RM-77-11

THE ANALYSIS OF MULTIVARIATE TIME SERIES WITH A VIEW TO APPLICATIONS IN HYDROLOGY

Johannes Ledolter

February 1977

PREFACE

During the past three years, the IIASA Water Project (presently the Water Group of the Resources and Environment Area) has concentrated on the methodology for planning, design and operation of water resource systems. The importance of streamflow generation methods for the design and operation of water resource systems has been recognized and a study comparing multi-site streamflow generation methods was conducted by J. Kindler and W. Zuberek (*On Some Multi-Site Multi-Season Streamflow Generation Models*, RM-76-76).

The present Research Memorandum discusses multivariate time series methodology and shows how multi-site streamflow generation models can be brought into this general framework. The present paper is an extension of RM-76-69 (*ARIMA Models and Their Use in Modelling Hydrologic Sequences*) which considers a general class of univariate time series methods for stream flow generation at a single site.

ABSTRACT

In this paper we discuss stochastic models for vector processes, in particular the class of multivariate autoregressive moving average models. Special cases of this class have been discussed in the literature on multi-site streamflow generation and it is shown how these can be brought into a general framework.

An iterative model building procedure, consisting of model specification--estimation--diagnostic checking is stressed. Results on model specification are given and it is shown how partial autocovariance matrices can be used to check whether multivariate autoregressive models provide adequate representation for (standardized) streamflow sequences. Furthermore, estimation of parameters in multivariate autoregressive moving average models is discussed and it is pointed out that moment estimators can be inefficient when moving average parameters are present. An approximate maximum likelihood estimation procedure is suggested.

In the concluding section, we summarize important practical implications for hydrologists.

The Analysis of Multivariate Time Series With a View to Applications in Hydrology

1. Introduction

In a previous paper [12] we discussed stochastic models for univariate time series data. Applications of stochastic modelling in water resource systems will, however, frequently demand the modelling of hydrologic sequences at a number of sites within a river basin. If the individual series are spatially uncorrelated, univariate models can be used at each of the sites within the system and no representation of spatial correlation is necessary. However, significant cross correlations will usually exist between historic observations measured at neighboring sites at the same (lag zero cross correlations) and lagged (lag k cross correlations) time points. Models have to be found which preserve cross correlations between sites in addition to the appropriate properties at individual sites.

2. Stationary stochastic vector processes

Let $z_i(t)$, $i = 1,2,\ldots,s$; $t = 0,\pm1,\pm2,\ldots$ be s real valued sequences of random variables where

$$E[z_i(t)] = \mu_i \qquad 1 \leq i \leq s \tag{2.1}$$

is a finite constant independent of t, and where the covariance function defined by

$$E[(z_i(t) - \mu_i)(z_j(t+k) - \mu_j)] = \gamma_{ij}(k) \qquad 1 \leq i,j \leq s \tag{2.2}$$

is finite and does not depend on t for all $k = 0,\pm1,\pm2,\ldots$.
Then we say that each of the series $z_i(t)$ is covariance stationary and that the series are mutually stationary correlated.

In vector notation $\underset{\sim}{z}(t)' = (z_1(t),\ldots,z_s(t))$ we call $\underset{\sim}{z}(t)$

satisfying (2.1) and (2.2) a s-variate covariance stationary (weakly stationary) time series.

The matrix $\Gamma(k) = \{\gamma_{ij}(k)\}$ defined by

$$\Gamma(k) = E[(\underset{\sim}{z}(t) - \underset{\sim}{\mu})(\underset{\sim}{z}(t+k) - \underset{\sim}{\mu})'] = \Gamma(-k)' \quad , \tag{2.3}$$

where $\underset{\sim}{\mu}' = (\mu_1, \ldots, \mu_s)$, will denote the autocovariance matrix at lag k.

The autocorrelation matrix at lag k, $P(k) = \{\rho_{ij}(k)\}$, can be obtained from the autocovariance matrix by

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{(\gamma_{ii}(0)\gamma_{jj}(0))^{\frac{1}{2}}} \quad . \tag{2.4}$$

The autocovariance generating function, $\Gamma(B)$, is defined as

$$\Gamma(B) = \sum_{k=-\infty}^{+\infty} \Gamma(k)B^k \quad , \tag{2.5}$$

and the non-normalized spectral density matrix $F(w)$ of the stationary process $\underset{\sim}{z}(t)$ is given by the discrete Fourier transform of its autocovariance function,

$$F(w) = \{f_{ij}(w)\}$$

where

$$f_{ij}(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \gamma_{ij}(k)e^{-iwk} \quad . \tag{2.6}$$

For discrete time series with sampling rate of unity w lies in the frequency range $-\pi$ to $+\pi$. (The range of w for continuous time series goes from $-\infty$ to $+\infty$. Sampling the process at discrete points, however, "folds" the spectrum into the range $(-\pi/\Delta t, \pi/\Delta t]$ where Δt is the sampling interval.)

The spectral density matrix has the following properties:

(i) $F(w)$ is Hermitian; i.e., $F(w) = F(w)^T$ where T denotes the operation of matrix transposition and complex conjugation.

(ii) $F(w)$ is non negative definite,

(iii) For the continuous process we have that $F(w) \to 0$ as $w \to \pm\infty$. Furthermore $\int_{-\infty}^{+\infty} f_{ij}(w)dw$, is finite $(1 \leq i,j \leq s)$.

We assume in the follwoing that $\Gamma(B)$ is rational, that is we suppose that each element of the $(s \times s)$ matrix $\Gamma(B)$ is a rational function of B (or in terms of the spectral density matrix we assume that each $f_{ij}(w)$ is a rational function of e^{-iw}). Furthermore we assume that $\det(\Gamma(B)) \neq 0$ almost everywhere i.e.: with probability 1 (or equivalently in terms of the spectral density matrix $\det(F(w)) \neq 0$ almost everywhere). This assumption eliminates from consideration those joint processes in which certain constraints hold among the series; for example, the situation in which two series are generated from one shock series).

Given the autocovariance generating function (2.5) (or equivalently the spectral density matrix (2.6)) the statistical characteristics of the stochastic process are completely specified up to second order moments (in the case of Normally distributed $\underset{\sim}{z}(t)$ all moments are specified). Yet, for practical applications involving prediction and control a difference equation representation of the time series proves to be more convenient. It is easily shown that given the autocovariance generating function (2.5) there is a multiplicity of difference equation representations to choose from.

The existence of a *canonical* factorization of the autocovariance generating function (spectral density matrix) of a weakly stationary time series was shown, for example, in Rozanov [18], Hannan [7]. The resulting difference equation form (moving average representation) which is said to be a canonical model for the s-variate covariance stationary time series (from now on it will be assumed that the mean has been subtracted from the time series) is given by

$$\underset{\sim}{z}(t) = \Psi(B)\underset{\sim}{a}(t) \tag{2.7}$$

where: i) $\Psi(B) = \sum_{k=0}^{\infty} \Psi(k)B^k$ is a matrix of elements which

are rational functions in the backshift operator B ($B^m \underset{\sim}{a}(t) = \underset{\sim}{a}(t-m)$)

ii) det ($\Psi(B)$) does not vanish inside the unit circle and the elements of $\Psi(B)$ are holomorphic in and on the unit circle.
The holomorphic (or analytic) condition means that the denominators of all elements of $\Psi(B)$ have zeros lying outside the unit circle.

iii) $\underset{\sim}{a}(t)$ is an s-variate uncross correlated white noise sequence

$$E[\underset{\sim}{a}(t)\underset{\sim}{a}(t+k)'] = \delta_0^k I$$

where δ_0^k is the Kronecker delta function

$$\delta_0^k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$$

and I is the [s×s] identity matrix.

