CHAPTER 14

PERIODIC MODELS

14.1 INTRODUCTION

As emphasized by authors such as Moss and Bryson (1974), seasonal hydrological and other types of time series exhibit an autocorrelation structure which depends on not only the time lag between observations but also the season of the year. Furthermore, within a given season, usually second order stationarity is preserved by natural time series. For example, at a location in the northern hemisphere the monthly temperature for January across the years may fluctuate with constant variance around an overall mean of -5° C. In addition, the manner in which the January temperature is correlated with December and November as well as the previous January may tend to remain the same over the years. As another illustration of seasonally or periodically varying correlation, consider the case of runoff from snowmelt in late winter or early spring in a northern region. If the snowmelt is an important factor in runoff which might occur in either March or April, the correlation between observed riverflows for these months may be negative whereas at other times of the year it is usually positive. To model appropriately the foregoing and similar types of time series, periodic models can be employed. These models are ideal, for instance, for describing the average monthly flows of the Saugeen River at Walkerton, Ontario, Canada, plotted in Figure VI.1.

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.

Since the early 1960's a considerable amount of research has been executed in the area of periodic modelling. This research includes contributions by authors such as Gladyshev (1961, 1963), Jones and Brelsford (1967), Tao and Delleur (1976), Croley and Rao (1977), McLeod and Hipel (1978), Pagano (1978), Troutman (1979), Dunsmuir (1981), Tiao and Gruppe (1980), Sakai (1982), Salas et al. (1985), Cipra (1985a,b), Vecchia (1985a,b), Thompstone et al. (1985a), Cipra and Tlusty (1987), Jimenez et al. (1989) and McLeod (1993), as well as the books on stochastic hydrology referred to in Section 1.6.3.

As is explained in Section 14.3, a comprehensive range of *model construction* tools are available for conveniently fitting PAR models to seasonal time series. Because the theory and application of the PAR family of models are well-developed, this class of flexible models is stressed in this chapter. Nonetheless, some interesting developments in building PARMA models are pointed out in Section 14.7.

Subsequent to presenting model construction tools for use with PAR models in Section 14.3, a PAR model is developed for describing the average monthly flows of the Saugeen River plotted in Figure VI.1. A potential drawback of using a periodic model in an application is that the model often requires the use of a substantial number of parameters. Salas et al. (1980)

propose a Fourier series approach to reduce the number of model parameters in PAR and PARMA models. Thompstone et al. (1985a) suggest a procedure for combining individual AR models for various adjacent seasons, to obtain a single model for all of the seasons in the group. After joining appropriate seasons into groups, the overall periodic model that is fitted to the resulting data is called the *parsimonious periodic autoregressive (PPAR) model*. Subsequent to defining the PPAR model and presenting model construction methods in Section 14.5, PPAR models as well as other periodic models are fitted to seasonal hydrological time series in Section 14.6. Finally, in Section 14.8, *simulation experiments* are carried out to demonstrate that PAR and PPAR models statistically preserve *critical period statistics* which are used in reservoir design.

14.2 DEFINITIONS OF PERIODIC MODELS

14.2.1 Introduction

The definitions of PAR and PARMA models can be made from two different points of view. Firstly, PAR and PARMA models can be thought of as the periodic extensions of the nonseasonal AR and ARMA models, respectively, defined in Chapter 3. In other words, a PAR model consists of having a separate AR model for each season of the year whereas a PARMA model contains an ARMA model for each season. For both theoretical and practical reasons, the PAR and PARMA families of models are defined in these fashions in this chapter. For example, comprehensive model building procedures are now available for use with PAR models (Section 14.3) while significant progress has been made in developing model construction methods for employment with PARMA models (Section 14.7).

The second approach for defining PAR and PARMA models is to consider them to be special types of the multivariate ARMA models defined in Section 20.2. However, this approach is not recommended for various reasons. From an intuitive viewpoint, when one is trying to capture the physical characteristics of a natural phenomenon as portrayed in its time series of observations, it is more instructive and sensible to think of a periodic model as an extension of its nonseasonal counterpart. Hence, one can separately build models for each season of the year and then join them together to create the overall periodic model. Also, one can demonstrate theoretically that PAR and PARMA models can be written as equivalent multivariate AR and ARMA models, respectively, defined in Section 20.2. Conversely, multivariate AR and ARMA models can be represented as PAR and PARMA models, respectively.

The PAR family of models and some associated theoretical properties are presented in the next section. Following this, PARMA models are defined in Section 14.2.3.

