D'Astous and Hipel, 1979], [Hipel, McLeod and Noakes, 1982]. This could be used to estimate missing data in some of the data files mentioned in Chapter 6. These data sets may then be used in identifying MISO models to represent the flow at Senchi. It is reasonable to assume that the inclusion of flow and / or rainfall data upstream of Senchi may help identify models which may be more representative of the monthly average flows at Senchi.

REFERENCES

Acres International Limited, Ghana Generation Planning Study, Update / Training Seminar, Documentation for GFEPM, Volta River Authority, 1986.

Acres International Limited, Ghana Generation Planning Study, Appendix A, Present Status of the Power System, Volta River Authority, 1985.

Acres International Limited, Ghana Generation Planning Study, Appendix C, Hydrology, Volta River Authority, 1984.

Box, G. E. & Jenkins, G. M., Time Series Analysis Forecasting and Control, Holden-Day, 1976.

D'Astous, F. & Hipel K. W., Analyzing Environmental Time Series, Journal of the Engineering Division, 1979.

Fay, D. M. & Watt, W. E. & Watts, D. G., A Stochastic real-time flood forecasting system for Carman, Manitoba, Canadian Journal of Civil Engineering, Vol. 14, #1, 1987

Government of Ghana, The Volta River Basin Hydrological Data Book, Part 1, Ministry of Works and Housing, Hydrological Services, 1966.

Hipel, K. W., Stochastic Research in Multivariate Analysis, Presented at Fourth International Hydrology Symposium on Multivariate Analysis of Hydrological Processes, Colorado State University, Fort Collins, Colorado, 1985.

Hipel, K. W. & McLeod A. I. & Noakes D. J., Fitting Dynamic Models to Hydrological Time Series, Time Series Methods in Hydrosciences, Elsevier Scientific Publishing Company, Amsterdam, 1982.

Hipel, K. W., Time Series in Perspective, Water resources Bulletin, Vol 24, #4, American Water Resources Association, 1985.

Hipel, K. W., Geophysical Model Discrimination Using the Akaike Information Criteria, IEEE Transactions on Automatic Control, Vol. AC-26, #2, April 1981.

Hipel, K. W. & McLeod A. I. & McBean E. A., Stochastic Modelling of the Effects of Reservoir Operation, Journal of Hydrology, Elsevier Scientific Publishing Company, Amsterdam, 1977.
Hipel, K. W., Stochastic Techniques in System Design, 1981.

Hipel, K. W. & McLeod A. I. & Lennox W. C., Advances in Box Jenkins Modeling, Part 1, Model Construction, Water Resources Research, Vol. 13, #3, American Geophysical Union, Amsterdam, 1977.

Hipel, K. W. & McLeod A. I. & Lennox W. C., Advances in Box Jenkins Modeling, Part 2, Applications, Water Resources Research, Vol. 13, #3, American Geophysical Union, Amsterdam, 1977.

Hipel, K. W. & McLeod A. I., Simulation Procedures for Box-Jenkins Models, Water Resources Research, Vol. 14, #5, American Geophysical Union, Amsterdam, 1978.

Kottegoda, N. T., Stochastic Water Resources Technology, MacMillan, 1980.

Lavender, S. T. & Anderson, J. E., "The recent African drought within the context of earlier African hydrological experience", presented at the Sixth Meeting of the Committee of Cooperation for Development in Africa: Energy Group, Rome, October 1984.

MacGregor, J. F., Discrete Time Process Identification: Theory and Applications, May 1986.

Noakes, D. J. & McLeod, A. I. & Hipel K. W., Forecasting Monthly Riverflow Time Series, International Journal of Forecasting, Elsevier Scientific Publishing Company, B.V. (North-Holland), 1985.

Olason, T. & Watt, W. E., Multivariate Transfer Function-Noise Model of River Flow for Hydropower Operation, Nordic Hydrology, 17, 1986.

Thompstone, R. M., Hipel, K. W. & McLeod, A. I., Grouping of Periodic Autoregressive Models, Time Series Analyses: Theory and Practice 6, Elsevier Science Publishers B. V. (North-Holland), 1985.