If in addition one assumes that det $\Psi(B)$ has no zeros on the unit circle one can expand $[\Psi(B)]^{-1}$ as

$$[\Psi(B)]^{-1} = \Pi(B) = \Pi(0) - \sum_{k=1}^{\infty} \Pi(k)B^k \quad ,$$

where $\Pi(0)$ is nonsingular and $\Pi(k)$ goes to a zero matrix O as $k \to \infty$. Thus (2.7) can be written as

$$\underset{\sim}{z}(t) = [\Pi(0)]^{-1}\underset{\sim}{a}(t) + \sum_{k=1}^{\infty} [\Pi(0)]^{-1}\Pi(k)\underset{\sim}{z}(t-k) \quad . \qquad (2.8)$$

The present observation vector is a weighted average of past observations and the random perturbation $\underset{\sim}{a}(t)$. Since the weighting function $\Pi(k)$ tends to the null matrix as $k \to \infty$ the observations $\underset{\sim}{z}(t)$ depend less and less on the remote past.

Specification that the representation be canonical does not guarantee uniqueness of the model. This can easily be seen since

$$\underset{\sim}{z}(t) = \Psi^*(B)\underset{\sim}{a}(t) \quad , \qquad (2.9)$$

where $\Psi^*(B) = \Psi(B)P$ and P is any orthonormal matrix $PP' = I$ will result in the same covariance generating function.

To show this, we first point out the following

Lemma: Suppose that the covariance stationary process $\underset{\sim}{z}(t)$ has the representation $\underset{\sim}{z}(t) = \Psi(B)\underset{\sim}{b}(t)$ where $\underset{\sim}{b}(t)$ is multivariate white noise with $E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k G$. Then the covariance generating function is given by

$$\Gamma(B) = \Psi(F)G\Psi(B)' \quad \text{where} \quad F = B^{-1} \quad . \tag{2.10}$$

Proof: $\Gamma(k) = E[(\Psi(B)\underset{\sim}{b}(t))(\Psi(B)\underset{\sim}{b}(t+k))']$

$$= E\left[\left(\sum_{\ell=0}^{\infty} \Psi(\ell)\underset{\sim}{b}(t-\ell)\right)\left(\sum_{m=0}^{\infty} \Psi(m)\underset{\sim}{b}(t+k-m)\right)'\right]$$

$$= \sum_{\ell}\sum_{m} \Psi(\ell)E[\underset{\sim}{b}(t-\ell)\underset{\sim}{b}(t+k-m)']\Psi(m)'$$

$$= \sum_{\ell} \Psi(\ell)G\Psi(\ell+k)' \quad . \tag{2.11}$$

since the only non zero terms in $E[\underset{\sim}{b}(t-\ell)\underset{\sim}{b}(t+k-m)']$ occur when $m = \ell + k$. Furthermore, (2.11) is the coefficient of B^k in $\Psi(F)G\Psi(B)'$ and thus it follows that

$$\Gamma(B) = \sum_{k=-\infty}^{+\infty} \Gamma(k)B^k = \Psi(F)G\Psi(B)' \quad .$$

q.e.d.

Using this result it follows that the autocovariance generating function of (2.9) with $E[\underset{\sim}{a}(t)\underset{\sim}{a}(t+k)'] = \delta_0^k I$ is given by

$$\Gamma^*(B) = \Psi^*(F)\Psi^*(B)' = \Psi(F)PP'\Psi(B)' = \Psi(F)\Psi(B)'$$

and coincides with the autocovariance generating function of (2.7).

However, suppose that after obtaining a canonical representation (2.7) we define $\Psi^*(B) = \Psi(B)[\Psi(0)]^{-1}$ and $G = \Psi(0)\Psi(0)'$ where $\Psi(0)$ is nonsingular because of the canonicalness of the factorization and since $\underset{\sim}{z}(t)$ is of full rank. Then we can write

$$\Gamma(B) = \Psi(F)\Psi(B)' = \Psi^*(F)G\Psi^*(B)' \qquad (2.12)$$

and the above factorization gives the model in the following form

$$\underset{\sim}{z}(t) = \Psi^*(B)\underset{\sim}{b}(t) \quad , \qquad (2.13)$$

where $\Psi^*(0) = I$ and $\underset{\sim}{b}(t)$ is a s-variate white noise sequence with $E[\underset{\sim}{b}(t)\underset{\sim}{b}(t)'] = G = \Psi(0)\Psi(0)'$. Uniqueness of the factorization in (2.12) ensures the uniqueness of the model. Model (2.13) is a canonical model with contemporaneously correlated white noise.

Although the conditions on $\Psi^*(B)$ in (2.13) provide a unique specification for $\underset{\sim}{z}(t)$ with covariance generating function $\Gamma(B)$ and noise covariance G, other specifications are possible.

By specifying $\overline{\Psi}(0)$ to be a lower (upper) triangular matrix a unique representation in terms of

$$\underset{\sim}{z}(t) = \overline{\Psi}(B)\underset{\sim}{a}(t)$$

where

$$E[\underset{\sim}{a}(t)\underset{\sim}{a}(t+k)'] = \delta_0^k I$$

can be achieved (Quenouilli [16]).

If one is interested in a unique representation in terms of contemporaneously uncorrelated white noise $\underset{\sim}{\bar{a}}(t)$ with $E[\underset{\sim}{\bar{a}}(t)\underset{\sim}{\bar{a}}(t+k)'] = \delta_0^k D$ where D is a diagonal matrix, it can be achieved by

$$\underset{\sim}{z}(t) = \overline{\overline{\Psi}}(B)\underset{\sim}{\bar{a}}(t)$$

where $\overline{\overline{\Psi}}(0)$ is a lower (upper) diagonal matrix with 1's in the diagonal (Haugh [9]).

3. Canonical autoregressive moving average representations

The elements of $\Psi^*(B)$ in (2.12) are rational functions in B; in the following we thus consider the factorization

$$\Psi^*(B) = \Phi^{-1}(B)\Theta(B)$$

where $\Phi(B)$ and $\Theta(B)$ are of the form

$$\Phi(B) = I - \sum_{j=1}^{p} \Phi_j B^j$$

$$\Theta(B) = I - \sum_{j=1}^{q} \Theta_j B^j \quad .$$

It is assumed that the roots of $\det \Phi(B) = 0$ all lie outside the unit circle (stationarity condition) and that the roots of $\det \Theta(B) = 0$ all lie on or outside the unit circle. Furthermore it is assumed that $\det \Phi(B) = 0$ has no common roots with $\det \Theta(B) = 0$. These conditions are necessary for an identified model in the economic sense, but are not sufficient to guarantee an identified model. Conditions to guarantee an identified model are given in Hannan [8].

If one additionally restricts $\det \Theta(B)$ to have no zeros on the unit circle then we have the conditions for invertibility as discussed in Box and Jenkins [3].

The model of $\underset{\sim}{z}(t)$ can be written in terms of the difference equation

$$\Phi(B)\underset{\sim}{z}(t) = \Theta(B)\underset{\sim}{b}(t) \tag{3.1}$$

$$\underset{\sim}{z}(t) - \Phi_1\underset{\sim}{z}(t-1) - \ldots - \Phi_p\underset{\sim}{z}(t-p) = \underset{\sim}{b}(t) - \Theta_1\underset{\sim}{b}(t-1) - \ldots - \Theta_q\underset{\sim}{b}(t-q) \tag{3.2}$$

with

$$E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k G \quad . \tag{3.3}$$

Model (3.1) is commonly known as the multivariate autoregressive moving average model (ARMA(p,q)). Representation (3.1) is

in terms of contemporaneously correlated white noise $\underset{\sim}{b}(t)$. Similarly, specifying the moving average parameter at $B^{0}\tilde{}$ to be a lower (upper) triangular matrix will result in a unique ARMA model with uncorrelated white noise $E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k I$.

A main motivation for the use of mixed autoregressive moving average models is to satisfy the principle of parsimony. Since stochastic models contain parameters whose values must be estimated from a record of observations, it is important, that one employs models with the smallest possible number of parameters for adequate representation. A combination of autoregressive and moving average terms represents a flexible class of models capable of approximating many stochastic models observed in practice.

