CHAPTER 21

CONTEMPORANEOUS AUTOREGRESSIVE-MOVING AVERAGE MODELS

21.1 INTRODUCTION

The *contemporaneous ARMA, or CARMA, family of models* is designed for modelling two or more time series that are statistically related to one another only at the same time, or simultaneously. For example, two riverflow series that are measured within the same climatic zone but not at locations where one station is upstream from the other, may be only correlated simultaneously with one another. As demonstrated by the applications given in Section 21.5 of this chapter, a CARMA model is the most appropriate type of multivariate model to describe this situation mathematically.

Because of the usefulness of CARMA modelling in water resources, this chapter is devoted entirely to presenting this interesting and simple model. As explained in Section 20.2.2 of the previous chapter, CARMA models actually form a special type of *general multivariate ARMA models*. Besides possessing far fewer parameters than the general multivariate ARMA models described in detail in Chapter 20, CARMA models can be conveniently fitted to multiple time series using well developed model construction techniques.

Another useful subset of models from the general multivariate ARMA family is the group of *TFN models* which includes the closely related *intervention models*. TFN models can be employed when a single output series is dependent upon one or more input series plus a noise component. If a single output series is affected by one or more external interventions and perhaps also some input series, an intervention model can parsimoniously describe this situation. Along with many interesting applications, TFN models are presented in Part VII while intervention models are discussed in detail in Part VIII and Section 22.4.

Descriptions of the historical development of multivariate models in water resources are presented in Sections 20.4 and 20.5 as well as in the papers by Salas et al. (1985) and Hipel (1986). In addition to other types of multivariate models, many references are listed at the end of Chapter 20 for previous research in CARMA modelling. Much of the material presented in this chapter is drawn from research completed by Camacho et al. (1985, 1986, 1987a,b,c) and Camacho (1984).

In the next section, two alternative approaches to deriving the equations for CARMA models are presented. Following this, a comprehensive set of *model construction tools* are described in Section 21.3. To avoid introducing bias into synthetic sequences, a correct method for *generating simulated data* from a CARMA model is presented in Section 21.4. The *practical applications* in Section 21.5 demonstrate how convenient and simple it is to use the building methods for properly describing both water quantity and quality time series.

21.2 DERIVING CARMA MODELS

21.2.1 Introduction

CARMA models can be defined using two distinct viewpoints. Firstly, as noted in Section 20.2.2, the CARMA group of models can be thought of as being a subset of the general multivariate ARMA family of models. Instead of going from a more general class of models to a more specific subset of models, the second approach for defining the CARMA group of models goes in the reverse direction. In particular, a CARMA model can be considered as a collection of, say, k univariate ARMA models with contemporaneously correlated innovations. This second interpretation is particularly useful for the development of model construction tools, especially computationally efficient estimation algorithms. The subset and concatenation definitions of the CARMA group of models are now presented.

21.2.2 Subset Definition

The mathematical definition for the general multivariate ARMA family of models is given in [20.2.1] and [20.2.2]. By constraining the AR and MA parameter matrices to be diagonal matrices, the CARMA subset of models is defined. More specifically, following the notation of Section 20.2.2, let k time series at time t be represented 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 $\mu = (\mu_1, \mu_2, \ldots, \mu_k)^T$. Assuming that the orders of the AR and MA components are p and q, respectively, the *CARMA(p,q) model* can be written as

$$(\mathbf{Z}_t - \mu) - \begin{bmatrix} \phi_{111} & & & & \\ & \phi_{221} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \phi_{kk1} \end{bmatrix} (\mathbf{Z}_{t-1} - \mu) - \begin{bmatrix} \phi_{112} & & & & \\ & \phi_{222} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \phi_{kk2} \end{bmatrix} (\mathbf{Z}_{t-2} - \mu)$$

$$- \cdots - \begin{bmatrix} \phi_{11p} & & & & \\ & \phi_{22p} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \phi_{kkp} \end{bmatrix} (\mathbf{Z}_{t-p} - \mu) \qquad [21.2.1]$$

$$= \mathbf{a}_t - \begin{bmatrix} \theta_{111} & & & & \\ & \theta_{221} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \theta_{kk2} \end{bmatrix} \mathbf{a}_{t-1} - \begin{bmatrix} \theta_{112} & & & & \\ & \theta_{222} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \theta_{kk2} \end{bmatrix} \mathbf{a}_{t-2} - \cdots$$

$$- \begin{bmatrix} \theta_{11q} & & & & \\ & \theta_{22q} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \theta_{kkq} \end{bmatrix} \mathbf{a}_{t-q}$$

where

$$\Phi_i = \begin{bmatrix} \phi_{11i} & & & & \\ & \phi_{22i} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \phi_{kki} \end{bmatrix}$$

is the AR parameter matrix for $i = 1,2, \ldots, p$, having zero entries for all the off diagonal elements;

$$\Theta_i = \begin{bmatrix} \theta_{11i} & & & & \\ & \theta_{22i} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \theta_{kki} \end{bmatrix}$$

is the MA parameter matrix for $i = 1,2, \ldots, q$, possessing zero values for all the non-diagonal elements; and

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

is the k dimensional vector of innovations for $\mathbf{Z}_t$ at time t. Notice that the model in [21.2.1] has the same form as the general multivariate ARMA model in [20.2.1] and [20.2.2], except for the fact that the AR and MA parameter matrices are diagonal.

After performing the matrix multiplications in [21.2.1], one obtains a set of k simultaneous difference equations. In particular, the ith difference equation for the variable Z_{ti} is

$$Z_{ti} - \mu_i - \phi_{ii1}(Z_{t-1,i} - \mu_i) - \phi_{ii2}(Z_{t-2,i} - \mu_i) - \cdots - \phi_{iip}(Z_{t-p,i} - \mu_i)$$
$$= a_{ti} - \theta_{ii1}a_{t-1,i} - \theta_{ii2}a_{t-2,i} - \cdots - \theta_{iiq}a_{t-q,i}, \text{ for } i = 1,2, \ldots, k \quad [21.2.2]$$

Notice from [21.2.2] that only the ith variable and ith innovation series appear in the equation. The simultaneous correlation among the k variables is incorporated into the CARMA model by allowing the innovations to be contemporaneously correlated. More precisely, the vector of innovations given by $\mathbf{a}_t$ are assumed to be IID vector random variables with a mean of zero and variance covariance matrix given by $\Delta = E[\mathbf{a}_t \cdot \mathbf{a}_t^T]$. For practical applications, the normality assumption is invoked and $\mathbf{a}_t \sim NID(\mathbf{0},\Delta)$.

The model in [21.2.2] can be more compactly written as

$$\phi_i(B)(Z_{ti} - \mu_i) = \theta_i(B)a_{ti}, \quad i = 1,2, \ldots, k \quad [21.2.3]$$

where

$$\phi_i(B) = 1 - \phi_{ii1}B - \phi_{ii2}B^2 - \cdots - \phi_{iip}B^p$$

is the ith AR operator of order p and

$$\theta_i(B) = 1 - \theta_{ii1}B - \theta_{ii2}B^2 - \cdots - \theta_{iiq}B^q$$

is the ith MA operator of order q. For the CARMA model to be stationary and invertible, the zeroes of the characteristic equations $\phi_i(B) = 0$ and $\theta_i(B) = 0$, respectively, must lie outside the unit circle.

Example: Consider a bivariate CARMA(1,1) model for connecting the two variables contained in the vector

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

having theoretical means given by

$$\mu = (\mu_1, \mu_2)^T$$

From [21.2.1], the bivariate CARMA(1,1) model is written as

$$\begin{pmatrix} Z_{t1} - \mu_1 \\ Z_{t2} - \mu_2 \end{pmatrix} - \begin{bmatrix} \phi_{111} & 0 \\ 0 & \phi_{221} \end{bmatrix} \begin{pmatrix} Z_{t-1,1} - \mu_1 \\ Z_{t-1,2} - \mu_2 \end{pmatrix} = \begin{pmatrix} a_{t1} \\ a_{t2} \end{pmatrix} - \begin{bmatrix} \theta_{111} & 0 \\ 0 & \theta_{221} \end{bmatrix} \begin{pmatrix} a_{t-1,1} \\ a_{t-1,2} \end{pmatrix} \quad [21.2.4]$$

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

$$Z_{t1} - \mu_1 - \phi_{111}(Z_{t-1,1} - \mu_1) = a_{t1} - \theta_{111}a_{t-1,1}$$
$$Z_{t2} - \mu_2 - \phi_{221}(Z_{t-1,2} - \mu_2) = a_{t2} - \theta_{221}a_{t-1,2} \quad [21.2.5]$$

The vector of innovations for the bivariate model is

$$\mathbf{a}_t = (a_{t1}, a_{t2})^T$$

where the variance covariance matrix for $\mathbf{a}_t$ is

$$\Delta = E[\mathbf{a}_t \cdot \mathbf{a}_t^T] = E\left[\begin{pmatrix} a_{t1} \\ a_{t2} \end{pmatrix}(a_{t1}\ a_{t2})\right]$$

$$= \begin{bmatrix} E[a_{t1}^2] & E[a_{t1}a_{t2}] \\ E[a_{t2}a_{t1}] & E[a_{t2}^2] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Because $\sigma_{21} = \sigma_{12}$, then Δ is a symmetric matrix. Under the normality assumption $\mathbf{a}_t \sim NID(\mathbf{0}, \Delta)$.