APPENDIX A

TYPICAL OUTPUT FROM PROGRAM DRUSMAIN

Volta River Flows at Senchi (1936-1980)

SARIMA(1, 0, 0)(0, 1, 1)12

TRANSFORMED DATA, ALAMDA = .000000D+00 CONS = .000000D+00

TREND COEFFICIENT	RESIDUAL VARIANCE
4.687846D-04	2.429629D-01

GENERALIZED AUTOGRESSIVE COEFFICIENTS

1: .656,	2: .000,	3: .000,	4: .000,	5: .000,	6: .000,
7: .000,	8: .000,	9: .000,	10: .000,	11: .000,	12: 1.000,
13: -.656,					

MOVING AVERAGE TERM

1: .000,	2: .000,	3: .000,	4: .000,	5: .000,	6: .000,
7: .000,	8: .000,	9: .000,	10: .000,	11: .000,	12: .916,

PIVOTAL VALUES FOR FORECASTING
FORECASTING ORIGIN = 540

OBSERVATIONS

Z(525)= 8.853D+00,	Z(526)= 8.746D+00,	Z(527)= 7.315D+00,	Z(528)= 5.694D+00
Z(529)= 4.585D+00,	Z(530)= 3.871D+00,	Z(531)= 3.584D+00,	Z(532)= 3.784D+00
Z(533)= 4.625D+00,	Z(534)= 5.727D+00,	Z(535)= 6.706D+00,	Z(536)= 7.555D+00
Z(537)= 8.488D+00,	Z(538)= 8.380D+00,	Z(539)= 6.950D+00,	Z(540)= 5.328D+00

DISTURBANCES

A(529)=-3.071D-01,	A(530)=-2.140D-02,	A(531)=-3.995D-02,	A(532)=-1.715D-01
A(533)= 9.502D-02,	A(534)= 1.684D-01,	A(535)= 2.270D-03,	A(536)=-6.098D-02
A(537)= 5.014D-02,	A(538)= 1.906D-01,	A(539)= 3.877D-01,	A(540)=-1.052D-01

Volta River Flows at Senchi (1981-1984)

OBSERVED VALUES ARE THE BOX-COX TRANSFORMATION OF THE ORIGINAL SERIES

TIME PERIOD	OBSERVED VALUE	1-STEP AHEAD FORECAST	INNOVATION
541	4.127134D+00	4.626786D+00	-4.996514D-01
542	3.433987D+00	3.590797D+00	-1.568099D-01
543	3.135494D+00	3.333643D+00	-1.981483D-01
544	3.332205D+00	3.647766D+00	-3.155615D-01
545	4.174387D+00	4.241714D+00	-6.732718D-02
546	5.267858D+00	5.277277D+00	-9.418457D-03
547	6.246107D+00	6.402782D+00	-1.566757D-01
548	7.095064D+00	7.310132D+00	-2.150680D-01
549	8.027150D+00	8.140161D+00	-1.130108D-01

550	7.919356D+00	7.903436D+00	1.592029D-02
551	6.489205D+00	6.292645D+00	1.965598D-01
552	4.867534D+00	5.122439D+00	-2.549046D-01
553	3.555348D+00	4.283383D+00	-7.280348D-01
554	2.833213D+00	3.202890D+00	-3.696767D-01
555	2.564949D+00	2.923257D+00	-3.583077D-01
556	2.772589D+00	3.247412D+00	-4.748236D-01
557	3.610918D+00	3.869270D+00	-2.583523D-01
558	4.700480D+00	4.907141D+00	-2.066603D-01
559	5.676754D+00	6.017780D+00	-3.410262D-01
560	6.527958D+00	6.918956D+00	-3.909978D-01
561	7.460490D+00	7.758984D+00	-2.984939D-01
562	7.352441D+00	7.533323D+00	-1.808815D-01
563	5.921578D+00	5.937454D+00	-1.587577D-02
564	4.304065D+00	4.729068D+00	-4.250027D-01
565	3.806662D+00	3.853217D+00	-4.655416D-02
566	3.135494D+00	3.337421D+00	-2.019268D-01
567	2.484907D+00	3.092188D+00	-6.072816D-01
568	2.639057D+00	3.155682D+00	-5.166251D-01
569	3.713572D+00	3.760518D+00	-4.694550D-02
570	5.117994D+00	4.957720D+00	1.602736D-01
571	6.222576D+00	6.263785D+00	-4.120837D-02
572	6.710523D+00	7.244998D+00	-5.344750D-01
573	7.198931D+00	7.854340D+00	-6.554086D-01
574	6.037871D+00	7.347014D+00	-1.309143D+00
575	4.060443D+00	5.073814D+00	-1.013371D+00
576	2.890372D+00	3.472527D+00	-5.821556D-01
577	2.564949D+00	2.921957D+00	-3.570075D-01
578	2.397895D+00	2.506057D+00	-1.081614D-01
579	2.639057D+00	2.557824D+00	8.123338D-02
580	3.465736D+00	3.214167D+00	2.515687D-01
581	3.850148D+00	4.299633D+00	-4.494854D-01
582	5.579730D+00	5.061215D+00	5.185146D-01
583	6.796824D+00	6.563859D+00	2.329646D-01
584	7.656810D+00	7.577711D+00	7.909909D-02
585	8.522778D+00	8.421129D+00	1.016490D-01
586	8.349484D+00	8.106993D+00	2.424909D-01
587	6.956545D+00	6.506795D+00	4.497501D-01
588	5.420535D+00	5.325144D+00	9.539063D-02