Several special cases of multivariate ARMA models have been considered in the literature on modelling of streamflow data recorded at different sites.

i) Matalas [13] uses the multivariate first order autoregressive model (multivariate Markovian model) to model and generate streamflow data at s different sites.

ii) O'Connel [14] considers the multivariate first order autoregressive model with correlated residuals (ARMA (1,1)) as well as the first order moving average model (MA(1)). Iterative procedures are given for the derivation of the parameter estimates from empirical autocovariance matrices.

Some special cases of multivariate ARMA models

a) Multivariate autoregressive model of order p

$$(I - \Phi_1 B - \ldots - \Phi_p B^p)\underset{\sim}{z}(t) = \underset{\sim}{b}(t) \qquad (3.4)$$

with

$$E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k G$$

or equivalently in terms of contemporaneously uncorrelated white noise

$$(I - \Phi_1 B - \ldots - \Phi_p B^p)\underset{\sim}{z}(t) = C\underset{\sim}{a}(t) \qquad (3.5)$$

with

$$E[\underset{\sim}{a}(t)\underset{\sim}{a}(t+k)'] = \delta_0^k I$$

and C is a lower (upper) triangular matrix; since CC' = G, the matrix C can be derived by lower (upper) triangularization of the matrix G.

From (3.4) it follows that

$$\Gamma(k) - \Gamma(k-1)\Phi_1' - \ldots - \Gamma(k-p)\Phi_p' = 0 \qquad k \geq 1 \tag{3.6}$$

and

$$G = \Gamma(0) - \Phi_1\Gamma(1) - \ldots - \Phi_p\Gamma(p) \quad . \tag{3.7}$$

Considering the first p equations and using the relation $\Gamma(k) = \Gamma(-k)'$ one can derive $\Phi_1, \ldots, \Phi_p$ and G in terms of the first p lag autocovariance matrices by solving

$$\begin{bmatrix} \Gamma(0) & \Gamma(1)' & \ldots & \Gamma(p-1)' \\ \Gamma(1) & \Gamma(0) & \ldots & \Gamma(p-2)' \\ & & & \\ \Gamma(p-1) & \Gamma(p-2) & \ldots & \Gamma(0) \end{bmatrix} \begin{bmatrix} \Phi_1' \\ \Phi_2' \\ \vdots \\ \Phi_p' \end{bmatrix} = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(p) \end{bmatrix} \tag{3.8}$$

and

$$G = \Gamma(0) - \Phi_1\Gamma(1) - \ldots - \Phi_p\Gamma(p) \quad . \tag{3.9}$$

For the first order autoregressive model this simplifies to

$$\Gamma(0)\Phi_1' = \Gamma(1) \quad \Rightarrow \quad \Phi_1 = \Gamma(1)'[\Gamma(0)]^{-1}$$

and

$$G = \Gamma(0) - \Phi_1\Gamma(1) = \Gamma(0) - \Gamma(1)'[\Gamma(0)]^{-1}\Gamma(1). \tag{3.10}$$

b) Multivariate moving average model of order q

$$\underset{\sim}{z}(t) = (I - \Theta_1 B - \ldots - \Theta_q B^q)\underset{\sim}{b}(t) \tag{3.11}$$

where $\underset{\sim}{b}(t)$ is white noise $E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k G$.

It is easily shown that

$$\Gamma(k) = \sum_{\ell=0}^{q-k} \Theta_\ell G \Theta'_{\ell+k} \qquad 0 \leq k \leq q \tag{3.12}$$

where

$$\Theta_0 \doteq I$$

and

$$\Gamma(k) = 0 \qquad j > q \quad .$$

Difficulties to determine the moving average parameters and the covariance matrix of the shocks $\underset{\sim}{b}(t)$ from this set of equations, however, occur, since there are $2^{\tilde{q}}$ schemes compatible with the set of covariance matrices (3.12).

From a prediction point of view, however, the choice doesn't matter and according to the invertibility condition we choose the one set of solution such that there are no roots of $\det(\Theta(B)) = 0$ inside the unit circle.

c) Multivariate autoregressive moving average model

$$(I - \Phi_1 B - \ldots - \Phi_p B^p)\underset{\sim}{z}(t) = (I - \Theta_1 B - \ldots - \Theta_q B^q)\underset{\sim}{b}(t) \tag{3.13}$$

$$E[\underset{\sim}{b}(t)\underset{\sim}{b}(t+k)'] = \delta_0^k G \quad .$$

From (3.13) it follows that from lag q+1 the autocovariance matrices follow the matrix difference equation

$$\Gamma(k) - \Gamma(k-1)\Phi_1' - \ldots - \Gamma(k-p)\Phi_p' = 0 \qquad \text{for } k \geq q+1 \quad .$$

These schemes behave similarly to ordinary autoregressive schemes in the relationships between their covariance matrices $\Gamma(k)$, except for the first few which depend on the extent of the moving average.

The following example of the multivariate ARMA(1,1) process will demonstrate this in more detail.

$$(1 - \Phi_1 B)\underset{\sim}{z}(t) = (1 - \Theta_1 B)\underset{\sim}{b}(t) \quad . \tag{3.14}$$

Then it can be shown that

$$\begin{cases} \Gamma(0) = \Phi_1\Gamma(1) + G - \Theta_1 G\Phi_1' + \Theta_1 G\Theta_1' \\ \Gamma(1) = \Gamma(0)\Phi_1' - G\Theta_1' \\ \Gamma(k) = \Gamma(k-1)\Phi_1' \qquad k \geq 2 \end{cases} \tag{3.15}$$

and iterative procedures can be found to solve these equations for Φ_1, Θ_1 and G (O'Connel [14]).

d) Extension to nonstationary models

Many time series encountered exhibit nonstationary behavior and in particular do not vary about a fixed mean (trend, periodicity). Nevertheless they exhibit homogeneity in the sense that apart from local level (trend, periodicity) one part of the series behaves very much like the other.

It is shown in great detail in [3] how simplifying operators (such as the ordinary differences $(1-B)^d$, seasonal differences $(1-B^s)^d$, or in general operators with roots on the unit circle) can be used to transform nonstationary series into stationary ones.

4. Models for individual series from multivariate AR(MA) processes

In the literature on stochastic modelling of univariate streamflow sequences the first order autoregressive model (applied to the standardized monthly logarithm of the observations) is frequently used to generate synthetic streamflow records. These models are subsequently extended to the case of observations at several sites within a watershed basin. A frequently considered extension of the univariate AR(1) model for observations at one site is the multivariate AR(1) model for observations recorded at several sites (Matalas [13]).

In this part of the paper we discuss the question whether individual series (subsets) from a multivariate AR(MA) process follow the same AR(MA) type process. It is shown below that individual series from a multivariate AR process follow a univariate

autoregressive model, but of higher order and with correlated residuals (ARMA model). Individual series from a multivariate MA process, however, are shown to follow again a moving average process of the same (or lower) order.

Let us first consider the case when the s-dimensional series $\underset{\sim}{z}(t)$ follows the multivariate AR(p) model.

$$\Phi(B)\underset{\sim}{z}(t) = (I - \Phi_1 B - \ldots - \Phi_p B^p)\underset{\sim}{z}(t) = \underset{\sim}{b}(t)$$

$$\begin{bmatrix} v_1(B) & \underset{\sim}{v}_2(B)' \\ \underset{\sim}{v}_3(B) & V(B) \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_s(t) \end{bmatrix} = \begin{bmatrix} b_1(t) \\ b_2(t) \\ \vdots \\ b_s(t) \end{bmatrix} \qquad (4.1)$$

We partition the [s×s] operator $\Phi(B)$. $v_1(B)$ is a scalar operator; $\underset{\sim}{v}_2(B)$ and $\underset{\sim}{v}_3(B)$ are vectors of order [s×1]; V(B) is of order [s-1 × s-1]. Then it follows that

$$[v_1(B) - \underset{\sim}{v}_2(B)'V(B)^{-1}\underset{\sim}{v}_3(B)]z_1(t) = b_1(t) - \underset{\sim}{v}_2(B)'V(B)^{-1}\underset{\sim}{b}^*(t) \qquad (4.2)$$

where

$$\underset{\sim}{b}^*(t)' = (b_2(t),\ldots,b_s(t)) \quad .$$

Multiplying each side with the determinant $|V(B)|$ reduces the operators in (4.2) to polynomials in B.

$$[|V(B)|v_1(B) - \underset{\sim}{v}_2(B)'V^A(B)\underset{\sim}{v}_3(B)]z_1(t) = |V(B)|b_1(t) - \underset{\sim}{v}_2(B)'V^A(B)\underset{\sim}{b}^*(t) \qquad (4.3)$$

where $V^A(B)$ is the adjoint of matrix V(B). From (4.3) one can see that the univariate series from a multivariate AR(p) process follows a complicated model. Since for a general AR(p) process $\underset{\sim}{v}_2(B)$ and $\underset{\sim}{v}_3(B)$ will have no common factors with V(B), any individual series (in our case we took the first one) of a s dimensional AR(p) process will follow an autoregressive model of order sp with correlated residuals.

