RM-76-69

ARIMA MODELS AND THEIR USE IN MODELLING HYDROLOGIC SEQUENCES

J. Ledolter

September 1976

PREFACE

The problem of how to model the time dependence of observations and of how to predict future observations using past data arises in many areas of application at IIASA. A very broad class of models which is capable of representing many real time series is the class of autoregressive integrated moving average (ARIMA) models.

ABSTRACT

In recent years there has been considerable interest in building models which preserve the autocorrelation structure of hydrologic sequences. In particular, Markov models, are frequently entertained to describe the time dependence of run-off sequences.

In this paper we follow a more general approach. Instead of restricting ourselves to Markov models, we consider the class of autoregressive integrated moving average (ARIMA) models. This broad class of models is capable of representing many time series observed in practice.

Since the distribution of the run-off sequences is frequently skewed, one has to transform the data. In this paper we give some thought to the questions of which transformation one should choose. The class of power transformations is discussed in detail.

ARIMA Models and Their Use in Modelling Hydrologic Sequences

1. Introduction

The design of water resource systems demands frequently the generation of synthetic sequences of various water quality and run-off sequences. Before being able to generate synthetic sequences and/or forecast future values models have to be found which describe past data adequately. Ideally, these models should preserve all the properties of the observed data. In practice, however, this cannot be achieved and most advisable criteria for evaluating the statistical resemblance between historic and generated hydrologic data have to be chosen. They should depend on the uses to which the generated sequences are to be put and the loss that accumulates from reaching the wrong decision. General applicable criteria include the preservation of

i) the correlation structure

ii) the distribution of the observations, in particular

   a) extreme values

   b) sums and ranges

iii) the Hurst coefficient

In recent years there has been considerable interest in building models which preserve the autocorrelation structure of the observations. Systematic study of the autocorrelation function (or spectral distribution function) has led to the specification of stochastic models which can be used for prediction and generation of hydrologic sequences ([7,8,11,14]). In particular, autoregressive (Markovian) models are frequently used to describe the time dependence (persistance) of hydrologic run-off sequences.

However, not only autoregressive models can be used to describe stochastic processes. In this paper, we will review and discuss the class of ARIMA (autoregressive integrated moving average) models, the autoregressive model being a special case of it. A three stage iterative modelling procedure, due to Box and Jenkins [6], is discussed showing how models can be specified, their parameters estimated and their validity checked.

Models in hydrology are frequently applied to operate on the logarithms of the record rather than working with the historical untransformed sequence. The logarithmic transformation is usually chosen to transform the skewed run-off distribution and achieve Normality. It has to be pointed out that:

1) investigations should be made whether the logarithmic transformation is appropriate;
2) a choice has to be made with respect to the increment that should be added to the historical sequence to avoid infinite logarithms for times of zero flows. This can give rise to serious problems since it is easily shown that, for example, the variance of the logarithmic series may be drastically changed depending on the increment added to the original series. Thus a great part of the appropriateness of the model can depend on the increment added.

In this paper we discuss methods which enable us to determine whether the logarithmic transformation is feasible or whether one should consider different transformations. Furthermore, these methods provide information on the choice of the increment to be

added. We discuss the class of power transformations and their application to models in hydrology. This class was first introduced into the statistical literature by Box and Cox [4], and has been applied in various other fields such as economics and business ([16,18]). It is reviewed how this class of transformations can be applied in the context of ARIMA models.

To illustrate the Box-Jenkins iterative model building procedure and the class of power transformations suggested by Box and Cox we consider an example using 40 years of monthly stream flow data.

2. The general linear process and its parsimonious versions. The class of ARIMA models:

In the early part of the 20th century, time series studies were based on models consisting of a non-stochastic trend term with added independent shocks (Schuster [15]). Early criticism, however, prompted the search for models which could better describe the time dependence of the data. This led to stochastic models based on Yule's idea [22], that time series in which successive values are highly dependent can be regarded as a linear aggregate of independent shocks a_t. Shocks or innovations are independent drawings from some fixed distribution, with mean zero and constant variance σ_a^2. A sequence of such independent random drawings from such a distribution is called white noise sequence.

Wold [19] proved the fact that any weakly stationary process z_t (i.e. a process with time invariant first and second order moments) can be uniquely represented as the sum of two process:

$$z_t = v_t + u_t \quad , \tag{2.1}$$

where i) the process u_t is uncorrelated with the process v_t

ii) the process v_t is deterministic

iii) u_t has a one sided moving average representation

$$u_t = \sum_{k=0}^{\infty} \psi_k a_{t-k} \quad \text{with } \psi_0 = 1 \quad \text{and} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty \quad .$$

Such a decomposition of stationary stochastic processes in terms of uncorrelated random variables is important since it provides a natural setting for predicting future values of the process from values obtained in the past.

One other extensively used decomposition of stationary stochastic processes is the spectral decomposition. This is an integral expansion with complex coefficients. There are several studies using spectral analysis to investigate the statistical structure of observed hydrological records such as stream flow and precipitation ([2,12,13]). It is easily shown that the autocorrelation function (time domain) contains the same information as the spectral density function (frequency domain).

Wold's decomposition states that any non-deterministic weakly stationary stochastic process can be represented as a linear aggregate of random shocks. The general linear process is a linear transformation of white noise,

$$z_t = \psi(B) a_t \quad , \tag{2.2}$$

where i) z_t is the difference between the original observations and some deterministic component v_t (e.g. deviations from the mean).

ii) $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$; $\psi_0 = 1$ and

B is the backshift operator $B^m z_t = z_{t-m}$.

iii) $\{a_t\}$ is a white noise sequence with

$$Ea_t = 0$$

$$Ea_t a_{t-k} = \begin{cases} 0 & \text{for } k \neq 0 \\ \sigma_a^2 & \text{for } k = 0 \end{cases}$$

Such a linear process is said to be stationary if $\psi(B)$, the generating function of the ψ-weights, converges for $|B| \leq 1$. This condition ensures that the ψ weights form a convergent series, the variance of the process is finite and the matrix of autocovariances is positive definite.

An equivalent representation of the stochastic process (2.2) is given by

$$\pi(B) z_t = a_t \quad , \tag{2.3}$$

where

$$\pi(B) = \psi^{-1}(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j \quad .$$

The linear process is said to be invertible if $\pi(B)$, the generating function for the π-weights, converges for $|B| \leq 1$. The invertibility condition is independent of the stationarity condition and is only necessary if one is interested in a particular direction of time. As far as defining a stationary stochastic process is concerned, it is easily shown that the representation (2.2) is not unique. For example, the two representations

$$z_t = (1 + \psi_1 B)a_t \quad ,$$

$$z_t = (1 + \psi_1 F)a_t = (1 + \psi_1^{-1} B)e_t \quad ,$$

where the forward shift operator $F = B^{-1}$ and $e_t = -\theta a_{t+1}$, define the same process (correlation function). The first representation is directed into the future and for $|\psi_1| < 1$ future values are derived as a convergent weighted sum of past observations. The second representation is directed into the past, and for $|\psi_1| < 1$ future values will be a linear combination of past values, but with divergent weights.

Parsimonious versions

In general, representations (2.2) and (2.3) could contain an infinite number of parameters ψ_j and π_j and they would not be useful if only a finite record of data is available to estimate the parameters.

The consideration of a ratio of two polynomials in B,

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)} \quad \text{where:}$$

$$\theta_q(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$$

and

$$\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \quad .$$

allows great flexibility with only few parameters.

(i) $\phi_p(B)z_t = \theta_q(B)a_t$ (2.3)

is called autoregressive moving average model of order p and q; ARMA (p,q).

(ii) $\phi_p(B)z_t = a_t$ (2.4)

is called autoregressive process of order p; AR(p).

(iii) $z_t = \theta_q(B)a_t$ (2.5)

is called moving average process of order q; MA(q).

It can be shown that the ARMA (p,q) process is stationary if the characteristic equation $\phi_p(B) = 0$ has all its roots outside the unit circle. It is invertible if the roots of $\theta_q(B) = 0$ lie outside the unit circle.

Yaglom ([20,21]), Box and Jenkins [6] extend this class of ARMA models to cover nonstationary series which in particular do not vary about a fixed level and/or trend and/or periodicity, but apart from this one part of the series behaves very much like the other. Stochastic processes which exhibit these characteristics have some roots of $\phi(B) = 0$ lying on the unit circle implying that the autoregressive operator will contain factors of the form $(1 - B)$, $(1 - B)^2$, $(1 - \sqrt{3}\,B + B^2)$, $(1 - B^{12})$.... Box and Jenkins refer to such as simplifying operators. Particular examples of the simplifying operators are the ordinary differences. The introduction of the operator $(1 - B)^d$ allows for nonstationarity of the original series and up to the $(d-1)^{st}$ difference, while all higher differences are stationary. For d = 1, the model allows for nonstationarity in terms of having no fixed level. For d = 2 it allows for nonstationarity in both level and slope. Processes of the form

$$\phi_p(B)(1 - B)^d z_t = \varphi_{p+d}(B)z_t = \theta_q(B)a_t \quad , \tag{2.6}$$

are called autoregressive integrated moving average processes; ARIMA (p,d,q).