AVERAGE MEAN SQUARE ERROR = 1.631481D-01
COEFFICIENT OF FORECASTIBILITY = 95.28%

BOX-TIAO TEST FOR COMPARISON OF FORECAST WITH ACTUALITY

CHI SQ	DF	S.L.
32.23	48	.96075

FORECASTS IN ORIGINAL DATA DOMAIN

TIME PERIOD	OBSERVED VALUE	MINIMUM MSE FORECAST
541	6.200000D+01	1.153842D+02
542	3.100000D+01	4.094700D+01
543	2.300000D+01	3.166221D+01
544	2.800000D+01	4.334743D+01
545	6.500000D+01	7.850764D+01
546	1.940000D+02	2.211316D+02
547	5.160000D+02	6.814780D+02
548	1.206000D+03	1.688530D+03
549	3.063000D+03	3.872449D+03
550	2.750000D+03	3.056169D+03
551	6.580000D+02	6.104074D+02
552	1.300000D+02	1.894112D+02
553	3.500000D+01	8.184804D+01
554	1.700000D+01	2.778153D+01
555	1.300000D+01	2.100454D+01
556	1.600000D+01	2.904638D+01
557	3.700000D+01	5.409554D+01
558	1.100000D+02	1.527224D+02
559	2.920000D+02	4.637110D+02
560	6.840000D+02	1.141887D+03
561	1.738000D+03	2.645104D+03
562	1.560000D+03	2.110762D+03
563	3.730000D+02	4.279197D+02
564	7.400000D+01	1.278106D+02
565	4.500000D+01	5.323404D+01
566	2.300000D+01	3.178208D+01
567	1.200000D+01	2.487018D+01
568	1.400000D+01	2.650051D+01
569	4.100000D+01	4.852111D+01
570	1.670000D+02	1.606458D+02
571	5.040000D+02	5.930425D+02
572	8.210000D+02	1.582055D+03
573	1.338000D+03	2.909747D+03
574	4.190000D+02	1.751969D+03
575	5.800000D+01	1.804214D+02
576	1.800000D+01	3.637962D+01
577	1.300000D+01	2.097725D+01
578	1.100000D+01	1.383966D+01
579	1.400000D+01	1.457497D+01
580	3.200000D+01	2.809660D+01
581	4.700000D+01	8.318895D+01
582	2.650000D+02	1.781626D+02
583	8.950000D+02	8.005833D+02
584	2.115000D+03	2.206565D+03
585	5.028000D+03	5.128713D+03
586	4.228000D+03	3.746116D+03
587	1.050000D+03	7.561780D+02
588	2.260000D+02	2.319741D+02

```
AVERAGE MEAN SQUARE ERROR       =   1.583376D+05
COEFFICIENT OF FORECASTIBILITY =          86.86%

TOTAL OBSERVED VOLUME MCM       =   8.167896D+04
TOTAL FORECASTED VOLUME MCM     =   1.013032D+05
ACCURACY OF VOLUME FORECAST     =          75.97%
```