To illustrate this result in more detail we consider a bivariate AR(1) model, i.e.:

$$(I - \Phi B)\underset{\sim}{z}(t) = \underset{\sim}{b}(t)$$

$$\begin{bmatrix} 1-\phi_{11}B & -\phi_{12}B \\ -\phi_{21}B & 1-\phi_{22}B \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} b_1(t) \\ b_2(t) \end{bmatrix} \tag{4.4}$$

Then

$$\left[(1 - \phi_{11}B) - \frac{\phi_{12}\phi_{21}B^2}{1 - \phi_{22}B}\right] z_1(t) = b_1(t) + \frac{\phi_{12}B}{1 - \phi_{22}B}\, b_2(t) \tag{4.5}$$

or

$$[1 - (\phi_{11} + \phi_{22})B - (\phi_{12}\phi_{21} - \phi_{11}\phi_{22})B^2]z_1(t) = (1 - \phi_{22}B)b_1(t) + \phi_{12}b_2(t-1) \tag{4.6}$$

It is seen that in general the univariate series $z_1(t)$ follows a second order autoregressive process with correlated residuals (ARMA model).

In the special case $\phi_{12} = 0$ (situation when no feedback from z_2 to z_1 present), the operator $(1 - \phi_{22}B)$ in (4.6) cancels on each side and the univariate series $z_1(t)$ follows a first order autoregressive model.

For the multivariate moving average process of order q it was shown in (3.12) that the covariance matrix function has a cut off after lag q. The individual series (let's say $z_1(t)$) has thus at most the first q autocovariances different from zero and application of Theorem 10 on page 63 in Hannan [7] implies that $z_1(t)$ has again a moving average representation of order at most q. (Note that the requirement of nonnegative spectral density for the individual series $z_1(t)$ is satisfied since the spectral density matrix for the multivariate process is nonnegative definite).

5. Specification of multivariate time series models

The objective of this section is to discuss the methodology of model building for multivariate time series. It was suggested by Box and Jenkins [3] that quantitative model building should incorporate the following steps:

(1) entertaining a class of models broad enough to cover a variety of situations;

(2) selection of specific member(s) of the entertained class for further analysis (specification or identification);

(3) estimation of unknown parameters of the chosen model(s) (estimation);

(4) Checking the validity of the entertained model(s) and possible respecification and reestimation (diagnostic checking).

In the pure time series approach each of the above steps depends on the data. For example, the choice of the selected model in step (2) depends on the information coming from the data or appropriate functions of the data.

System analysts might want to argue that there is usually a considerable theory about the data generation mechanism and that the assumption of a-priori ignorance is rarely true. However, one frequently encounters model building problems about which theory has little or nothing to say (for example the specification of error terms). In such instances when ignorance about the underlying theoretical mechanism is admitted, time series methods play an important role.

It has been pointed out [3] that autocorrelation and partial autocorrelation function provide useful tools to tentatively specify (identify) univariate ARIMA models (i.e. deciding the order of moving average, autoregressive and simplifying operators.

Autocovariance function

If we consider multivariate sequences (for example previously standardized and transformed streamflow sequences of various sites in a watershed basin) we observe matrices of autocovariances as defined in (2.3). Estimates of lag k autocovariance and autocorrelation matrices $\Gamma(k)$ and $P(k)$ are given by

$$C(k) = \{c_{ij}(k)\} \qquad c_{ij}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (z_i(t) - \bar{z}_i)(z_j(t+k) - \bar{z}_j) \tag{5.1}$$

where $\bar{z}_i$ and $\bar{z}_j$ are the sample means of the observed sequences. Division by n (compared to n-k) will lead to a biased estimator of $\gamma_{ij}(k)$ which only becomes unbiased as n tends to infinity. As with the estimation of autocovariances it can be argued that the divisor n is preferable to n-k since the estimator has smaller mean square error.

Furthermore

$$R(k) = \{r_{ij}(k)\} \qquad r_{ij}(k) = \frac{c_{ij}(k)}{\{c_{ii}(0)c_{jj}(0)\}^{\frac{1}{2}}} \tag{5.2}$$

Partial autocovariance function

The usefulness of the partial autocorrelation function in deciding the order of stochastic processes has already been pointed out, for example, by Box and Jenkins [3]. The partial autocorrelation function is a device which exploits the fact that whereas the AR(p) process has an autocorrelation function which is infinite in extent, the partial autocorrelations will be zero after lag p.

For multivariate stochastic processes the concept of partial autocovariance matrices can be used in the specification stage of model building. Useful statistics and their distribution were first described by Quenouille [16] and later extended by Hannan [7].

A stochastic process with autocovariance generating function $\Gamma(B)$ can be defined in two different ways depending on the deriction of time it is aimed at. For example, the multivariate AR(p) process with autocovariance function $\Gamma(B)$ has equivalently

(i) a forward representation

$$\underset{\sim}{z}(t) - \Phi_1\underset{\sim}{z}(t-1) -\ldots- \Phi_p\underset{\sim}{z}(t-p) = \underset{\sim}{b}(t) \tag{5.3}$$

with

$$\Gamma(k)' - \Phi_1\Gamma(k-1)'-\ldots-\Phi_p\Gamma(k-p)' = 0 \qquad k \geq 1$$

and

$$E\underset{\sim}{b}(t)\underset{\sim}{b}(t)' = G = \Gamma(0) - \Phi_1\Gamma(1) -\ldots- \Phi_p\Gamma(p)$$

(ii) a backward representation

$$\underset{\sim}{z}(t) - \Phi_1^*\underset{\sim}{z}(t+1) -\ldots- \Phi_p^*\underset{\sim}{z}(t+p) = \underset{\sim}{b}^*(t) \tag{5.4}$$

with

$$\Gamma(k) - \Phi_1^*\Gamma(k-1) -\ldots- \Phi_p^*\Gamma(k-p) = 0 \qquad k \geq 1$$

and

$$E\underset{\sim}{b}^*(t)\underset{\sim}{b}^*(t)' = G^* = \Gamma(0) - \Gamma(1)\Phi_1^{*\prime} -\ldots- \Gamma(p)\Phi_p^{*\prime} \quad .$$

In the univariate case the parameters ϕ and ϕ^* coincide.
It is easily shown that for the AR(p) process

$$E\underset{\sim}{b}(t+k)\underset{\sim}{b}^*(t)' = 0 \qquad \text{for } k > p \quad .$$

$E\underset{\sim}{b}(t+k)\underset{\sim}{b}^*(t)'$ is the partial covariance matrix of lag k. It is the covariance matrix of $\underset{\sim}{z}(t+k)$ and $\underset{\sim}{z}(t)$ after removing the effect of the intermediate $\underset{\sim}{z}(t+k-1),\ldots,\underset{\sim}{z}(t+1)$; more mathematically speaking, it is the covariance matrix of the projection errors of $\underset{\sim}{z}(t+k)$ and $\underset{\sim}{z}(t)$ after projecting it on the linear subspace generated by $\{\underset{\sim}{z}(t+k-1),\ldots,\underset{\sim}{z}(t+1)\}$.