Forecasts for this class of models are easily derived using general prediction theory as discussed by Kolmogorov [9], Wiener [17], and Wold [19]. The minimum mean square error forecast of $z_{n+\ell}$, denoted by $\hat{z}_n(\ell)$, is given by

$$\hat{z}_n(\ell) = \sum_{j=0}^{\infty} \psi_{\ell+j} a_{n-j} = [z_{n+\ell}] \quad , \tag{2.7}$$

where $[z_{n+\ell}]$ is the conditional expectation of $z_{n+\ell}$ given knowledge of all the observations z_t up to time n. Forecasts can be expressed in different equivalent forms:

(i) forecasts from the difference equation

$$\hat{z}_n(\ell) = \varphi_1[z_{n+\ell-1}] + \ldots + \varphi_{p+d}[z_{n+\ell-p-d}] \tag{2.8}$$

$$+ [a_{n+\ell}] - \theta_1[a_{n+\ell-1}] - \ldots - \theta_q[a_{n+\ell-q}] \quad ,$$

where

$$[z_{n-j}] = \begin{cases} E(z_{n-j} \mid z_n, z_{n-1}, \ldots) & \text{for } j < 0 \\ z_{n-j} & \text{for } j \geq 0 \end{cases}$$

$$[a_{n-j}] = \begin{cases} 0 & \text{for } j < 0 \\ a_{n-j} & \text{for } j \geq 0 \end{cases}$$

Forecasts are updated easily from one observation to the other:

$$\hat{z}_{n+1}(\ell) = \hat{z}_n(\ell+1) + \psi_\ell a_{n+1} \quad ,$$

where a_{n+1} is one step ahead forecast error;

$$a_{n+1} = z_{n+1} - \hat{z}_n(1) \quad .$$

(ii) Another way to represent minimum mean square error forecasts is over the eventual forecast function. The eventual forecast function is the solution of the difference equation

$$\varphi_{p+d}(B)\hat{z}_n(\ell) = 0 \qquad \text{for } \ell > q$$

which is given by

$$\hat{z}_n(\ell) = b_1(n)f_1(\ell) + \ldots + b_{p+d}(n)f_{p+d}(\ell) \qquad \text{for } \ell > q-p-d \qquad (2.9)$$

$f_1(\ell),\ldots,f_{p+d}(\ell)$ are functions of the lead time ℓ and depend on the autoregressive part of the model only. In general these functions can be polynomials, exponentials, sines, cosines and combinations of these functions.

For a given forecast origin t the coefficients $\underset{\sim}{b}(n) = [b_1(n),\ldots b_{p+d}(n)]'$ are constants and are the same for all lead times ℓ. They change, however, from one forecast origin to the next and shown by Box and Jenkins [6] are updated by

$$\underset{\sim}{b}(n) = L\underset{\sim}{b}(n-1) + \underset{\sim}{g}[z_n - \hat{z}_{n-1}(1)] \qquad (2.10)$$

where

$$L = F_\ell^{-1}F_{\ell+1}$$

$$F_\ell = \begin{bmatrix} f_1(\ell) & \ldots & f_{p+d}(\ell) \\ \vdots & & \vdots \\ f_1(\ell+p+d-1) & \ldots & f_{p+d}(\ell+p+d-1) \end{bmatrix}$$

and

$$\underset{\sim}{g} = F_\ell^{-1}\psi_\ell \qquad \text{with} \quad \underset{\sim}{\psi}_\ell = [\psi_\ell,\psi_{\ell+1},\ldots,\psi_{\ell+p+d-1}]'$$

for any $\ell > q - p - d$.

(iii) Forecasts as weighted average of previous observations

$$\hat{z}_n(\ell) = [z_{n+\ell}] = \sum_{j=1}^{\infty} \pi_j [z_{n+\ell-j}] + [a_{n+\ell}] = \sum_{j=1}^{\infty} \pi_j^{(\ell)} z_{n+1-j} \qquad (2.11)$$

where

$$\pi_j^{(\ell)} = \pi_{j+\ell-1} + \sum_{k=1}^{\ell-1} \pi_k \pi_j^{(\ell-k)} \quad .$$

Seasonal Models

The fundamental fact about seasonal time series with period s is that observations which are s intervals apart are similar. Therefore one might expect that the operation $B^s z_t = z_{t-s}$ would play an important role in the analysis of seasonal series. Furthermore, if nonstationarity is to be expected in the series z_t, the simplifying operation $(1 - B^s)z_t = z_t - z_{t-s}$ might be useful. This nonstationary simplifying operator has s zeros $e^{\frac{i2\pi k}{s}}$ (k=0,1,...,s-1) evenly spaced on the unit circle.

For monthly observations with a yearly seasonal pattern, a complete set of sinusoids would be generated by the operator $(1 - B^{12})$. This operator can be factored into operators corresponding to different components of sines and cosines.

Factor	Root	Period	Frequency in Cycles per Year
$1 - B$	1	Constant	
$1 - \sqrt{3}B + B^2$	$\frac{1}{2}(\sqrt{3} \pm i)$	12	1
$1 - B + B^2$	$\frac{1}{2}(1 \pm i)$	6	2
$1 + B^2$	$\pm i$	4	3
$1 + B + B^2$	$\frac{1}{2}(-1 \pm i)$	3	4
$1 + \sqrt{3}B + B^2$	$\frac{1}{2}(-\sqrt{3} \pm i)$	12/5	5
$1 + B$	-1	2	6

When we analyze a series which exhibits seasonal behaviour with known periodicity (let's say monthly run-off data with yearly seasonal pattern) one realizes that in periodic data there are not one, but two time intervals of importance. One expects relationships to occur

(i) between observations for the same month in successive years;

(ii) between observations for successive months in the same year.

Since in seasonal time series the observation z_t, let's say for the month of August, is related to previous August values, we might be able to link these observations by

$$\Phi(B^{12})(1-B^{12})^D z_t = \Theta(B^{12})\alpha_t \qquad (2.12)$$

where

$$\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \ldots - \Phi_P B^{12P}$$

$$\Theta(B^{12}) = 1 - \Theta_1 B^{12} - \ldots - \Theta_Q B^{12Q} \quad .$$

Usually it will be reasonable to assume that the parameters Φ and Θ contained in these models are approximately the same for each month. If this assumption is too restrictive, additional parameters have to be introduced.

The error components $\alpha_t, \alpha_{t-1}, \ldots$ in these models, however, would in general not be uncorrelated. For example, the observation for August would be related to the observation of July of the same year. To take care of such relationships, we introduce the second model

$$\phi(B)(1-B)^{d}\alpha_t = \theta(B)a_t \quad , \tag{2.13}$$

where a_t is a white noise process and

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \quad ,$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \quad .$$

Combining the models in (2.12) and (2.13) we obtain the multiplicative model as it is discussed in Box and Jenkins ([5,6])

$$\phi(B)\Phi(B^{12})(1-B)^{d}(1-B^{12})^{D}z_t = \theta(B)\Theta(B^{12})a_t \quad . \tag{2.14}$$

The resulting multiplicative model is said to be of order $(p,d,q) \times (P,D,Q)_{12}$.

To get better understanding of these seasonal stochastic difference equation models it is best to consider some examples.

Model I: The multiplicative $(0,1,1) \times (0,1,1)_{12}$ model

$$(1-B)(1-B^{12})z_t = (1-\theta B)(1-\Theta B^{12})a_t \quad . \tag{2.15}$$

The eventual forecast function satisfies the difference equation

$$(1-B)(1-B^{12})\hat{z}_n(\ell) = 0 \qquad \text{for } \ell > 13 \quad .$$

Writing $\ell = (r,m)$ $r = 0,1\ldots$; $m = 1,2,\ldots,12$ to represent a lead time of r years and m months it can be readily shown that the solution of (2.15) is given by

$$\hat{z}_n(\ell) = \hat{z}_n(r,m) = b_{0,m}(n) + b_1(n)r \quad . \tag{2.16}$$

The forecast function contains 13 adjustable coefficients $b_{0,1}(n),\ldots,b_{0,12}(n),b_1(n)$ representing 12 monthly and one yearly contribution, determined by the first 13 forecasts.

The ψ-weights of model (2.15) are derived by equating

$$(1 - \theta B)(1 - \Theta B^{12}) = \psi(B)(1 - B)(1 - B^{12}),$$

and are given by

$$\begin{aligned}
&\psi_1 = \psi_2 = \ldots = \psi_{11} = \lambda && \psi_{12} = \lambda + \Lambda \\
&\psi_{13} = \psi_{14} = \ldots = \psi_{23} = \lambda(1+\Lambda) && \psi_{24} = \lambda(1+\Lambda) + \Lambda \\
&\psi_{25} = \psi_{26} = \ldots = \psi_{35} = \lambda(1+2\Lambda) && \psi_{36} = \lambda(1+2\Lambda) + \Lambda \quad .
\end{aligned}$$

$$\ldots\ldots$$

Representing ψ_j as $\psi_{r,m}$ where r refers to years and m to months

$$\psi_j = \psi_{r,m} = \lambda(1+r\Lambda) + \delta\Lambda \quad , \tag{2.17}$$

where

$$\delta = \begin{cases} 1 & \text{for } m = 12 \\ 0 & \text{for } m \neq 12 \end{cases}$$

and

$$\lambda = 1 - \theta \quad ; \qquad \Lambda = 1 - \Theta \quad .$$

The π-weights are derived by equating coefficients in

$$(1 - \theta B)(1 - \Theta B^{12})\pi(B) = (1 - B)(1 - B^{12}) \quad ,$$

and are given by

$$\begin{aligned}
\pi_j &= (1-\theta)\theta^j \quad , && j = 1,2,\ldots,11 \\
\pi_{12} &= (1-\theta)\theta^{11} + (1-\Theta) \\
\pi_{13} &= (1-\theta)\theta^{12} - (1-\theta)(1-\Theta) \\
\pi_j &= \theta\pi_{j-1} + \Theta\pi_{j-12} - \theta\Theta\pi_{j-13} \quad , && j \geq 14
\end{aligned} \tag{2.18}$$

The one step ahead forecast for model I is an exponentially weighted moving average taken over previous months modified by

a second exponentially weighted moving average of discrepancies found between similar monthly exponentially weighted moving averages and actual performaces in previous years.

