PART IX

MULTIPLE INPUT-MULTIPLE OUTPUT MODELS

As explained in Parts VII and VIII, in many natural systems, a single output or response variable is caused by one or more input or covariate variables. For example, riverflows are caused by physical variables which include precipitation and temperature. To model a system for which one or more variables cause another but not vice versa, the **TFN model** of Part VII can be employed. When one or more external interventions have modified the behaviour of the output series, the **intervention model** of Chapter 19 and Section 22.4 can be used. The intervention model is, in fact, a special type of TFN model.

When there is **feedback** in a system for which one variable causes another and vice versa, one must use a **multivariate model** to describe this situation. According to the definition used in the statistical literature, a multivariate time series model that is designed for handling feedback contains both multiple input and multiple output series. Although feedback is not as common as one way causality in hydrological systems, feedback can sometimes occur. A large lake, for instance, may affect local climatic conditions and thereby create precipitation which in turn increases the water level of the lake. In socio-economic systems, the phenomenon of feedback is very common. For example, unemployment may cause inflation which in turn increases unemployment. A detailed discussion of how to statistically detect various kinds of causality, including feedback, using the **residual** CCF (cross-correlation function) is presented in Section 16.2.2.

To formally model a system that contains feedback, a multivariate model must be used. In particular, Part IX of the book focuses upon the **general multivariate ARMA model** and simplifications thereof. Qualitatively, a general multivariate ARMA model can be written as

Multiple Outputs = Multiple Inputs + Multiple Noise

When interventions affect one or more of the series, the multivariate model can be easily extended to handle that situation. Furthermore, because the multivariate model describes how series influence one another over time, it is a **dynamic model**.

In Chapter 20, the general multivariate ARMA model is defined and other kinds of multivariate models that have been used in hydrology and environmental engineering are described and compared. Because the general multivariate ARMA model contains a large number of parameters, it is too cumbersome and overly complex for modelling most water resources systems problems. Nevertheless, the **TFN** and **contemporaneous ARMA (CARMA) models** constitute **two important subsets** of the general multivariate ARMA model that are well designed for effectively modelling water resources systems. As demonstrated by extensive applications for the TFN models in Chapters 17 and 18, and for the intervention models in Chapter 19 and Section 22.4, these models have widespread applicability in the environmental sciences.

The **CARMA** model is designed for modelling situations where two or more series affect one another at the same time or simultaneously. Because of this, the model is called the contemporaneous ARMA or simply CARMA model. Although CARMA models are not used as often as TFN models in water resources, they are still indispensible for modelling many types of problems. For example, two riverflow series measured at sites in two different rivers neither of which is upstream from the other may be related contemporaneously because the two measuring sites fall within the same general climatic region. Rather than separately model each of the two time series using an ARMA model, a CARMA model can more efficiently model both series together within a single mathematical framework. In Chapter 21, the CARMA model is defined and flexible model building procedures are presented. Both **water quantity and quality applications** confirm the great utility of CARMA models in water resources and environmental engineering.

Figure IX.1 depicts the hierarchical relationships among the dynamic models described in the book. Additionally, the figure contains the locations in the book where the definitions, model construction tools and applications for these dynamic models can be found.

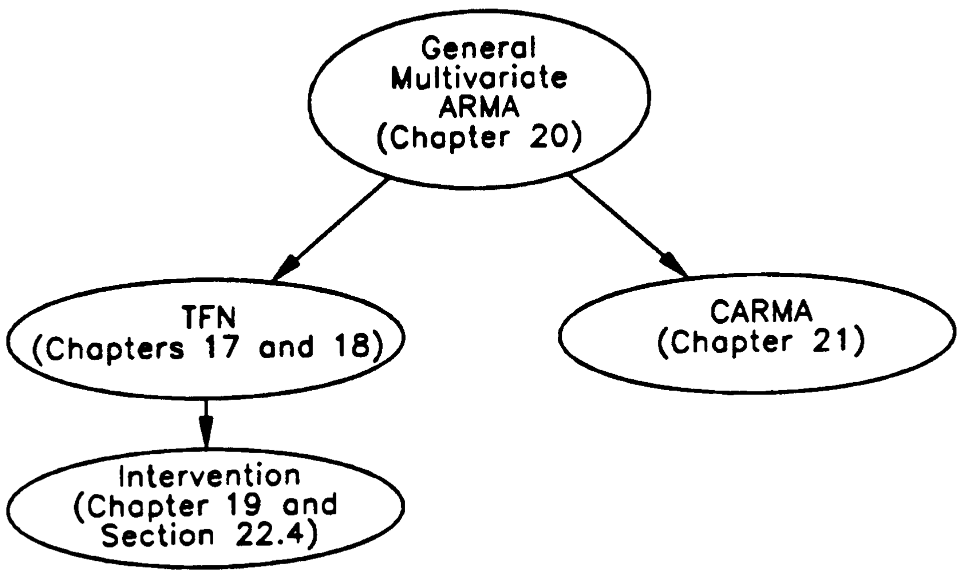

Figure IX.1. Hiearchy of dynamic models.

CHAPTER 20

GENERAL MULTIVARIATE AUTOREGRESSIVE MOVING AVERAGE MODELS

20.1 INTRODUCTION

The term *multivariate* possesses a number of related interpretations that are used commonly by both practitioners and researchers. For example, many people consider the word multivariate to indicate that multiple variables in a system have been measured and, consequently, a multivariate model is needed to model the system. Under this definition, the TFN models of Part VII would be classified as multivariate models because the model statistically describes how one or more input variables affect the behaviour of a single output variable. Likewise, any deterministic model or mixed deterministic-stochastic model that formally describes the relationships among at least two physical variables can be thought of as being a multivariate model.

Because this book deals mainly with stochastic or time series models, the statistical definition of multivariate models is utilized. In particular, as noted in the preface to Part IX, a *general multivariate ARMA model* is a model that statistically describes how multiple outputs are influenced by multiple inputs and multiple noise terms. According to this statistical definition, the TFN models of Part VII and the related intervention models of Chapter 19 and Section 22.4 are not multivariate models. As a matter of fact, since these models possess a single output variable they are statistically classified as being univariate models.

The many time series applications presented throughout this book firmly establish the fact that the scientific community clearly recognizes the importance of time series modelling in water resources and environmental engineering. Indeed, as the demand for water continues to increase and more and more of the natural environment is altered due to industrialization and other land use changes, greater emphasis will be placed upon using more flexible systems sciences methodologies to assist decision makers in water resources (see Sections 1.2 to 1.5). To better understand how man's activities affect the environment, extensive measurements will have to be taken of a wide range of water quality variables, riverflows and lake levels, meteorological phenomena, as well as many other kinds of variables. The resulting vast amounts of data will have to be stored, processed and transferred using extensive computer networks. This in turn means that the need for having comprehensive multivariate models for describing multiple time series will continue to expand. In fact, the foregoing scenario is what Prof. V. Yevjevich considers to be the *major challenge for hydrological research* (personal communication from Prof. V. Yevjevich during the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes held at Colorado State University, Fort Collins, Colorado, July 15 to 17, 1985). Due to the current and expanding importance of multivariate analysis in hydrology, the organizers of the Fourth International Hydrology Symposium selected the theme of their conference to be *Multivariate Analysis of Hydrologic Processes*. Within this text, Chapters 20 and 21 put multivariate modelling into proper perspective and present attractive kinds of stochastic multivariate models for use in practical applications.

Stochastic or time series models are not the only type of multivariate models that can be used in water resources for modelling more than one physical variable at the same time. A *conceptual model* constitutes a *deterministic model* that is specifically designed to mathematically simulate the physical processes involved in the hydrological cycle. When studying a given problem, a scientist should employ a type of multivariate model which he or she feels is most useful and realistic. In some cases, a scientist may utilize a physically based (i.e., conceptual) model which he thinks can explain certain deterministic aspects of a natural system. After removing the portion of the data which can be explained using a physical model, the scientist can then model what is left over using a stochastic model. The overall model is referred to as a *mixed deterministic-stochastic model*. Applications of stochastic, deterministic and mixed deterministic-stochastic models to hydrological systems are given in the Proceedings of the Fourth International Hydrology Symposium (Shen et al., 1986). Within the Proceedings, a keynote paper on stochastic research in multivariate analysis is presented by Hipel (1986). In a specially edited Monograph on *Time Series Analysis in Water Resources* (Hipel, 1985b), Salas et al. (1985) review and compare alternative approaches for modelling multiple water resources time series.

Conceptual models can possess a number of common problems. In particular, they are often very complex and have a large number of parameters related to physical phenomena, all of which must be calibrated (Tong et al., 1985). Furthermore, due to the great complexity of natural systems, the conceptual models are, like other models, only rough approximations to reality. As demonstrated by a case study in Section 18.3 and also by Thompstone et al. (1985a), a simple stochastic TFN model forecasts more accurately than a cumbersome conceptual model which is very expensive to maintain and calibrate.

Even though most time series models were not originally designed to reflect the behaviour of physical phenomena, a *physical basis* to these models can often be justified. For instance, as explained by Salas and Smith (1981) and also in Section 3.6, a particular conceptual model of a watershed leads to ARMA streamflows and ARMA groundwater storage. Further discussions regarding physically based models are give by Klemes (1978). Yevjevich and Harmancioglu (1985) stress the importance of linking stochastic models with physically consistent properties of any particular water resources time series.

Many time series analysis approaches to multivariate modelling fall within the general framework of multivariate ARMA models. Consequently, in the next section the general ARMA multivariate is defined, while in Section 20.3 model construction is discussed and modelling limitations are clearly pointed out. Subsequent to this, an historical overview of the development of multivariate ARMA time series modelling in water resources is presented. Part of this historical evolution leads to the conclusion that the contemporaneous ARMA (CARMA) and TFN models constitute the two subclasses of the general family of multivariate ARMA models that are suitable for use in practical applications. Accordingly, as shown in Figure IX.1, TFN, intervention, and CARMA models are studied in depth in this book in Chapters 17 and 18, Chapter 19 and Section 22.4, and Chapter 21, respectively. Following the historical summary, of the development of multivariate ARMA models, other families of multivariate models are discussed in Section 20.5. In Section 20.5.2, the ongoing debate regarding the relative usefulness and philosophical foundations of disaggregation and aggregation models is described. Additional classes of models referred to in Section 20.5 include nonGaussian, nonlinear, fractional differencing, frequency domain, pattern recognition, and nonparametric models, in Sections 20.5.3 to 20.5.8, respectively. In the conclusions, a wide variety of challenging problems are suggested for future research projects in multivariate time series modelling in water resources and environmental engineering.

20.2 DEFINITIONS OF MULTIVARIATE ARMA MODELS

20.2.1 Introduction

After defining the general family of multivariate ARMA models, the TFN and CARMA classes of models are defined as subsets of this general family. As described in Section 20.3.2, some rather cumbersome model construction techniques are available for use with general multivariate ARMA models. However, limitations on using general multivariate ARMA models in water resources applications are clearly pointed out in Section 20.3.1. To overcome these drawbacks, subfamilies of multivariate ARMA models are suggested for use in hydrology. In particular, the CARMA model described in detail in Chapter 21 is recommended for modelling multiple time series when the series are contemporaneously correlated with one another at a given time but not at lags other than zero. To model a single response series which is driven by one or more covariate series plus a noise component, the TFN family of models of Part VII can be used. As explained in Section 20.4, it is interesting to note how the development of multivariate modelling in water resources converged over a period of two decades to the conclusion that CARMA and TFN models are the most appropriate kinds of multivariate ARMA models to use in practical hydrological applications.

20.2.2 Definitions

General Multivariate ARMA model

Let a set of k time series be represented at time t by the vector

$$\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tk})^T$$

where the vector of the theoretical means for $\mathbf{Z}_t$ is given by $\mu = (\mu_1, \mu_2, \ldots, \mu_k)^T$ and the superscript T stands for the transpose of a vector. If the AR (autoregressive) order is p and the MA (moving average) order is q, the *general k-dimensional multivariate ARMA(p,q) model* can be conveniently written as

$$\begin{aligned}(\mathbf{Z}_t - \mu) - \Phi_1(\mathbf{Z}_{t-1} - \mu) - \Phi_2(\mathbf{Z}_{t-2} - \mu) - \cdots - \Phi_p(\mathbf{Z}_{t-p} - \mu) \\ = \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1} - \Theta_2\mathbf{a}_{t-2} - \cdots - \Theta_q\mathbf{a}_{t-q}\end{aligned} \qquad [20.2.1]$$

where

$$\Phi_i = \begin{bmatrix} \phi_{11i} & \phi_{12i} & \cdots & \phi_{1ki} \\ \phi_{21i} & \phi_{22i} & \cdots & \phi_{2ki} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \phi_{k1i} & \phi_{k2i} & \cdots & \phi_{kki} \end{bmatrix}$$

is the ith AR parameter matrix of order $k \times k$ for $i = 1, 2, \ldots, p$;

$$\Theta_i = \begin{bmatrix} \theta_{11i} & \theta_{12i} & \dots & \theta_{1ki} \\ \theta_{21i} & \theta_{22i} & \dots & \theta_{2ki} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{k1i} & \theta_{k2i} & \dots & \theta_{kki} \end{bmatrix}$$

is the ith MA parameter matrix of order $k \times k$ for $i = 1,2,\dots,q$; $\mathbf{a}_t = (a_{t1},a_{t2},\dots,a_{tk})^T$, is the k-dimensional vector of innovations for $\mathbf{Z}_t$ at time t. Because the vectors have the form $(\mathbf{Z}_t - \mu)$ in [20.2.1], the multivariate ARMA model is often referred to as a vector ARMA model. When man-induced or natural interventions affect one or more of the multiple series in [20.2.1], the model can be easily extended to handle multiple interventions (Abraham, 1980).