In large samples the projections are derived by substituting the empirical autocovariance matrices C(j) into the first p equations of the forward and backward representations and solving

$$\begin{bmatrix} C(0) & C(1)' & \dots & C(p-1)' \\ C(1) & C(0) & \dots & C(p-2)' \\ \vdots & & & \\ C(p-1) & C(p-2) & \dots & C(0) \end{bmatrix} \begin{bmatrix} \hat{\Phi}_1' \\ \hat{\Phi}_2' \\ \vdots \\ \hat{\Phi}_p' \end{bmatrix} = \begin{bmatrix} C(1) \\ C(2) \\ \vdots \\ C(p) \end{bmatrix} \tag{5.5}$$

and

$$\begin{bmatrix} C(0) & C(1) & \dots & C(p-1) \\ C(1)' & C(0) & \dots & C(p-2) \\ \vdots & & & \\ C(p-1)' & C(p-2)' & \dots & C(0) \end{bmatrix} \begin{bmatrix} \hat{\Phi}_1^{*\prime} \\ \hat{\Phi}_2^{*\prime} \\ \vdots \\ \hat{\Phi}_p^{*\prime} \end{bmatrix} = \begin{bmatrix} C(1)' \\ C(2)' \\ \vdots \\ C(p)' \end{bmatrix} . \tag{5.6}$$

$$\hat{\underset{\sim}{b}}(t) = \underset{\sim}{z}(t) - \hat{\Phi}_1 \underset{\sim}{z}(t-1) - \dots - \hat{\Phi}_p \underset{\sim}{z}(t-p) \tag{5.7}$$

$$\hat{\underset{\sim}{b}}^*(t) = \underset{\sim}{z}(t) - \hat{\Phi}_1^* \underset{\sim}{z}(t+1) - \dots - \hat{\Phi}_p^* \underset{\sim}{z}(t+p) \tag{5.8}$$

$$\hat{G} = C(0) - \hat{\Phi}_1 C(1) - \dots - \hat{\Phi}_p C(p) \tag{5.9}$$

$$\hat{G}^* = C(0) - C(1)\hat{\Phi}_1^{*\prime} - \dots - C(p)\hat{\Phi}_p^{*\prime} . \tag{5.10}$$

Ignoring end effects (assuming large n) it can be shown that

$$\frac{1}{n} \sum_t \hat{\underset{\sim}{b}}(t+p+1)\hat{\underset{\sim}{b}}^*(t)' =$$

$$= \frac{1}{n} \sum_t [\underset{\sim}{z}(t+p+1) - \hat{\Phi}_1 \underset{\sim}{z}(t+p) - \dots - \hat{\Phi}_p \underset{\sim}{z}(t+1)] [\underset{\sim}{z}(t) - \hat{\Phi}_1^* \underset{\sim}{z}(t+1) - \dots - \hat{\Phi}_p^* \underset{\sim}{z}(t+p)]'$$

$$= C(p+1)' - [\hat{\Phi}_1 \dots \hat{\Phi}_p] \begin{bmatrix} C(p-1)' & \dots & C(1)' & C(0) \\ C(p-2)' & \dots & C(0) & C(1) \\ \vdots & & & \\ C(0) & \dots & C(p-2) & C(p-1) \end{bmatrix} \begin{bmatrix} \hat{\Phi}_1^{*\prime} \\ \vdots \\ \hat{\Phi}_p^{*\prime} \end{bmatrix}$$

$$= C(p+1)' - \sum_{i=1}^{p} \hat{\Phi}_i C(p+1-i)' \quad \text{by using the result in (5.6).} \tag{5.11}$$

Hannan [7; page 398] proves the result that if the $\underset{\sim}{z}(t)$ are generated by an AR(p) process the s^2 elements of

$$\sqrt{n}\ V(p+1) = \sqrt{n}\ \hat{G}^{-\frac{1}{2}}(C(p+1)' - \sum_{i=1}^{p} \hat{\Phi}_i C(p+1-i)')\hat{G}^{*-\frac{1}{2}} \qquad (5.12)$$

have an asymptotic joint Normal distribution with meanvector $\underset{\sim}{0}$ and covariance matrix I, where I is the $[s^2 \times s^2]$ identity matrix.

It thus follows that

$$n\ tr\ [V(p+1)V(p+1)'] \qquad (5.13)$$

is asymptotically χ^2 with s^2 degrees of freedom.

Computationally the derivation of (5.13) is simplified by the following result:

Theorem:

$$tr[V(p+1)V(p+1)'] = tr[\hat{\Phi}^{*'}_{p+1,p+1}\hat{\Phi}_{p+1,p+1}] \qquad (5.14)$$

where $\hat{\Phi}_{p+1,p+1}$ and $\hat{\Phi}^{*}_{p+1,p+1}$ are the estimates of the last parameter in a $(p+1)^{st}$ order forward (backward) autoregressive process. $\hat{\Phi}^{*}_{p+1,p+1}$ is the solution of

$$\begin{bmatrix} C(0) & C(1) & \dots & C(p) \\ C(1)' & C(0) & \dots & C(p-1) \\ \vdots & & & \\ C(p)' & C(p-1)' & \dots & C(0) \end{bmatrix} \begin{bmatrix} \hat{\Phi}^{*'}_{p+1,1} \\ \hat{\Phi}^{*'}_{p+1,2} \\ \vdots \\ \hat{\Phi}^{*'}_{p+1,p+1} \end{bmatrix} = \begin{bmatrix} C(1)' \\ C(2)' \\ \vdots \\ C(p+1)' \end{bmatrix} \qquad (5.15)$$

and $\hat{\Phi}_{p+1,p+1}$ is the solution of

$$\begin{bmatrix} C(0) & C(1)' & \dots & C(p)' \\ C(1) & C(0) & \dots & C(p-1)' \\ \vdots & & & \\ C(p) & C(p-1) & \dots & C(0) \end{bmatrix} \begin{bmatrix} \Phi'_{p+1,1} \\ \Phi'_{p+1,2} \\ \vdots \\ \Phi'_{p+1,p+1} \end{bmatrix} = \begin{bmatrix} C(1) \\ C(2) \\ \vdots \\ C(p+1) \end{bmatrix} \qquad (5.16)$$

The proof of this equivalence is given in the Appendix.

Estimated autocovariance and partial autocovariance matrices can thus provide more insight into the type of model worth considering. Especially the result in (5.13) and (5.14) should be important for applications in hydrology. The multivariate AR(1) model is frequently considered for standardized streamflow sequences at several sites. The test in (5.13) could indicate whether the AR(1) process is in reality sufficient or whether more complicated models of higher order should be considered.

Specifying the model for multivariate time series is an extremely difficult task and no simple solutions exist. Various approaches have been put forward how to specify multivariate ARMA models. Parzen [15] points out that for premathematical statistical investigations of the specification of models to be fitted it may be essential to first model each component separately (prewhiten the individual series). A similar strategy is adopted by Haugh [9] suggesting a two-stage identification procedure. The basic idea involved is to identify the relationship between the series by first characterizing separately each of their univariate models and secondly modelling the relationship between the two residual series driving each univariate model. The task at the second stage is made more tractable by the fact that one is cross-correlating two white noise series and hence the sample cross correlation function is easier to interpret. A similar approach is adopted by Granger and Newbold [6] and Jenkins [11].

A somewhat different approach is stressed by Zellner and Palm [21] and Wallis [19]. They use the fact that ARMA models in (3.1) can be written as

$$|\Phi(B)|\underset{\sim}{z}(t) = \Phi^{A}(B)\Theta(B)\underset{\sim}{b}(t) \qquad (5.17)$$

where $\Phi^{A}(B)$ is the adjoint of $\Phi(B)$. This representation implies that the autoregressive operator is the same for every series.