14.2.2 PAR Models

Definition

For convenience, an observation in a time series is written in the same way as it is in Section 13.2.2 for deseasonalized models. When one is considering a time series having s seasons per year ($s = 12$ for monthly data) over a period of n years, let $z_{r,m}$ represent a time series observation in the rth year and mth season where $r = 1,2, \ldots, n$, and $m = 1,2, \ldots, s$. If required, the given data may be transformed by the Box-Cox transformation in [13.2.1] to form the transformed series denoted by $z_{r,m}^{(\lambda)}$. The purpose of the Box-Cox transformation is to correct problems such as heteroscedasticity and/or non-normality in the residuals of the PAR or PARMA model fitted to the time series.

In essence, a PAR model is formed by defining an AR model for each season of the year. The *PAR model* of order $(p_1,p_2, \ldots ,p_s)$ is defined for season m as

$$z_{r,m}^{(\lambda)} - \mu_m = \sum_{i=1}^{p_m} \phi_i^{(m)}(z_{r,m-i}^{(\lambda)} - \mu_{m-i}) + a_{r,m} \quad [14.2.1]$$

where μ_m is the mean of the series $z_{r,m}^{(\lambda)}$ for the mth season, $\phi_i^{(m)}$ is the AR coefficient for season m and ith lag, and $a_{r,m}$ is the innovation or white noise disturbance. The innovation series $a_{r,m}$ where $r = 1,2, \ldots , n$, is assumed to have an expected value of zero and a covariance defined by

$$cov(a_{r,m},a_{r,m-i}) = \begin{cases} \sigma_m^2, & i = 0, \\ 0, & i \neq 0 \ \text{for}\ i = 1,2, \ldots ,s \end{cases} \quad [14.2.2]$$

Hence, the $a_{r,m}$ disturbances are distributed as $\text{IID}(0,\sigma_m^2)$. By utilizing the backshift operator B, where $B^k z_{r,m}^{(\lambda)} = z_{r,m-k}^{(\lambda)}$, the model in [14.2.1] can be more succinctly written as

$$\phi^{(m)}(B)(z_{r,m}^{(\lambda)} - \mu_m) = a_{r,m}, \quad m = 1,2, \ldots ,s \quad [14.2.3]$$

where

$$\phi^{(m)}(B) = 1 - \phi_1^{(m)}B - \phi_2^{(m)}B^2 - \cdots - \phi_{p_m}^{(m)}B^{p_m}$$

is the AR operator of order p_m for season m in which $\phi_i^{(m)}$ is the ith AR parameter. For stationarity in season m, the roots of the seasonal characteristic equation $\phi^{(m)}(B) = 0$ must lie outside the unit circle. A necessary and sufficient condition for stationarity for a PAR model is given in [14.2.26].

Some authors recommend deseasonalizing the data using [13.2.3] before fitting a PAR or PARMA model to the time series [see, for example, Tao and Delleur (1976) and Croley and Rao (1977)]. However, when using the PAR model in [14.2.1] or [14.2.3], this step can easily be shown to be unnecessary, thereby reducing the number of model parameters. For example, suppose for the mth season that only one AR parameter were required and hence $p_m = 1$. From [14.2.1] or [14.2.3], this model is written as

$$z_{r,m}^{(\lambda)} - \mu_m = \phi_1^{(m)}(z_{r,m-1}^{(\lambda)} - \mu_{m-1}) + a_{r,m} \quad [14.2.4]$$

which can be equivalently given as

$$\frac{z_{r,m}^{(\lambda)} - \mu_m}{\sqrt{\gamma_0^{(m)}}} = \phi_1^{(m')} \left(\frac{z_{r,m-1}^{(\lambda)} - \mu_{m-1}}{\sqrt{\gamma_0^{(m-1)}}} \right) + a'_{r,m} \quad [14.2.5]$$

where

$$\gamma_0^{(m)} = var(z_{r,m}^{(\lambda)}), \quad \text{for}\ m = 1,2, \ldots ,s;$$

$$\phi_1^{(m')} = \left[\frac{\gamma_0^{(m-1)}}{\gamma_0^{(m)}}\right]^{0.5} \phi_1^{(m)};$$

and

$$a'_{r,m} = \left[\frac{\gamma_0^{(m-1)}}{\gamma_0^{(m)}}\right]^{0.5} a_{r,m} .$$