To satisfy the stationarity assumption, the roots of

$$\phi_1(B) = 1 - \phi_{111}B = 0$$

and

$$\phi_2(B) = 1 - \phi_{221}B = 0$$

must lie outside the unit circle. Consequently, $|\phi_{111}| < 1$ and $|\phi_{221}| < 1$. For invertibility, the roots of

$$\theta_1(B) = 1 - \theta_{111}B = 0$$

and

$$\theta_2(B) = 1 - \theta_{221}B = 0$$

must lie outside the unit circle. Hence, $|\theta_{111}| < 1$ and $|\theta_{221}| < 1$ for satisfying the invertibility condition.

21.2.3 Concatenation Definition

The clue to discovering the second approach to defining a CARMA model is given by the form of the component equation in [21.2.2] and [21.2.3]. Notice that the model in [21.2.3] for the ith variable is in fact an ARMA model and is identical to the ARMA model defined in [3.4.3] and [3.4.4]. Accordingly, one can consider the CARMA model to consist of a concatenation of k ARMA models where there is a separate ARMA model to describe each of the k series. In general, the orders of the AR and MA operators may vary across the k models. Therefore, the ARMA model for Z_{ti} can be written more precisely as

$$\phi_i(B)(Z_{ti} - \mu_i) = \theta_i(B)a_{ti}\ ,\ i = 1,2, \ldots, k \qquad [21.2.6]$$

where

$$\phi_i(B) = 1 - \phi_{ii1}B - \phi_{ii2}B^2 - \cdots - \phi_{iip_i}B^{p_i}$$

is the ith AR operator of order p_i and

$$\theta_i(B) = 1 - \theta_{ii1}B - \theta_{ii2}B^2 - \cdots - \theta_{iiq_i}B^{q_i}$$

is the ith MA operator of order q_i. The chain that links the k ARMA models together in terms of contemporaneous correlation is the variance covariance matrix, Δ, for the innovations $\mathbf{a}_t = (a_{t1}, a_{t2}, \ldots, a_{tk})^T$ where $\mathbf{a}_t \sim NID(\mathbf{0}, \Delta)$. When using the notation CARMA(p,q) to stand for the overall model, one sets $p = \max(p_1, p_2, \ldots, p_k)$ and $q = \max(q_1, q_2, \ldots, q_k)$. By constraining appropriate parameters to be zero in the subset definition of the CARMA(p,q) model in [21.2.1], one can also allow the orders of the AR and MA operators to vary when using this equivalent definition.

In summary, from [21.2.6], the CARMA model can be thought of as a set of k univariate ARMA models for which the innovations are contemporaneously correlated. This contemporaneous correlation is modelled using the variance covariance matrix, Δ, which has a typical entry denoted by σ_{ij}. For the situation where none of the series are contemporaneously correlated with each other, $\sigma_{ij} = 0$ for $i \neq j$ and the multivariate CARMA model collapses into a collection of k independent univariate ARMA models. Consequently, one can interpret the CARMA model as a natural extension of the univariate ARMA model. Alternatively, under the subset definition in Section 21.2.2, the CARMA model can be considered to be a special case of a more general family of models.

21.3 CONSTRUCTING CARMA MODELS

21.3.1 Introduction

Because flexible and simple model construction procedures are now available for fitting CARMA models to a data set, it is currently possible for practitioners to conveniently employ these models in practical applications. Some of the techniques used at the three stages of model construction have naturally evolved from the concatenation interpretation of the CARMA model presented in [21.2.6]. Consequently, construction methods used for fitting univariate ARMA models have been cleverly extended for use with CARMA models for which there are contemporaneous correlations among the innovation series. Specific details and a comprehensive list of references regarding the available procedures for use in model fitting can be found in papers by authors such as Camacho et al. (1985, 1986, 1987a,b,c), Salas et al. (1985), Hipel (1986) and Jenkins and Alavi (1981). In this section, some of the most useful model construction tools are described.

21.3.2 Identification

A sound *physical understanding* of a given problem in conjunction with a thorough appreciation of the capabilities of the various types of multivariate ARMA models, are of utmost importance in model identification. For instance, when riverflows from different river basins are controlled by the same general climatic conditions within an overall region, a CARMA model may be appropriate to use with this multisite data. The *residual CCF* is a very useful statistical tool for ascertaining statistically whether or not a CARMA model is needed to fit to two or more time series and also to decide upon the orders of the AR and MA parameters. In Section 16.2.2, the theoretical and sample residual CCF functions are defined in [16.2.5] and [16.2.6], respectively, and it is explained how the residual CCF can be used to determine the type of causality existing between two series and thereby confirm in a statistical manner what one may suspect a priori from a physical understanding of the problem. A summary of the use of the residual CCF in causality studies is given in Table 16.2.1 and also in Section 20.3.2 under the heading Causality.

Suppose that one has a set of k time series given by

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

where each series has n equally spaced observations that are available at the same time as the other series. For the ith time series, the data set is given as $[Z_{1i}, Z_{2i}, \ldots, Z_{ni}]$. Using the sample residual CCF for model identification involves the following two steps.

Step 1 - Fitting Univariate ARMA Models: Using the ARMA model construction procedures of Part III, the most appropriate ARMA model is fitted separately to each of the data sets $\left\{Z_{1i}, Z_{2i}, \ldots, Z_{ni}\right\}$, $i = 1,2,\ldots,k$. This step produces a residual series

$$\left\{\bar{a}_{ti}\right\} = \left\{\bar{a}_{1i}, \bar{a}_{2i}, \ldots, \bar{a}_{ni}\right\}$$

for each series $i = 1,2,\ldots,k$, for the univariate ARMA model in [21.2.6] or [3.4.3]. Obtaining the residuals of an ARMA model fitted to a given series is referred as prewhitening in Sections 16.2.2 and 20.3.2. Besides the residuals, a vector of parameter estimates given by

$$\bar{\boldsymbol{\beta}}_i = (\bar{\phi}_{ii1}, \bar{\phi}_{ii2}, \ldots, \bar{\phi}_{iip_i}, \bar{\theta}_{ii1}, \bar{\theta}_{ii2}, \ldots, \bar{\theta}_{iiq_i})^T$$

is found for each of the series $i = 1,2,\ldots,k$. The bar above a variable or parameter means that the variable or parameter has been estimated using an efficient univariate estimation procedure from Section 6.2.3 or Appendix A6.1.

Step 2 - Analysis of the Residual CCF: As explained in Section 16.2.2, to determine statistically the type of causality existing between two series Z_{ti} and Z_{tj}, one examines the *residual CCF* which is calculated for the residuals series $[\bar{a}_{ti}]$ and $[\bar{a}_{tj}]$. Following [16.2.6], the residual CCF is determined for lag l as

$$\bar{r}_{ij}(l) = \bar{c}_{ij}(l) \,/\, [\bar{c}_{ii}(0)\bar{c}_{jj}(0)]^{1/2}$$

where

$$\bar{c}_{ij}(l) = \begin{cases} n^{-1}\sum_{t=1}^{n-l} \bar{a}_{ti}\bar{a}_{t+l,j}\,, & \text{for } l \geq 0 \\ n^{-1}\sum_{t=1-l}^{n} \bar{a}_{ti}\bar{a}_{t+l,j}\,, & \text{for } l < 0 \end{cases} \qquad [21.3.1]$$

is the estimated cross covariance function at lag l between the two residual series, and $\bar{c}_{ii}(0)$ and $\bar{c}_{jj}(0)$ are the estimated variances of the ith and jth residual series, respectively.