Volta River Flows at Senchi (1936-1984)

SARIMA(1, 0, 0)(0, 1, 1)12

LENGTH OF THE INPUT TIME SERIES = 588

LENGTH OF DIFFERENCED TIME SERIES = 576

MODIFIED SUM OF SQUARES ESTIMATION

SUM OF SQUARES	RESIDUAL VARIANCE	R-SQ
1.07084102D+07	2.35436711D-01	92.95%

AIC	BIC
5.78829421D+03	-2.90290056D+03

BOX-COX TRANSFORMATION PARAMETERS

LAMDA	SE(LAMDA)	CONS
.0000	.0000	.000000D+00

DETERMINISTIC COMPONENT

SERIES MEAN	S.E.	TREND TERM	S.E.
-8.092909D-03	5.669448D-03	-2.463993D-03	1.751303D-03

ESTIMATED BETA PARAMETERS

BETA	SE(BETA)
.6955	.0299
.9146	.0168

CORRELATION MATRIX OF BETA

1.000	
.005	1.000

--------------------RESIDUAL ANALYSIS--------------------------------------

SKEWNESS / KURTOSIS

G1	S.L.	G2	S.L.
.0948	.343515	2.2220	.000000

TEST FOR HETEROSCEDASTICITY DEPENDING ON THE CURRENT LEVEL

CHI	SE(CHI)
.004771	.031085

TEST FOR TRENDS IN THE VARIANCE OVER TIME

CHI SE(CHI)

.000896 .000354

RESIDUAL AND SQUARED-RESIDUAL AUTOCORRELATIONS

LAG	RA	SE	QA	RAA	QAA
1	.03563	.02898	.73	.05142	1.53
2	-.07310	.03609	3.83	.18297	20.95
3	.01438	.03907	3.95	-.01984	21.18
4	.02145	.04043	4.22	.00178	21.18
5	.00183	.04107	4.22	-.01618	21.33
6	.00631	.04138	4.25	-.01509	21.46
7	-.01606	.04153	4.40	-.05588	23.29
8	.00374	.04160	4.41	.03114	23.86
9	-.00113	.04163	4.41	-.06552	26.38
10	.02438	.04165	4.76	.00280	26.39
11	.09314	.04166	9.87	-.00083	26.39
12	.01114	.03811	9.94	.01333	26.49
13	-.02554	.04166	10.33	.05332	28.17
14	-.08271	.04167	14.38	.02251	28.47
15	.03445	.04167	15.09	-.03197	29.08
16	.09437	.04167	20.38	-.02713	29.52
17	-.00234	.04167	20.38	-.07062	32.49
18	-.03745	.04167	21.22	-.02013	32.73
19	.04002	.04167	22.18	.02227	33.02
20	.04066	.04167	23.17	-.01645	33.19
21	.03354	.04167	23.84	-.04858	34.60
22	-.01029	.04167	23.91	-.01902	34.82
23	.01035	.04167	23.97	-.01929	35.04
24	-.04167	.03871	25.02	.09248	40.20
25	.03231	.04167	25.65	.01254	40.30
26	.04918	.04167	27.11	.08476	44.65
27	.04414	.04167	28.29	-.02528	45.03
28	.02273	.04167	28.61	.01941	45.26
29	-.01272	.04167	28.71	.02032	45.51
30	-.04689	.04167	30.05	-.01209	45.60
31	.06296	.04167	32.47	.00033	45.60
32	.02061	.04167	32.73	.00411	45.61
33	-.06093	.04167	35.00	.00867	45.66
34	.04228	.04167	36.10	.02641	46.09
35	-.02093	.04167	36.37	.07720	49.75
36	-.03023	.03921	36.94	.07423	53.15
37	.05350	.04167	38.70	-.00661	53.18
38	-.02610	.04167	39.13	.04566	54.47
39	.02721	.04167	39.58	.00546	54.49
40	-.04220	.04167	40.69	-.03258	55.15
41	-.03279	.04167	41.36	-.04509	56.41
42	.05219	.04167	43.06	-.04337	57.58
43	-.01631	.04167	43.22	-.00630	57.61
44	.00195	.04167	43.23	-.04844	59.08
45	-.02652	.04167	43.67	-.02441	59.45
46	.01911	.04167	43.90	-.02884	59.97