Stationarity

In Chapter 20, the general multivariate ARMA model is defined. As explained by authors such as Rose (1977), Newton (1982), Vecchia (1985a,b), Obeysekera and Salas (1986), Haltiner and Salas (1988), Bartolini et al. (1988) and Ula (1990), a PAR model can be equivalently written as a special case of the multivariate ARMA model. Because the stationarity conditions for a multivariate ARMA are known, they are also available for a PAR model. Consider, for example, the case of a PAR model in [14.2.4] for which there is one AR parameter for each of the s seasons. The stationarity requirement for this model is (Obeysekera and Salas, 1986)

$$\left| \prod_{m=1}^{s} \phi_1^{(m)} \right| < 1 \qquad [14.2.6]$$

Periodic Autocorrelation Function

The theoretical ACF for the PAR model in [14.2.1] or [14.2.3] for season m can be found by following a similar procedure to that used for obtaining the theoretical ACF for the nonseasonal AR model in Section 3.2.2. First, however, it is necessary to formulate some definitions. For season m, the theoretical *periodic autocovariance function* at lag k is defined for $z_{r,m}^{(\lambda)}$ as

$$\gamma_k^{(m)} = E[(z_{r,m}^{(\lambda)} - \mu_m)(z_{r,m-k}^{(\lambda)} - \mu_{m-k})] \qquad [14.2.7]$$

for $m = 1,2, \ldots, s$, where μ_m and μ_{m-k} are the theoretical means for seasons m and $m-k$, respectively. When $k = 0$, the periodic autocovariance is simply the variance, $\gamma_0^{(m)}$, of the random variable representing the observations in season m.

A standardized variable that is more convenient to deal with than $\gamma_k^{(m)}$, is the theoretical *periodic ACF* which is defined for season m at lag k as

$$\rho_k^{(m)} = \frac{\gamma_k^{(m)}}{\sqrt{\gamma_0^{(m)}\gamma_0^{(m-k)}}} \qquad [14.2.8]$$

Due to the form of [14.2.8], the theoretical periodic ACF is dimensionless and, hence, independent of the scale of measurement. Furthermore, the possible values of $\rho_k^{(m)}$ range from -1 to 1, where $\rho_k^{(m)}$ has a magnitude of unity at lag zero.

Given the above definitions of periodic linear dependence, one can find the theoretical periodic ACF for the PAR model in [14.2.1] or [14.2.3]. For season m, multiply [14.2.1] by $z_{r,m-k}^{(\lambda)} - \mu_{m-k}$ and take expected values to obtain

$$\gamma_k^{(m)} = \phi_1^{(m)}\gamma_{k-1}^{(m-1)} + \phi_2^{(m)}\gamma_{k-2}^{(m-2)} + \cdots + \phi_{p_m}^{(m)}\gamma_{k-p_m}^{(m-p_m)} + E[(z_{r,m-k}^{(\lambda)} - \mu_{m-k})a_{r,m}] \quad [14.2.9]$$

for $k \geq 0$ and $m = 1,2,\ldots,s$. The last term on the right hand side of [14.2.9] is zero for $k > 0$ because $z_{r,m-k}^{(\lambda)}$ is only a function of the disturbances $a_{r,m}$ up to time $m-k$ and $a_{r,m}$ is independent of these shocks. Hence, for $k > 0$ [14.2.9] becomes

$$\gamma_k^{(m)} = \phi_1^{(m)}\gamma_{k-1}^{(m-1)} + \phi_2^{(m)}\gamma_{k-2}^{(m-2)} + \cdots + \phi_{p_m}^{(m)}\gamma_{k-p_m}^{(m-p_m)} \quad [14.2.10]$$

By using the periodic AR operator given in [14.2.3], one can rewrite [14.2.10] for season m as

$$\phi^{(m)}(B)\gamma_k^{(m)} = 0 \quad \text{for } k > 0 \quad [14.2.11]$$

where B operates on the subscript k and the superscript (m) in $\gamma_k^{(m)}$. The relationship in [14.2.11] is valid for each season $m = 1,2,\ldots,s$. Because of the form of [14.2.10] and [14.2.11], the theoretical autocovariance function attenuates for a PAR process in season m when $p_m > 0$.