A more compact format for writing the model in [20.2.1] is given by

$$\Phi(B)(\mathbf{Z}_t - \mu) = \Theta(B)\mathbf{a}_t \qquad [20.2.2]$$

where B is the backward shift operator defined by $B^h\mathbf{Z}_t = \mathbf{Z}_{t-h}$, $\Phi(B) = \mathbf{I} - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p$ is the AR operator of order p where $\mathbf{I}$ is the identity matrix of order $k \times k$, and $\Theta(B) = \mathbf{I} - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q$ is the MA operator of order q.

There are a number of assumptions underlying the linear multivariate ARMA(p,q) model given in [20.2.1] or [20.2.2]. To start with, it is assumed that the innovations given by $\mathbf{a}_t$ are identically independently distributed (IID) vector random variables with a mean of zero and variance covariance matrix Δ. In order to obtain MLE's (maximum likelihood estimates) of the parameters and also design sensitive diagnostic checks, for practical applications it is necessary to invoke the normality assumption so that the innovations are normally independently distributed (NID) and hence $\mathbf{a}_t \approx \text{NID}(\mathbf{0},\Delta)$. Finally, to permit the model in [20.2.1] or [20.2.2] to be stationary and invertible, the zeroes of the determinant equations $|\Phi(B)| = 0$ and $|\Theta(B)| = 0$, respectively, must lie outside the unit complex circle.

Example: Consider a multivariate ARMA(1,1) model possessing two variables contained in the vector

$$\mathbf{Z}_t = (Z_{t1},Z_{t2})^T$$

having theoretical means given by $\mu = (\mu_1,\mu_2)^T$. Because there are two variables, the multivariate model used to describe mathematically the relationship between the two variables is called a bivariate model. Following [20.2.1], the bivariate ARMA(1,1) model is written as

$$(\mathbf{Z}_t - \mu) - \Phi_1(\mathbf{Z}_{t-1} - \mu) = \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1}$$

where

$$\Phi_1 = \begin{bmatrix} \phi_{111} & \phi_{121} \\ \phi_{211} & \phi_{221} \end{bmatrix}$$

is the AR parameter matrix;

$$\Theta_1 = \begin{bmatrix} \theta_{111} & \theta_{121} \\ \theta_{211} & \theta_{221} \end{bmatrix}$$

is the MA parameter matrix; $\mathbf{a}_t = (a_{t1}, a_{t2})$ is vector of innovations containing IID random variables. Substituting the AR and MA matrices into the bivariate ARMA model produces

$$\begin{bmatrix} Z_{t1} - \mu_1 \\ Z_{t2} - \mu_2 \end{bmatrix} - \begin{bmatrix} \phi_{111} & \phi_{121} \\ \phi_{211} & \phi_{221} \end{bmatrix} \begin{bmatrix} Z_{t-1,1} - \mu_1 \\ Z_{t-1,2} - \mu_2 \end{bmatrix} = \begin{bmatrix} a_{t1} \\ a_{t2} \end{bmatrix} - \begin{bmatrix} \theta_{111} & \theta_{121} \\ \theta_{211} & \theta_{221} \end{bmatrix} \begin{bmatrix} a_{t-1,1} \\ a_{t-1,2} \end{bmatrix}$$

After matrix multiplication, the two component equations of the bivariate model are

$$Z_{t1} - \mu_1 - \phi_{111}(Z_{t-1,1} - \mu_1) - \phi_{121}(Z_{t-1,2} - \mu_2) = a_{t1} - \theta_{111}a_{t-1,1} - \theta_{121}a_{t-1,2}$$

$$Z_{t2} - \mu_2 - \phi_{211}(Z_{t-1,1} - \mu_1) - \phi_{221}(Z_{t-1,2} - \mu_2) = a_{t2} - \theta_{211}a_{t-1,1} - \theta_{221}a_{t-1,2}$$

TFN Model

Because of the great importance of TFN models in water resources, these models are studied in depth in Chapters 17 and 18 while the closely related intervention models are entertained in Chapter 19 and Section 22.4. As noted earlier, the TFN model is a subset of the multivariate ARMA model in [20.2.1]. In particular, when the AR and MA parameter matrices in [20.2.1] are either all upper or else lower triangular, the model is called a TFN model. For the case where the matrices are lower triangular, the *TFN model* is defined following [20.2.1] as

$$\begin{aligned} (\mathbf{Z}_t - \mu) - \Phi_1(\mathbf{Z}_{t-1} - \mu) - \Phi_2(\mathbf{Z}_{t-2} - \mu) - \cdots - \Phi_p(\mathbf{Z}_{t-p} - \mu) \\ = \mathbf{a}_t - \Theta_1\mathbf{a}_{t-1} - \Theta_2\mathbf{a}_{t-2} - \cdots - \Theta_q\mathbf{a}_{t-q} \end{aligned} \qquad [20.2.3]$$

where the ith AR parameter matrix is

$$\Phi_i = \begin{bmatrix} \phi_{11i} & 0 & 0 & \cdots & 0 \\ \phi_{21i} & \phi_{22i} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \phi_{k1i} & \phi_{k2i} & \phi_{k3i} & \cdots & \phi_{kki} \end{bmatrix}$$

and the ith MA parameter matrix is written as

$$\Theta_i = \begin{bmatrix} \theta_{11i} & 0 & 0 & \cdots & 0 \\ \theta_{21i} & \theta_{22i} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \theta_{k1i} & \theta_{k2i} & \theta_{k3i} & \cdots & \theta_{kki} \end{bmatrix}$$

Salas et al. (1985) refer to the TFN model in [20.2.3] as a triangular relationship because Z_{t1} only depends on its own past, Z_{t2} depends on its own past plus the present and past of Z_{t1}, Z_{t3} is dependent on its own past and on the present and past of Z_{t1} and Z_{t2}, and so on. In the final component equation in [20.2.3], Z_{tk} depends on its own past and on the present and past of $Z_{t1}, Z_{t2}, \ldots, Z_{tk-1}$. This means that the single output Z_{tk} depends upon the input variables

$Z_{t1}, Z_{t2}, \ldots, Z_{tk-1}$. Recall from Section 17.5.2 that this definition constitutes, in fact, a TFN model. By appropriate algebraic manipulations, one can easily demonstrate that the TFN model in [20.2.3] is equivalent to the more convenient form for writing the TFN model given in [17.5.3]. Comprehensive model building procedures for constructing TFN models and numerous water resources applications are presented in Chapters 17 and 18. Because a single output variable is dependent upon multiple input variables, a TFN model can be statistically classified as a univariate model.

Camacho et al. (1986) provide a simple example to demonstrate when a TFN model would clearly be selected over a general multivariate ARMA model. More specifically, when dealing with unregulated multisite hydrological systems, it can be argued that the general multivariate ARMA model would never be required to model the data and that a TFN model with only upper or lower triangular parameters will always be appropriate. To illustrate this fact, consider for simplicity the three-station riverflow system shown in Figure 20.2.1. It is clear from the nature of the system that only flows located upstream of any given station will influence the flows at that station. Therefore, if the vector of flows at time t is $\mathbf{Z}_t = (Z_{t1}, Z_{t2}, Z_{t3})^T$, the parameter matrices of the model in [20.2.1] or [20.2.2] will contain only (possible) nonzero elements at entries (1,1), (2,2), (3,1), (3,2) and (3,3). No other entry in the matrix should be allowed to be different from zero. For example, if the (1,3) element of a matrix were permitted to be nonzero, it would imply that flows at Station 1, Z_{t1} would be written as a linear combination of past values of Z_{t1}, past values of Z_{t3} and some error terms. This, of course, would not have any physical meaning. It is easy to see that the resulting matrices of the model are lower triangular. The same argument can be extended to more complex systems.

The simple example presented above shows that when the physical restrictions of the system are taken into consideration in the formulation of the model, it is possible to substantially reduce the number of parameters. The benefits of such a reduction can be appreciated by looking at the precision of the parameters estimates. Suppose, for example, that the bivariate series $\mathbf{Z}_t = (Z_{t1}, Z_{t2})$ is modelled as a general multivariate ARMA(1,0) when in fact a CARMA (1,0) would suffice. It is shown by Camacho et al. (1985a) and also in Section 21.5 that the variances of the estimated parameters obtained using the CARMA model are always smaller than the ones obtained using the full multivariate model and that such reductions may be well over 50%.

CARMA Model

A *CARMA model* is obtained from [20.2.1] or [20.2.2] when all of the parameter matrices are diagonal. Consequently, for the AR and MA matrices given by $\mathbf{\Phi}_i$ and Θ_i, respectively, the elements $\phi_{jmi} = 0$ and $\theta_{jmi} = 0$ for $j \neq m$ and $i = 1, 2, \ldots, k$. In Chapter 21, the CARMA (p,q) model is written out in full in [21.2.1] as well as [21.2.4]. Because this parsimonious model implies a contemporaneous relationship among the concurrent multivariate observations or, equivalently, the multivariate innovations which occur at the same time t, it is referred to as a contemporaneous ARMA model. Furthermore, since a CARMA model contains multiple output series, it constitutes a multivariate model. Chapter 21 of this book is entirely devoted to this useful and interesting class of models. A comprehensive set of model construction techniques given in Section 21.3 allows CARMA models to be conveniently applied to practical problems and the water resources applications in Section 21.5 demonstrate the utility of these models.

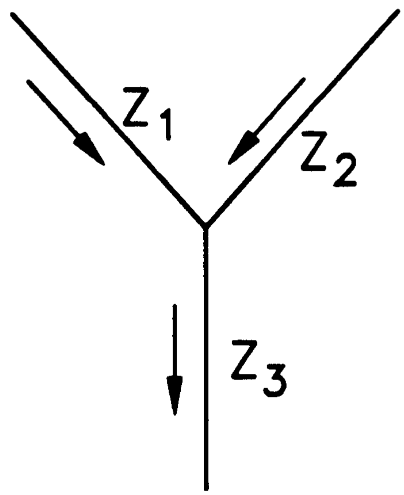

Figure 20.2.1. A three-station riverflow system where Z_i represents measurements at station i.

20.3 CONSTRUCTING GENERAL MULTIVARIATE ARMA MODELS

20.3.1 Limitations

As noted by authors such as Salas et al. (1985), Camacho et al. (1985a, 1986, 1987a,b,c) and Hipel (1986), there are two major drawbacks for using the general multivariate ARMA in [20.2.1] and [20.2.2] for applications in water resources. First of all, because the number of parameters increases exponentially with the dimensionality of the model, the multivariate ARMA model is very complicated and possesses too many parameters. Secondly, a comprehensive set of operational and simple model building techniques are not available for constructing multivariate ARMA models by following the identification, estimation and diagnostic check stages of model construction. As a result, one cannot assume the most general form of the multivariate ARMA model to begin with and employ construction techniques to identify an appropriate model to parsimoniously describe the data set under consideration.

To overcome the foregoing problems, CARMA and TFN models can be employed. Both of these subfamilies of models contain far fewer parameters than the cumbersome general multivariate ARMA model. Additionally, a wide range of flexible model building techniques are now readily available for use with each of these two classes of models. As described in detail in Chapters 17 and 19, comprehensive model building techniques are available for conveniently constructing TFN models and the closely related intervention models, respectively. In Chapter 21, flexible model construction methods are presented for building CARMA models.

As argued in this book as well as by authors such as Salas et al. (1980), Salas et al. (1985), Camacho et al. (1985, 1986, 1987a,b,c) and Hipel (1986), the physical properties of hydrological systems often dictate that TFN and CARMA models are the proper types of multivariate models to use in practice. Nonetheless, in some water resources applications which may, for example, require the use of socio-economic data, it may be necessary to employ a multivariate ARMA model that does not fall within the TFN or CARMA categories. Consequently, the purpose of this section and Appendix A20.1 is to outline some of the model construction methods that can be used for building general multivariate ARMA models. However, the reader should bear in mind that due to the complexity of the vector ARMA model, the model building methods are unwieldy and are not as flexible or comprehensive as those currently available for building TFN or CARMA models.

20.3.2 Model Construction

Introduction

As is the case for all of the families of models entertained in this book, constructing multivariate ARMA models is effected by adhering to the three iterative stages consisting of identification, estimation and diagnostic checking. Because the concept of causality described in Section 16.2 is closely linked to the identification of not only TFN models (Section 17.3.1) and CARMA models (Section 21.3), but also general multivariate ARMA models, this idea is briefly discussed next. Following the summary given by Camacho et al. (1986), a review of recent model building procedures is then given. When modelling a set of seasonal time series using a multivariate ARMA model, one of the approaches outlined in Section 20.3.3 may be used.

Causality

As explained in Section 16.2.1, Granger (1969) defined causality between two time series in terms of predictability. In particular, a variable X causes another variable, Y, with respect to a given universe or information set that included X and Y, if the present Y can be better predicted by using past values of X than by not doing so, all other relevant information (including the past of Y) being used in either case. Causality from Y to X can be defined in the same way. Feedback occurs when X causes Y and Y also causes X.