An extensive treatment in the literature is given in the case when there is no feedback present (series $z_1(t)$ is influenced by $z_2(t)$, but in turn doesn't influence the latter). Box and

Jenkins [3] give identification procedure for this special case. They first prewhiten the input series $z_2(t)$ and apply the same prewhitening transformation to the output $z_1(t)$. This approach is different (compared to the procedures in the feedback situation) since the same prewhitening transformation is applied to both series.

6. Estimation of parameters in multivariate time series models

After identifying (specifying) the underlying ARMA model (determining the order of autoregressive, moving average and simplifying operators), one has to estimate the parameters. From now on we will assume that the distribution of the shocks $\underset{\sim}{b}(t)$ has a multivariate Normal distribution. This, however, does not appear to be a restrictive assumption since transformations can be used to achieve Normality. For example, the log transformation will make the skewed distribution of standardized run off sequences approximately Normal. One can even go one step further and estimate the transformation from the data. The class of power transformations as considered by Box and Cox [2] represents a particularly useful parameterization and we showed in a previous paper [12] how this methodology can be used for hydrologic sequences.

Hannan [7] gives an extensive treatment of inference in the frequency domain. He gives methods for estimating pure autoregressive and pure moving average models in the multidimensional case. In theory this method can be extended to cover mixed autoregressive moving average models; computational complications, however, become extensive.

In this paper we consider estimation in the time domain only. Before we discuss maximum likelihood procedures for estimating the parameters in model (3.1), we elaborate shortly on an estimation procedure suggested in the hydrologic literature.

O'Connel [14] derives estimates for special cases of multivariate ARMA models by solving low lag autocovariance equations such as (3.15) in terms of the parameters of the process. For models with moving average parameters the so derived estimates

are asymptotically unbiased; however they are inefficient and more efficient estimators, such as maximum likelihood estimators, can be found.

To illustrate the above assertion we consider the special case of a first order moving average process in s = 1 dimensions.

$$\text{MA(1)} \qquad z(t) = (1 - \theta B)b(t) \tag{6.1}$$

where b(t) is a Normally distributed white noise sequence with $E[b(t)] = 0$ and $E[b(t)^2] = \sigma^2$.

It is easily shown that the autocorrelation function is given by

$$\rho_1 = \frac{-\theta}{1 + \theta^2} \qquad \text{and} \qquad \rho_k = 0 \quad \text{for} \quad k > 1 \tag{6.2}$$

The solution of the above equation satisfying the invertibility condition (zeros of $1 - \theta B$ lie outside the unit circle) is given by

$$\theta = \frac{-2\rho_1}{1 + \sqrt{1-4\rho_1^2}} \quad . \tag{6.3}$$

Using the estimated lag one autocorrelation r_1 in this equation, the estimate of θ in terms of the estimated autocorrelation is given by

$$\tilde{\theta} = \frac{-2r_1}{1 + \sqrt{1-4r_1^2}} \quad . \tag{6.4}$$

It is shown in Hannan ([7], page 373) that $\tilde{\theta}$ is asymptotically unbiased, and that the asymptotic variance of $\tilde{\theta}$ is given by

$$\sigma_a^2(\tilde{\theta};\theta) = n^{-1}\left[\frac{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8}{(1 - \theta^2)^2}\right] \quad . \tag{6.5}$$

On the other hand, Box and Jenkins [3] derive the maximum likeli-

hood estimator $\hat{\theta}$ and show that the asymptotic variance is given by

$$\sigma_a^2(\hat{\theta};\theta) = n^{-1}(1 - \theta^2) \quad . \tag{6.6}$$

The asymptotic efficiency of $\tilde{\theta}$ relative to $\hat{\theta}$ is given by the ratio of the two asymptotic variances and is given by

$$Ef_a(\tilde{\theta},\hat{\theta};\theta) = \frac{(1 - \theta^2)^3}{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8} \tag{6.7}$$

Unless $|\theta|$ is quite small the efficiency of $\tilde{\theta}$ is unacceptably low. The table below gives the asymptotic efficiency for several selected values of θ.

$\lvert\theta\rvert$	Asymptotic efficiency of $\tilde{\theta}$ relative to $\hat{\theta}$
.75	.03
.50	.28
.25	.76
.10	.96

But low efficiency of these estimates directly derived as functions of the autocovariances (autocorrelations) are not the only disadvantage of this estimation procedure. For multivariate autoregressive schemes with correlated residuals (ARMA) the solutions of autoregressive and moving average parameters in terms of the covariance matrices become very complicated, as it can, for example, be seen from equations (3.15).

A convenient procedure to derive maximum likelihood estimates for the case of Normal errors thus has to be found. Wilson [20] presents a practical iterative method for estimating parameters in mixed autoregressive moving average models. This method is a generalization of the iterative estimation procedure suggested by Box and Jenkins [3] for the univariate case.

The multivariate ARMA model satisfying stationarity and

invertibility conditions (roots of $\det\Phi(B) = 0$ and $\det\Theta(B) = 0$ outside the unit circle) is given by

$$(I - \Phi_1 B - \ldots - \Phi_p B^p)\underset{\sim}{z}(t) = (I - \Theta_1 B - \ldots - \Theta_q B^q)\underset{\sim}{b}(t) \quad . \quad (6.8)$$

The unknown parameters $\Phi_1, \Phi_2, \ldots, \Phi_p, \Theta_1, \ldots, \Theta_q$, which are for convenience arranged in a columnvector $\underset{\sim}{\beta}$, and the elements of the covariance matrix of the white noise sequence $\underset{\sim}{b}(t)$, G, are to be estimated from the available data $\underset{\sim}{z}(1), \underset{\sim}{z}(2), \ldots, \underset{\sim}{z}(n)$.

Assuming joint Normality for $\underset{\sim}{b}(t)$, and neglecting the effect of starting values, the likelihood of the parameters $\underset{\sim}{\beta}$ and G is given by

$$\ell(\underset{\sim}{\beta}, G | \underset{\sim}{z}(1), \ldots, \underset{\sim}{z}(n)) \propto |G|^{-\frac{n}{2}} \exp\{-\frac{1}{2} \sum_{t=1}^{n} \underset{\sim}{b}(t)' G^{-1} \underset{\sim}{b}(t)\} \quad , \quad (6.9)$$

where $\underset{\sim}{b}(t)$ is a function of the parameter vector $\underset{\sim}{\beta}$

$$\underset{\sim}{b}(t) = \underset{\sim}{z}(t) - \Phi_1 \underset{\sim}{z}(t-1) - \ldots - \Phi_p \underset{\sim}{z}(t-p) + \Theta_1 \underset{\sim}{b}(t-1) + \ldots + \Theta_q \underset{\sim}{b}(t-q) \quad . \quad (6.10)$$

The log likelihood function is given by

$$L(\underset{\sim}{\beta}, G | \underset{\sim}{z}(1), \ldots, \underset{\sim}{z}(n)) \propto -\frac{n}{2}\left[\log |G| + \frac{1}{n} \sum_{t=1}^{n} \underset{\sim}{b}(t)' G^{-1} \underset{\sim}{b}(t)\right] \quad . \quad (6.11)$$

The objective function to be minimized with respect to the elements in $\underset{\sim}{\beta}$ and G is thus given by

$$F(\underset{\sim}{\beta}, G) = \log |G| + \frac{1}{n} \sum_{t=1}^{n} \underset{\sim}{b}(t)' G^{-1} \underset{\sim}{b}(t) \quad . \quad (6.12)$$

This objective function is motivated by the Normality assumption. It may, however, also be used to derive (generalized) least squares estimates, when the assumption of Normality is not valid.