Periodic Yule-Walker Equations

Following the approach used for a nonseasonal AR model in Section 3.2.2, one can find the theoretical Yule-Walker equations for a PAR model. Specifically, by setting $k = 1,2,\ldots,p_m$, in [14.2.10], one obtains the *periodic Yule-Walker equations* for season m as:

$$\begin{array}{lllllll}
\gamma_1^{(m)} & = & \phi_1^{(m)}\gamma_0^{(m-1)} & + & \phi_2^{(m)}\gamma_1^{(m-2)} & + \cdots + & \phi_{p_m}^{(m)}\gamma_{p_m-1}^{(m-p_m)} \\
\gamma_2^{(m)} & = & \phi_1^{(m)}\gamma_1^{(m-1)} & + & \phi_2^{(m)}\gamma_0^{(m-2)} & + \cdots + & \phi_{p_m}^{(m)}\gamma_{p_m-2}^{(m-p_m)} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\
\gamma_{p_m}^{(m)} & = & \phi_1\gamma_{p_m-1}^{(m-1)} & + & \phi_2^{(m)}\gamma_{p_m-2}^{(m-2)} & + \cdots + & \phi_{p_m}^{(m)}\gamma_0^{(m-p_m)}
\end{array} \quad [14.2.12]$$

By writing the periodic Yule-Walker equations in [14.2.12] in matrix form, the relationship for expressing the AR parameters for season m is

$$\underline{\phi}^{(m)} = \left[\underline{\Gamma}_{p_m}^{(m)}\right]^{-1}\underline{\gamma}^{(m)} \quad [14.2.13]$$

where

$$\underline{\phi}^{(m)} = \begin{bmatrix} \phi_1^{(m)} \\ \phi_2^{(m)} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{p_m}^{(m)} \end{bmatrix}, \underline{\gamma}^{(m)} = \begin{bmatrix} \gamma_1^{(m)} \\ \gamma_2^{(m)} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_m^{(m)} \end{bmatrix}, \underline{\Gamma}_{p_m}^{(m)} = \begin{bmatrix} \gamma_0^{(m-1)} & \gamma_1^{(m-2)} & \cdots & \gamma_{p_m-1}^{(m-p_m)} \\ \gamma_1^{(m-1)} & \gamma_0^{(m-2)} & \cdots & \gamma_{p_m-2}^{(m-p_m)} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot \quad \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \gamma_{p_m-1}^{(m-1)} & \gamma_{p_m-2}^{(m-2)} & \cdots & \gamma_0^{(m-p_m)} \end{bmatrix}$$

By setting $k = 0$ in [14.2.9], the expression for the variance $\gamma_0^{(m)}$ is

$$\gamma_0^{(m)} = \phi_1^{(m)}\gamma_1^{(m-1)} + \phi_2^{(m)}\gamma_2^{(m-2)} + \cdots + \phi_{p_m}^{(m)}\gamma_{p_m}^{(m-p_m)} + \sigma_m^2 \qquad [14.2.14]$$

where $E[(z_{r,m}^{(\lambda)} - \mu_m)(a_{r,m})] = \sigma_m^2$ since $z_{r,m}^{(\lambda)}$ is only correlated with $a_{r,m}$ due to the most recent shock $a_{r,m}$. As is explained in Section 14.3.3, the periodic Yule-Walker equations in [14.2.12] or [14.2.13] provide a means of obtaining efficient moment estimates for the parameters of the PAR model in [14.2.1] or [14.2.3].

Periodic Partial Autocorrelation Function

Since the periodic autocorrelation function of a PAR model in season m for which $p_m > 0$ attenuates and does not truncate at a specified lag, it would be useful for identification purposes to define a function which cuts off. To accomplish this one can define the periodic PACF for a PAR model in a manner similar to that done in Section 3.2.2 for a nonseasonal AR model.

For season m, the *periodic PACF* is defined as the last AR parameter of an AR model of order p_m. Therefore, in the Yule-Walker equations in [14.2.12], $\phi_{p_m}^{(m)}$ is by definition the periodic PACF at lag p_m. By setting p_m to values of 1,2, . . . , in [14.2.12], one can define the periodic PACF in season m for lags 1,2, . . . , respectively. Because of the definition of the theoretical periodic PACF, it must be equal to zero after lag p_m in season m when the order of the AR model in this season is p_m. Furthermore, the possible values of the theoretical PACF fall between -1 and +1.

Markov Model

For a Markov model in season m the order is $p_m = 1$. A Markov model for season m is written in [14.2.4]. When the PAR is Markov for each of the s seasons, the stationarity condition for the overall Markov PAR model is the one given in [14.2.6].