This can be seen from (2.15), since the model can be written as

$$\left[1 - \frac{\lambda B}{1 - \theta B}\right]\left[1 - \frac{\Lambda B^{12}}{1 - \Theta B^{12}}\right] z_{n+1} = a_{n+1} \quad .$$

Thus

$$\hat{z}_n(1) = \frac{\lambda}{1 - \theta B} z_n + \frac{\Lambda}{1 - \Theta B^{12}} \left[z_{n-11} - \frac{\lambda}{1 - \theta B} z_{n-12}\right] \quad . \tag{2.19}$$

The one step ahead prediction for August, let's say, is an exponentially weighted average of previous months up to July. However, it is also corrected by a second exponentially weighted average over previous discrepancies between actual August values and the corresponding monthly moving averages taken over previous months in those years.

The adaptive nature of the forecast function can be seen by considering the updating relation

$$\hat{z}_{n+1}(\ell) = \hat{z}_n(\ell+1) + \psi_\ell a_{n+1} \quad .$$

If $m \neq 12$

$$b_{0,m}(n+1) + r b_1(n+1) = b_{0,m+1}(n) + r b_1(n) + \lambda(1+r\Lambda) a_{n+1} \quad .$$

Equating coefficients in r

$$\begin{cases} b_{0,m}(n+1) = b_{0,m+1}(n) + \lambda a_{n+1} \quad , \\ b_1(n+1) = b_1(n) + \lambda\Lambda a_{n+1} \quad . \end{cases} \tag{2.20}$$

Similarily if m = 12

$$\begin{cases} b_{0,12}(n+1) = b_{0,1}(n) + b_1(n) + (\lambda+\Lambda)a_{n+1} \ , \\ b_1(n+1) = b_1(n) + \lambda\Lambda a_{n+1} \ . \end{cases} \tag{2.21}$$

Model II: The multiplicative $(1,0,0) \times (0,1,1)_{12}$ model.

$$(1-\phi B)(1-B^{12})z_t = (1-\Theta B^{12})a_t \ . \tag{2.22}$$

The eventual forecast function for (2.22) is the solution of $(1-\phi B)(1-B^{12})\hat{z}_n(\ell) = 0$ for $\ell > 12$ and is given by

$$\hat{z}_n(r,m) = b_{0,m}(n) + b_1(n)\phi^m(1-\phi^{12r}) \tag{2.23}$$

Proof:

$$(1-\phi B)(1-B^{12})\hat{z}_n(r,m) = 0 \qquad (r,m) \geq (1,1) \ .$$

Defining

$$u_n(r,m) = (1-B^{12})\hat{z}_n(r,m)$$

we see that

$$(1-\phi B)u_n(r,m) = 0$$

or

$$u_n(r,m) = \phi u_n(r,m-1) = \ldots = \phi^{12(r-1)+m}\, u_n(0,12) \ .$$

Furthermore

$$\begin{cases} u_n(r,m) = \hat{z}_n(r,m) - \hat{z}_n(r-1,m) = \phi^{12(r-1)+m}\, u_n(0,12) \\ u_n(r-1,m) = \hat{z}_n(r-1,m) - \hat{z}_n(r-2,m) = \phi^{12(r-2)+m}\, u_n(0,12) \\ \vdots \\ u_n(1,m) = \hat{z}_n(1,m) - \hat{z}_n(0,m) = \phi^{m}\, u_n(0,12) \ . \end{cases}$$

Adding the r equations we get

$$\hat{z}_n(r,m) - \hat{z}_n(0,m) = \phi^m \sum_{k=0}^{r-1} \phi^{12k} u_n(0,12)$$

or

$$\hat{z}_n(r,m) = \hat{z}_n(0,m) + \phi^m \frac{1 - \phi^{12r}}{1 - \phi^{12}} u_n(0,12) \quad .$$

Interpreting $\hat{z}_n(0,m)$ as $b_{0,m}(n)$ and $\frac{u_n(0,12)}{1 - \phi^{12}}$ as $b_1(n)$ the result (2.23) follows.

The forecast function contains 13 adjustable coefficients $b_{0,}(n),\ldots,b_{0,12}(n),b_1(n)$. These coefficients are determined by z_n and the first 12 forecasts and are updated as new observations become available. The adjustment of the coefficients depends on the ψ-weights of the model.

Adaptive stochastic difference equation versus models with time as extraneous variable.

Models with time as extraneous variable

Time dependent observations are often modelled as

$$z_t = f(t) + a_t \tag{2.24}$$

where f(t) is a deterministic function of time and a_t are independent shocks. For example,

$$z_t = \beta_0 + a_t \tag{2.25a}$$

$$z_t = z_{r,m} = \beta_{0,m} + a_t \tag{2.25b}$$

$$z_t = z_{r,m} = \beta_{0,m} + \beta_1 r + a_t \quad . \tag{2.25c}$$

Model (2.25a) specifies a fixed mean, model (2.25b) a monthly seasonal pattern and model (2.25c) a monthly seasonal pattern

and yearly trend. These models suppose that the seasonal and trend components are fixed and that the same projection will apply at all times. The models (2.25) can be shown inadequate for two reasons:

i) the noise is modelled inadequately, since it generally does **not** account for correlation among the errors

ii) the models assume a functional form which remains constant over time. A great deal of flexibility might be added to the predictive capacity of such models if the β coefficients were adaptive and could change as the time series evolves.

Both these deficiencies can be corrected by defining the models in terms of stochastic difference equations (ARIMA models).

Adaptive stochastic difference equation models.

It was pointed out before that from the time of Yule (1927) it has been realized that models in which the noise may be correlated and in which the basic parameters are functions of time can be achieved by finding the model not in explicit terms as in (2.24), but in terms of stochastic difference equations. It was pointed out, for example, in (2.9), (2.16) and (2.23) that stochastic difference equations lead to eventual forecast functions of the form (2.24). The basic parameters of the forecast function, however, are adaptive.

The difference between adaptive difference equation models and models with time as extraneous variable is best brought out by considering the following simple example:

Example:

The first order integrated model; ARIMA (0,1,1)

$$(1-B)z_t = (1-\theta B)a_t \quad . \tag{2.26}$$

It is easily shown that

$$\hat{z}_n(\ell) = (1-\theta) \sum_{j=0} \theta^j z_{n-j} = b_0(n) \tag{2.27}$$

and

$$b_0(n+1) = b_0(n) + (1-\theta)a_{n+1} \quad . \tag{2.28}$$

(2.28) shows that the forecast is adaptive in the sense that the coefficient $b_0(n)$ is updated with each new observation. However, if one would forecast with the model (2.25a) $z_t = \beta_0 + a_t$, we get $\hat{z}_n(\ell) = \beta_0$. (2.25a) implies that the added data leaves the basic parameter unchanged.

. . . . .

Fears have sometimes been expressed that the stochastic difference equation (ARIMA) approach may lead to unnecessarily complicated models when models of simpler type (e.g. deterministic component and additive independent errors) are adequate. One additional advantage of the stochastic difference equation models besides allowing for general noise structure and adaptive basic parameters in the forecast function, however, is that they encapsulate within themselves the traditional models like (2.24) (for detailed discussions of this point see Abraham [1]). For example, this can be seen in (2.26)-(2.28) by letting $\theta \to 1$. In this case $b_0(n+1) = b_0(n)$ and the model (2.28) coincides with (2.25a).

Another example which illustrates this fact is given below. It will be shown in section 5 that the $(1,0,0) \times (0,1,1)_{12}$ model provides a suitable stochastic difference equation for representing monthly run-off sequences.

$$(1-\phi B)(1-B^{12})z_t = (1-\Theta B^{12})a_t \quad . \tag{2.29}$$

If $\Theta \to 1$, it can be seen that the model in (2.29) approaches

$$(1-\phi B)(z_t - f(t)) = a_t \tag{2.30}$$

where $f(t)$ satisfies $(1-B^{12})f(t)$ and thus, in general, can be represented by the 12 monthly means. If $\Theta \to 1$, (2.29) is equivalent to a first order autoregressive process of the observations corrected by their monthly means.

3. The philosophy of iterative model building

If the form of stochastic model is known and the values of the parameters are given, synthetic sequences can be generated. In practice, however, the structure of the process is rarely, if ever, known and one has to use past observations to derive adequate models and estimate their parameters.

Box and Jenkins [6] develop a three stage iterative procedure "Identification-Fitting-Diagnostic Checking" to find adequate model(s) describing the correlation structure of past data. The iterative procedure is shown in the diagram.