To determine the type of causality relationship that exists between X and Y, the properties of the residual CCF are examined. Following the more detailed explanation given in Section 16.2.2, let the sequences of the observations for two variables be represented by the time series X_t and Y_t, respectively. These series can be prewhitened by fitting ARMA models to the series and obtaining the white noise residuals u_t and v_t in [16.2.3] and [16.2.4] for X_t and Y_t, respectively. When required for rectifying problems with non-normality and/or heteroscedasticity in the ARMA model residuals, X_t or Y_t can be transfromed using the Box-Cox transformation in [3.4.30] prior to prewhitening. Subsequent to prewhitening, the residual CCF, written as $\rho_{uv}(k)$ at lag k between u_t and v_t can be considered by using [16.2.5]. In addition to reflecting the type of linear dependence between u and v and consequently between X and Y, $\rho_{uv}(k)$ gives the kind of causality relationship between these variables for linear systems. As explained by Pierce and Haugh (1977) and summarized in Table 16.2.1, there are many possible types of causal interactions between X and Y which can be characterized by the properties of $\rho_{uv}(k)$. If X and Y are independent, $\rho_{uv}(k) = 0$ for all k and, hence, it would not be appropriate to develop any kind of multivariate ARMA model to link these two variables. When there is unidirectional causality such that X causes Y and Y does not cause X, it can be proven that $\rho_{uv}(k) \neq 0$ for some $k > 0$, and $\rho_{uv}(k) = 0$ for either all $k < 0$ or else all $k \leq 0$. For this situation, the most appropriate kind of multivariate ARMA model to link the input or covariate series X_t with the output or response Y_t is a TFN model defined in [17.2.5]. If X and Y are only related instantaneously, $\rho_{uv}(0) \neq 0$ and $\rho_{uv}(k) = 0$ for all $k \neq 0$. When this is the case, a CARMA model in [21.2.1] or [21.2.4] can be used to mathematically describe the contemporaneous linear dependence between X_t and Y_t. Finally, if there is feedback and hence X causes Y and vice versa, $\rho_{uv}(k) \neq 0$ for some $k > 0$ and for some $k < 0$. Feedback between two or more variables can be modelled using the multivariate ARMA model in [20.2.1] or [20.2.2].

In practical applications, one examines the sample residual CCF in [16.2.6] for X_t and Y_t to identify the type of linear dependence between X and Y. Besides detecting causality between two series, the sample residual CCF can be used for identifying time series models to mathematically describe the dynamic linkage between X_t and Y_t. In addition to output from other identification techniques, this information can then be used for identifying the appropriate kind of multivariate model to fit to the series. The manner in which the sample residual CCF can be used to design TFN and CARMA models is explained in Sections 17.3.1 and 21.3, respectively. In the next three subsections, model construction methods are presented for building general multivariate ARMA models.

Identification

As shown by applications in Chapter 5, selecting the order of a simple univariate ARMA model can sometimes be challenging. For the multivariate ARMA case, model identification is far more difficult (Tiao and Tsay, 1983a; Tjostheim and Paulsen, 1982; Jenkins and Alavi, 1981; Tiao and Box, 1981). Different identification procedures have been advocated in the literature. For example, Tiao and Box (1981) and Jenkins and Alavi (1981) have extended the use of the sample CCF and the PACF to identify the order of the AR and MA operators of a multivariate process. Tiao and Tsay (1983a,b) have proposed the use of the extended sample cross correlation function (ESCCF) to identify the order of general multivariate ARMA(p,q) models. Newbold and Hotopp (1984) have expanded a two-step procedure given by Hannan and Rissanen (1982) to the multivariate case in order to identify and model. These procedures are based on the calculation of simple sample statistics and stepwise regressions. Another technique based on the estimation of heavily parameterized ARMA models and the use of a consistent information criterion (Quinn, 1980; Hannan and Quinn, 1979) has also been proposed. The difficulty of applying such procedures clearly lies in the high computational costs.

In Appendix A20.1, the following three identification methods are described:

1. sample CCF matrix,
2. sample PACF matrix,
3. ESCCF matrix.

Because the theoretical CCF matrix cuts off for pure multivariate MA processes, the sample CCF matrix can be used to identify when a multivariate MA model is required (Tiao and Box, 1981; Jenkins and Alavi, 1981). This is similar to the manner in which the sample ACF in [2.5.9] is employed for detecting when a pure MA univariate model is needed (see Section 5.3.4). The theoretical PACF matrix truncates for a pure multivariate AR process and therefore the sample PACF matrix can be used for finding out when a multivariate AR model is required to model a given data set (Tiao and Box, 1981; Jenkins and Alavi, 1981). This is similar to the way in which the sample PACF in Section 3.2.2 can be used for designing a pure univariate AR model (see Section 5.3.5). When both MA and AR parameters are needed in a multivariate ARMA model, the ESCCF matrix can be utilized for ascertaining the orders of the MA and AR parameter matrices (Tiao and Tsay, 1983a,b; Tsay and Tiao, 1984).

Estimation

The likelihood function of the general multivariate ARMA(p,q) model has been given by Nicholls and Hall (1979), Hillmer and Tiao (1979), and Wilson (1973). Conditional and exact likelihood estimators have been proposed, by these authors. It has been shown by Hillmer and Tiao (1979) that if the determinant of the moving average operator has one or more zeroes close to the unit circle, the exact likelihood should be employed. Algorithms to evaluate the likelihood function have been proposed by Hall and Nicholls (1980) and by Ansley and Kohn (1983), who discussed the use of the Kalman filter to incorporate the case of missing or aggregated data. Shea (1989) provided a computer program for calculating the exact likelihood of a multivariate ARMA(p,q) model. For estimating the parameters of TFN and CARMA models, algorithms which are more computationally efficient can be employed. When obtaining MLE's for the parameters of TFN and intervention models, the estimator given in appendix A17.1 can be utilized. For CARMA models, Camacho (1984) and Camacho et al. (1985a, 1987a,b) have developed a computationally efficient algorithm to estimate the parameters of the model. They have also extended the algorithm to include the case of CARMA models with unequal sample sizes. Their algorithm is described in Section 21.3.3.

Diagnostic Checking

As pointed out in Section 20.2.2 with equations [20.2.1] and [20.2.2], the innovations of the general multivariate ARMA model are assumed to be NID. Diagnostic checks should be executed to insure that the key residual assumptions are satisfied. In particular, to determine if the residuals are white one can employ the sample CCF and sample PACF described in Appendix A20.1. Alternatively, one can use the modified Portmanteau test of Li and McLeod (1981) for whiteness checks. To verify that the normality assumption is satisfied, one can utilize the multivariate normality tests proposed by Royston (1983). Finally, one could develop multivariate extensions of the constant variance tests of Section 7.5 to make sure that the residuals are not heteroscedastic.

If the residuals are not white, the multivariate model must be appropriately redesigned. When the residuals are not approximately normally distributed and/or homoscedastic, one may wish to transform one or more of the variables using an appropriate transformation such as the Box-Cox transformation in [3.4.30]. Following this, the parameters of the multivariate model can be estimated for the transformed series.

20.3.3 Seasonality

The general multivariate ARMA model given in [20.2.1] and [20.2.2] is defined for handling nonseasonal time series. Likewise, the model construction techniques of Section 20.3.2 are explained for the nonseasonal case. When a set of seasonal time series are to be modelled using a multivariate model, one of the two useful approaches described below can be used.

Deseasonalized Multivariate Model

A commonly used procedure for modelling seasonal data is to first deseasonalize each time series using the *deseasonalization* methods defined in [13.2.2] or [13.2.3]. As explained in Section 13.3.3, to decrease the number of parameters required for deseasonalization a Fourier series approach can be utilized. Subsequent to deseasonalization, the most appropriate type of nonseasonal multivariate model in [20.2.1] can be fit to the set of deseasonalized series by following the model construction procedures presented in Section 20.3.2 as well as Appendix A20.1.

Periodic Multivariate Model

An assumption underlying the deseasonalized model is that the correlation structure among seasons is the same throughout the year. To allow for a seasonally varying correlation structure, *periodic models* can be employed. As described in depth in Chapter 14, two popular periodic models are the PAR (periodic autoregressive) and PARMA (periodic ARMA) models. When fitting a PAR model to a single seasonal series, a separate AR model is designed for each season of the year. In a similar manner, a PARMA model consists of having a separate ARMA model for each season of the year. Within hydrology, PAR modelling dates back to the research of Thomas and Fiering (1962) who proposed a specialized type of PAR model whereby the order of the AR operator for each season is fixed at unity. More recently, authors such as Salas et al. (1980) and Thompstone et al. (1985a,b) have suggested that the order of the AR operator for each season be properly identified. Model construction techniques that can be employed with PAR models are presented in Sections 14.3 and 14.5.3 while PARMA modelling methods are discussed in Section 14.7. Because there is a separate model for each season of the year, periodic models can be considered to be special types of multivariate models (Salas et al., 1985; Vecchia et al., 1983; Vecchia, 1985a,b).

PAR and PARMA models can be considered as the periodic extensions of nonseasonal AR and ARMA models, respectively, for modelling seasonal data. Similar to the PAR and PARMA models described in Chapter 14, a periodic multivariate model would essentially consist of having a separate multivariate model for each season of the year. Salas and Pegram (1978) define the periodic version of multivariate ARMA(p,0) models while, Salas et al. (1980) present the periodic extension of the general multivariate ARMA(p,q) models given in [20.2.1] and [20.2.2]. Bartolini and Salas (1986), Haltiner and Salas (1988), Bartolini et al. (1988) and Ula (1990) investigate the statistical properties of multivariate PARMA(1,1) processes. Additionally, Salas and Abdelmohsen (1993) devise an initialization procedure for generating univariate and multivariate PAR(1), PAR(2) and PARMA(1,1) processes. Their approach is, in fact, the PARMA version of the WASIM2 simulation algorithm presented in Section 9.4 for simulating with ARMA models. In Section 19.6.3 and also in the paper by Hipel and McLeod (1981), specific kinds of periodic TFN and intervention models are proposed.

Because a periodic multivariate ARMA model possesses many more parameters than the complex nonperiodic version, one must devise ways to decrease the number of model parameters. Following the approach of Thompstone et al. (1985a) described in Section 14.5, one method to reduce the size of a periodic model is to divide the year into groups of seasons where consecutive seasons having similar correlation structures are put into the same group. The periodic multivariate ARMA model used to fit to the grouped data would only have parameters that preserve the correlation relationships among the groups of data rather than the original seasons. Secondly, as proposed by Salas et al. (1980) for PAR and PARMA models, a Fourier series approach could be utilized to reduce the number of parameters required in a periodic multivariate model. A final approach to economize on the number of parameters is to adopt the procedure suggested for periodic intervention models in Section 19.6.3. Depending upon the statistical characteristics of the multiple time series being modelled, only specified components of the multivariate ARMA model would be permitted to have a periodic structure. In fact, this is what is done for the intervention models developed in Sections 19.2.5 and 22.4.2 for modelling seasonal riverflows, as well as in Section 22.4.2 for describing water quality times series that have been impacted by external interventions.

20.4 HISTORICAL DEVELOPMENT

Recently, Salas et al. (1985) presented a comprehensive review of various approaches to multivariate modelling in hydrology. A significant portion of their paper deals with research closely related to multivariate ARMA modelling of hydrological time series. Camacho et al. (1986) also put research on multivariate ARMA modelling from the hydrological and statistical literature into proper perspective. The section follows closely the historical survey given by Hipel (1986), which was presented at the Fourth International Hydrology Symposium held at Colorado State University from July 15 to 17, 1985 (Shen et al., 1986).

Research in multivariate modelling in water resources goes back to the early 1960's when researchers such as Maas et al. (1962) introduced systems sciences techniques into the field of water resources. Much of this research dealt with proposing fairly simple multivariate models, most of which are either subsets of or else closely related to the multivariate ARMA model in [20.2.1] and [20.2.2]. In the earlier research, often the exact form of the model used for fitting to a data set was specified prior to model construction. For instance, some researchers suggested using a multivariate AR(1) model while others proposed employing a multivariate ARMA(1,1) model. This type of procedure may result in using a model that does not fit the data well. Because of this, Finzi et al. (1975) found that synthetic data generated by prespecified models were inadequate in several applications. Another disadvantage of this approach is that it may cause inefficient estimation of the model parameters (Camacho et al., 1985).

In a pioneering paper, Fiering (1964) proposed a two station multivariate model to link two series, X_t and Y_t. The model of Fiering was later modified by Kahan (1974) and Lawrance (1976).

Matalas (1967) suggested a multisite AR(1) model for use in hydrology. His model preserves both the lag zero and lag one cross covariance matrices. He also pointed out that the model could be simplified by having a diagonal AR matrix and he described a parameter estimation procedure based on the method of moments. Kuczera (1987) explained how to obtain MLE's for the parameters of a multivariate AR(1) model using the EM algorithm of Dempster et al. (1977) when there are missing observations. To take into account seasonality in a time series, Young and Pisano (1968) suggested first deseasonalizing the multivariate series before fitting the Matalas model. Furthermore, they designed improved estimation procedures and suggested transformations for removing skewness in the data.

For modelling monthly multivariate data, Bernier (1971) considered a monthly multivariate AR(1) model. His model is actually a combination of the Fiering and Matalas model.

For a multivariate AR(p) model, Pegram and James (1972) proposed a moment estimation procedure to estimate the parameters and when using the model for streamflow generation they gave reasons for diagonalizing the AR matrices. A general multivariate AR(p) model with seasonally varying parameters was designed by Salas and Pegram (1978) who suggested both the methods of moments and maximum likelihood to estimate the parameters. When their model has diagonal AR matrices, it forms a periodic contemporaneous AR(p) model.

As explained in Section 10.4, the FGN (fractional Gaussian noise) model defined in [10.4.2] was developed to model long term persistence and thereby provide an explanation for the Hurst phenomenon described in Section 10.3.1. To model the multivariate version of long term persistence, Matalas and Wallis (1971) considered the multivariate fractional Gaussian noise (FGN) model for which each of the series is modelled by a univariate FGN model with contemporaneously correlated innovations. O'Connell (1974) proposed a vector ARMA(1,1) model to describe long term persistence. Canfield and Tseng (1979) studied the same model with diagonal AR and MA matrices while Lettenmaier (1980) suggested improved estimation procedures for the vector ARMA(1,1) model.