Conditional estimation of G:

The derivative of $F(\underset{\sim}{\beta}, G)$ with respect to elements of $G^{-1} = \{g^{ij}\}$ is proportional to

$$-g_{ij} + \frac{1}{n}\sum_{t=1}^{n} b_i(t)b_j(t) \tag{6.13}$$

since

i) $$\frac{d \log |G|}{dg^{ij}} = -\frac{d \log |G^{-1}|}{dg^{ij}} = -\frac{1}{|G^{-1}|}\frac{d|G^{-1}|}{dg^{ij}} = \begin{cases} -2g_{ij} & i \neq j \\ -g_{ij} & i = j \end{cases}$$

according to Theorem 7 in Appendix 1 of Anderson [1]

ii) $$\frac{d\sum_{t=1}^{n} \underset{\sim}{b}(t)'G^{-1}\underset{\sim}{b}(t)}{dg^{ij}} = \sum_{t=1}^{n} \operatorname{tr}\left(\frac{dG^{-1}}{dg^{ij}}\underset{\sim}{b}(t)\underset{\sim}{b}(t)'\right) = \begin{cases} 2\sum_{t=1}^{n} b_i(t)b_j(t) & i \neq j \\ \sum_{t=1}^{n} b_i(t)b_j(t) & i = j \end{cases}$$

Thus for given values of $\underset{\sim}{\beta}$ the estimate of the elements of $G = \{g_{ij}\}$ is given by

$$g_{ij} = \frac{1}{n}\sum_{t=1}^{n} b_i(t)b_j(t) \quad . \tag{6.14}$$

Conditional estimation of $\underset{\sim}{\beta}$*:*

In order to derive the conditional estimate of $\underset{\sim}{\beta}$ given the value of G one has to minimize the second factor in (6.12) or equivalently

$$\sum_{t=1}^{n} \underset{\sim}{b}(t)'G^{-1}\underset{\sim}{b}(t) = \sum_{t=1}^{n} \underset{\sim}{h}(t)'\underset{\sim}{h}(t) = \sum_{t=1}^{n}\sum_{j=1}^{s} h_j(t)^2 \tag{6.15}$$

where $\underset{\sim}{h}(t) = \underset{\sim}{b}(t)'P$

and $PP' = G^{-1}$ or $P'G^{-1}P = I$.

It is easily seen from Corollary 4 in Appendix 1 of Anderson [1], for example, that

$$P = HD^{-\frac{1}{2}}$$

where H is the matrix of normalized characteristic vectors of G^{-1} and D is the diagonal matrix with the corresponding characteristic roots in the diagonal.

Nonlinear regression (optimization) routines can be used to derive the estimates in $\underset{\sim}{\beta}$ such that (6.15) becomes minimal. An introductory exposition of nonlinear regression methods is given in Draper and Smith [5].

Simultaneous estimation of $\underset{\sim}{\beta}$ *and G*

The strategy to estimate the parameters $\underset{\sim}{\beta}$ and G is to apply the conditional estimation schemes alternately

$$\begin{cases} G_n = G(\underset{\sim}{\beta}_n) \\ \underset{\sim}{\beta}_{n+1} = \underset{\sim}{\beta}(G_n) \end{cases} \qquad n = 0,1,2,... \qquad (6.16)$$

Since each of the steps is a conditional minimization, the above parameter estimates must converge to the overall minimum $\hat{\underset{\sim}{\beta}}, \hat{G}$.

Wilson also investigates the distribution of the parameter estimates. The estimates $\hat{\underset{\sim}{\beta}}$ and $\hat{G}$ derived by minimizing (6.12) are consistent and asymptotically uncorrelated. The asymptotic distribution of $\hat{\underset{\sim}{\beta}}$ is Normal.

The following comments about this estimation procedure come to mind:

(1) The maximum likelihood procedure, as described above, is conditional on starting values $\underset{\sim}{z}(0),\underset{\sim}{z}(1),...,\underset{\sim}{z}(-p+1)$ and $\underset{\sim}{b}(0),\underset{\sim}{b}(1),...,\underset{\sim}{b}(-q+1)$. In practice the starting values for the shocks are set equal to their expectation, which is zero, and $\underset{\sim}{z}(0),...,\underset{\sim}{z}(-p+1)$ are the first p observation vectors of the series.

The assumption of fixed starting values is in most cases reasonable, since for invertible models the contribution of the starting values will be of negligible importance (see equations (2.8)). Some care, however, has to

be given to the case when the moving average parameters approach the non invertibility region (roots of det $\Theta(B) = 0$ approach the unit circle).

Hillmer [10] illustrates that for multivariate time series one can sometimes expect roots of det $\Theta(B) = 0$ on or near the unit circle. This can occur in seasonal data where the seasonality is nearly deterministic or in situations when there are only a few nonstationary components which affect all of the series. He illustrates that in such cases the traditional methods of estimation which assume that the starting value contribution to the likelihood is negligible do not perform adequately. In these situations Hillmer suggests estimates based on the exact likelihood (unconditional on starting values).

(2) Implementation of this estimation procedure requires a nonlinear regression routine and matrix routines to find eigenvalues and eigenvectors of symmetric positive definite matrices.

(3) In the context of iterative non linear regression routines it is easy to incorporate restrictions on the parameter vector β (such as setting certain elements equal to zero a priori).

7. Diagnostic checking

After specification of the model and estimation of its parameters, diagnostic checks have to be applied to see whether there is serious inadequacy of the model. Diagnostic checks must be such that they place the model in jeopardy, thus being sensitive to discrepancies which are likely to occur. Examples of such model inadequacies are misspecifications resulting in not preserving the autocorrelation structure, missing transformations of the data, time varying parameters, etc.

(1) One useful method to check a model is to overfit (i.e. estimate the parameters of a slightly extended model than the one supposed to be true.) This procedure assumes that we know

about the direction in which the model is likely to be violated (for example not enough autoregressive or moving average terms in the model).

(2) Another useful type of diagnostic checks looks at the residuals (observed minus estimated values). If both, the model were correctly specified and its parameters exactly known, the shocks $\underset{\sim}{b}(t)$ would be independently distributed with mean zero and covariance matrix G. It is known then that the estimated autocorrelations $r_{ii}(1),\ldots,r_{ii}(K)$ of $b_i(t)$ are asymptotically independent and Normally distributed with mean zero and variance n^{-1} (where n is the number of observations) (see [3], page 290). Furthermore (see [9]) the same result holds for the estimated crosscorrelations $r_{ij}(1),\ldots,r_{ij}(K)$ between the independent series $b_i(t)$ and $b_j(t)$.

These facts could be used to assess the statistical significance of departures of the autocorrelations and crosscorrelations from zero and thus detect lack of fit. This can be achieved by plotting and comparing the correlations $r_{ij}(1),\ldots,r_{ij}(K)$ with confidence bands at $\pm 2n^{-\frac{1}{2}}$.

Another useful statistics, a portmanteau lack of fit test ([3],[11]), uses the property that the sum of K squared standard Normal deviates follows a χ^2 distribution with K degrees of freedom. Thus under the null hypothesis (no lack of fit)

$$n \sum_{k=1}^{K} r_{ij}^2(k) \quad \text{is} \quad \chi_K^2 \qquad (\text{for } 1 \leq i, j \leq s) \tag{7.1}$$

The above procedure is, however, not immediately applicable, since the parameters of the model, and thus $\underset{\sim}{b}(t)$, are not known. In practice we only observe parameter estimates and residuals $\underset{\sim}{b}(t)$. It thus has to be investigated how this affects the above procedure.

Box and Pierce [4] discuss the univariate case and show that the large sample variances for the autocorrelations of the (univariate) residuals b(t) can be less than n^{-1}, especially at low lags. They thus conclude that in such cases the use of $n^{-\frac{1}{2}}$ as

standard error for the estimated autocorrelations of the residuals would underestimate the statistical significance of apparent departures from zero and that $n^{-\frac{1}{2}}$ should be considered as an upper bound, especially at low lags.

General results on the effect of parameter estimation errors on the distribution of autocorrelations and crosscorrelations of the residuals are still lacking. Nevertheless, Haugh's [9] results on the crosscorrelations of two independent, individually prewhitened series indicate that the use of estimated parameters does not seem to significantly impair these diagnostic checking procedures.

8. Conclusion and Recommendations

After discussing multivariate time series methodology in a statistical context we want to summarize the important practical implications for the hydrologist who wants to use these methods on some actual data sets.