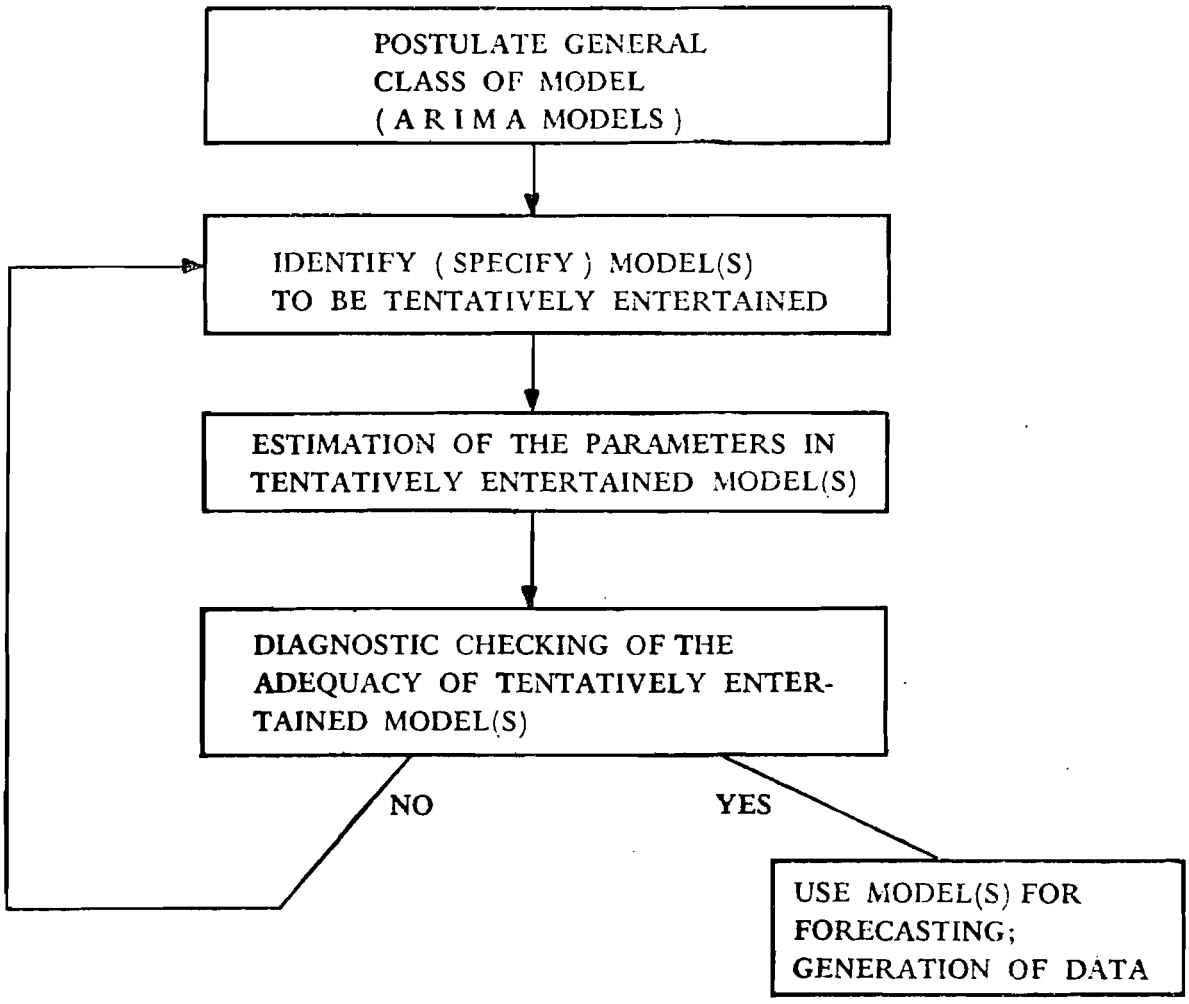

DIAGRAM: THREE STAGE ITERATIVE MODELLING PROCEDURE: "IDENTIFICATION - ESTIMATION - DIAGNOSTIC CHECKING."

Model Identification

At the first stage of the iterative approach to model building, the identification (specification) stage, one uses the data and the knowledge of the system to suggest an appropriate parsimonious subclass of models which may be tentatively entertained. At this stage one has to decide about the particular metric to present the data in, whether the data are modelled in their original form or whether transformations have to be considered. Quite frequently it is useful modelling the random component in a different metric achieving a simpler structure, constancy of error variances and normality of distribution. It will be discussed in the fourth part of this paper how transformations can be parameterized and the value of the parameters estimated from observed data.

At this stage, the basic tools for modelling the time dependence of the data are plots of the data and sample autocorrelation and partial autocorrelation function. The properties of theoretical autocorrelation and partial autocorrelation functions provide a convenient reference table for the model(s) to be chosen. It must be pointed out that at this stage statistically inefficient methods must necessarily be used since no precise formulation of the problem is yet available.

Model Estimation

At this stage efficient use of the data is made by making inferences about the parameters conditionally on the adequacy of the entertained model. For the case of Normally distributed shocks, the likelihood function of ARIMA models can be obtained

and the limiting properties of these estimates derived. For the case of Normally distributed errors the maximum likelihood estimators can be approximated by the least squares estimates and nonlinear regression routines can be utilized to derive the estimates. For the case of other error distributions, however, the maximum likelihood estimates can differ from the usual least squares estimates. Some discussion of this point for the class of symmetric exponential power distributions is given in Ledolter [10]. Since it can become very laborious to derive maximum likelihood estimates in the case of other non-Normal distributions it appears to be a better approach to first transform the data to achieve approximately Normal error distributions.

Diagnostic Checking

After fitting the tentatively entertained model(s) to the observed data one has to check the fitted model in its relation to the observed data with intent to reveal model inadequacies and to achieve improvement. The residuals from the fitted model contain all the information about the adequacy of the fitted model. Inspection of sample autocorrelation function of the residuals indicate whether the entertained model is adequate to describe the correlation structure for the data set under study, or if and how the model should be revised. Furthermore, a plot of the histogram of the residuals shows whether the Normality assumption of the errors is justified.

Model identification, estimation and diagnostic checking are important in the iterative model building procedure. After diagnostic checks satisfy the critic as of the adequacy of the model he originally sponsored, the derived model can be used for forecasting and generating purposes.

4. The class of power transformation applied to time series models

Box and Cox [4] consider the following family of power transformations:

(i) the one parameter family:

$$z_t^{(\lambda)} = \begin{cases} \dfrac{z_t^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \log z_t & (\lambda = 0) \end{cases} \qquad (4.1)$$

(ii) the two parameter family:

$$z_t^{(\lambda_1,\lambda_2)} = \begin{cases} \dfrac{(z_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & (\lambda_1 \neq 0) \\ \log(z_t + \lambda_2) & (\lambda_1 = 0) \end{cases} \qquad (4.2)$$

The transformation in (i) is valid for $z_t > 0$; the transformation in (ii) for $z_t > -\lambda_2$. This class of transformations has proved useful in transforming the data to achieve a simple structure of the model as well as constant error variances and Normality of the error distribution.

Suppose we observe a time series of n observations $\underset{\sim}{z}' = (z_1, z_2, \ldots, z_n)$ from a ARIMA model

$$\varphi(B)\, z_{t\sim}^{(\lambda)} = \theta(B)\, a_t \quad , \qquad (4.3)$$

where $\underset{\sim}{\lambda}' = (\lambda_1, \lambda_2)$ as in (4.2) above and a_t is a Normal white noise sequence. We thus assume that for some unknown $\underset{\sim}{\lambda}$ the transformed observations can be represented as ARIMA model satisfying the usual Normal assumptions of the shocks a_t.

Conditional on a set of starting values $\underset{\sim}{z}_0^{(\lambda)}$ and $\underset{\sim}{a}_0$ which can be set equal to their conditional expectation (which is zero) the joint probability function of the transformed data is given by

$$p(\underset{\sim}{z}^{(\lambda)} | \underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda}, \sigma^2) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} a_t^2(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda})\} \quad , \qquad (4.4)$$

where

$$\underset{\sim}{z}^{(\lambda)\prime} = (\underset{\sim}{z}_1^{(\lambda)}, \underset{\sim}{z}_2^{(\lambda)}, \ldots, \underset{\sim}{z}_n^{(\lambda)})$$

$$\underset{\sim}{\phi}' = (\phi_1, \phi_2, \ldots, \phi_p)$$

$$\underset{\sim}{\theta}' = (\theta_1, \theta_2, \ldots, \theta_q)$$

$$a_t(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda}) = \theta^{-1}(B)\varphi(B)\underset{\sim}{z}_t^{(\lambda)} \quad .$$

In terms of the original observations $\underset{\sim}{z}' = (z_1, \ldots, z_n)$ the likelihood of $(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda}, \sigma^2)$ is given by

$$\ell(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda}, \sigma^2 | \underset{\sim}{z}) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} a_t^2(\underset{\sim}{\phi}, \underset{\sim}{\theta}, \underset{\sim}{\lambda})\} \times J(\underset{\sim}{\lambda}; \underset{\sim}{z}) \quad , \qquad (4.5)$$

where the Jacobian of the transformation is

$$J(\underset{\sim}{\lambda}; \underset{\sim}{z}) = \prod_{i=1}^{n} \left| \frac{dz_t^{(\lambda)}}{dz_t} \right| = \prod_{t=1}^{n} (z_t + \lambda_2)^{\lambda_1 - 1} \quad .$$