The residual CCF can be calculated for negative, zero and positive lags for all possible pairs of series. If a CARMA model is adequate for modelling the data, only the residual CCF at lag zero should be significantly different from zero. If this is not the case, a more complicated model such as a TFN model (see Chapter 17) or a general multivariate ARMA model (see Chapter 20) may be needed. Under the hypothesis that the CARMA model is adequate for describing the data, the quantities $\pm 2/n^{1/2}$ can be considered as approximate 95% confidence limits to decide whether a value of the residual CCF is significant or not. The test for the significance of the cross correlations can be easily performed by plotting the residual CCF

$$\bar{r}_{ij}(l),\ l = 0, \pm 1, \pm 2, \ldots, \pm m$$

where $m < n/4$ together with the 95% confidence limits for each distinct pair of residual series. If a CARMA model is appropriate for modelling the series, only the residual CCF at lag zero will be significantly different from zero for all pairs of series.

An alternative to plotting the residual CCF's is to summarize the significance of each value of the residual CCF in the *residual CCF matrix* denoted by $\bar{\mathbf{R}}(l) = [\bar{r}_{ij}(l)]$. Because there are k series, the dimension of $\bar{\mathbf{R}}(l)$ is $k \times k$ where the (i,j) entry gives the result for the ith and jth series. Also, since $\bar{r}_{ij}(l) = -\bar{r}_{ji}(l)$, one only has to determine the residual CCF matrices $\bar{\mathbf{R}}(l)$ for zero and positive lags so that $l = 0, 1, 2, \ldots, m$. For convenience in detecting significant values in $\bar{\mathbf{R}}(l)$, each $\bar{r}_{ij}(l)$ entry can be replaced by a "+" to indicate a value greater than $2n^{-1/2}$ or or by a "-" to point out a value smaller than $-2n^{-1/2}$ or by a "." to indicate a value falling between $-2n^{-1/2}$ and $2n^{1/2}$. Thus, "+", "-" and "." stand for values significantly greater, significantly less and not significantly different from zero, respectively. If the approximate 95% confidence interval given by $(-2n^{1/2}, 2n^{1/2})$ is not considered to be accurate enough, exact confidence intervals could be calculated (Li and McLeod, 1981), although this is not usually necessary.

In summary, from Step 1, one knows the number of AR and MA parameters required to model each series and one has univariate estimates of these parameters. If the residual CCF calculated in Step 2 for each pair of series is only significantly different from zero at lag zero, then a CARMA model is the most appropriate type of multivariate model to fit to the k series. An advantage of using the residual CCF as an identification technique is that it may indicate the direction of departure from the CARMA model, if this model is not adequate to fit the data. For example, if the residual CCF were significantly different from zero for lag zero and also a few positive lags but not significant for any negative lags, this may indicate that a TFN model is required (see Part VII). Although one could also use the model identification techniques described in Appendix A20.1 of the previous chapter, these techniques are not as convenient to use as the residual CCF, especially when one suspects from a physical viewpoint that a CARMA model is needed.

21.3.3 Estimation

After a tentative model has been identified, the next step is to estimate the parameters of the model. General multivariate ARMA estimation procedures based upon maximum likelihood, such as the methods of Hillmer and Tiao (1979) and Nicholls and Hall (1979) referred to in Section 20.3.2, could be employed to estimate the parameters of the CARMA model. It should be pointed out, however, that these algorithms are not computationally efficient for the estimation of the parameters of the CARMA model, and efficient algorithms can be readily obtained, as explained below. On the other hand, the univariate estimates $\bar{\beta}_i$ obtained for each of the series Z_{ti}, $i = 1,2,\ldots,k$ in Step 1 in Section 21.3.2, do not provide statistically efficient estimators of the parameters in the overall CARMA model. This is because the variance of the estimated AR and MA parameters contained in $\bar{\beta}_i$ for the ith time series may be quite high. Camacho (1984) and Camacho et al. (1987a,b) show theoretically that the variances of the univariate estimators $\bar{\beta}_i$, $i = 1,2,\ldots,k$, are greater than the variances of the estimators obtained using the joint multivariate estimation algorithm described below. In some cases, the univariate estimators are much less efficient than the joint multivariate estimators.

To overcome the aforementioned inefficiencies of the estimation techniques, Camacho et al. (1987a,b) developed an algorithm to obtain efficient MLE's of the model parameters. As is also assumed at the identification stage, let $\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tk})^T$ for $t = 1,2,\ldots,n$ be a sample of n consecutive observations for the k time series Z_{ti}, $i = 1,2,\ldots,k$. Hence, for the ith time series the set of observations is given as

$$\{Z_{ti}\} = \{Z_{1i}, Z_{2i}, \ldots, Z_{ni}\}$$

Let the parameters of the CARMA model for the ith series be contained in the vector

$$\beta_i = (\phi_{ii1}, \phi_{ii2}, \ldots, \phi_{iip_i}, \theta_{ii1}, \theta_{ii2}, \ldots, \theta_{iiq_i})^T$$

Consequently, the vector of parameters for the complete CARMA model is written as

$$\beta = (\beta_1, \beta_2, \ldots, \beta_k)^T$$

The CARMA estimation algorithm consists of the following steps:

1. For each series Z_{ti}, $i = 1,2,\ldots,k$, obtain univariate estimates of the ARMA model parameters using univariate ARMA estimation techniques such as those by Newbold (1974), Ansley (1979), Ljung and Box (1979), and McLeod (1977) referred to in Section 6.2.3. The ARMA estimator of McLeod is described in Appendix A6.1. In Step 1 of the identification stage of the previous section, the univariate estimates $\bar{\beta}_i$, $i = 1,2,\ldots,k$, are already found in order to produce the prewhitened series for each fitted ARMA model. Recall that a bar written above a vector indicates that an efficient univariate estimator has been employed to obtain estimates of the parameters contained in the vector.

2. Calculate

$$\beta^* = \bar{\beta} - V(\bar{\beta})(\partial S/\partial \beta)\Big|_{\beta=\bar{\beta}} \qquad [21.3.2]$$

where $\bar{\beta} = (\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_k)^T$ is the vector of univariate estimates for which $\bar{\beta}_i$ is the vector of univariate estimates of the ARMA model for the ith series;

$$V(\bar{\beta})^{-1} = plim(\partial^2 S/\partial\beta\partial\beta^T)\Big|_{\beta=\bar{\beta}}$$

is the inverse of the variance-covariance matrix for the parameters contained in $\bar{\beta}$, which is the information matrix;

$$(\partial S/\partial \beta)|_{\beta=\bar{\beta}}$$

denotes the vector of partial derivates of the sum of squares function, S, with respect to the CARMA parameter β, evaluated at the point $\beta=\bar{\beta}$. The sum of squares function S, is defined as

$$S = \sum_{t=1}^{n} \mathbf{a}_t^T \Delta^{-1} \mathbf{a}_t / 2n$$

where

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

$$\Delta = (\sigma_{ij})$$

is the variance covariance matrix of $\mathbf{a}_t$, and a_{ti} is defined in

$$a_{ti} = Z_{ti} - \phi_{ii1} Z_{t-1,i} - \phi_{ii2} Z_{t-2,i} - \cdots - \phi_{iip_i} Z_{t-p_i,i}$$

$$+ \theta_{ii1} a_{t-1,i} + \theta_{ii2} a_{t-2,i} + \cdots + \theta_{iiq_i} a_{t-q_i,i}\,, \quad t \geq p_i$$

Initial values for a_{ti} can be calculated using the algorithm given by McLeod and Sales (1983) or can be set equal to zero. To calculate the information matrix or, equivalently, $V(\beta)^{-1}$ which is the inverse of the variance-covariance matrix of β, the algorithm given by Ansley (1980) and Kohn and Ansley (1982) can be employed. Camacho (1984) proves that β^* is asymptotically efficient. Using $\bar{\beta}$ as the initial point, the estimation procedure corresponds to one iteration of the Gauss Newton optimization scheme. To obtain the maximum likelihood estimator, $\hat{\beta}$, for the complete CARMA model, iterations can be continued until convergence is reached.

Camacho et al. (1985, 1987a) extend their estimation algorithm for CARMA models to include the situation where the multiple time series have unequal sample sizes. In this way, the modeller can take full advantage of all the available data and none of the observations in any of the series have to be omitted from the analysis. This estimation algorithm is outlined in Appendix A21.1.