(1) It is pointed out that the multivariate Markovian streamflow generation model as discussed by Matalas [13] and the streamflow generation model proposed by O'Connell [14] are special cases of the class of multivariate autoregressive moving average models. It is suggested that members of this general class of models are considered candidates for multi-site streamflow generation models.

(2) Our objective is to derive models possessing maximum simplicity and the minimum number of parameters consonant with representational adequacy.

(3) The process of model building is concerned with relating the class of multivariate autoregressive moving average models to actual observed streamflow data and involves much more than data fitting. Instead of restricting ourselves a-priori to special cases of this class of models (such as AR(1) in Matalas'

approach), we let the data speak for themselves and develop identification procedures which are designed to suggest what particular kind of model might be worth considering. It is shown how autocovariances and partial autocovariances can help us in this decision.

(4) The specification process leads to a tentative formulation of the model; we then need to obtain efficient estimates of the parameters. Estimates of the parameters should have the property of consistency (i.e.: for large sample size they should converge to the true values); furthermore, the asymtotic variance (i.e. the variation of the estimate from the true value) should be small as small as possible. It can be shown that maximum likelihood estimates possess these properties. In particular it can be shown that if moving average terms are present, the maximum likelihood estimates have smaller asymptotic variance than estimates derived by solving low order autocovariance matrix identities as proposed in the hydrologic literature [14].

(5) It has to be emphasized that estimation (model fitting) is not the last step in any model building procedure. Diagnostic checks have to be applied to detect possible model inadequacy.

The difference of observed and fitted observations (residuals) gives insight whether and how the model ought to be changed.

APPENDIX

Proof of the Identity in Theorem of Section 5

$$\operatorname{tr}(V(p+1)V(p+1)') = \operatorname{tr}\left[\hat{G}^{-1}\left(C(p+1)' - \sum_{i=1}^{p}\hat{\Phi}_i C(p+1-i)'\right)\hat{G}^{*-1}\left(C(p+1)' - \sum_{i=1}^{p}\hat{\Phi}_i C(p+1-i)'\right)'\right]$$

In order to show the result we prove that

$$\hat{G}^{-1}\left(C(p+1)' - \sum_{i=1}^{p}\hat{\Phi}_i C(p+1-i)'\right) = \hat{\Phi}^{*\prime}_{p+1,p+1} \tag{A1}$$

From (5.14) and result 2.7 on page 29 of Rao [17] it follows that

$$\hat{\Phi}^{*\prime}_{p+1,p+1} = \left\{C(0) - \left[C(p)' \ldots C(1)'\right] C_*^{(p)^{-1}} \begin{bmatrix} C(p) \\ \vdots \\ C(1) \end{bmatrix}\right\}^{-1} \left\{C(p+1)' - \left[C(p)' \ldots C(1)'\right] C_*^{(p)^{-1}} \begin{bmatrix} C(1)' \\ \vdots \\ C(p)' \end{bmatrix}\right\} \tag{A2}$$

where

$$C_*^{(p)} = \begin{bmatrix} C(0)C(1) & \ldots & C(p-1) \\ C(1)'C(0) & \ldots & C(p-2) \\ & & \\ C(p-1)'C(p-2)' & \ldots & C(0) \end{bmatrix} \tag{A3}$$

Furthermore

$$C_*^{(p)} = FC^{(p)}F \tag{A4}$$

where

$$F = \underset{[ps \times ps]}{\begin{bmatrix} 0 & \ldots & 0I \\ 0 & \ldots & I0 \\ & & \\ I & \ldots & 00 \end{bmatrix}}$$

and

$$C^{(p)} = \begin{bmatrix} C(0)C(1)' & \dots & C(p-1)' \\ C(1)C(0) & \dots & C(p-2)' \\ \\ C(p-1)C(p-2) & \dots & C(0) \end{bmatrix}$$

It follows that

$$C_*^{(p)^{-1}} = FC^{(p)^{-1}}F \tag{A5}$$

and

$$C(0) - [C(p)' \dots C(1)']\, C_*^{(p)^{-1}} \begin{bmatrix} C(p) \\ \vdots \\ C(1) \end{bmatrix} = C(0) - [C(1)' \dots C(p)']C^{(p)^{-1}} \begin{bmatrix} C(1) \\ \vdots \\ C(p) \end{bmatrix}$$

$$= C(0) - [C(1)' \dots C(p)'] \begin{bmatrix} \hat{\Phi}_1' \\ \vdots \\ \hat{\Phi}_p' \end{bmatrix} = \hat{G} \tag{A6}$$

using the relations in (5.5) and (5.10).

Furthermore,

$$C(p+1)' - [C(p)' \dots C(1)']C_*^{(p)^{-1}} \begin{bmatrix} C(1)' \\ \vdots \\ C(p)' \end{bmatrix} = C(p+1)' - [C(1)' \dots C(p)']C^{(p)^{-1}} \begin{bmatrix} C(p)' \\ \vdots \\ C(1)' \end{bmatrix}$$

$$= C(p+1)' - \sum_{i=1}^{p} \hat{\Phi}_i C(p+1-i)' \tag{A7}$$

using the relation in (5.5).

Substituting (A6) and (A7) into (A2) shows relation (A1).

In the same way one can show that

$$\left(C(p+1)' - \sum_{i=1}^{p} \hat{\Phi}_i C(p+1-i)' \right) \hat{G}^{*-1} = \hat{\Phi}'_{p+1,p+1} \tag{A8}$$

thus showing the theorem (5.14).

q.e.d.

REFERENCES

[1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958.

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

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

[4] Box, G.E.P. and D.A. Pierce, Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models, *Journal American Stat. Assoc.*, 65, (1970), 1509-1526.

[5] Draper, N.R. and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.

[6] Granger, C.W.J. and P. Newbold, Identification of Two-Way Causal Systems, in M.D. Intriligator, ed., *Frontiers of Quantitative Economics III*, North-Holland, Amsterdam, 1976.

[7] Hannan, E.J., *Multiple Time Series*, Wiley, New York, 1970.

[8] Hannan, E.J., The Identification Problem for Multiple Equation Systems with Moving Average Errors, *Econometrica*, 39, (1971), 577-591.

[9] Haugh, L.D., *The Identification of Time Series Interrelationship with Special Reference to Dynamic Regression*, Ph.D. thesis, University of Wisconsin, Madison, 1972.

[10] Hillmer, S.C., *Time Series: Estimation, Smoothing and Seasonal Adjusting*, Ph.D. thesis, Department of Statistics, University of Wisconsin, Madison, 1976.

[11] Jenkins, G.M., *The Interaction Between the Muskrat and Mink Cycles in North America*, paper presented at The International Biometric Conference, Constanta, Roumania, Aug. 25-30, 1974.

[12] Ledolter, J., *ARIMA Models and Their Use in Modelling Hydrologic Sequences*, RM-76-69, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

[13] Matalas, N.C., Mathematical Assessment of Synthetic Hydrology, *Water Resources Research*, 3, (1967], 937-945.

[14] O'Connell, E.P., *Stochastic Modelling of Long-Term Persistence in Stream Flow Sequences*, Ph.D. thesis, University of London, London, 1974.

[15] Parzen, E., Multiple Time Series Modelling, in P. Krishnaiah, ed., *Multivariate Analysis II*, Academic Press, New York, 1969.

[16] Quenouille, M.H., *The Analysis of Multiple Time Series*, Griffin's Statistical Monographs and Courses, Hafner, New York, 1957.

[17] Rao, C.R., *Linear Statistical Inference and Its Applications*, Wiley, New York, 1965.

[18] Rozanov, Y.A., *Stationary Random Processes*, Holden-Day, San Francisco, 1967.

[19] Wallis, K.F., *Seasonal Adjustment and Multiple Time Series Analysis*, paper presented to the European Meeting of the Econometric Society, Helsinki, August, 1976.

[20] Wilson, G., The Estimation of Parameters in Multivariate Time Series Models, *Journal of Royal Stat. Soc., Series B*, 35, (1973), 76-85.

[21] Zellner, A. and F. Palm, Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics*, 2, (1974), 17-54.