We find the maximum likelihood estimates in two steps. For given $\underset{\sim}{\lambda}$ the likelihood in (4.5) is, except for the constant $J(\underset{\sim}{\lambda}; \underset{\sim}{z})$, the likelihood of the ARIMA model in $\underset{\sim}{z}_t^{(\lambda)}$. The maximum likelihood estimates of $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}$ can be derived using a non-linear regression analysis program. For fixed $\underset{\sim}{\lambda}$, the estimate of σ^2, $\sigma^2(\underset{\sim}{\lambda})$, is given by

$$\hat{\sigma}^2(\underset{\sim}{\lambda}) = \frac{SS(\underset{\sim}{\lambda})}{n} = \frac{\sum_{t=1}^{n} a_t^2(\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}}, \underset{\sim}{\lambda})}{n} \quad ,$$

and the maximized log likelihood is, except for a constant, given by

$$L_{max}(\underset{\sim}{\lambda}) \propto -\frac{1}{2} n \hat{\sigma}^2(\underset{\sim}{\lambda}) + \log J(\underset{\sim}{\lambda};\underset{\sim}{z}) \quad . \tag{4.6}$$

The result simplifies considerably if one works with normalized variables

$$\dot{z}_t^{(\underset{\sim}{\lambda})} = z_t^{(\underset{\sim}{\lambda})} [J(\underset{\sim}{\lambda},\underset{\sim}{z})]^{-\frac{1}{n}} \quad . \tag{4.7}$$

In the case of the one parameter transformation $\dot{z}_t^{(\lambda)}$ is given by

$$\dot{z}_t^{(\lambda)} = \frac{z_t^{\lambda} - 1}{\lambda_1 \{gm(z)\}^{\lambda-1}} \quad , \tag{4.8}$$

and in the case of the two parameter transformation

$$\dot{z}_t^{(\underset{\sim}{\lambda})} = \frac{(z_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1 \{gm(z + \lambda_2)\}^{\lambda_1 - 1}} \tag{4.9}$$

where $gm(\cdot)$ denotes the sample geometric mean. It is easily seen that (4.6) simplifies to

$$L_{max}(\underset{\sim}{\lambda}) = -\frac{n}{2} \log \hat{\dot{\sigma}}^2(\underset{\sim}{\lambda}) \quad , \tag{4.10}$$

where

$$\hat{\dot{\sigma}}^2(\underset{\sim}{\lambda}) = \frac{\sum_{t=1}^{n} \dot{a}_t^2(\hat{\underset{\sim}{\phi}},\hat{\underset{\sim}{\theta}},\underset{\sim}{\lambda})}{n}$$

and

$$\dot{a}_t = \varphi^{-1}(B)\theta(B)\dot{z}_t^{(\underset{\sim}{\lambda})}$$

and $\hat{\underset{\sim}{\phi}}$ and $\hat{\underset{\sim}{\theta}}$ are the maximum likelihood estimates for the ARIMA model in $\dot{z}_t^{(\underset{\sim}{\lambda})}$.

A plot of the maximized likelihood function $L_{max}(\lambda)$ for a trial series of λ values (in the one parameter transformation) will be informative. For the two parameter family contours of $L_{max}(\underset{\sim}{\lambda})$ can be derived combining the points $\underset{\sim}{\lambda}' = (\lambda_1, \lambda_2)$ of equal likelihood.

From these plots the maximizing value $\hat{\lambda}$ (in the one parameter transformaiton) or $\underset{\sim}{\hat{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2)$ (in the two parameter transformation) may be read off. Furthermore, approximate $100(1-\alpha)$ percent confidence regions can be obtained from

$$L_{max}(\underset{\sim}{\hat{\lambda}}) - L_{max}(\underset{\sim}{\lambda}) \leq \frac{1}{2}\chi_i^2(\alpha) \qquad (4.11)$$

where $\chi_i^2(\alpha)$ is the α per cent cut off point of a Chi-square distribution with degrees of freedom given by $i = 1(2)$ for the one (two) parameter transformation.

Equivalently one can also use a Bayesian approach to derive the posterior distribution of $\underset{\sim}{\lambda}$. Assuming a non informative a priori distribution which is flat over the region of appreciable likelihood will result in point estimates close to the maximum likelihood estimates. The Bayesian approach, however, can also be used to incorporate a priori information into the analysis.

5. An example:

To illustrate the Box-Jenkins approach to model building and the use of the family of power transformations, we consider a record of 40 years of monthly run-off data collected at a Carpathian river near the border of Czechoslovakia and Poland. The data together with their monthly means and standard deviations are given in Table 1; a plot of the data is given in Fig. 1.

Inspection of the data immediately reveals that:

i) the observations show a strong yearly seasonal pattern and thus seasonal models are called for.

ii) higher measures of variation (monthly standard deviations) are observed at higher levels of run-off, thus calling for a transformation to transform the data to a Normal distribution with constant variance.

Seasonal ARIMA model

Identification.

To tentatively identify the form of the model, we use the estimated autocorrelation and partial autocorrelation function. We first consider the logarithm of the original data, since it appears that the monthly means are roughly proportional to their standard deviations. Later in the estimation stage, however, we will parameterize the transformation and estimate the transformation simultaneously with the parameters in the ARIMA model.

The estimated autocorrelation function (ACF) for $\log z_t$ and $(1-B^{12})\log z_t$ are given in Fig. 2. It can be seen that the ACF of $\log z_t$ at lag 12,24,36,... decays almost linearly and fails to die out quickly. This is an indication of nonstationarity and leads to consider a model which includes the seasonal simplifying operator $(1-B^{12})$. The ACF for $(1-B^{12})\log z_t$, given in Fig. 2, shows an exponential decay in the non-seasonal part and significant autocorrelations around lag 12. In the Appendix, it is shown that a model which generates such a ACF can be given by

$$(1-\phi B)(1-B^{12})\log z_t = (1-\Theta B^{12})a_t \quad .$$

The partial ACF strengthens this tentative conclusion.

Estimation.

After tentatively entertaining the above model, we have to estimate its parameters. Since there are no months of zero flow we can use the family of one parameter power transformations

$$z_t^{(\lambda)} = \begin{cases} \dfrac{z_t^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log z_t & \lambda = 0 \end{cases} .$$

Estimates for the parameters in

$$(1-\phi B)(1-B^{12})z_t^{(\lambda)} = (1-\theta B^{12})a_t \qquad (5.1)$$

have to be derived. We use the estimation approach as outlined in section 4 and use a non-linear regression subroutine to derive approximate maximum likelihood estimates (conditional on starting values).

The plot of $L_{max}(\lambda)$ as given in (4.10) for a series of λ values can be found in Fig. 3. The maximum is reached at

$$\hat{\lambda} = -.17$$
$$\hat{\phi} = .60$$
$$\hat{\theta} = .87$$
$$\hat{\sigma}_a^2 = .269$$

Using the Normal theory approximation (4.11) approximate 95% confidence limits for λ are given by (-.26,-.08). The 95% confidence intervals for ϕ and θ are given by (.53,.67) and (.82,.92) respectively.

Diagnostic checking.

Before adapting the model (5.1) and using it for forecasting, data generation or control, its adequacy has to be checked. Most

important are the residuals at this stage, since they provide the information whether the model is inadequate and how it should be changed. Plots of the ACF of the residuals are given in Fig. 4. If the time dependence in the observations is adequately represented by the model, the residuals should be white noise and the autocorrelations of the residuals $r_k(\hat{a})$ should, for large samples, be independent and approximately Normal with mean zero and variance $\frac{1}{n}$ (where n is the number of available observations). Fig. 4 does not reveal any shortcomings of the model; only one autocorrelation (at lag 18) exceeds the 2σ limits. A portmaneau lack of fit test reaches the same conclusion. It is shown in Box and Jenkins [6] that under the null hypothesis (no lack of fit) $n \sum_{k=1}^{K} r_k^2(\hat{a})$ follows a χ^2 distribution with K- (numbers of estimated AR and MA parameters) degrees of freedom. For the present example, the test statistic is 38.06. Compared to a χ^2 distribution with 34 (36-2 estimated parameters) degrees of freedom, it is insignificant at the .1 level.

Further tests to check whether there are differences in the monthly means of the residuals (analysis of variance) and whether the residuals are normally distributed (Chi-square test) failed to put model (5.1) in jeopardy.