Camacho et al. (1986) consider the effect on the estimation of the parameters when a bivariate series $\mathbf{Z}_t = (Z_{t1}, Z_{t2})^T$ is incorrectly modelled as a general multivariate AR(1) model using [20.2.1] when a CARMA(1,0) model from [21.2.1] would suffice. As pointed out in Section 20.4, the general multivariate AR(1) model has been proposed for utilization in hydrology. Using simulation studies, they show that the loss in efficiency of the parameter estimates obtained using the full multivariate model can be very substantial and in many cases can be well over 50%.

21.3.4 Diagnostic Checks

After obtaining efficient estimates for the model parameters, possible inadequacies in the fitted model can be found and subsequently corrected by examining the statistical properties of the residuals. As explained in Section 20.3.2, a range of tests are available for ascertaining whether or not the residuals are *white* (Li and McLeod, 1981), *homoscedastic* (see Section 7.5.2) and *normally distributed* (Royston, 1983).

For detecting misspecifications in the model, the *residual CCF* is both informative and sensitive. In addition to the joint estimates for the model parameters, one can obtain the model residuals $\hat{a}_{ti}$, $i = 1,2,\ldots,k$, using the efficient estimation procedure of Section 21.3.3. To calculate the residual CCF, $\hat{r}_{ij}(l)$, at lag l between two residual series, one simply replaces $\bar{a}_{ti}$ and $\bar{a}_{tj}$ by $\hat{a}_{ti}$ and $\hat{a}_{tj}$, respectively, in [21.3.1]. Each entry in the *residual CCF matrix*, $\hat{\mathbf{R}}(l)$, should not be significantly different from zero for $l > 1$. As is done at the identification stage, it is convenient to use the symbols "+", "-", and "." in $\hat{\mathbf{R}}(l)$ to indicate entries that are significantly larger, significantly smaller, and not significantly different from zero, respectively.

Based upon the work of Li and McLeod (1981), Camacho et al. (1985) suggest a *modified Portmanteau test* statistic to test for the independence of the residuals. As in Section 21.3.3, let $\Delta = E[\mathbf{a}_t \cdot \mathbf{a}_t^T]$ be the variance-covariance matrix of $\mathbf{a}_t = (a_{t1}, a_{t2}, \ldots, a_{tk})^T$ and let

$$\hat{r}(l) = (\hat{r}_{11}(l), \hat{r}_{21}(l), \ldots, \hat{r}_{k1}(l), \hat{r}_{12}(l), \hat{r}_{22}(l), \ldots, \hat{r}_{k2}(l), \ldots, \hat{r}_{kk}(l))^T$$

The modified Portmanteau test statistic is then written as

$$Q_{L_m} = n\sum_{l=1}^{L} \hat{r}(l)^T (\hat{\Delta}^{-1} \otimes \hat{\Delta}^{-1}) \hat{r}(l) + k^2 L(L+1)/2 \qquad [21.3.3]$$

Under the assumption that the residuals are white noise, Q_{L_m} is approximately χ^2 distributed with $k^2L - k(p+q)$ degrees of freedom for large values of L and n.

If the residual CCF possesses significantly large values at lags other than zero, the CARMA model must be appropriately redesigned. Perhaps it may be only necessary to add additional AR and MA parameters to the CARMA model. If the CARMA class of models itself is not adequate, a more complex family of multivariate models, such as the TFN set of models, may have to be considered. When the residuals are not approximately normally distributed and/or homoscedastic, it may be required to transform one or more of the series using an appropriate transformation such as the Box-Cox transformation in [3.4.30]. Subsequent to this, the parameters of the CARMA can be estimated again using the algorithm in [21.3.3].

21.3.5 Seasonality

The CARMA(p,q) model presented in Sections 21.2.2 and 21.2.3 is defined for handling nonseasonal series and the model construction techniques of this section are explained for the nonseasonal case. When one wishes to fit a CARMA model to seasonal data, the two approaches described in more detail in Section 20.3.3 can be used. In particular, one can first *deseasonalize* each series using a technique from Section 13.2.2 and then fit a nonseasonal CARMA model to the deseasonalized data. An alternative approach is to employ a *periodic version of the CARMA model* to fit directly to the seasonal data.

As explained in Part VI for univariate models, usually deseasonalized (Chapter 13) or periodic (Chapter 14) models are the most appropriate types of seasonal models to fit to natural time series. This is because data within a given season for a natural time series are usually stationary across the years. However, when the data within seasons are nonstationary over the years, it may be appropriate to seasonally and perhaps also nonseasonally difference the data to remove the nonstationarity (see Chapter 12). For example, a seasonal economic time series may possess an upward trend which causes the overall level of the series to increase over the years. After appropriately differencing the series, a seasonal ARMA model can be fitted to the resulting stationary data in order to obtain the parameter estimates for the seasonal ARIMA model.

In a manner similar to that for the univariate seasonal ARIMA model of Chapter 13, a *CARIMA model* containing differencing operators can be easily defined. To accomplish this, one simply introduces seasonal and nonseasonal differencing operators along with seasonal AR and MA operators into [21.2.7]. A model containing differencing operators could also be defined for the general multivariate ARMA models of the previous chapter.

21.4 SIMULATING USING CARMA MODELS

21.4.1 Introduction

Comprehensive techniques for generating synthetic sequences using ARMA and ARIMA models were developed by McLeod and Hipel (1978b) and are presented in detail in Chapter 9. To avoid the introduction of systematic bias into the simulated series by employing fixed starting values, the simulation methods described in Sections 9.3 and 9.4 are designed such that random realizations of the underlying model are used for starting values. The simulation techniques developed for the univariate ARMA and ARIMA models can be extended for use with CARMA models.

Originally, McLeod (1979) suggested a simulation algorithm for use with CARMA models possessing no MA parameters while Camacho (1984) presented the algorithm for the general case. Simulation experiments which employ this new algorithm are given by McLeod (1979), Camacho (1984), and Camacho et al. (1985, 1986). A similar type of simulation algorithm can be developed for use with the general multivariate ARMA models of Chapter 20.

21.4.2 Simulation Algorithm

Overall Algorithm

Suppose that there are k time series and at time t the vector of time series is denoted by

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

For the ith time series, let the order of the AR and MA parameters needed in [21.2.6] be p_i and q_i, respectively. Now, define

$$\mathbf{Z}_{p_i,i} = (Z_{1i}, Z_{2i}, \ldots, Z_{p_i,i})^T$$

and

$$\mathbf{a}_{q_i,i} = (a_{p_i-q_i+1,i}, a_{p_i-q_i+2,i}, \ldots, a_{p_i,i})^T$$

for series $i = 1,2,\ldots,k$. Then, the values contained in the vectors $\mathbf{Z}_{p_i,i}$ and $\mathbf{a}_{q_i,i}$ represent the starting values for the ith series where $i = 1,2,\ldots,k$.

Suppose that it is required to generate N synthetic observations for the CARMA model in [21.2.6]. Without loss of generality, it is assumed that the mean of each of the k series is zero. The following algorithm provided by Camacho (1984, pp. 57-68) is used to obtain simulated values $\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_N$ where $\mathbf{Z}_t=(Z_{t1},Z_{t2},\ldots,Z_{tk})^T$. Moreover, this algorithm is exact in the sense that it is not subject to inaccuracies associated with fixed initial values.

1. Determine the lower triangular matrix $\mathbf{M}$ by Cholesky decomposition such that (Ralston, 1965)

$$\Delta=\mathbf{M}\mathbf{M}^T \qquad [21.4.1]$$

where Δ is the variance-covariance matrix for $\mathbf{a}_t=(a_{t1},a_{t2},\ldots,a_{tk})^T$ in [21.2.6].

2. Obtain the vectors of initial values $\mathbf{Z}_{p_i,i},\mathbf{a}_{q_i,i}$, $i=1,2,\ldots,k$. (See next subsection for the method used to calculate the initial values.)

3. Following the two steps given next, generate $\mathbf{a}_{p+1},\mathbf{a}_{p+2},\ldots,\mathbf{a}_N$ which is a sequence of $N-p$ vectors each of which has dimension k and is NID($\mathbf{0}$,Δ). As in [21.2.6], the $p=\max(p_1,p_2,\ldots,p_k)$.

   (i) Simulate $\mathbf{e}_{t1},\mathbf{e}_{t2},\ldots,\mathbf{e}_{tk}$, which is a sequence of $N-p$ vectors each of which has dimension k and is distributed as NID($\mathbf{0}$,$\mathbf{1}$) where $\mathbf{0}$ is a $k\times1$ vector consisting of k zeroes, $\mathbf{1}$ is diagonal matrix of dimension $k\times k$ having entries of unity along the main diagonal.