Franchini et al. (1986) developed a type of multivariate AR model which has the ability to preserve long term persistence and to reproduce the statistical properties of the seasonal flows at more than one station situated in a given river basin. They pointed out that their model is capable of maintaining the time-space correlations at the seasonal level as well as the properties of the flow volumes at the annual level.

In 1974, Mejia et al. considered the situation where the generation of synthetic hydrological sequences are obtained from a mixture of distributions. To reproduce the historical moments in the simulated data, they proposed a transformation of the moments of the historical data to be used in the estimation of the parameters of the model. This procedure appears to have little statistical justification according to Stedinger (1981) who shows that direct estimation of the moments of the transformated historical data can result in significantly better estimates of the true cross correlations.

Kottegoda and Yevjevich (1977) compared the preservation of the correlation in the generated samples of four kinds of existing two station models. Because the models produced essentially equivalent results, they concluded that one has "to apply the simplest model with the best physical justification".

Stedinger (1981) compared different approaches for estimation of the correlations in multivariate streamflow models. He concluded that "there appears to be little statistical justification to the idea that one should select a streamflow model's parameters so as to reproduce exactly the observed correlations of the flows themselves ... and perhaps the most important lesson to be learned ... is that estimates of many streamflow model parameters are inaccurate". Therefore, it "is very reasonable to use statistically efficient parameter estimation which may not exactly reproduce the observed means, variances and correlations of the historical flows" (Stedinger and Taylor (1982a).

By generalizing the methodology of Vicens et al. (1975), Valdes et al. (1977) developed a Bayesian procedure to generate synthetic streamflows for multivariate AR models. An advantage of this procedure is that it takes into account parameter uncertainty. However, Davis (1977) and McLeod and Hipel (1978c) mention drawbacks to the simulation approach of Vicens et al. (1975) for handling parameter uncertainty. As explained in Section 9.7, the method of McLeod and Hipel (1978) for simulating, using univariate ARMA models, correctly takes into account parameter uncertainty and could easily be extended to the multivariate case. As pointed out Stedinger and Taylor (1982b), to include uncertainty in the parameters of the model is very important for obtaining realistic and honest estimates of system reliability.

In 1978, Ledolter proposed that the general class of multivariate ARMA models be used in hydrology. Salas et al. (1980) suggested the CARMA models with constant or periodic parameters constitute parsimonious models that reflect the physical reality of hydrological systems. They also proposed procedures for use in model construction. Further advances in identification, estimation and diagnostic checking of CARMA models were given by Camacho et al. (1985, 1986, 1987a,b,c). Other research related to contemporaneous modelling is given by authors including Hannan (1970), Wilson (1973), Granger and Newbold (1977), Wallis (1977), Chan and Wallis (1978), Hillmer and Tiao (1979), Nicholls and Hall (1979), Risager (1980, 1981), Tiao and Box (1981), and Jenkins and Alavi (1981).

Along with model construction procedures, Cooper and Wood (1982a,b) propose the multiple input-output model. This class of models is actually equivalent to the multivariate ARMA family in [20.2.1] and [20.2.2]. The mathematical and statistical properties of the multivariate models considered by Cooper and Wood (1982a,b) were studied by Hannan and Kavalieres (1984).

As was the case for the CARMA class of models, the space-time ARMA or STARMA family of models was designed to overcome the problem of too many parameters in multivariate ARMA models (Deutsch and Ramos, 1984, 1986; Pfeifer and Deutsch, 1980; Deutsch and Pfeifer, 1981). However, Camacho et al. (1986) argued that the parameter restrictions incorporated into the STARMA model may be too severe and thereby limit the applicability of the model. Nonetheless, Adamowski et al. (1986) found the STARMA model useful for modelling eleven raingage sites located in a watershed in Southern Ontario, Canada.

Kelman et al. (1986) devised a multivariate version of a model proposed by Kelman (1980) for separately modelling the rising and falling limbs of daily hydrographs. The multivariate extension of the model follows the approach suggested by Matalas (1967).

Srikanthan (1986) proposed a multivariate model for simulating daily climatic data. Daily rainfall was simulated using a multistate first order Markov model and the remaining climatic variables were simulated using a multistate type model (Matalas, 1967; Richardson, 1981). Nasseri (1986) utilized a multivariate AR model of order one to generate hourly rainfall for a network of raingages.

Venugopal et al. (1986) used a multisite model for simulating flows of the Narmada River system in India. In particular, the HEC-4 (Feldman, 1981) and disaggregation models were employed for the synthetic generation of riverflows.

As pointed out in Section 20.2.2 the TFN and intervention group of models is actually a subset of the general multivariate ARMA family of models in [20.2.1] and [20.2.2] when the AR and MA parameter matrices are either all upper or lower triangular. The use of TFN modelling in water resources dates back to the time of Fiering (1964) who proposed a bivariate TFN model which was later modified by Lawrance (1976). In fact, because of the great importance of TFN modelling and intervention analysis in water resources and environmental engineering, Parts VII and VIII of this book deal exclusively with TFN and intervention modelling, respectively.

References regarding the theory and practice of TFN modelling are listed at the ends of Chapters 16 to 18 while references for intervention modelling are given in the final parts of Chapter 19 and 22. At the Fourth International Hydrology Sympoisum on Multivariate Analysis of Hydrologic Processes held at Colorado State University from July 15 to 17, 1985, a number of research papers were concerned with TFN modelling (Shen et al., 1986). In particular, Nicklin (1986) employed a TFN model to mathematically formulate the dependence between nonstationary irrigation diversion and return flows. In order to identify an appropriate TFN model, he suggested novel identification procedures designed to use with his particular type of problem. Delleur (1986) developed a model consisting of a mixed model for forecasting real-time flash floods. The model consisted of a nonlinear conceptual submodel for transforming the observed rainfalls into effective precipitation followed by a TFN model relating the effective rainfall to the observed flood.

20.5 OTHER FAMILIES OF MULTIVARIATE MODELS

20.5.1 Introduction

The previous section on the historical development of statistical multivariate models in hydrology dealt mainly with models closely related to the general multivariate ARMA family of models in [20.2.1] and [20.2.2]. Other classes of statistical models have also been used in hydrology. For example, Fiering (1964) introduced multivariate analysis for generating multisite streamflows using *principal component analysis*. Numerous authors, have developed and employed *regression analysis* models for use in applications such as environmental impact assessment, data filling, and synthetic streamflow generation. In fact, within Chapter 24 of this book, ways in which regression analysis can be employed for both exploratory and confirmatory data analysis purposes are explained and illustrated. In Section 24.3, a trend analysis methodology, which uses techniques such as regression analysis and nonparametric tests (Chapter 23), is presented for detecting and modelling trends in water quality time series measured in rivers.

In a perceptive paper, Yevjevich and Harmancioglu (1985) discuss the past and future of time series analysis in water resources. Some of the challenging research projects that these authors feel should be actively pursued include the proper treatment of nonGaussian, nonlinear and multivariate time series. Besides defining new univariate and multivariate models for handling the foregoing and other problems, any new models must be made fully operational by developing appropriate model construction techniques. In a sequence of nineteen invited papers written by statisticians, hydrologists and other scientists, these as well as many other challenges are met head on by many original research contributions (Hipel, 1985b). As pointed out in Section 1.1, because of the great importance of time series analysis as well as other statistical techniques in the environmental sciences, the new field of *environmetrics* has evolved into a promising new discipline. Moreover, many journals and books in which environmetrics research is published are mentioned in Section 1.6.3. In this section, various types of recently designed multivariate models, many of which are currently under development, are now discussed. Because of the controversy surrounding the disaggregation model, this family of models is entertained first.

20.5.2 Disaggregation Models

A special class of multivariate models is the *disaggregation* family of models. This class allows one to break down a series for which there are longer time units separating values into a sequence of values separated by shorter time units. For instance, an annual series can be disaggregated into a monthly series. The major reason why Valencia and Schaake (1973) proposed the disaggregation model was to insure that relevant statistics at both the annual and seasonal levels are consistent with one another. Annual flows, for example, could be generated by a short or long memory model and these annual flows could then be disaggregated to the seasonal level. As noted by Salas et al. (1985), disaggregation can be used for not only disaggregating variables in time, but for disaggregating in space as well. For example, precipitation over an area may be disaggregated into precipitation over sub-areas (Salas et al., 1980). Frevert and Lane (1986) presented a technique for accomplishing two level spatial disaggregation in a single run of their computer programs for disaggregation.

As discovered in a discussion with V. Klemes on May 30, 1985, in Tucson, Arizona, the basic idea of disaggregation is relatively old. In earlier research, the idea of disaggregation was utilized for addressing problems related to storage (Savarenskiy, 1940; Gould, 1961; Svanidze, 1962, 1980; Klemes, 1963, 1981). Woolhiser and Osborn (1986) devised a special kind of model for disaggregating storms into seasonal and regional components. Valencia and Schaake (1973) proposed a disaggregation model for obtaining seasonal flows from riverflows simulated at the annual level. As explained by Salas et al. (1985), since 1973 there have been numerous papers suggesting improvements to the original disaggregation model of Valencia and Schaake (1973) as well as related models developed thereafter. For example, Mejia and Rouselle (1976) put forward enhancements for the original disaggregation model of Valencia and Schaake (1973). Lee (1986) developed a multisite, multiseason synthetic flow generation model within a disaggregation framework. Other contributions to research in disaggregation are provided by authors including Tao and Delleur (1976), Stedinger and Vogel (1984), and Grygier and Stedinger (1988).

In a conversation held with V. Yevjevich on May 30, 1985, in Tucson, Arizona, he stated that two questions should be satisfactorily answered in order to adequately justify the use of disaggregation models in hydrology. The first question is whether there is information in annual measurements which is not contained in the seasonal observations. If there is not more information contained in the annual series, a better procedure may be to aggregate rather than disaggregate. When aggregating, a seasonal time series is modelled directly using a process such as a PARMA model and then it is aggregated to produce a compatible model for the aggregated series, usually at the annual level. Vecchia et al. (1983) presented a convincing argument which favours the concept of *aggregation* over disaggregation. They proved that if the original seasonal data follow a PARMA model in [14.2.15] or [14.2.16] with one moving average and one autoregressive parameter (i.e., PARMA(1,1)) then the aggregated annual data must be an ARMA model in [3.4.3] or [3.4.4] with one AR parameter (i.e., ARMA(1,0)) or else an ARMA model with one AR and one moving average parameter (i.e., ARMA(1,1)). Furthermore, there is significant gain in parameter estimation efficiency at the aggregated level when the seasonal data and their model are used rather than the aggregated (i.e., annual) data and their model. Rao et al. (1985) derived similar results for the situation where the seasonal data follow a PAR model in

[14.2.1] or [14.2.3]. In addition, they showed theoretically that the aggregated data can be more accurately predicted by using a valid model of the aggregated data. Moreover, in a related topic, aggregation of forecasts are discussed in Section 15.6 of this book. Further research regarding aggregation is presented by authors such as Kavvas et al. (1977), Obeysekera and Salas (1982, 1986) and Bartolini et al. (1988). Finally, Eagleson (1978) employs the principle of aggregation when he derives the distribution of annual precipitation from observed storm sequences.

A second issue raised by V. Yevjevich was whether the large number of parameters in a disaggregation model can be significantly reduced. In addition to other approaches, Stedinger et al. (1985) proposed a more parsimonious disaggregation model which is designed in a fashion which is analogous to the CARMA model described in detail in Chapter 21. Koutsoyiannis (1992) developed a multivariate dynamic disaggregation model, having a reduced parameter set, as a stepwise approach to disaggregation problems.

In certain situations, a practitioner or researcher may feel that it is appropriate to employ disaggregation models. A well-tested set of computer programs for implementing disaggregation models are available for use on both main frame and personal computers (Lane and Frevert, 1990).

20.5.3 Gaussian and NonGaussian Variables

As noted by Lewis (1985), simple linear models, such as the family of ARMA models, are not necessarily defined as having Gaussian variates but are simplest to use as such because linear operations on Gaussian variates preserves Gaussianity or normality. Furthermore, model construction procedures, based on the assumption of Gaussianity, are well developed. As a result, theoretical research regarding the development of stochastic models which can explicitly handle variables which are nonGaussian and therefore do not follow a normal distribution, has only been initiated recently. Among others, Tong et al. (1985) point out the fact that hydrological data are often not normally distributed and procedures are required to effectively handle this problem.

When the data are nonGaussian, one approach for obtaining data which are approximately normally distributed is to transform the original data using a transformation such as a Box-Cox transformation (Box and Cox, 1964) in [3.4.30]. This will produce a transformed series which is approximately Gaussian. A model based upon the Gaussian assumption can then be fitted to the transformed series. An alternative approach is not to assume Gaussianity in the first place but to select a distribution that the original data actually follow. Li and McLeod (1988) present results on estimation and diagnostic checking for ARMA models having nonGaussian innovations. Lewis (1985) describes a range of new models developed for use with continuous variate nonGaussian time series. The nonGaussian distributions he considers are the Exponential, Gamma, Weibull, Laplace, Beta and Mixed Exponential distributions. McKenzie (1985) presents a variety of models similar to Markov chains for describing discrete variate time series that follow various distributions The distributions which he entertains are the Poisson, Geometric, Negative Binomial and Binomial distributions. Finally, Brillinger (1985) develops procedures for fitting finite parameter models to nonGaussian series via bispectral fitting. The foregoing and other univariate developments in *nonGaussian modelling*, can be defined for modelling multiple time series. For example, Lewis (1985) mentions that his nonGaussian models can be extended to the multivariate case and that a periodic version of these models can be devised for modelling seasonal data.