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
24.00	7.57	3.65	5.61	16.50	14.10	7.84	2.48	1.47	4.53	11.20	5.30
8.32	3.90	8.45	2.05	5.47	14.60	4.14	3.57	1.74	1.62	0.80	1.67
2.51	2.66	1.89	7.60	4.76	2.28	2.24	1.30	1.68	1.33	0.72	0.81
1.50	1.89	2.10	3.25	12.30	5.49	2.90	1.83	1.07	1.06	3.29	1.96
2.98	3.57	3.04	9.24	12.80	13.80	5.25	2.75	2.11	1.28	1.38	4.82
5.36	11.00	10.20	7.00	8.40	5.10	4.00	7.70	2.80	10.30	2.70	3.80
6.51	2.75	1.35	3.52	39.70	13.80	4.33	2.29	3.76	5.43	15.20	2.69
2.86	7.92	9.23	5.09	18.60	5.71	4.21	2.89	2.36	8.90	17.00	1.69
3.60	5.70	6.80	7.90	16.20	20.50	36.20	10.00	16.70	15.80	7.60	15.10
16.30	11.90	1.80	2.40	19.80	8.88	17.70	7.43	4.00	5.05	9.40	6.30
16.10	2.90	3.48	14.40	28.00	30.80	23.30	7.90	5.90	7.75	4.90	4.25
4.90	6.20	2.50	1.55	10.70	22.20	11.50	3.90	2.00	1.05	1.25	1.05
1.08	1.03	1.07	1.90	1.65	3.19	1.65	2.50	1.24	1.05	1.20	0.85
1.03	0.93	1.30	2.60	3.15	15.50	4.51	2.20	5.28	7.60	1.10	2.00
11.00	11.20	2.30	4.32	15.40	7.78	12.10	3.90	3.30	1.51	1.10	6.51
11.90	7.46	5.01	13.90	12.90	5.67	2.50	2.61	2.77	1.52	1.55	1.39
3.66	3.38	3.07	2.14	14.10	7.16	0.94	0.79	0.86	0.85	0.74	0.70
1.44	5.37	17.20	11.00	10.50	4.46	2.34	4.40	2.03	2.09	1.42	0.91
1.90	0.98	1.95	3.01	11.10	16.60	9.58	3.28	2.93	2.85	0.98	0.61
2.47	7.10	3.95	8.77	6.62	7.41	5.28	1.04	0.98	1.07	1.57	1.50
9.10	7.44	6.46	7.61	11.90	9.20	15.60	2.66	1.51	1.33	0.79	0.81
1.33	1.82	2.04	5.42	5.26	32.50	5.45	3.25	1.30	0.96	2.94	3.54
12.70	5.92	7.57	15.40	12.10	9.10	6.35	3.12	2.62	1.37	1.33	1.23
1.08	0.95	0.74	0.73	4.60	8.12	8.70	2.93	13.00	2.35	1.13	1.20
1.45	3.43	5.49	7.18	11.80	10.80	9.05	3.70	5.68	3.88	2.10	1.53
1.85	3.98	3.77	2.17	10.60	19.10	14.00	3.78	2.45	2.43	2.04	2.20
4.70	7.38	4.38	7.18	9.46	5.35	2.33	2.43	7.66	4.18	7.48	4.00
2.65	4.33	3.20	19.20	5.59	17.40	10.80	10.00	15.00	3.25	2.10	5.68
4.23	8.25	5.72	2.65	7.80	4.81	3.25	1.88	2.50	3.62	1.10	0.90
1.00	1.70	2.92	3.85	10.10	4.13	4.55	5.60	6.98	4.18	5.18	4.45
3.10	5.40	2.65	3.75	7.52	8.68	2.55	3.00	1.80	1.45	1.09	1.64
2.65	4.15	3.58	3.66	11.10	33.90	16.30	9.85	2.50	1.15	1.00	0.94
2.63	2.08	1.57	0.92	6.26	16.30	9.25	2.92	1.69	1.15	2.42	2.33
4.95	3.51	1.58	1.85	6.90	8.38	2.54	2.90	2.85	4.32	1.83	6.42
9.05	6.40	4.40	2.88	21.00	19.20	15.80	22.60	7.58	3.70	2.05	1.63
2.15	5.51	3.93	22.50	7.65	10.00	4.70	2.80	13.80	6.35	3.12	2.40
3.92	6.28	2.95	13.00	23.30	17.70	11.70	5.28	4.40	4.13	5.75	3.00
4.11	7.30	6.40	5.35	10.70	9.40	4.12	9.58	3.38	8.58	6.28	4.43
6.60	2.47	1.75	2.30	6.40	17.70	5.18	3.45	4.28	1.43	1.15	0.92
1.81	2.00	1.97	1.38	10.00	36.50	7.45	3.44	4.33	8.02	3.28	2.80
5.26	4.89	4.10	6.15	11.72	13.08	8.05	4.50	4.26	3.76	3.48	2.90
5.06	2.90	3.14	5.18	7.20	8.69	6.92	3.86	3.94	3.26	3.87	2.65

Monthly Mean

Monthly Standard Deviation

Table 1: Listing of 40 years of monthly run-off data (in m^3/sec.); Carpathian River.

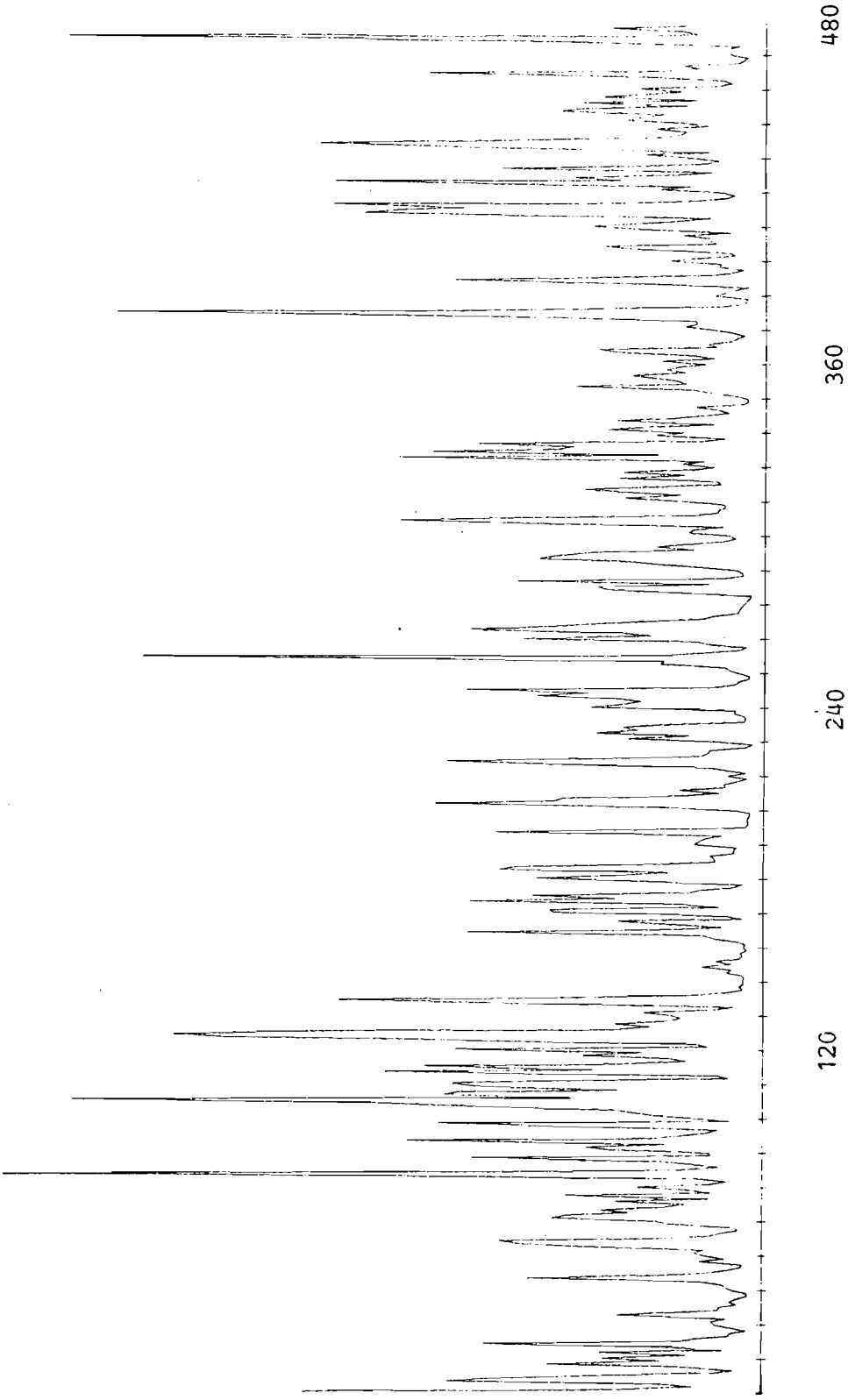

Figure 1: Plot of 40 years of monthly run-off data; Carpathian River.