   (ii) Calculate

$$a_{ti}=\sum_{j=1}^{i}m_{ij}e_{tj} \qquad [21.4.2]$$

   for $i=1,2,\ldots,k$ and $t=p+1,\ldots,N$.

4. Obtain $\mathbf{Z}_{p+1},\mathbf{Z}_{p+2},\ldots,\mathbf{Z}_N$, where each vector of observations at a given time has k entries, by using

$$\begin{aligned}Z_{ti}=\ &\phi_{ii1}Z_{t-1,i}+\phi_{ii2}Z_{t-2,i}+\cdots+\phi_{iip_i}Z_{t-p_i,i}+a_{ti}\\&-\theta_{ii1}a_{t-1,i}-\theta_{ii2}a_{t-2,i}-\cdots-\theta_{iiq_i}a_{t-q_i,i}\end{aligned} \qquad [21.4.3]$$

   for $i=1,2,\ldots,k$ and $t=p+1,p+2,\ldots,N$.

5. If another series of length N is required return to step 2.

The above algorithm is described for simulating stationary series having no Box-Cox transformations. If the original set of series were differenced and also were transformed using Box-Cox transformations, the techniques of Sections 9.5 and 9.6, respectively, could be employed in conjunction with the algorithm of this section to obtain synthetic sequences in the original untransformed domain.

Calculation of the Initial Values

The joint distribution of $\mathbf{Z}_{p_i,i}$ and $\mathbf{a}_{q_i,i}$, $i = 1,2,\ldots,k$, is used to generate the starting values for the simulation algorithm for a CARMA model. As demonstrated by Camacho (1984, p. 59), the joint distribution of

$$\upsilon = (\mathbf{Z}_{p_1,1},\mathbf{Z}_{p_2,2},\ldots,\mathbf{Z}_{p_k,k},\mathbf{a}_{q_1,1},\mathbf{a}_{q_2,2},\ldots,\mathbf{a}_{q_k,k})^T$$

is multivariate normal having a mean of zero and variance covariance matrix given by

$$V = \begin{pmatrix} \gamma_{gh}(i-j) & \sigma_{gh}\psi_g(i-j) \\ Symm & \Delta\otimes\mathbf{I}_{q\times q} \end{pmatrix} \qquad [21.4.4]$$

where

$$\gamma_{gh}(r) = <Z_{t,g}Z_{t+r,h}>\,,\quad g,h = 1,2,\ldots,k$$

$$\phi_i(B)\cdot\psi_i(B) = \theta_i(B)\,,\quad i = 1,2,\ldots,k \qquad [21.4.5]$$

Ansley (1980) and Kohn and Ansley (1982) provide an algorithm to obtain the theoretical autocovariance function of the general multivariate ARMA model. This algorithm could be employed to calculate the terms $\gamma_{gh}(i-j)$ in [21.4.4]. However, due to the diagonal structure of the CARMA model, Camacho (1984, p. 61-62), has developed a computationally efficient algorithm for the calculation of the theoretical autocovariance function of the CARMA model.

The following algorithm can be used to obtain the initial values required in step 2 of the overall algorithm for simulating using the CARMA model given in Section 21.4.2.

1. Calculate $\psi_g(s)$, $g = 1,\ldots,k$; $s = 0,1,\ldots,\max\{p,q\}$ from [21.4.5].

2. Calculate the theoretical autocovariance functions $\gamma_{gh}(r)$,

   $$r = 1-p,\ldots,0,\ldots,p-1\,,\quad 1 < g < h < k$$

3. Form the variance-covariance matrix $\mathbf{V}$ of υ given by [21.4.4] and obtain the lower triangular matrix $\mathbf{L}$ by Cholesky decomposition such that

   $$\mathbf{V} = \mathbf{L}\mathbf{L}^T$$

4. Generate $e_1e_2,\ldots,e_{k(p+q)}$, a sequence of $k(p+q)$ NID (0,1) random variables and determine the vector of initial values by:

   $$U_j = \sum_{i=1}^{j} l_{ji}e_i\,,\quad j = 1,2,\ldots,k(p+q)$$

   Note that if another series is required only step 4 is needed.

21.5 PRACTICAL APPLICATIONS

21.5.1 Introduction

In order to clearly demonstrate the usefulness of CARMA modelling in water resources and environmental engineering, three case studies are presented. The first and third applications involve water quantity data while the second one deals with water quality time series. All three applications show how the model construction techniques of Section 21.3 can be conveniently used in practice to obtain models that adequately describe the series and possess efficient parameter estimates. In the third example where one series has more data points than the other, the estimation algorithm of Appendix A21.1 is employed so that all of the measurements can be used for efficiently estimating the CARMA model parameters. These three applications were originally presented by Camacho et al. (1985).

21.5.2 Fox and Wolf Rivers

Average annual riverflows in m^3/s for the Fox River near Berlin, Wisconsin, and the Wolf River near London, Wisconsin, are available from Yevjevich (1963) and also the hydrological data tapes of Colorado State University at Fort Collins, for the years from 1899 to 1965. A plot of the data is given in Figure 21.5.1, where the overall shapes and dependencies of the data can be compared. In order to facilitate these comparisons, the y-axes have been purposely deleted and the data have been scaled so that each of the two series takes up half the graph (these considerations were also taken into account to produce the plots given in Figures 21.5.3 and 21.5.5).

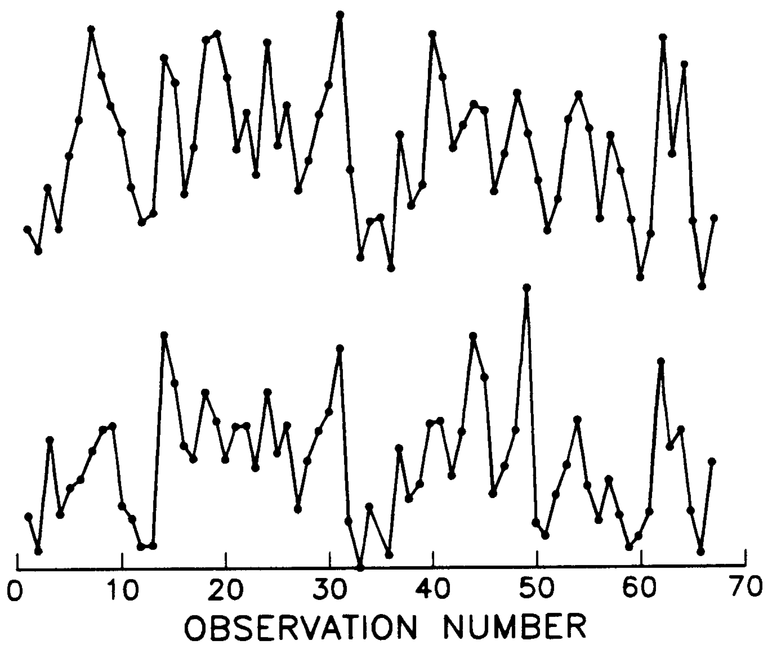

Figure 21.5.1. Annual riverflows for the Fox and Wolf Rivers (m^3/s).

Because the Fox and Wolf Rivers lie within the same geographical and climatic region of North America, a priori one may expect from a physical viewpoint that a CARMA model would be more appropriate to use than separate univariate ARMA models. Subsequent to taking a natural logarithmic transformation of the observations in both time series, univariate identification results from Chapter 5 suggest that it may be adequate to fit a MA model of order one (i.e., MA(1)) given in [3.3.1] to each data set. After prewhitening each series using the calibrated MA(1) model, the residual CCF for each series is calculated using [21.3.1] with the prewhitened

Fox and Wolf riverflows in order to obtain the graph of the residual CCF shown in Figure 21.5.2, along with the 95% confidence limits. Because the residual CCF in this figure is only significantly different from zero at lag zero, this indicates that a CARMA model could be fitted to the logarithms of the bivariate series. Additionally, the fact that each series can adequately be described by a univariate MA(1) model suggests that the following CARMA(0,1) model should be used.

$$\log Z_{ti} - \mu_i = (1 - \theta_{ii1})a_{ti}, \quad i = 1,2 \qquad [21.5.1]$$

where $i = 1$, and $i = 2$ refer to the Fox and Wolf logarithmic riverflows, respectively, μ_i is the theoretical mean of the logarithmic series for Z_{ti}, and the general definitions of all parameters and variables follow [21.2.7].