20.5.4 Linear and Nonlinear Models

When a model is linear, it is a linear function of the variables in the model. In a *nonlinear model*, there is at least one term where the variables and/or innovations appear as products or are raised to powers. Lewis (1985) gives a brief discussion regarding alternative definitions for linearity and nonlinearity. Generally speaking, as the time interval between observations becomes smaller, the nonlinearities present in the data become more pronounced. For instance, average daily riverflow data may have to be modelled using a model containing nonlinear terms because of the nonlinear relationship between runoff and precipitation over a small time scale. On the other hand, a linear model may be sufficient to model mean annual riverflows for which the nonlinearities have been "averaged out".

A variety of interesting nonlinear stochastic models are now available (Tong, 1990). For example, Tong et al. (1985), Ozaki (1985), as well as Brillinger (1985) and Gallant (1987) describe threshold, discrete time storage, and nonlinear regression models, respectively, which are capable of modelling various kinds of nonlinearities which may be present in natural time series. Li (1992) derives the asymptotic standard errors of residual autocorrelations in nonlinear time series models for use in diagnostic checking while, in 1993, Li presents a statistical test for discriminating among different nonlinear time series models. As noted by Tong et al. (1985), the threshold model (Tong, 1983) can be easily defined for the multivariate case. Further, Brillinger (1985) defines a spatial-temporal process for multivariate modelling.

20.5.5 Multivariate Fractional Autoregressive-Moving Average (FARMA) Models

Short and long memory models are defined in Section 2.5.3 using [2.5.7]. A special class of long memory models is the FARMA family of models defined in [11.2.4]. Because the FARMA model is a generalized type of ARIMA (autoregressive integrated moving average) model given in [4.3.4] for which the differencing operator can assume real values, the FARMA model can be easily written for the multivariate case. However, a great deal of research is required to develop model construction techniques for use with multivariate FARMA models.

20.5.6 Time and Frequency Domains

Time series models such as the nonGaussian models of Lewis (1985) or the general multivariate ARMA model in [20.2.1] or [20.2.2] are defined in terms of discrete time variables. In order to fit a time series model to a data set, various techniques are available for use at the three stages of model construction. If a given method or statistic, such as the sample ACF which can be used for model identification, is expressed directly in terms of the time variable, it is said to be expressed in the *time domain*. Alternatively, one can work in the *frequency domain* by entertaining Fourier transforms. As explained in Sections 2.6 and 3.5, the Fourier transform of the autocovariance function produces the spectrum which expresses the distribution of the variance of the series with frequency (Jenkins and Watts, 1968). Although it is more common in water resources to execute univariate and multivariate time series modelling in the time domain rather than the frequency domain, sometimes it is advantageous to work in the latter domain. For instance, Brillinger (1985) presents interesting results regarding Fourier inference. Canfield and Bowles (1986) devise a method for conveniently generating multivariate series from the spectrum. Ghani and Metcalfe (1986) employ a spectral approach for predicting the probability of the peak flow exceeding a given level during a specified time period. A host of other applications of spectral methods to environmental problems can be found in journals and books referred to in 1.6.3, although only some of the published spectral research deals with multivariate problems.

20.5.7 Pattern Recognition

In order to model multivariate hydrological time series, MacInnes and Unny (1986) extend to the multivariate level the univariate approach of Panu and Unny (1980) and Unny et al. (1981) for modelling time series from a *pattern recognition* viewpoint. They apply their pattern recognition model to familiar multistation streamflow problems and discuss both the advantages and disadvantages of using pattern recognition-based models.

20.5.8 Nonparametric Tests

In order to lessen the number of underlying assumptions required for testing a hypothesis such as the presence of a specific kind of trend in a data set, researchers developed nonparametric procedures for use in hypothesis testing. Due to the great importance of nonparametric testing in environmental impact assessment, Chapter 23 of this book is entirely devoted to this type. Some of the nonparametric tests, such as the partial rank correlation tests of Section 23.3.6, can either be used or else extended for use with multi-site and/or multiple variable data sets. As a result, nonparametric tests are very useful in multivariate analysis, especially when the data are very *messy*. Part X of this book explains how intervention analysis, nonparametric tests and regression analysis can be used for modelling *messy environmental data*.

20.6 CONCLUSIONS

The general multivariate ARMA model is defined in [20.2.1] and [20.2.2] while model construction procedures are presented in Section 20.3.2 and Appendix A20.1. Within this general family of time series models, the CARMA and TFN models are of particular importance in the field of water resources and environmental engineering. As explained in this chapter and Section 21.1 and also by Salas et al. (1985) and Camacho et al. (1985a, 1986), the physical constraints dictated by a given hydrological system negate the need for using the general form of the multivariate ARMA model and, therefore, usually some type of CARMA or TFN model is all that is required in a practical application. Additionally, as described in depth in Chapters 17 and 21, model construction techniques are now fully developed for employment with TFN and CARMA models, respectively. The CARMA model constitutes a parsimonious version of the general multivariate ARMA model for describing multiple time series that are contemporaneously correlated with one another. Besides being able to model the impacts of interventions upon the mean level of a series and estimate missing observations (see Chapter 19), the TFN model can describe the mathematical relationships between a single response variable and any number of covariate series (see Chapter 17). The most general form of the TFN or intervention model is defined in [19.5.8].

In Section 20.4 the historical development of multivariate ARMA modelling in hydrology is outlined while other kinds of families of multivariate models are referred to in Section 20.5. These additional types of multivariate models include various classes of disaggregation, nonGaussian, nonlinear and long memory models. For some of these models, such as the nonGaussian models of Lewis (1985) and the nonlinear threshold models of Tong et al. (1985), further research is required for developing model construction techniques for use in modelling multiple time series.

For classifying the capabilities of a family of time series models, Hipel (1985b) suggests a list of twenty-five criteria that reflect the main statistical characteristics that could be modelled. This list includes linear, nonlinear Gaussian, nonGaussian, long memory and short memory criteria. Using these criteria, a given multivariate model, such as the CARMA model, can be categorized as being linear, Gaussian and short memory. By understanding the modelling capabilities of each family of models, as well as the main physical and statistical characteristics of the time series being studied, a practitioner can decide upon which classes of models are most appropriate to consider for modelling the key statistical properties of his or her data set. Subsequent to exploratory data analysis referred to in Section 1.2.4 and described in detail in Section 22.3, the practitioner can select the best specific time series model at the confirmatory data analysis stage by following the three stages of model construction.

An obvious extension to the work completed thus far in time series analysis is to develop more comprehensive families of models that can simultaneously handle a wider variety of the criteria. For example, it may be possible to design a multivariate model that can take care of both nonlinear and nonGaussian characteristics of the data. However, any new class of models should be designed to be as simple as possible and thereby not have too many parameters, as well as provide a good statistical fit to the data. Whenever possible, researchers are encouraged to incorporate both the physical and statistical aspects of the problem into the basic model design. Subsequent to the design, appropriate algorithms are required for use at the three stages of model construction. A continuous dialogue among the statisticians, water resources engineers and other scientists should increase the probability of designing new models that will be welcomed by the practitioners for solving pressing water resources problems.

Besides designing new models, existing models should be rigorously compared from both theoretical and empirical viewpoints in order to ascertain which families of models are most appropriate to use in practice. For instance, thorough scientific comparisons of the disaggregation and aggregation approaches to time series modelling are long overdue. "To disaggregate or not to disaggregate", that is the nagging question haunting both practitioners and theoreticians alike in hydrology.

Due to the continued and growing abuse of the natural environment by man-induced activities such as industrialization and agricultural development, there will continue to be a great demand for having flexible multivariate models for use in environmental impact assessment. Future research in the time series aspects of environmental impact assessment will probably entail developing more nonparametric tests for handling a wider variety of situations in trend detection and evaluation, rigorously comparing the capabilities of both parametric and nonparametric approaches, and providing guidelines for optimally designing sampling schemes for water quality variables.

As is also emphasized in Section 3.6 and elsewhere in the book, Yevjevich and Harmancioglu (1985) stress the importance of linking stochastic models with physical consistent properties of any particular water resources time series. In some circumstances, it may be possible to employ purely stochastic models to accomplish this goal. Alternatively, it may be necessary to employ a combination of deterministic and stochastic models for realistically modelling certain kinds of water resources systems.

The authors maintain that the development and use of multivariate models, in general, and multivariate time series models in particular, will significantly increase in the future within the realm of water resources and environmental engineering. Besides hydrological time series, multivariate models will be used more and more for modelling other kinds of water related time series such as water quality (physical, chemical and biological), water demand, water pricing and meteorological time series.

APPENDIX A20.1

IDENTIFICATION METHODS FOR GENERAL MULTIVARIATE ARMA MODELS

As advocated by Tiao and Box (1981), the sample CCF and PACF matrices are especially useful for identifying pure multivariate MA and AR models, respectively. When the set of time series follow a multivariate ARMA(p,q) process, the extended sample cross correlation function (ESCCF) matrix of Tsay and Tiao (1984) and Tiao and Tsay (1983a,b) can also be used. Section 20.3 describes how model construction is carried out for general multivariate ARMA models.

A20.1.1 Sample CCF Matrix

Suppose k time series of length n are represented at time t by the vector

$$\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tk})^T$$

For lag l, where $l = 1,2,\ldots,$ the *theoretical CCF matrix* of order $k \times k$ is written as $\rho(l)$. A typical entry, $\rho_{ij}(l)$, in the matrix is theoretically defined as

$$\rho_{ij}(l) = Cov(Z_{ti}, Z_{t+l,j})/[Var(Z_{ti}) \cdot Var(Z_{tj})]^{1/2} \qquad [A20.1.1]$$

The *sample CCF matrix* at lag l is denoted by $\mathbf{R}(l)$. Each element $r_{ij}(l)$ in the matrix is calculated using

$$r_{ij}(l) = \sum_{t=1}^{n-l}(Z_{ti} - \bar{Z}_i)(Z_{t+l,j} - \bar{Z}_j)/[\sum_{t=1}^{n}(Z_{ti} - \bar{Z}_i)^2 \sum_{t=1}^{n}(Z_{tj} - \bar{Z}_j)^2]^{1/2} \qquad [A20.1.2]$$

where $\bar{Z}_i$ and $\bar{Z}_j$ are the sample means of the ith and jth series, respectively.

For a pure multivariate MA process of order q, the theoretical CCF matrix vanishes after lag q (Tiao and Box, 1981; Jenkins and Alavi, 1981). Hence, if $\mathbf{Z}_t$ follows a MA(q) process, the entries in the sample CCF matrix are not significantly different from zero for $l > q$. Because the asymptotic distribution of $r_{ij}l$ is N(0,1/n), the 95% confidence limits given approximately by $\pm\frac{2}{n^{1/2}}$ can be used to decide whether or not the estimated value is significantly different from zero. If each entry in $\mathbf{R}(l)$ falls within these limits, it can be assumed that the sample CCF matrix has cutoff at lag l and, consequently, l can be considered as the order of the multivariate MA model. If the series $\mathbf{Z}_t$ follows a multivariate AR model of order p or a general multivariate

ARMA(p,q) model, no cut-off will appear in $\mathbf{R}(l)$. Tiao and Box (1981) suggest summarizing the numerical values of each $r_{ij}(l)$ with "+" to indicate a value greater than $\frac{2}{n^{1/2}}$, with "-" to indicate a value less than $\frac{-2}{n^{1/2}}$ and with "." to indicate a value between $\frac{-2}{n^{1/2}}$ and $\frac{2}{n^{1/2}}$. This is a very convenient way to interpret the entries in $\mathbf{R}(l)$. Similar to the situation for the identification of univariate ARMA models in Chapters 3 and 5, one must define another statistic for detecting cutoff in pure multivariate AR(p) models.

A20.1.2 Sample PACF Matrix

For a pure multivariate AR process of order p, denoted by AR(p), the theoretical PACF matrix, $\mathbf{P}(l)$, for lag l, where $l = 1,2,\ldots,$ can be defined. Let $\mathbf{\Phi}_{lj}$ be the jth AR matrix in a multivariate AR process of order l so that $\mathbf{\Phi}_{ll}$ is the last matrix. The *theoretical PACF matrix*, $\mathbf{P}(l)$, is defined as $\mathbf{\Phi}_{l,l}$ where $\mathbf{\Phi}_{l1},\mathbf{\Phi}_{l2},\ldots,\mathbf{\Phi}_{ll}$, are the solutions of the system of equations

$$\sum_{i=1}^{l}\mathbf{\Phi}_{li}\cdot\Gamma(j-i)=\Gamma(-j),\ j=1,2,\ldots,l \qquad [A20.1.3]$$

where $\Gamma(j) = Cov[\mathbf{Z}_t,\mathbf{Z}_{t+j}]$. This equation constitutes a multivariate generalization of the Yule-Walker equations in [3.2.17] for AR models in univariate time series analysis. If the process is multivariate AR(p) then by definition $\mathbf{\Phi}_{pp} = \mathbf{\Phi}_p$ in [20.2.1] and $\mathbf{\Phi}_{p+j,p+j} = \mathbf{0}$ for $j > 0$. Consequently, for a pure multivariate AR(p) process, $\mathbf{P}(l) = \mathbf{0}$ for $l > p$. If the process is MA(q) or ARMA(p,q), $\mathbf{\Phi}_{ll}$ will not cutoff but rather decay to zero.