ACF for $\log z_t$

1 - 12	.62	.32	.15	.04	-.14	-.12	-.07	.01	.07	.17	.31	.38
13 - 34	.31	.14	.04	-.05	-.15	-.19	-.18	-.10	-.01	.08	.24	.32
25 - 36	.25	.09	-.02	-.13	-.25	-.32	-.26	-.12	-.08	-.01	.17	.26

ACF for $(1 - B^{12}) \log z_t$

1 - 12	.51	.28	.16	.09	.11	.06	.08	.07	.01	-.05	-.20	-.44
13 - 24	-.20	-.11	-.03	.01	-.00	.06	-.03	-.07	-.01	.00	.02	.01
25 - 36	.01	-.03	-.08	-.10	-.12	-.17	-.08	.00	-.06	-.09	-.04	-.03

Table 2: Empirical ACF for $\log z_t$ and $(1 - B^{12}) \log z_t$

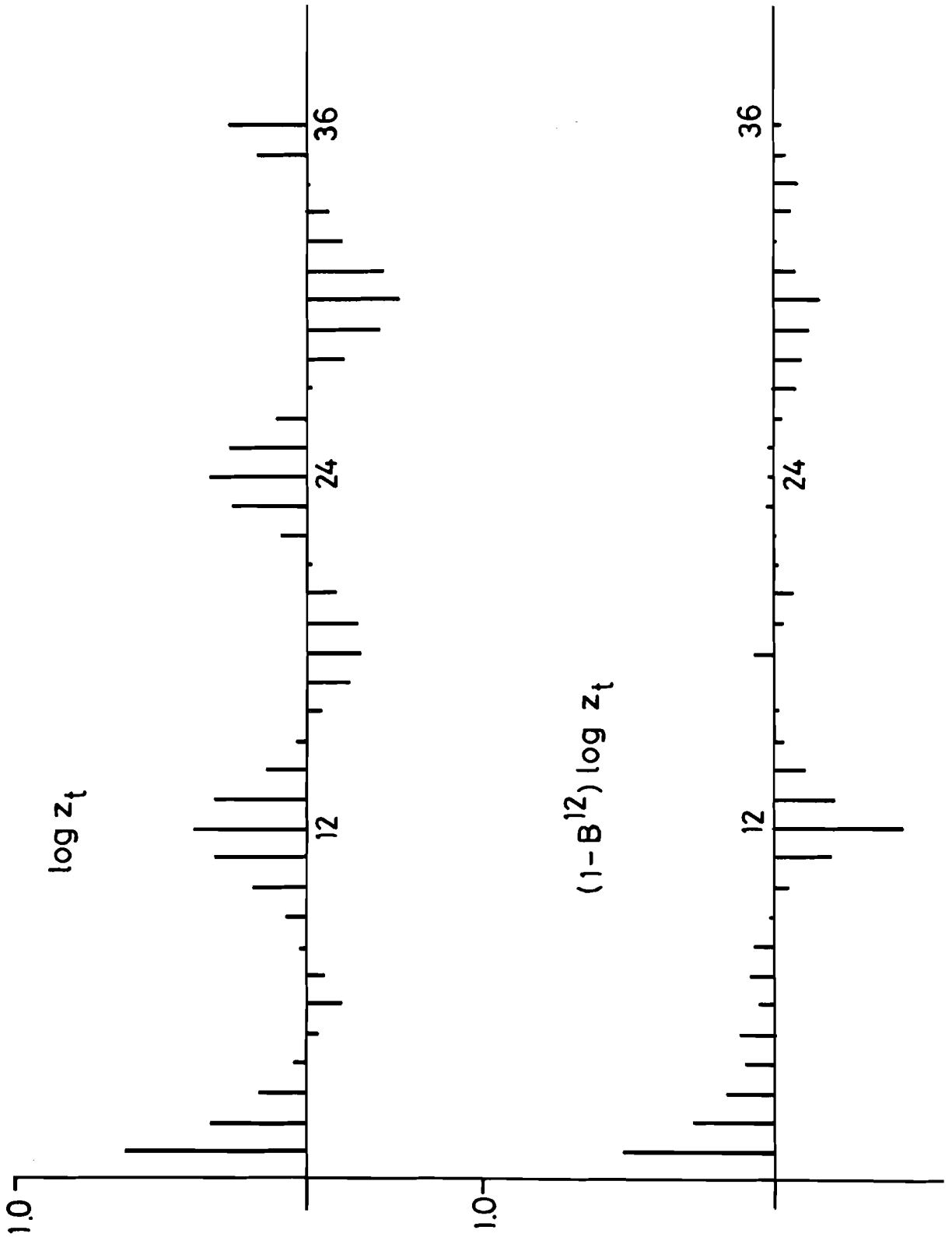

Figure 2: Plot of Empirical ACF for log z_t and $(1 - B^{12})$ log z_t

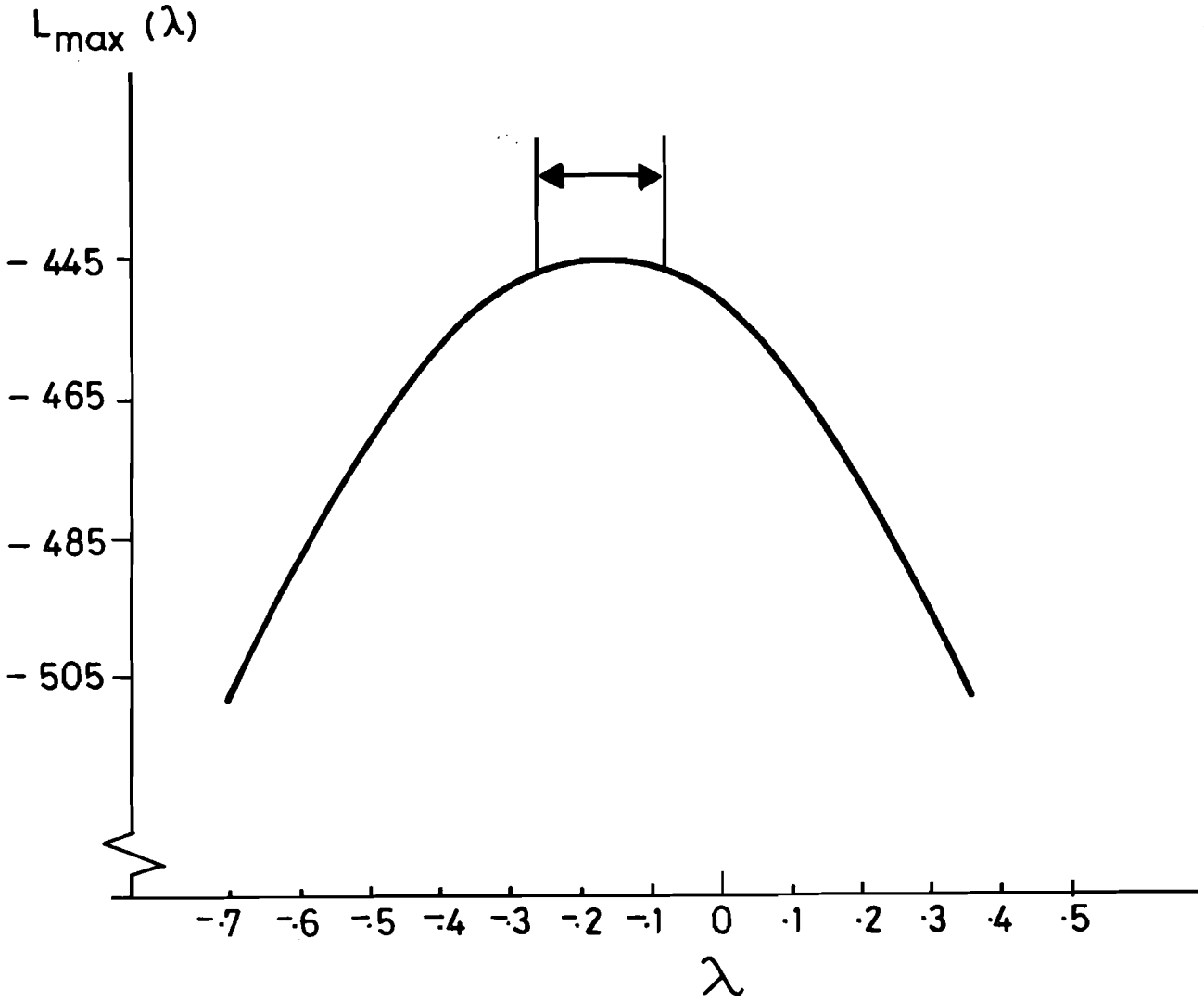

Figure 3: Plot of the function $L_{max}(\lambda)$ for unstandardized run-off data. Arrows show approximate 95 percent confidence interval for λ.

1 - 12	.00	-.01	-.01	-.02	.10	.03	.07	.02	.01	.04	.05	.01
13 - 24	.09	-.02	.01	-.01	-.02	.12	-.03	-.04	-.01	-.02	.07	-.01
25 - 36	.05	.02	-.03	-.05	-.01	-.07	-.04	-.08	-.04	-.07	-.04	-.02

Table 3: ACF of The Residuals $\hat{a}_t$ of Model (5.1)