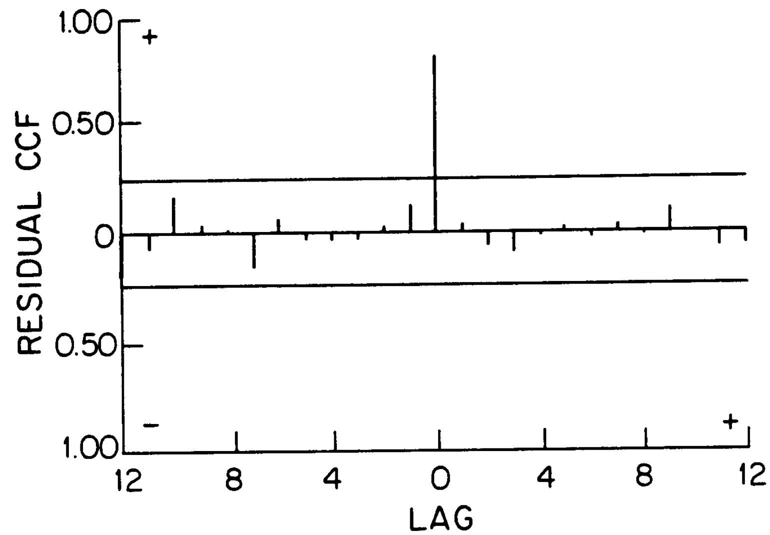

Figure 21.5.2. Residual CCF for the Fox and Wolf Rivers.

Table 21.5.1 lists the parameter estimates along with their standard errors appearing in brackets, using the univariate approach (Appendix A6.1) and the joint estimation algorithm described in Section 21.3.3. As can be observed in Table 21.5.1, there is a significant reduction in the variance of the parameter estimates when the joint estimation is employed. This in turn means that the relative efficiency of the univariate estimates with respect to the joint multivariate estimator is much less than unity. This relative efficiency is calculated using

$$eff = var(\hat{\theta}_{ii1})/var(\bar{\theta}_{ii1}) \qquad [21.5.2]$$

where $\hat{\theta}_{ii1}$ and $\bar{\theta}_{ii1}$ are the joint and univariate estimates, respectively, for the parameter θ_{ii1}. The correlation between $\hat{a}_{t1}$ and $\hat{a}_{t2}$ is calculated to be 0.78. When the residuals of the CARMA(0,1) are subjected to residual checking, no misspecifications of the fitted model are detected.

Table 21.5.1 Parameter estimates for the CARMA model and univariate models for the Fox and Wolf Rivers.

	Fox River	Wolf River
Univariate	-0.483	-0.411
Estimates of θ_{ii1}	(0.110)	(0.111)
Joint	-0.626	-0.543
Estimates of θ_{ii1}	(0.075)	(0.080)
Efficiency of		
Univariate Estimator	0.465	0.519
Mean of Log Z_{ti}	3.39	3.84
	(0.037)	(0.042)
Residual Variance	5.52×10^{-2}	7.5×10^{-2}

21.5.3 Water Quality Series

In the second example, two series corresponding to different measurements of the concentration of nitrogen in the Middle Fork Creek near Seebe, located in the Province of Alberta, Canada, are modelled. The series represent monthly measurements of total nitrogen and nitrogen Kjeldahl from 1972 to 1979 and are part of an overall data set that are studied using both exploratory and confirmatory data analysis tools in Sections 22.3 and 22.4, respectively. The seasonal adjustment algorithm of Section 22.2 was used to obtain the monthly means of the series from data available at irregular time intervals. A plot of the estimated monthly series is given in Figure 21.5.3.

Following Chapter 5 and Section 12.3.2, univariate identification techniques suggest that an adequate model for describing the natural logarithms of the total nitrogen series, Z_{t1} is a seasonal $AR(1)_6$ model of the form

$$(1 - \phi_{116}B^6)(logZ_{t1} - \mu_1) = a_{t1} \qquad [21.5.3]$$

where $B^6\log Z_{t1} = \log Z_{t-6,1}$. An appropriate model to fit to the nitrogen kjeldahl series, Z_{t2} is an AR(1) model of the form

$$(1 - \phi_{221}B)(\log Z_{t2} - \mu_2) = a_{t2} \qquad [21.5.4]$$

The univariate estimated parameters and their SE's given in brackets are are listed in Table 21.5.2.

A perusal of the residual CCF for the fitted models from [21.5.3] and [21.5.4], shows that only the CCF at lag zero is significantly different from zero. This identification result implies that a CARMA model is appropriate for fitting to the bivariate series. The specific parameters required in the two component equations of the overall CARMA model are the same as those

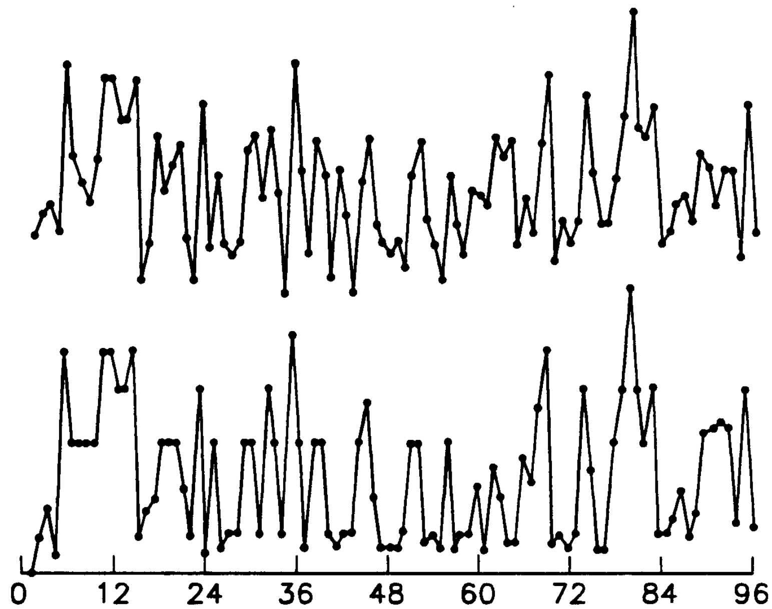

Figure 21.5.3. Concentration of total nitrogen and nitrogen kjeldahl (mg/l) for the Middle Fork Creek, near Seebe, Alberta, Canada.

used in [21.5.3] and [21.5.4]. Following the joint estimation procedure of Section 21.3.3, the efficient estimators for the CARMA model are calculated and displayed in Table 21.5.2. The reduction in the variances of the joint estimators compared with the variances of the univariate estimators is quite substantial. If only univariate series were used to estimate the parameters of the model for each one of the series, it would be necessary to increase the sample size of the series by a factor of four in order to obtain the same reduction in the variances of the parameters estimates. This increase in the sample size of the series is very expensive and in some cases infeasible. Consequently, this demonstrates that the CARMA model could also be employed to increase the accuracy of the parameters of the univariate models. The correlation at lag zero between $\hat{a}_{t1}$ and $\hat{a}_{t2}$ for the models given in [21.5.3] and [21.5.4], respectively, is found to be 0.88.

21.5.4 Two Riverflow Series Having Unequal Sample Sizes

As an example of two riverflow time series possessing unequal numbers of observations, consider the French Broad River at Asheville, North Carolina and the French Broad River near Newport, Tennessee, which have average annual flows from 1896 to 1965 and 1921 to 1965, respectively. As is the case for the application in Section 21.5.2, these flows are available from Yevjevich (1963) and also the hydrological data tapes of Colorado State University.

A plot of the 70 observations of the flows at Asheville and the 45 observations of the flows near Newport are displayed in Figure 21.5.5. Univariate MA(1) models like the ones in [21.5.1] were found to be adequate to fit the logarithms of the series. A plot of the residual CCF is given in Figure 21.5.6 Although the flows near Newport are measured downstream from the flows at

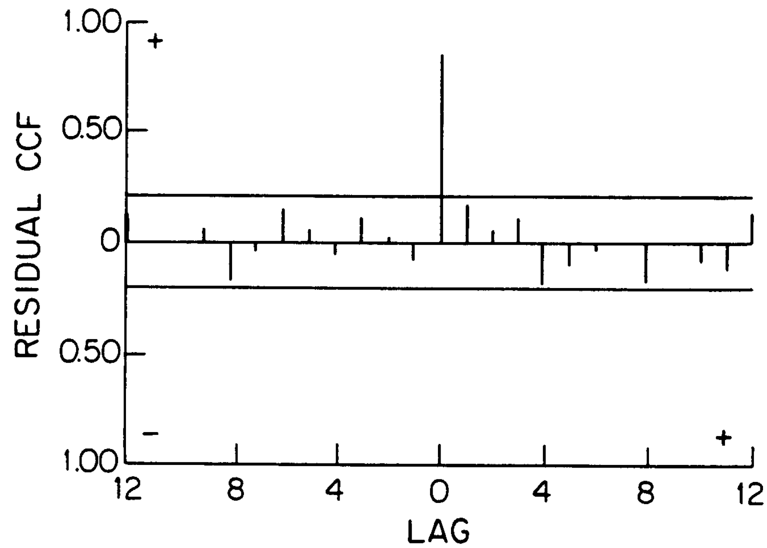

Figure 21.5.4. Residual CCF for the total nitrogen and nitrogen kjeldahl.