The *sample PACF matrix* at lag l where $l = 1,2,\ldots,$ is defined as the $\hat{\mathbf{P}}(l) = \hat{\mathbf{\Phi}}_{ll}$ matrix in the solution of the system of equations

$$\sum_{i=1}^{l}\hat{\mathbf{\Phi}}_{li}\mathbf{R}(j-i)=\mathbf{R}(-j) \qquad [A20.1.4]$$

The sample PACF matrix, $\hat{\mathbf{P}}(l)$ at lag l, is calculated by fitting a multivariate AR(l) model using

$$\mathbf{Z}_t = \mathbf{c} + \mathbf{\Phi}_{l1}\mathbf{Z}_{t-1} + \mathbf{\Phi}_{l2}\mathbf{Z}_{t-2} + \cdots + \mathbf{\Phi}_{l,l}\mathbf{Z}_{t-l} + \mathbf{a}_t^{(l)} \qquad [A20.1.5]$$

and setting $\hat{\mathbf{P}}(l) = \mathbf{\Phi}_{ll}$, where $\mathbf{c}$ is a vector of k constants that are recursively estimated along with the other model parameters using standard multivariate least squares (Tiao and Box, 1981). Asymptotically, the distribution for each entry in $\hat{\mathbf{P}}(l)$ is $N(0,\frac{1}{n})$. To denote whether an entry in $\hat{\mathbf{P}}(l)$ is greater than, less than, or falls within the approximate 95% confidence limits given by $\pm\frac{2}{n^{1/2}}$, one can employ "+", "-" or ".", respectively. Using the foregoing symbols rather than numerical values makes it easier to detect the important identification information contained in $\hat{\mathbf{P}}(l)$.

Basically, the sample PACF matrices are determined by fitting AR models of order $l = 1,2,\ldots,$ in [A20.1.5]. By analyzing the variance-covariance matrices corresponding to successive AR fittings, one can ascertain how much the statistical fit improves as the order l is increased.

For any given l, $l = 1,2,\ldots,$ a formal test of hypothesis for testing the null hypothesis: $\mathbf{P}(l)=\mathbf{0}$ against the alternative $\mathbf{P}(l) \neq \mathbf{0}$ can be performed. The likelihood ratio statistic is given by

$$U = |S l|/|S(l-1)| \quad \text{[A20.1.6]}$$

where

$$S(l) = \sum_{t=1}^{n} \hat{\mathbf{a}}_t^{(l)} \cdot \hat{\mathbf{a}}_t^{(l)T}$$

for which $\hat{\mathbf{a}}_t^{(l)}$ are the residuals of the fitted model in [A20.1.5].

Using Barlett's (1938) approximation, the statistic

$$\chi(l) = -(N - 1/2 - lk)\cdot log[|S(l)|/|S(l-1)|] \quad \text{[A20.1.7]}$$

will have a chi-square distribution with k^2 degrees of freedom when $l > p$. Now, for MA(q) or general ARMA(p,q) models, the sample PACF matrices do not have a cutoff and, therefore, they are expected to obtain significant values of $\chi(l)$ even for large lags. Examples of the use of the sample ACF and PACF matrices in the identification of hydrologic time series are given by Camacho et al. (1987c).

The sample ACF and PACF matrices are very useful for identifying pure MA and pure AR models, respectively. In practice, the difficulty arises in the identification of mixed ARMA(p,q) models when both p and q are larger than zero. In these cases, the ESCCF can be employed to help in the identification.

A20.1.3 ESCCF Matrix

The *ESCC (extended sample cross correlation) matrix* was proposed by Tsay and Tiao (1984) and Tiao and Tsay (1983a,b) to help in the identification of the order of mixed multivariate ARMA models. The main idea of the technique is to calculate consistent estimates of the AR matrices $\Phi_1, \ldots, \Phi_p$, say, and then use the properties of the transformed series

$$\mathbf{W}_t = \mathbf{Z}_t - \sum_{l=1}^{p} \hat{\Phi}_l \mathbf{Z}_{t-l} \quad \text{[A20.1.8]}$$

to identify the order of the process. If $\mathbf{Z}_t$ follows a multivariate ARMA(p,q) process and the estimated parameters are consistent, then $\mathbf{W}_t$ follow a multivariate MA(q) process, so that the sample CCF matrices of $\mathbf{W}_t$ should be able to identify the proper order of the model. Tiao and Tsay (1983a) have shown that when the order of the model is known, it is possible to obtain consistent estimates for the Φ's using a process of iterated regressions. However, in practice the order of the model is not known in advance, and therefore, it is necessary to study the properties of the iterated regression estimates and their associated transformed series $\mathbf{W}_t$.

The following algorithm can be used to obtain the iterated regression estimates where further details are given by Tiao and Tsay, (1983a):

Step 1: For $m = 1,2,\ldots,M+J+1$ fit an AR(m) regression

$$\mathbf{Z}_t = \sum_{l=1}^{m} \Phi_{l(m)}^{(o)} \mathbf{Z}_{t-l} + emt^{(o)}, \quad t = m+1, \ldots, n \qquad [A20.1.9]$$

using ordinary least squares. Denote the estimated parameters as $\hat{\Phi}_{l(m)}^{(o)}$. The superscript (o) indicates the ordinary least square estimates and the subscript (m) indicates the order of the AR fit.

Step 2: For $j = 1,2,\ldots,J$, and $m = 1,\ldots,M$, recursively compute the AR(m) jth iterated estimates $\hat{\Phi}_{l(m)}^{(j)}$, $l = 1,\ldots,m$ as:

$$\hat{\Phi}_{l(m)}^{(j)} = \hat{\Phi}_{l(m+1)}^{(j-1)} - \hat{\Phi}_{m+1(m+1)}^{(j-1)}[\hat{\Phi}_{m(m)}^{(j-1)}]^{-1}\hat{\Phi}_{l-1(m)}^{(j-1)} \qquad [A20.1.10]$$

where $\hat{\Phi}_{o(m)}^{(j-1)} = -\mathbf{I}$.

Now if $\mathbf{Z}_t$ follows a multivariate ARMA(p,q) and no linear combination of $\mathbf{Y}_{pt} = (\mathbf{Z}_t^T, \ldots, \mathbf{Z}_{t-p+1})^T$ follows a MA model with order less than q, then the matrices $\hat{\Phi}_{l(m)}^{(j)}$ are consistent estimators for Φ_l when (i) $m \geq p$ and $j = q$, or, (ii) $m = p$ and $j > q$ (Tiao and Tsay, 1983a). Also, for $j > q$, the transformed series

$$\mathbf{W}_{p,t}^{(j)} = \mathbf{Z}_t - \sum_{l=1}^{m} \hat{\Phi}_{l(p)}^{(j)} \mathbf{Z}_{t-l} \qquad [A20.1.11]$$

will approximately follow a multivariate MA(q) model. Therefore, the sample CCF matrix of $W_{p,t}^{(j)}$ will have a cutoff after lag q. In particular, the lag j CCF matrix of $\mathbf{W}_{p,t}^{(j)}$, $\hat{\rho}_{(p)}(j)$ will be such that

$$\hat{\rho}_{(p)}(j) \xrightarrow{p} \begin{cases} \mathbf{C}, & j \leq q \\ \mathbf{0}, & j > q \end{cases} \qquad [A20.1.12]$$

where $\mathbf{C}$ is a generic symbol for a matrix whose elements are not necessarily all zero.

The mth ESCCF matrix is now defined as $\hat{\rho}_{(m)}(j)$ is the lag j sample CCF matrix of $\mathbf{W}_{m,t}^{(j)}$. For a general multivariate ARMA(p,q) process, $\hat{\rho}_m(j)$ has the following asymptotic property:

$$\hat{\rho}_{(m)}(j) \xrightarrow{p} \begin{cases} \mathbf{0}, & 0 \leq m-p < j-q \\ \mathbf{C}, & \text{otherwise} \end{cases} \qquad [A20.1.13]$$

This property of the ESCCF can now be exploited in the following way to help in the identification of the order (p,q) of an ARMA model:

Stage 1: Arrange the ESCCF matrices $\hat{\rho}_{(m)}(j)$ in a block matrix as is shown in Table A20.1.1. The rows numbered 0,1,2, . . . , signify the AR order and, similarly, the columns signify the MA order. To investigate how to use this table, suppose that the true model for $\mathbf{Z}_t$ is a multivariate ARMA(2,1) model. The pattern of the asymptotic behaviour of the ESCCF is shown in Table A20.1.2.

It can be observed from Table A20.1.2 that there is a triangle of asymptotic zero matrices and that the vertex of the triangle is located at the entry (2,1). For a general multivariate ARMA(p,q) model the same pattern is expected and the vertex of such a triangle will always be located at entry (p,q). This observation suggests the second idea in the identification of the process.

Table A20.1.1. The ESCCF table.

AR	MA					
	0	1	2	3	·	·
0	$\hat{\rho}_{1(0)}$	$\hat{\rho}_{2(0)}$	$\hat{\rho}_{3(0)}$	$\hat{\rho}_{4(0)}$	·	·
1	$\hat{\rho}_{1(1)}$	$\hat{\rho}_{2(1)}$	$\hat{\rho}_{3(1)}$	$\hat{\rho}_{4(1)}$	·	·
2	$\hat{\rho}_{1(2)}$	$\hat{\rho}_{2(2)}$	$\hat{\rho}_{3(2)}$	$\hat{\rho}_{4(2)}$	·	·
3	$\hat{\rho}_{1(3)}$	$\hat{\rho}_{2(3)}$	$\hat{\rho}_{3(3)}$	$\hat{\rho}_{4(3)}$	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·

Table 20.1.2 The asymptotic ESCCF matrix table of an ARMA(2,1) model where C and **0** denote, respectively, a nonzero and a zero matrix.

AR	MA					
	0	1	2	3	4	5
0	C	C	C	C	C	C
1	C	C	C	C	C	C
2	C	**0**	**0**	**0**	**0**	**0**
3	C	C	**0**	**0**	**0**	**0**
4	C	C	C	**0**	**0**	**0**

Stage 2: Look for the vertex with entry (d_1,d_2) of a triangle of asymptotic zero matrices with boundary lines $m = d_1$ and $j-m = d_2 \geq 0$ and tentatively specify the order of the model as $\rho = d_1$ and $q = d_2$. As a crude approximation, the value $(n-m-j)^{-1}$ can be used for the variance of the elements of $\hat{\rho}_{(m)}(j)$ under the hypothesis that the transformed series $\mathbf{W}_{m,t}^{(j)}$ is a white noise process. As an informative summary, (+, ·, -) signs can be used to replace numerical values of the elements of $\hat{\rho}_{m(j)}$, where a plus sign (+) is used to indicate a value greater than two times the standard deviation, a minus sign (-) to indicate a value less than minus two times the standard deviation, and a period (·) for in-between values.

Simulated and hydrological applications of using the ESCCF for multivariarte ARMA(p,q) model identification are given by Camacho et al. (1986). Consider, in particular, the simulated example. On hundred realizations of an ARMA(1,1) model with parameter matrices

$$\Phi = \begin{bmatrix} .8 & .0 \\ .5 & .7 \end{bmatrix} \quad \Theta = \begin{bmatrix} .5 & .0 \\ .0 & .5 \end{bmatrix} \quad \Delta = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix} \qquad [A20.1.14]$$

were generated using the simulation algorithm of Camacho (1984) which is similar to the one described in Section 21.4.2 for CARMA models. The plot of the data is given in Figure A20.1.1 and the pattern of the ESCCF matrices is shown in Table A20.1.3. As can be observed, the vertex of a triangle of zero matrices is located at entry (1,1), indicating that an ARMA(1,1) model would be adequate to fit the data.

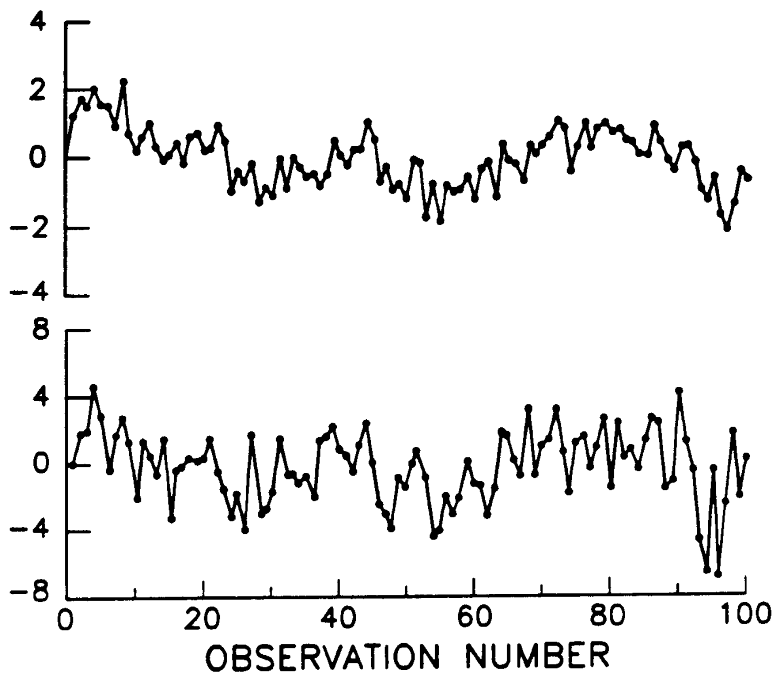

Figure A20.1.1. 100 simulated observations of an ARMA(1,1) model with parameter matrices given in [A20.1.14].

Table A20.1.3. ESCCF matrix for the simulated ARMA(1,1) data.

	MA					
AR	0	1	2	3	4	5
0	+ + + +	+ + + +	• + + +	• + • +	• + • +	• • • +
1	+ + • •	• • • •	• • • •	• • • •	• • • •	• • • •
2	+ − + −	• + • •	• • • •	• • • •	• • • •	• • • •
3	• − • •	− − • •	• − • −	• • • •	• • • •	• • • •
4	• + • +	• • • •	+ • + •	• • − •	• • • •	• • • •

PROBLEMS

20.1 The general multivariate ARMA model is defined in [20.2.1] and [20.2.2]. Write down the following vector ARMA(p,q) models using both the matrix notation and the full length description when matrices and vectors are not used.