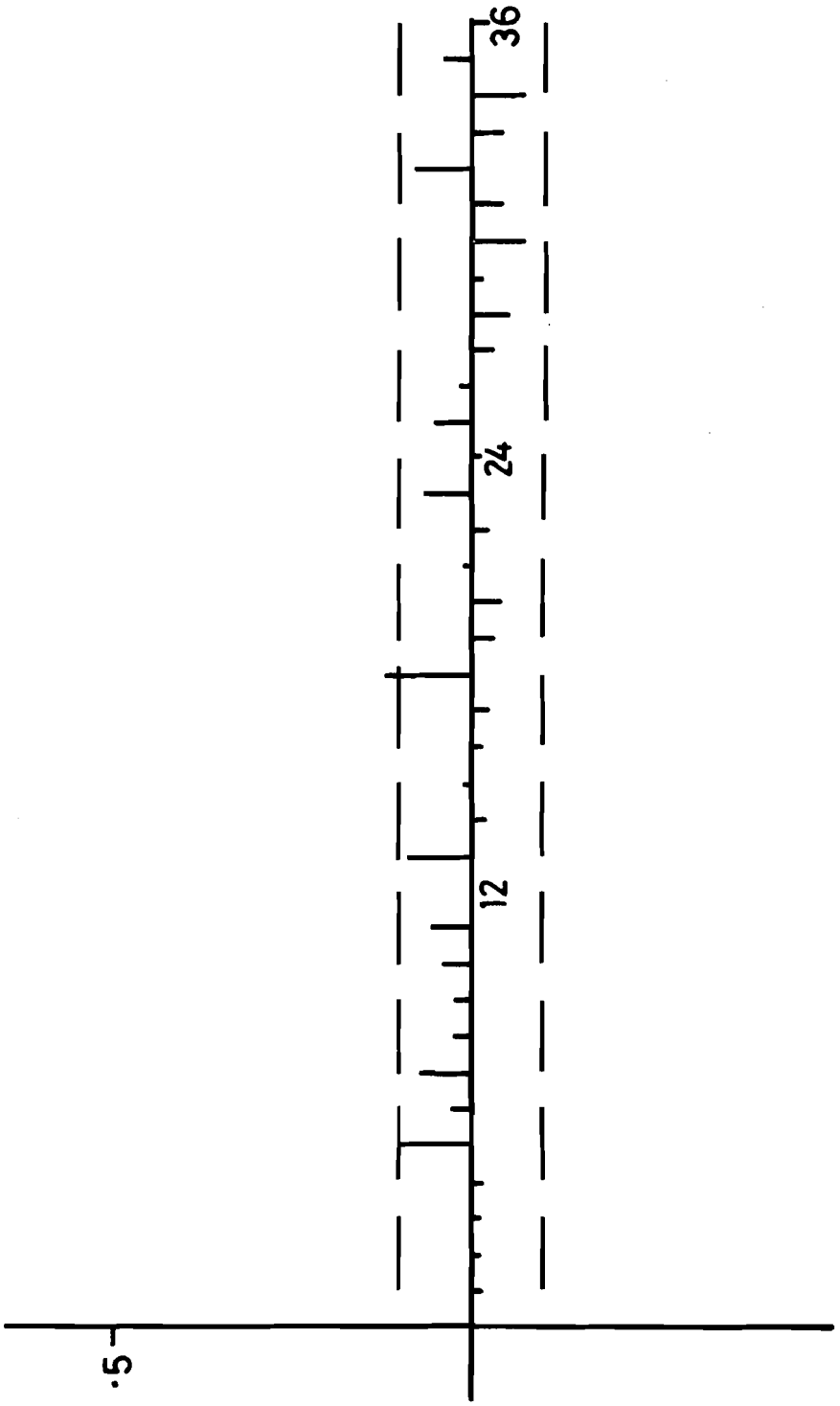

Figure 4: Plot of the ACF of the Residuals and $\pm 2\sigma$ limits of model (5.1).

APPENDIX:

USE OF AUTOCORRELATION FUNCTIONS IN IDENTIFICATION.

Since the class of ARIMA models is too extensive to be conveniently fitted directly to the data, rough methods for identifying parsimonious subclasses have to be found. The empirical autocorrelation functions of the original series and its differences are important tools for the indentification of parsimonious models

$$r_k = \sum_{t=1}^{n-k} (z_t-\bar{z})(z_{t+k}-\bar{z}) / \sum_{t=1}^{n} (z_t-\bar{z})^2 \quad , \qquad (A.1)$$

where $\bar{z}$ is the sample mean of n available observations. The empirical autocorrelations r_k are for any stationary process consistent estimators of the theoretical autocorrelations ρ_k

$$\rho_k = \rho_{-k} = E(z_t-\mu)(z_{t+k}-\mu)/E(z_t-\mu)^2 \quad . \qquad (A.2)$$

The strategy in determining the order of the autoregressive, moving average and differencing operators, is first to difference the observations until stationarity is achieved. It can be shown that the autocorrelations of any ARMA (p,q) process statisfy

$$\phi(B)\rho_k = 0 \qquad \text{for } k > q \quad .$$

The solution of this difference equation (assuming distinct roots for convenience) is given by

$$\rho_k = A_1 G_1^{\,k} + \ldots + A_p G_p^{\,k} \qquad \text{for } k > q - p \qquad (A.3)$$

where A_i are the inverse of the roots of $\phi(B) = 0$. For a stationary model, the ACF will die out quickly as k increases. Whenever a root, however, approaches the unit circle

(nonstationarity) the ACF will fall off slowly and very nearly linearly. This situation occurs, for example, in Fig. 2, where the autocorrelations of the log of the observations 12 steps apart fail to die out quickly. This prompted us to consider $(1-B^{12})$ log z_t to achieve stationarity.

After achieving stationarity, the orders of the autoregressive and moving average operations have to be determined. It can be shown that the ACF of a MA(q) process is given by only q non-zero autocorrelations, while the autocorrelations for higher lags (greater than q) vanish. The ACF of autoregressive or mixed autoregressive moving average process is infinite in extent and, as can be seen from (A.3), is dominated by exponentials and damped sine waves.

As far as our example in section 5 is concerned, it can be seen from Fig. 2 that the autocorrelations of $(1-B^{12})$ log z_t decay roughly exponentially from lag 1 onwards. However, apart from this correlation pattern, which can be explained as coming from a first order autoregressive process, one can still notice a seasonal effect at lag 12. Since the seasonal effect only results in significant correlation at lag 12, but not at lag 24,36,... we include a seasonal moving average operator into the model. Therefore we tentatively identify the model as in (5.1).

Another useful tool for the identification of models is the partial autocorrelation function. It is a device which exploits the fact that whereas an autoregressive process of order p has an autocorrelation function which is infinite in extent, it can, by its very nature, be described in terms of p non-zero functions of the autocorrelations.

REFERENCES

[1] Abraham, B. and G.E.P. Box. *Linear Models, Time Series and Outliers 3: Stochastic Difference Equation Models*, Technical Report 438, Department of Statistics, University of Wisconsin-Madison, 1975.

[2] Adamowski, K. Spectral Density of a River Flow Time Series, *Journal of Hydrology*, 14, (1971), pp. 43-52.

[3] Bartlett, M.S. On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series, *Journal Royal Stat. Soc., Series B*, 8, (1946), pp.27.

[4] Box, G.E.P. and D.R. Cox. An Analysis of Transformations, *Journal Royal Stat. Soc., Series B*, 26, (1964), pp.211-252.

[5] Box, G.E.P. and G.M. Jenkins. *Models for Prediction and Control; VIII Forecasting Seasonal Time Series*, Technical Report 104, Department of Statistics, University of Wisconsin-Madison, 1967.

[6] Box, G.E.P. and G.M. Jenkins. *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA., 1970.

[7] Carlson, R.F., MacCormick, A.J.A. and D.G. Watts. Application of Linear Random Models to Four Annual Flow Series, *Water Resources Res.*, 6, (1970), pp. 1070-1078.

[8] Kareliotis, S.J. and V.T. Chow. Analysis of Residual Hydrologic Stochastic Process, *Journal of Hydrology*, 15, (1972), pp. 113-130.

[9] Kolomogorov, A.N. Interpretation and Extrapolation von Stationären Zufälligen Folgen, *Bull. Acad. Sci. (Nauk) USSR, Ser. Math.*, 5, (1941), pp. 3-14.

[10] Ledolter, J. *Topics in Time Series Analysis*, Ph.D. Thesis, (1975) Department of Statistics, University of Wisconsin-Madison.

[11] McKerchar, A.I. and J.W. Delleur. Application of Seasonal Parametric Linear Stochastic Models to Monthly Flow Data, *Water Resources Res.*, 10, (1974), pp. 246-255.

[12] Quimpo, R.G. Stochastic Analysis of Daily River Flows, *Journal of the Hydraulics Division, ASCE*, 94, (1968) pp. 43-57.

[13] Quimpo, R.G. Autocorrelation and Spectral Analysis in Hydrology, *Journal of Hydraulics Division, ASCE*, 94, (1968) pp. 363-373.

[14] Rodriguez-Irtube, I., Mejia, J.M. and D.R. Dawdy. Streamflow Simulation. 1 - A New Look at Markovian Models, Fractional Gaussian Noise and Crossing Theory, *Water Resources Res.*, 8, (1972), pp. 921-930.

[15] Schuster, A. On the Periodicities of Sunspots, *Phil. Trans. A*, 206, (1906), pp.69-100.

[16] Tintner, G. and G. Kadekodi. A Note on the Use of Transformations and Differences in the Estimation of Econometric Relations, *Sankhya, B*, 35, (1971), pp.263-277.

[17] Wiener, N. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley, New York, N.Y., 1949.

[18] Wilson, T.G. Discussion of paper by Chatfield and Prothero, "Box and Jenkins Seasonal Forecasting : Problems in a Case Study," *Journal Royal Stat. Soc., Series A*, 136, (1973), pp. 315-319.

[19] Wold, H.O. *A Study in the Analysis of Stationary Time Series*, Almquist and Wicksell, Upsala, 2nd Edition, 1954.

[20] Yaglom, A.M. *An Introduction to the Theory of Random Functions*, English Translation by A. Silverman, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[21] Yaglom, A.M. The Correlation Theory of Processes Whose n'th Difference Constitutes a Stationary Process, *Matem. Sb.*, 37, (1955), pp. 141.

[22] Yule, G.U. On a Method of Investigating Periodicities in Disturbed Series, With Special Reference to Wölfer's Sunspot Numbers, *Phil. Trans., A*, 226, (1927), pp. 267-298.