47	-.01865	.04167	44.12	-.00346	59.98
48	.05924	.03962	46.33	.08610	64.66
49	-.05704	.04167	48.38	-.00247	64.66
50	-.00447	.04167	48.40	-.05508	66.58
51	-.05641	.04167	50.41	-.03799	67.50
52	.00672	.04167	50.44	-.05445	69.38
53	-.02773	.04167	50.93	-.03155	70.01
54	.01646	.04167	51.11	.00143	70.01
55	.01318	.04167	51.22	-.03259	70.69
56	.05702	.04167	53.30	-.02622	71.13
57	-.00244	.04167	53.30	.00836	71.18
58	-.00580	.04167	53.32	.00266	71.18
59	.06025	.04167	55.66	.04810	72.67
60	-.04106	.03996	56.75	.04953	74.25

Q(60) = 56.75 ON 58 DF IS NOT SIGNIFICANT AT THE 5 PERCENT LEVEL

Q(60) = 74.25 ON 60 DF IS NOT SIGNIFICANT AT THE 5 PERCENT LEVEL

RESIDUAL AND SQUARED-RESIDUAL SEASONAL AUTOCORRELATIONS

	LAG	RAS	QAS	RAAS	QAAS
1	12	.01114	.07	.01333	.10
2	24	-.04167	1.12	.09248	5.26
3	36	-.03023	1.68	.07423	8.66
4	48	.05924	3.90	.08610	13.33
5	60	-.04106	4.98	.04953	14.92
6	72	.02441	5.38	.05283	16.76
7	84	-.06875	8.58	.03396	17.54
8	96	.06140	11.19	.01627	17.73
9	108	-.06188	13.92	-.00521	17.74
10	120	-.01424	14.06	.01794	17.98
11	132	.01937	14.35	-.00592	18.01
12	144	-.01249	14.47	.01693	18.23
13	156	-.08658	20.41	.05084	20.28
14	168	-.03323	21.31	.01650	20.50
15	180	-.02588	21.87	.02012	20.84
16	192	.02496	22.41	.01662	21.08
17	204	-.03281	23.38	.05210	23.51
18	216	-.03667	24.62	-.00524	23.53
19	228	.02864	25.40	-.00936	23.62
20	240	-.00371	25.42	-.01069	23.73

Q(20) = 25.42 ON 20 DF IS NOT SIGNIFICANT AT THE 5 PERCENT LEVEL

Q(20) = 23.73 ON 20 DF IS NOT SIGNIFICANT AT THE 5 PERCENT LEVEL

Volta River Flows at Senchi (1936-1984)

SARIMA(1, 0, 0)(0, 1, 1)12

TRANSFORMED DATA, ALAMDA = .000000D+00 CONS = .000000D+00

TREND COEFFICIENT	RESIDUAL VARIANCE
-2.463993D-03	2.354367D-01

GENERALIZED AUTOGRESSIVE COEFFICIENTS

1:	.696,	2:	.000,	3:	.000,	4:	.000,	5:	.000,	6:	.000,
7:	.000,	8:	.000,	9:	.000,	10:	.000,	11:	.000,	12:	1.000,
13:	-.696,										

MOVING AVERAGE TERM

1:	.000,	2:	.000,	3:	.000,	4:	.000,	5:	.000,	6:	.000,
7:	.000,	8:	.000,	9:	.000,	10:	.000,	11:	.000,	12:	.915,

COEFFICIENTS IN THE MOVING AVERAGE EXPANSION

1:	.696,	2:	.484,	3:	.336,	4:	.234,	5:	.163,	6:	.113,
7:	.079,	8:	.055,	9:	.038,	10:	.026,	11:	.018,	12:	.098,
13:	.068,	14:	.048,	15:	.033,	16:	.023,	17:	.016,	18:	.011,
19:	.008,	20:	.005,	21:	.004,	22:	.003,	23:	.002,	24:	.087,
25:	.060,	26:	.042,	27:	.029,	28:	.020,	29:	.014,	30:	.010,
31:	.007,	32:	.005,	33:	.003,	34:	.002,	35:	.002,	36:	.086,
37:	.060,	38:	.042,	39:	.029,	40:	.020,	41:	.014,	42:	.010,
43:	.007,	44:	.005,	45:	.003,	46:	.002,	47:	.002,	48:	.086,