(a) ARMA(4,0)

(b) ARMA(0,3)

(c) ARMA(2,2)

(d) ARMA(3,1)

20.2 Demonstrate that the two equations for writing a TFN model given in [20.2.3] and [17.5.3] are equivalent.

20.3 Explain why the PARMA model given in [14.2.15] can be considered as a special case of the multivariate ARMA model in [20.2.1].

20.4 By referring to [20.2.1] give the matrix equations as well as the equations where matrices and vectors are not employed for the following CARMA(p,q) models:

(a) CARMA(4,0)

(b) CARMA(0,3)

(c) CARMA(2,2)

(d) CARMA(3,1)

20.5 The general multivariate ARMA model for fitting to a set of seasonal time series is defined in [20.2.1] and [20.2.2]. Assuming that there are 5 seasons per year, define the general multivariate deseasonalized ARMA model using both the matrix notation and the full length notation when matrices and vectors are not used. Explain how you would fit this model to a set of seasonal time series by following the three stages of model construction. In your explanation be sure to mention specific graphical methods and algorithms that you would employ.

20.6 Repeat the instructions of problem 20.5 for the case of a general multivariate PAR model.

20.7 Follow the instructions of problem 20.5 for the case of a general multivariate PARMA model.

20.8 In appendix A20.1, three procedures are presented for identifying general multivariate ARMA models. By referring to this appendix and also the original references, compare the advantages and drawbacks of these approaches. Explain how they could be expanded for use with seasonal data.

20.9 Using equations when necessary, outline the approach of Hillmer and Tiao (1979) for estimating the parameters of a general multivariate ARMA(p,q) model.

20.10 Employing equations when needed, summarize the procedure of Ansley and Kohn (1983) for calibrating a general multivariate ARMA model.

20.11 Explain how the sample CCF and sample PACF described in Appendix A20.1 can be employed for checking the whiteness assumption about the residuals of a calibrated general multivariate ARMA model.

20.12 Define the modified Portmanteau test of Li and McLeod (1981) and explain how it can be utilized for testing the whiteness assumption of the residuals of a fitted general multivariate ARMA model.

20.13 Using equations when necessary, explain how multivariate normality tests proposed by Royston (1983) can be employed for testing the normality assumption for the residuals of a calibrated general multivariate ARMA model.

20.14 Select two nonseasonal time series which you suspect should be modelled using some type of multivariate ARMA model. Using the residual CCF approach presented in detail in Section 16.3.2 and outlined in Section 20.3.2, determine what kind of multivariate ARMA model could be fitted to this data.

20.15 For the two nonseasonal time series examined in problem 20.14, calibrate the most appropriate multivariate ARMA(p,q) model.

20.16 Execute the instructions of problem 20.14 for the case of two seasonal time series.

20.17 Follow the instructions of problem 20.15 for the seasonal time series examined in problem 20.16.

20.18 Mathematically define the input-output class of models put forward by Cooper and Wood (1982a,b). Explain the assets and drawbacks of this group of models, especially with respect to the general multivariate ARMA family of models. How are the input-output models mathematically related to the multivariate ARMA models?

20.19 The STARMA (space-time ARMA) class of models is discussed in Section 20.4. By referring to appropriate references, mathematically define this group of models, summarize its advantages and drawbacks, and compare it to the multivariate ARMA family of models in [20.2.1] and [20.2.2].

20.20 Define mathematically the multivariate FGN model of Matalas and Wallis (1971) referred to in Section 20.4. Is this a realistic model to employ in practice? Justify your response. You may wish to refer to the discussions on FGN given in Section 10.5 as well as appropriate references listed at the end of Chapter 10.

20.21 Within the hydrological literature there has been an ongoing and heated debate about whether one should employ disaggregation or aggregation models in hydrology. By referring to appropriate research work that is referenced in Section 20.5.2, mathematically define a disaggregation model and also an aggregation model based upon ARMA processes. Compare these two categories of models according to their relative advantages and disadvantages. Which class of models would you employ in hydrological applications?

20.22 Using mathematical equations, extend the nonGaussian model of Lewis (1985) mentioned in Section 20.5.3 to the multivariate case. Discuss any implementation problems that you may encounter when applying this mathematical model.

20.23 Define mathematically the multivariate version of the threshold model of Tong et al. (1985) referred to in Section 20.5.4. Describe the types of data to which you think this model could be applied and discuss potential model construction techniques.

20.24 Mathematically define the multivariate FARMA model of Section 20.5.5. Describe the types of model construction tools that would have to be developed to apply this model in practice.

REFERENCES

CARMA MODELS

Camacho, F. (1984). Contemporaneous CARMA Modelling with Applications. Ph.D. thesis, Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario.

Camacho, F., McLeod, A. I., and Hipel, K. W. (1985a). Contemporaneous autoregressive-moving average (CARMA) modelling in hydrology. *Water Resources Bulletin*, 21(4):709-720.

Camacho, F., McLeod, A. I., and Hipel, K. W. (1986). (July 15-17, 1985). Developments in multivariate ARMA modelling in hydrology. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes*, Fort Collins, Colorado. Engineering Research Center, Colorado State University, pages 178-197.

Camacho, F., McLeod, A. I., and Hipel, K. W. (1987a). Contemporaneous bivariate time series. *Biometrika*, 74(1):103-113.

Camacho, F., McLeod, A. I., and Hipel, K. W. (1987b). Multivariate contemporaneous ARMA models with hydrological applications. *Stochastic Hydrology and Hydraulics*, 1:141-154.

Camacho, F., McLeod, A. I., and Hipel, K. W. (1987c). The use and abuse of multivariate time series models in hydrology. In MacNeill, I. B. and Umphrey, G. J., editors, *Advances in the Statistical Sciences*, Festschrift in Honor of Prof. V. M. Joshi's 70th Birthday, Volume IV, Stochastic Hydrology, pages 27-44. D. Reidel, Dordrecht, The Netherlands.

Chan, W. T. and Wallis, K. F. (1978). Multiple time series modelling: Another look at the mink-muskrat interaction. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 27(2):168-175.

Granger, C. W. J. and Newbold, P. (1977). *Forecasting Economic Time Series*. Academic Press, New York.

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley, New York.

Risager, F. (1980). Simple correlated autoregressive process. *Scandinavian Journal of Statistics*, 7:49-60.

Risager, F. (1981). Model checking of simple correlated autoregressive processes. *Scandinavian Journal of Statistics*, 8:137-153.

Wallis, K. F. (1977). Multiple time series analysis and the final form of econometric models. *Econometrica*, 45(6):1481-1497.

CONCEPTUAL MODELS

Thompstone, R. M., Hipel, K. W. and McLeod, A. I. (1985). Forecasting quarter-monthly riverflow. *Water Resources Bulletin*, 21(5):731-741.

DISAGGREGATION AND AGGREGATION

Eagleson, P. S. (1978). Climate, soil, and vegetation, 2. The distribution of annual precipitation derived from observed storm sequences. *Water Resources Research*, 14(5):713-721.

Frevert, D. K. and Lane, W. L. (1986). Development and application of a two level spatial disaggregation procedure for the LAST statistical hydrology package. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes*, July 15-17, 1985, Fort Collins, Colorado. Engineering Research Center, Colorado State University, pages 198-216.

Gould, B. W. (1961). Statistical methods for estimating the design capacity of dams. *Journal of the Institute of Engineers in Australia*, 33(12):405-416.

Grygier, J. C. and Stedinger, J. R. (1988). Condensed disaggregation procedures and conservation corrections for stochastic hydrology. Water Resources Research, 24(10):1574-1584.

Kavvas, M. L., Cote, L. J., and Delleur, J. W. (1977). Time resolution of the hydrologic time-series models. *Journal of Hydrology*, 32:347-361.

Klemes, V. (1963). Seasonal component of storage in statistical methods of streamflow control (in Czechoslovakian). *Vodohospid. Cas.*, 11(3):232-247, 11(4):341-360.

Klemes, V. (1981). Applied stochastic theory of storage in evolution. *Advances in Hydrosciences*, Academic Press, New York, 12:79-141.

Koutsoyiannis, D. (1992). A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series. *Water Resources Research*, 28(12):3175-3191.

Lane, W. L. and Frevert, D. K. (1990). *Applied Stochastic Techniques, User Manual, LAST Personal Computer Package Version 5.2*. United States Department of the Interior, Bureau of Reclamation, Denver Office, P. O. Box 25007, Denver Federal Center, Denver, Colorado 80225-0007, Tel (303) 236-3809.

Lee, H.-L. (1986). Preserving cross-year correlations among seasonal flow series using disaggregation procedures. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes*, July 15-17, 1985, Colorado State University, Fort Collins, Colorado, pages 595-604.

Mejia, J. M. and Rousselle, J. (1976). Disaggregation models in hydrology revisited. Water Resources Research, 12(2):185-186.

Obeysekera, J. T. B. and Salas, J. D. (1982). On the aggregation and disaggregation of streamflow time series. *American Geophysical Union EOS*, 63(18):321.

Obeysekera, J. T. B. and Salas, J. D. (1986). Modeling of aggregated hydrologic time series. *Journal of Hydrology*, 86:197-219.

Rao, A. R., Rao, S. G., and Kashyap, R. L. (1985). Stochastic analysis of time-aggregated hydrologic data. *Water Resources Bulletin*, 21(5):757-771.

Savarenskiy, A. D. (1940). A method for streamflow control computation. *Gidrotekhn. Stroit.*, 2:24-28. first released in 1938 as a report of Kuybyshevskiy Stroitelnyy Instit., Kuybyshev, (in Russian).

Stedinger, J. R., Lettenmaier, D. P., and Vogel, R. M. (1985). Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. *Water Resources Research*, 21(4):497-507.

Stedinger, J. R. and Vogel, R. M. (1984). Disaggregation procedures for generating serially correlated flow vectors. *Water Resources Research*, 20(1):47-56.

Svanidze, G. G. (1962). Simulation of hydrologic series with regard for intra-annual distribution of runoff (fragments method). *Dokl. Gidrol. Konf. CHSSR* (Reports of the Hydrological Conference of the Czechoslovak Socialist Republic), pages 411-418. (See also Tr. In-ta Energetiki AN GSSR, Vol. 17, 1963, pages 273-282, and VODOHOSPODARSKY CHASOPIS, No. 6, 1963, pages 138-149.).

Svanidze, G. G. (1980). *Mathematical Modeling of Hydrologic Series for Hydroelectric and Water Resources Computations*. Water Resources Publications, Littleton, Colorado. (Translated from the Russian edition, Gidrometeoizaat, Leningrad, USSR, 1977.).

Tao, P. C. and Delleur, J. W. (1976). Multistation, multiyear synthesis of hydrologic time series by disaggregation. *Water Resources Research*, 12(6):1303-1312.

Valencia, D. R. and Schaake, J. C. (1973). Disaggregation processes in stochastic hydrology. *Water Resources Research*, 9(3):580-585.

Woolhiser, D. A. and Osborn, H. B. (1986). Point storm disaggregation - seasonal and regional effects. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes*, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes, July 15-17, 1985, Fort Collins, Colorado. Colorado State University, pages 105-120.

LONG MEMORY MODELS

Franchini, M., Todini, E., Giuliano, G., and O'Connell, P. E. (1986). A multivariate multiseasonal model for long term persistent hydrological series. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes*, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes, July 15-17, 1985, Fort Collins, Colorado. Colorado State University, pages 243-262.

Matalas, N. C., and Wallis, J. R. (1971). Statistical properties of multivariate fractional noise processes. *Water Resources Research*, 7(6):1460-1468.

O'Connell, P. E. (1974). *Stochastic Modelling of Long-Term Persistence in Stream Flow Sequences*. Ph.D. Thesis, Civil Engineering Dept., Imperial College, London.

MULTIVARIATE ARMA MODELS

Abraham, B. (1980). Intervention analysis and multiple time series. *Biometrika*, 67(1):73-78.

Ansley, C. F. and Kohn, R. (1983). Autoregressive-moving average process with missing or aggregated data. *Biometrika*, 70:275-278.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, 34:33-40.

Bernier, J. (1971). Modeles probabilistes a variables hydrologiques multiples et hydrologie synthetique. *Mathematical Models in Hydrology*, IAHS-AISH Publication No. 100.

Canfield, R. V. and Tseng, L. (September 1979). Homogeneous ARMA processes. Paper presented at the American Water Resources Association Conference, Las Vegas, Nevada.

Cooper, D. M. and Wood, E. F. (1982a). Identification of multivariate time series and multivariate input-output models. *Water Resources Research*, 18(4):937-946.

Cooper, D. M. and Wood, E. F. (1982b). Parameter estimation of multiple input-output time series models: Application to rainfall-runoff processes. *Water Resources Research*, 18(5):13 52-1364.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, 39:1-38.

Fiering, M. B. (1964). Multivariate techniques for synthetic hydrology. *Journal of the Hydraulics Division*, ASCE, 90(HY5):43-60.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424-438.

Hall, A. D. and Nicholls, D. F. (1980). The evaluation of exact maximum likelihood estimates for VARMA models. *Journal of Statistical Computation and Simulation*, 10:251-262.

Hannan, E. J. and Kavalieris, L. (1984). Multivariate linear time series models. *Advances in Applied Probability*, 16:492-561.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, Series B, 41(2):190-195.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69:81-94.

Hillmer, S. C. and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association*, 74(367):652-660.

Hipel, K. W. (1985a). Time series analysis in perspective. *Water Resources Bulletin*, 21(4)609-624.