Table 21.5.2. Parameter estimates for the CARMA model and univariate models for the total nitrogen and nitrogen kjeldahl series for the Middle Fork Creek.

	Total Nitrogen	Nitrogen Kjeldahl
Univariate Estimates of ϕ_{116} and ϕ_{221}	0.310 (0.097)	0.294 (0.097)
Joint Estimates of ϕ_{116} and ϕ_{221}	0.141 (0.049)	0.141 (0.049)
Efficiency of Univariate Estimator	0.255	0.255
Mean of Log Z_{ti}	-1.33 (0.084)	-1.59 (0.104)
Residual Variance	0.131	0.152

Asheville, implying that a TFN model (see Chapter 17) may be required to model the bivariate series, it is observed from the plot of the residual CCF that a CARMA model would suffice, (only the residual CCF at lag zero is significantly different from zero). This is due to the fact that annual riverflows are being considered and this *temporal aggregation* of the data, by its very nature, incorporates some of the lagged relationships, which would be expected to hold in the model of the system (Granger and Newbold, 1977). If monthly data or less temporal aggregated data were considered, a TFN model would probably be required to model the data. The algorithm given in Appendix A21.1 is used to jointly estimate the parameters of the model. These estimates are given in Table 21.5.3. The significant reductions in the variances of the estimators compared with the univariate estimates can be observed. The correlation at lag zero between the residuals of the two series is calculated to be 0.91.

21.6 CONCLUSIONS

As illustrated by the practical applications of the previous section, the CARMA family of models can be used to model efficiently hydrological and other types of environmental series. When taking the physical characteristics of the system being modelled into account along with output from the identification methods of Section 21.3.2, the CARMA class of models is often found to be the most appropriate type of multivariate model to use. The application of Section 21.5.4 shows that the CARMA model can be ideal for modelling time series formed by temporal aggregation. Another attractive feature of fitting this kind of model is that well developed, yet simple, model construction tools are currently available for use in practical applications. For example, when estimating the parameters of time series having equal and unequal sample sizes, the estimation procedures presented in Section 21.3.3 and Appendix A21.1, respectively, can be utilized. Furthermore, the flexible algorithm described in detail in Section 21.4.2 can be used for simulating synthetic sequences from a CARMA model.

Besides environmental series, the CARMA class of models has been successfully employed to model and forecast economic time series. Umashankar and Ledolter (1983), Moriarity and Salomon (1980) and Nelson (1976) used CARMA models to increase the efficiency of the estimated parameters and to improve the accuracy of the forecasts. Risager (1980) fitted CARMA models to mean annual ice core measurements. Research related to the development and application of CARMA models in hydrology was referred to throughout this chapter as well as Section 20.4.

Research in CARMA modelling can be extended in a variety of directions. For instance, as mentioned in Sections 21.3.5 and 20.3.3, model construction methods could be developed for various kinds of periodic CARMA models. Camacho (1984, Section 2.4) defines a contemporaneous TFN model in which the innovations among a set of k TFN models are contemporaneously correlated. If practical applications dictate the need for this rather sophisticated type of contemporaneous model, appropriate model construction methods could be developed.

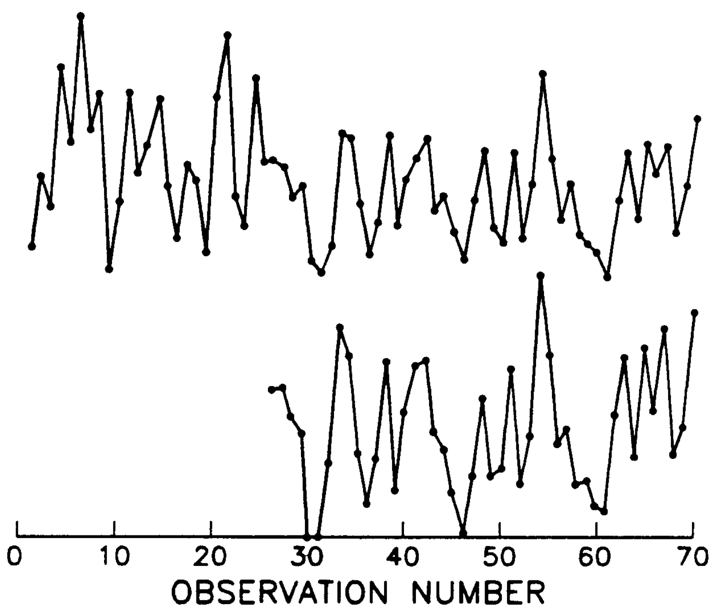

Figure 21.5.5. Annual riverflows for the French River at Asheville and near Newport (m^3/s).

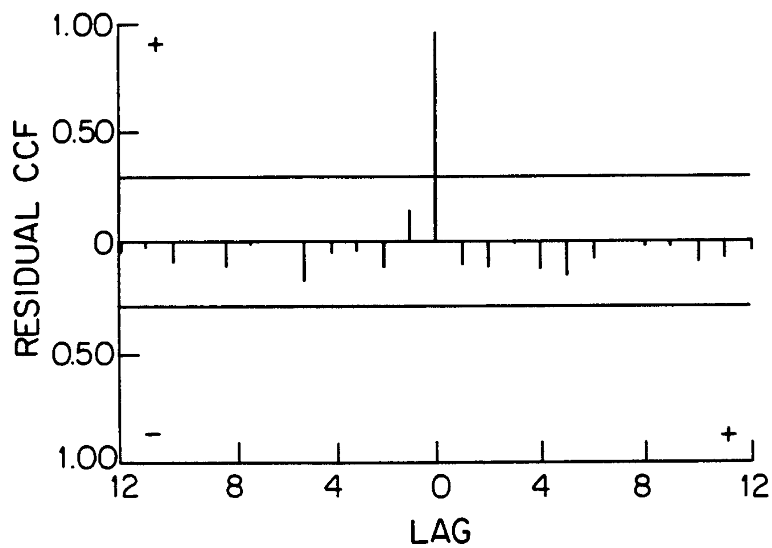

Figure 21.5.6. Residual CCF for the French Broad River at Asheville and near Newport.

Table 21.5.3. Parameter estimates for the CARMA model and univariate models for the French Broad River at Asheville and near Newport.

	At Asheville $n = 70$	Near Newport $n = 45$
Univariate Estimates of θ_{ii1}	-0.283 (0.115)	-0.469 (0.131)
Joint Estimates of θ_{ii1}	-0.170 (0.087)	-0.470 (0.081)
Efficiency of Univariate Estimator	0.572	0.382
Mean of Log Z_{ti}	4.01 (0.040)	4.36 (0.048)
Residual Variance	6.72×10^{-2}	5.79×10^{-2}

APPENDIX A21.1

ESTIMATOR FOR CARMA MODELS HAVING UNEQUAL SAMPLE SIZES

Within this appendix, an estimator is presented for obtaining maximum likelihood estimates for the parameters of a CARMA model [21.2.1] or [21.2.6] when the k time series used to calibrate the model do not have the same lengths. This algorithm was originally developed by Camacho et al. (1985). The CARMA estimator to be used with samples having the same number of observations over the same time period is given in Section 21.3.3.

When fitting models to multivariate hydrological data, it is common to find series with unequal numbers of observations. What is customary in this circumstance is to eliminate the additional information available in the longer series so that all the series end up with an equal number of observations. For example, Risager (1980) considered the modelling of a bivariate time series of mean annual ice core measurements for which data were available for the years 1861-1974 and 1169-1975, respectively. In his analysis, only data for the common period 1861-1974 could be used to jointly estimate the parameters of the model. Another possibility is to consider some of the observations of the shorter series as missing and use a procedure such as that given by Ansley and Kohn (1983) to estimate the parameters of the model. This approach, although sensible, is not computationally efficient for a large number of missing observations or for series having a large sample size. Another disadvantage of this procedure is the introduction of many additional parameters to be estimated, which reduce the accuracy of estimators. If a CARMA model is sufficient to fit to the data, the estimator described below can be employed for estimating the parameters of the CARMA model using all the available information in a very efficient way.