PIVOTAL VALUES FOR FORECASTING
FORECASTING ORIGIN = 540

OBSERVATIONS

Z(525)= 8.853D+00, Z(526)= 8.746D+00, Z(527)= 7.315D+00, Z(528)= 5.694D+00
Z(529)= 4.585D+00, Z(530)= 3.871D+00, Z(531)= 3.584D+00, Z(532)= 3.784D+00
Z(533)= 4.625D+00, Z(534)= 5.727D+00, Z(535)= 6.706D+00, Z(536)= 7.555D+00
Z(537)= 8.488D+00, Z(538)= 8.380D+00, Z(539)= 6.950D+00, Z(540)= 5.328D+00

DISTURBANCES

A(529)=-3.027D-01, A(530)= 4.593D-03, A(531)=-1.225D-02, A(532)=-1.423D-01
A(533)= 1.352D-01, A(534)= 2.025D-01, A(535)= 2.829D-02, A(536)=-3.352D-02
A(537)= 8.528D-02, A(538)= 2.248D-01, A(539)= 4.154D-01, A(540)=-9.266D-02

LEAD TIME	FORECAST	S.D.
1	4.604894D+00	4.852182D-01
2	3.878396D+00	5.910451D-01
3	3.597266D+00	6.359518D-01
4	3.921396D+00	6.565750D-01
5	4.594304D+00	6.663228D-01

6	5.517883D+00	6.709877D-01
7	6.531954D+00	6.732329D-01
8	7.462772D+00	6.743164D-01
9	8.342889D+00	6.748399D-01
10	8.071185D+00	6.750930D-01
11	6.352702D+00	6.752154D-01
12	4.994822D+00	6.752746D-01
13	4.370778D+00	6.769535D-01
14	3.713096D+00	6.777642D-01
15	3.479830D+00	6.781561D-01
16	3.837250D+00	6.783456D-01
17	4.533314D+00	6.784372D-01
18	5.472998D+00	6.784815D-01
19	6.498271D+00	6.785030D-01
20	7.436880D+00	6.785134D-01
21	8.322416D+00	6.785184D-01
22	8.054481D+00	6.785208D-01
23	6.338620D+00	6.785220D-01
24	4.982563D+00	6.785226D-01
25	4.359788D+00	6.798236D-01
26	3.702988D+00	6.804520D-01
27	3.470336D+00	6.807559D-01
28	3.828183D+00	6.809028D-01
29	4.524543D+00	6.809739D-01
30	5.464433D+00	6.810083D-01
31	6.489850D+00	6.810249D-01
32	7.428559D+00	6.810330D-01
33	8.314165D+00	6.810368D-01
34	8.046278D+00	6.810387D-01
35	6.330450D+00	6.810396D-01
36	4.974417D+00	6.810401D-01
37	4.351658D+00	6.823318D-01
38	3.694869D+00	6.829559D-01
39	3.462225D+00	6.832576D-01
40	3.820077D+00	6.834035D-01
41	4.516441D+00	6.834741D-01
42	5.456334D+00	6.835082D-01
43	6.481753D+00	6.835247D-01
44	7.420463D+00	6.835327D-01
45	8.306070D+00	6.835366D-01
46	8.038183D+00	6.835384D-01
47	6.322356D+00	6.835393D-01
48	4.966323D+00	6.835398D-01