Hipel, K. W., editor (1985b). *Time Series Analysis in Water Resources*. The American Water Resources Association, Bethesda, Maryland.

Hipel, K. W. (1986). Stochastic research in multivariate analysis. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, Keynote Address, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes,* July 15-17, 1985, Fort Collins, Colorado. Engineering Research Center, Colorado State University, pages 2-50.

Jenkins, G. M. and Alavi, A. S. (1981). Some aspects of modeling and forecasting multivariate time series. *Journal of Time Series Analysis,* 2(1):1-47.

Kahan, J. P. (1974). A method for maintaining cross and serial correlations and the coefficient of skewness under generation in a linear bivariate regression model. *Water Resources Research,* 10:1245-1248.

Kuczera, G. (1987). On maximum likelihood estimators for the multisite lag-one streamflow model: Complete and incomplete data cases. *Water Resources Research,* 23(4):641-645.

Lawrance, A. J. (1976). A reconsideration of the fiering two-station model. *Journal of Hydrology,* 29:77-85.

Ledolter, J. (1978). The analysis of multivariate time series applied to problems in hydrology. *Journal of Hydrology,* 36:327-352.

Lettenmaier, D. P. (1980). Parameter estimation of multivariate streamflow synthesis. In *Proceedings of the Joint Automatic Control Conference,* August 13-15, 1980, paper number FA6-D, San Francisco.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society,* Series B, 43(2):231-239.

Maas, A., Hufschmidt, M. M., Dorfman, R., Thomas Jr., H. A., Marglin, S. A., and Fair, M. G. (1962). *Design of Water-Resource Systems.* Harvard University Press, Cambridge, Massachusetts.

Matalas, N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research,* 3(4):937-945.

Newbold, P. and Hotopp, S. M. (1984). Multiple time series model building and testing for causality. Working paper, Department of Economics, University of Illinois, Urbana-Champaign.

Nicholls, D. F. and Hall, A. D. (1979). The exact likelihood of multivariate autoregressive-moving average models. *Biometrika,* 66:259-264.

Pierce, D. A. and Haugh, L. D. (1977). Causality in temporal systems. *Journal of Econometrics,* 5:265-293.

Quinn, B. G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society,* Series B, 42:182-185.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Journal of the Royal Statistical Society,* Series C (Applied Statistics), 32(2):121-133.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). *Applied Modelling of Hydrologic Time Series.* Water Resources Publications, Littleton, Colorado.

Salas, J. D. and Pegram, G. G. S. (1978). A seasonal multivariate multilag autoregressive model in hydrology. In Morel-Seytoux, H., Salas, J. D., Sanders, T. G., and Smith, R. E., editors, *Modeling Hydrologic Processes*. Water Resources Publications, Littleton, Colorado.

Salas, J. D., Tabios III, G. Q., and Bartolini, P. (1985). Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin*, 21(4).

Shea, B. L. (1989). AS242 the exact likelihood of a vector autoregressive moving average model. *Applied Statistics*, 38(1):161-184.

Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G. (1986). *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium* held at Colorado State University, Fort Collins, Colorado, July 15-17, 1985, Fort Collins, Colorado. Engineering Research Center, Colorado State University.

Stedinger, J. R. (1981). Estimating correlations in multivariate streamflow models. *Water Resources Research*, 17(1):200-208.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76:802-816.

Tiao, G. C. and Tsay, R. S. (1983a). Multiple time series modeling and extended cross-correlations. *Journal of Business and Economic Statistics*, 1:43-56.

Tiao, G. C. and Tsay, R. S. (1983b). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics*, 11:856-871.

Tjostheim, D. and Paulsen, J. (1982). Empirical identification of multiple time series. *Journal of Time Series Analysis*, 3(4):265-282.

Tsay, R. S. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, 79:84-96.

Wilson, G. T. (1973). The estimation of parameters in multivariate time series models. *Journal of the Royal Statistical Society*, Series B, 35(1):76-85.

Young, G. K. and Pisano, W. C. (1968). Operational hydrology using residuals. *Journal of the Hydraulics Division*, ASCE, 94(HY4):909-923.

NONGAUSSIAN MODELS

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26:211-252.

Brillinger, D. R. (1985). Fourier inference: Some methods for the analysis of array and nonGaussian series data. *Water Resources Bulletin*, 21(5):743-756.

Lewis, P. A. W. (1985). Some simple models for continuous variate time series. *Water Resources Bulletin*, 21(4):635-644.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21(4):645-650.

Li, W. K. and McLeod, A. I. (1988). ARMA modelling with nonGaussian innovations. *Journal of Time Series Analysis*, 9(2):155-168.

NONLINEAR MODELS

Gallant, A. R. (1987) *Nonlinear Statistical Models*. Wiley, New York.

Li, W. K. (1992). On the asymptotic standard errors residual autocorrelations in nonlinear time series modelling. *Biometrika*, 79(2):435-437.

Li, W. K. (1993). A simple one degree of freedom test for non-linear time series model discrimination. *Statistica Sinica*, 3(1):245-254.

Ozaki, T. (1985). Statistical identification of storage models with application to stochastic hydrology. *Water Resources Bulletin*, 21(4).

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag, New York.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Tong, H., Thanoon, B., and Gudmundsson, G. (1985). Threshold time series modelling of two Icelandic riverflow systems. *Water Resources Bulletin*, 21(4):651-661.

PATTERN RECOGNITION

MacInnes, C. D. and Unny, T. E. (1986). Multivariate hydrologic time series and pattern recognition. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, held at Colorado State University, Fort Collins, Colorado, July 15-17, 1985, Engineering Research Center, Colorado State University, pages 228-242.

Panu, U. S. and Unny, T. E. (1980). Stochastic synthesis of hydrologic data based on concepts of pattern recognition, Parts I, II and III. *Journal of Hydrology*, 46:5-34, 197-217, and 219-237.

Unny, T. E., Panu, U. S., MacInnes, C. D., and Wong, A. K. C. (1981). Pattern analysis and synthesis of time-dependent hydrologic data. *Advances in Hydroscience*, 12:195-295.

PERIODIC MODELS

Bartolini, P. and Salas, J. D. (1986). Properties of multivariate periodic ARMA(1,1) processes. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, Fort Collins, Colorado, July 15-17, 1985, Engineering Research Center, Colorado State University, pages 264-282.

Bartolini, P., Salas, J. D., and Obeysekera, J. T. B. (1988). Multivariate periodic ARMA(1,1) processes. *Water Resources Research*, 24(8):1237-1246.

Haltiner, J. P. and Salas, J. D. (1988). Development and testing of a multivariate, seasonal ARMA(1,1) model. *Journal of Hydrology*, 104:247-272.

Hipel, K. W. and McLeod, A. I. (1981). Box-Jenkins modelling in the geophysical sciences. In Craig, R. G. and Labovitz, M. L., editors, *Future Trends in Geomathematics*, pages 65-86. Prion, United Kingdom.

Salas, J. D. and Abdelmohsen, M. W. (1993). Initialization for generating single-site and multisite low-order periodic autoregressive and moving average processes. *Water Resources Research*, 29(6):1771-1776.

Salas, J. D. and Pegram, G. G. S. (1978). A seasonal multivariate multilag autoregressive model in hydrology. In Morel-Seytoux, H., Salas, J. D., Sanders, T. G., and Smith, R. E., editors, *Modelling Hydrologic Processes*. Water Resources Publications, Littleton, Colorado.

Thomas, H. A. and Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In Maass, A., Hufshmidt, M. M., Dorfman, R., Thomas Jr., H. A., Marglin, S. A., and Fair, M. G., editors, *Design of Water Resources Systems*, Harvard University Press, Cambridge, Massachusetts, pages 459-493.

Thompstone, R. M., Hipel, K. W., and McLeod, A. I. (1985a). Forecasting quarter-monthly riverflow. *Water Resources Bulletin*, 21(5):731-741.

Thompstone, R. M., Hipel, K. W., and McLeod, A. I. (1985b). Grouping of periodic autoregressive models. In Anderson, O. D., Ord, J. K., and Robinson, E. A., editors, *Time Series Analysis: Theory and Practice* 6. North-Holland, Amsterdam, pages 35-49.

Ula, T. A. (1990). Periodic covariance stationarity of multivariate periodic autoregressive moving average processes. *Water Resources Research*, 26(5):855-861.

Vecchia, A. V. (1985a). Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin*, 21(5):721-730.

Vecchia, A. V. (1985b). Maximum likelihood estimation for periodic autoregressive-moving average models. *Technometrics*, 27:375-384.

Vecchia, A. V., Obeysekera, J. T., Salas, J. D., and Boes, D. C. (1983). Aggregation and estimation for low-order periodic ARMA models. *Water Resources Research*, 19(5):1297-1306.

PHYSICAL BASIS OF MODELS

Klemes, V. (1978). Physically based stochastic hydrologic analysis. In Chow, V. T., editor, *Advances in Hydroscience*, 11:285-356.

Salas, J. D. and Smith, R. A. (1981). Physical basis of stochastic models of annual flows. *Water Resources Research*, 17(2):428-430.

Yevjevich, V. and Harmancioglu, N. B. (1985). Past and future of analysis of water resources time series. *Water Resources Bulletin*, 21(4):625-633.

SIMULATION

Davis, D. R. (1977). Comment on Bayesian generation of synthetic streamflows by G. J. Vicens, I. Rodriguez-Iturbe and J. C. Schaake Jr. *Water Resources Research*, 13(5):853-854.

Feldman, A. D. (1981). HEC models for water resources system simulation: Theory and experience. *Advances in Hydroscience*, 12:297-423.

Finzi, G., Todini, E., and Wallis, J. R. (1975). Comment upon multivariate synthetic hydrology. *Water Resources Research*, 11(6):844-850.

Kelman, J. (1980). A stochastic model for daily streamflow. *Journal of Hydrology*, 47:235-249.

Kelman, J., Damazio, J. M., and Costa, J. P. (1986). A multivariate synthetic daily streamflow generator. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes*, July 15-17, 1985, Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 121-140.

Kottegoda, N. T. and Yevjevich, V. (1977). Preservation of correlation in generated hydrologic samples through two-station models. *Journal of Hydrology*, 33:99-121.

McLeod, A. I. and Hipel, K. W. (1978c). Simulation procedures for Box-Jenkins models. *Water Resources Research*, 14(5):969-975.

Mejia, J. M., Rodriquez-Irturbe, I., and Cordova, J. R. (1974). Multivariate generation of mixtures of normal and log normal variables. *Water Resources Research*, 10(4):691-693.

Nasseri, I. (1986). Multistation stochastic model of hourly rainfall. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985, Fort Collins, Colorado, Engineering Research Center, Colorado State University, pages 531-541.

Pegram, G. G. S. and James, W. (1972). Multilag multivariate autoregressive model for the generation of operational hydrology. *Water Resources Research*, 8(4):1074-1076.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature and solar radiation data. *Water Resources Research*, 17(1):182-190.

Srikanthan, R. (1986). Stochastic simulation of daily climatic data. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985, Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 542-554.

Stedinger, J. R. and Taylor, M. R. (1982a). Synthetic streamflow generation, 1, Model verification and validation. *Water Resources Research*, 18(4):909-918.

Stedinger, J. R. and Taylor, M. R. (1982b). Synthetic streamflow generation, 2, Parameter uncertainty. *Water Resources Research*, 18(4):919-924.

Valdes, J. B., Rodriguez-Iturbe, I., and Vicens, G. J. (1977). Bayesian generation of synthetic streamflows, 2, Multivariate case. *Water Resources Research*, 13(2):291-295.

Venugopal, K., Ranganathan, T., Babu Rao, T., Raghavendran, R., and Sakthivadivel, R. (1986). Stochastic multisite modelling of Narmada River system in India. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Symposium*, July 15-17, 1985, Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 664-677.

Vicens, G. J., Rodriguez-Iturbe, I., and Schaake Jr., J. C. (1975). Bayesian generation of synthetic stream-flows. *Water Resources Research*, 11(6):827-838.

SPECTRUM

Canfield, R. V. and Bowles, D. S. (1986). Multivariate series generation from the spectral density function. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985, Engineering Research Center, University of Colorado, Fort Collins, Colorado, pages 572-581.

Ghani, A. A. A. and Metcalfe, A. V. (1986). Spectral predictions of flood risk. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985, Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 744-754.

Jenkins, G. M. and Watts, D. G. (1968). Spectral Analysis and its Applications. Holden-Day, San Francisco.

STARMA MODELS

Adamowski, K., Mohamed, F. B., Dalezios, N. R. and Birta, L. G. (1986). Space-time ARIMA modelling of precipitation time series. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985. Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 217-227.

Deutsch, S. J. and Pfeifer, P. E. (1981). Space-time ARMA modelling with contemporaneously correlated innovations. *Technometrics*, 23(4):401-409.

Deutsch, S. J. and Ramos, J. A. (1984). Space time evaluation of reservoir regulation policies. Technical report, School of Civil Engineering, Georgia Institute of Technology.

Deutsch, S. J. and Ramos, J. A. (1986). Space-time modeling of vector hydrologic sequences. *Water Resources Bulletin*, 22(6):967-981.

Pfeifer, P. E. and Deutsch, S. J. (1980). Identification and interpretation of first-order space-time ARMA models. *Technometrics*, 22:397-408.

TFN MODELS

Delleur, J. W. (1986). Recursive parameter identification for flashflood forecasting. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985. Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 154-177.

Nicklin, M. E. (1986). Bivariate transfer function modeling of irrigation return flows. In Shen, H. W., Obeysekera, J. T. B., Yevjevich, V., and DeCoursey, D. G., editors, *Multivariate Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, July 15-17, 1985. Engineering Research Center, Colorado State University, Fort Collins, Colorado, pages 620-637.