Suppose that the set of observations available for the series Z_{ti}, $i = 1,2,\ldots,k$, is given by

$$\{Z_{ti}\} = \{Z_{1-m_i,i},\ldots,Z_{0,i},Z_{1,i},\ldots,Z_{n,i}\} \quad \text{for } i = 1,2,\ldots,k,$$

where $t = 1-m_i, 1-m_i+1,\ldots,0,1,2,\ldots,n$, are the times at which the m_i+n observations in series i occur, $t = 1,2,\ldots,n$, are the common times for which all k series have measurements and hence n is the number of common observations across all k series. Although it is assumed that all the series go up to the same time n, it is possible to extend the procedures given below to include the case where not all the series end at the same time.

As in Section 21.3.3, let the parameters of the CARMA model for the ith series be contained in the vector

$$\beta_i = (\phi_{ii1},\phi_{ii2},\ldots,\phi_{iip_i},\theta_{ii1},\theta_{ii2},\ldots,\theta_{iiq_i})^T$$

Hence, the vector of parameters for the complete CARMA model is written as

$$\beta = (\beta_1,\beta_2,\ldots,\beta_k)^T$$

An approximate log-likelihood function of the CARMA model in [21.2.1] or [21.2.6] is

$$l(\beta,\delta) = -\frac{n}{2}\log\delta - \sum_{i=1}^{k}\frac{m_i}{2}\log\sigma_{ii} + S - \frac{1}{2}\sum_{i=1}^{k}\frac{S_{0i}}{\delta_{ii}}$$

where

$$S = \frac{1}{2}\sum_{t=1}^{n}\mathbf{a}_t^T\delta^{-1}\mathbf{a}_t$$

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

$$\delta = (\sigma_{ij}) \text{ and}$$

$$S_{i0} = \frac{1}{2}\sum_{t=1-m_i}^{0} a_{ti}^2$$

Using this approximation, it is possible to modify the algorithm given in Section 21.3.3 to estimate the parameters of a CARMA model when an equal number of observations are available for each series, to handle the case where the sample sizes are unequal.

The algorithm is as follows:

1. For each series Z_{ti}, $i = 1,2,\ldots,k$, obtain MLE's of the ARMA model parameters in [21.2.6] using an appropriate univariate ARMA estimation technique, such as one of those given by Newbold (1974), Ansley (1979), Ljung and Box (1979) or McLeod (1977) referred to in Section 6.2.3, with the complete set of observations $\{Z_{ti}\}$, $t = 1-m_i,1-m_i+1,\ldots,0,1,\ldots,n$. Let the vector of univariate estimates be given by

$$\bar{\beta} = (\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_k)^T$$

and set

$$\beta_0^0 = (\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_k)^T$$

2. Estimate $\delta = (\sigma_{ij})$ by solving the system of nonlinear equations

$$n\sigma_{ij} + \sum_{h=1}^{k} \frac{m_h}{\sigma_{hh}} \sigma_{ih}\sigma_{jh} = SS_{ij} + \sum_{h=1}^{k} \frac{S_{0h}}{\sigma_{hh}} \sigma_{ih}\sigma_{jh}$$

where

$$SS_{ij} = \sum_{t=1}^{n} a_{ti}a_{tj}$$

3. Calculate

$$\beta^{r+1} = \beta^r + \mathbf{V}(\delta l/\delta\beta)\big|_{\beta=\beta^r}$$

where $\mathbf{V}$ is obtained as follows: Let

$$n\mathbf{I}_{gh} = plim(\partial^2 S/\partial\beta_q \partial\beta_h)$$

Then

$$\mathbf{V}^{-1} = [\sigma^{gh}\mathbf{I}_{gh}] + Diag\left[\frac{m_1}{n\sigma_{11}}\mathbf{I}_{11} + \frac{m_2}{n\sigma_{22}}\mathbf{I}_{22} + \cdots + \frac{m_k}{n\sigma_{kk}}\mathbf{I}_{kk}\right]$$

where $\delta^{-1} = (\sigma^{gh})$ and $Diag[\cdots]$ indicates a block diagonal matrix. The $[\sigma^{gh}\mathbf{I}_{gh}]$ can be determined using the algorithm provided by Ansley (1980) and Kohn and Ansley (1982). Iterations of the algorithm are continued until convergence is reached for giving the approximate MLE's of β. An application of the estimator in this appendix for fitting a CARMA model to two annual riverflow series having unequal sample sizes is furnished in Section 21.5.4.

PROBLEMS

21.1 In Sections 21.2.2 and 21.2.3, the subset and concatenation definitions are given for CARMA(p,q) models. For the following CARMA(p,q) models, write down the subset and concatenation definitions, the stationarity and invertibility conditions, and the entries in the variance covariance matrix for the innovations:

(a) CARMA(3,0)

(b) CARMA(0,4)

(c) CARMA(2,2)

(d) CARMA(4,3)

21.2 Select two annual time series that you think could be adequately modelled using a CARMA model. Follow the identification procedure of Section 21.3.2 to ascertain whether or not your supposition is justified.

21.3 Carry out the instructions of problem 21.2 for the situation when you have three time series.

21.4 Describe in detail how the estimation algorithm of Section 21.3.3 works for the following bivariate CARMA models:

(a) CARMA(1,0)

(b) CARMA(0,2)

(c) CARMA(1,1)

21.5 Find two annual time series that are only contemporaneously correlated with one another as indicated by the residual CCF. Fit a CARMA to these series and check that the calibrated model provides an adequate fit.

21.6 Carry out the instructions of problem 21.5 for the case when you have three time series.

21.7 Suppose that in a set of k seasonal time series, each time series has s seasons per year. Using both the subset and concatenation definitions of CARMA models from Sections 21.2.2 and 21.3.2, write down the equations for the periodic CARMA model.

21.8 Carry out the instructions of problem 21.7 for the case of a seasonal CARIMA model.

21.9 Suppose that you wish to simulate 10 values for a bivariate CARMA model. Using the algorithm of Section 21.4.2, explain in detail how these are calculated for the following bivariate CARMA models

(a) CARMA(1,0)

(b) CARMA(0,1)

(c) CARMA(1,1)

21.10 Select a CARMA which is of direct interest to you. After setting the model parameters at some reasonable values or else using a model that you have already calculated, simulate three synthetic series of lengths 100, 500 and 1,000. Now fit a CARMA model to each of these series. Compare your modelling results for the three sets of simulated sequences and draw appropriate conclusions.

21.11 Explain how you would calculate minimum mean square error forecasts for a CARMA model.

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. (1985). 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). 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*, July 15-17, 1985, 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.

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.

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.

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.

McLeod, A. I. (1979). Distribution of the residual cross correlation in univariate ARMA time series models. *Journal of the American Statistical Association*, 74(368):849-855.

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

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

DATA SETS

Yevjevich, V. M. (1963). Fluctuation of wet and dry years, 1, Research data assembly and mathematical models. Hydrology paper no. 1, Colorado State University, Fort Collins, Colorado.

FORECASTING

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

Moriarty, M. and Salomon, G. (1980). Estimation and forecast performance of a multivariate time series model of sales. *Journal of Market Research*, 17:558-564.

Nelson, C. R. (1976). Gains in efficiency from joint estimation of systems of autoregressive-moving average processes. *Journal of Econometrics*, 4:331-348.

Umashankar, S. and Ledolter, J. (1983). Forecasting with diagonal multiple time series models: An extension of univariate models. *Journal of Market Research*, 20:58-63.

ESTIMATORS

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, 66(1):59-65.

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

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.

Ljung, G. M. and Box, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models. *Biometrika*, 66(2):265-270.

McLeod, A. I. (1977). Improved Box-Jenkins estimators. *Biometrika*, 64(3):531-534.

McLeod, A. I. and Salas, P. R. H. (1983). An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 32:211-223.

Newbold, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika*, 61(3):423-426.

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

SIMULATION

Ansley, C. F. (1980). Computation of the theoretical autocovariance function for a vector ARMA process. *Journal of Statistical Computation and Simulation*, 12:15-24.

Kohn, F. and Ansley, C. F. (1982). A note on obtaining theoretical autocovariances of an ARMA process. *Journal of Statistical Computing and Simulation*, 15:273-283.

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

Ralston, A. (1965). *A First Course in Numerical Analysis*. McGraw-Hill, New York.