FORECASTS IN UNTRANSFORMED DOMAIN

ORIGIN TIME	OBSERVED
540	2.06000D+02

LEAD TIME	NAIVE FORECAST	90% PROBABILITY INTERVAL	MMSE FORECAST
1	9.99724D+01	(4.50054D+01, 2.22073D+02)	1.12462D+02
2	4.83466D+01	(1.82875D+01, 1.27814D+02)	5.75736D+01
3	3.64983D+01	(1.28228D+01, 1.03888D+02)	4.46781D+01
4	5.04708D+01	(1.71402D+01, 1.48616D+02)	6.26110D+01
5	9.89193D+01	(3.30593D+01, 2.95984D+02)	1.23507D+02
6	2.49107D+02	(8.26164D+01, 7.51114D+02)	3.11997D+02
7	6.86739D+02	(2.26918D+02, 2.07833D+03)	8.61414D+02
8	1.74197D+03	(5.74571D+02, 5.28127D+03)	2.18664D+03
9	4.20021D+03	(1.38420D+03, 1.27451D+04)	5.27426D+03
10	3.20089D+03	(1.05443D+03, 9.71680D+03)	4.02009D+03
11	5.74042D+02	(1.89062D+02, 1.74294D+03)	7.21015D+02
12	1.47647D+02	(4.86230D+01, 4.48338D+02)	1.85456D+02
13	7.91052D+01	(2.59791D+01, 2.40872D+02)	9.94755D+01
14	4.09805D+01	(1.34405D+01, 1.24950D+02)	5.15617D+01
15	3.24542D+01	(1.06373D+01, 9.90173D+01)	4.08447D+01
16	4.63977D+01	(1.52027D+01, 1.41603D+02)	5.84006D+01
17	9.30665D+01	(3.04896D+01, 2.84076D+02)	1.17150D+02
18	2.38173D+02	(7.80225D+01, 7.27052D+02)	2.99815D+02
19	6.63992D+02	(2.17508D+02, 2.02699D+03)	8.35854D+02
20	1.69745D+03	(5.56033D+02, 5.18193D+03)	2.13681D+03
21	4.11509D+03	(1.34797D+03, 1.25626D+04)	5.18026D+03
22	3.14787D+03	(1.03114D+03, 9.60987D+03)	3.96268D+03
23	5.66015D+02	(1.85407D+02, 1.72794D+03)	7.12526D+02
24	1.45848D+02	(4.77747D+01, 4.45248D+02)	1.83600D+02
25	7.82406D+01	(2.55741D+01, 2.39367D+02)	9.85800D+01
26	4.05683D+01	(1.32467D+01, 1.24242D+02)	5.11363D+01
27	3.21475D+01	(1.04918D+01, 9.85020D+01)	4.05303D+01
28	4.59789D+01	(1.50022D+01, 1.40916D+02)	5.79741D+01
29	9.22537D+01	(3.00975D+01, 2.82773D+02)	1.16327D+02
30	2.36142D+02	(7.70363D+01, 7.23854D+02)	2.97769D+02
31	6.58424D+02	(2.14791D+02, 2.01835D+03)	8.30267D+02
32	1.68338D+03	(5.49145D+02, 5.16033D+03)	2.12274D+03
33	4.08128D+03	(1.33137D+03, 1.25111D+04)	5.14649D+03
34	3.12215D+03	(1.01848D+03, 9.57092D+03)	3.93704D+03
35	5.61409D+02	(1.83138D+02, 1.72100D+03)	7.07939D+02
36	1.44664D+02	(4.71912D+01, 4.43468D+02)	1.82422D+02
37	7.76070D+01	(2.52626D+01, 2.38410D+02)	9.79490D+01
38	4.02403D+01	(1.30856D+01, 1.23746D+02)	5.08095D+01
39	3.18878D+01	(1.03643D+01, 9.81092D+01)	4.02715D+01
40	4.56077D+01	(1.48200D+01, 1.40355D+02)	5.76043D+01
41	9.15094D+01	(2.97322D+01, 2.81647D+02)	1.15585D+02
42	2.34237D+02	(7.61013D+01, 7.20974D+02)	2.95872D+02
43	6.53115D+02	(2.12185D+02, 2.01032D+03)	8.24977D+02
44	1.66981D+03	(5.42482D+02, 5.13981D+03)	2.10922D+03
45	4.04837D+03	(1.31521D+03, 1.24613D+04)	5.11371D+03
46	3.09698D+03	(1.00613D+03, 9.53287D+03)	3.91197D+03
47	5.56884D+02	(1.80917D+02, 1.71416D+03)	7.03431D+02
48	1.43498D+02	(4.66188D+01, 4.41706D+02)	1.81261D+02