The periodic Yule-Walker equations for a PAR model are written in [14.2.12]. By setting $\phi_2^{(m)}$ to $\phi_p^{(m)}$ equal to zero, this equation becomes

$$\gamma_1^{(m)} = \phi_1^{(m)}\gamma_0^{(m-1)}$$

$$\gamma_2^{(m)} = \phi_1^{(m)}\gamma_1^{(m-1)}$$

$$\gamma_3^{(m)} = \phi_1^{(m)}\gamma_2^{(m-1)}$$

In general,

$$\gamma_k^{(m)} = \phi_1^{(m)}\gamma_{k-1}^{(m-1)}$$

Hence, the theoretical periodic autocovariance function attenuates for increasing lag k. However, by definition the theoretical periodic PACF cuts off and is exactly equal to zero after lag one for a Markov model.

14.2.3 PARMA Models

Definition

As is also the case for the PAR model in Section 14.2.2, let $z_{r,m}^{(\lambda)}$ be an observation in the rth year and mth season for $r = 1,2,\ldots,n$, and $m = 1,2,\ldots,s$, where the exponent λ indicates that the observation may be transformed using the Box-Cox transformation in [13.2.1]. A PARMA model is created by defining a separate ARMA model for each season of the year. The PARMA model of order $(p_1,q_1;p_2,q_2;\cdots;p_s,q_s)$ is defined for season m as

$$\phi^{(m)}(B)(z_{r,m}^{(\lambda)} - \mu_m) = \theta^{(m)}(B)a_{r,m}, \quad m = 1,2,\ldots,s \qquad [14.2.15]$$

where μ_m is the mean for series $z_{r,m}^{(\lambda)}$ for the mth season, $\phi^{(m)}(B) = 1 - \phi_1^{(m)}B - \phi_2^{(m)}B^2 - \cdots - \phi_{p_m}^{(m)}B^{p_m}$, is the AR operator of order p_m for season m in which $\phi_i^{(m)}$ is the ith AR parameter, and $\theta^{(m)}(B) = 1 - \theta_1^{(m)}B - \theta_2^{(m)}B^2 - \cdots - \theta_{q_m}^{(m)}B^{q_m}$, is the MA operator of order q_m for season m in which $\theta_i^{(m)}$ is the ith MA parameter. The innovation series $a_{r,m}$ where $r = 1,2,\ldots,n$, for each m is assumed to be distributed as $\mathrm{IID}(0,\sigma_m^2)$ which is the same as that for the PAR model in [14.2.1].

Using the AR and MA operators to define the PARMA model in [14.2.15] provides an economical and convenient format for writing this model. Also, the operator format in [14.2.15] can be easily manipulated for mathematical purposes. Nonetheless, one could also write the PARMA model for season m without the operator notation as

$$z_{r,m}^{(\lambda)} - \mu_m = \sum_{i=1}^{p_m}\phi_i^{(m)}(z_{r,m-i} - \mu_{m-i}) + a_{r,m} - \sum_{i=1}^{q_m}\theta_i^{(m)}a_{r,m-i}, \quad m = 1,2,\ldots,s \qquad [14.2.16]$$

Stationarity and Invertibility

The PARMA model given in [14.2.15] can be equivalently written as a particular case of the general multivariate ARMA model presented in Chapter 20. Since the stationarity and invertibility conditions for the general multivariate ARMA model are available, they are, of course, also known for the PARMA model (Rose, 1977; Vecchia, 1985a,b; Obeysekera and Salas, 1986; Bartolini et al., 1988; Ula, 1990). As an example of how these conditions are written for a specific PARMA model, consider a PARMA model from [14.2.15] for which there is one AR and one MA parameter for each of the m seasons. The stationarity restriction for this model is given in [14.2.6] while the invertibility requirement is

$$\left|\prod_{m=1}^{s}\theta_1^{(m)}\right| < 1 \qquad [14.2.17]$$

Periodic Autocorrelation Function

In Section 3.4.2, it is explained how the theoretical autocovariance function or, equivalently, the theoretical ACF can be determined for a nonseasonal ARMA(p,q) model. A similar approach can be followed to derive the system of equations for solving for the periodic autocovariance function in [14.2.7] for a PARMA model.

The steps required for accomplishing this are now described. For a given season m, multiply both sides of [14.2.15] by $z_{r,m-k}^{(\lambda)} - \mu_{m-k}$ and take the expected values to obtain

$$\gamma_k^{(m)} - \phi_1^{(m)}\gamma_{k-1}^{(m-1)} - \phi_2^{(m)}\gamma_{k-2}^{(m-2)} - \cdots - \phi_{p_m}^{(m)}\gamma_{k-p_m}^{(m-p_m)}$$
$$= \gamma_{za}^{(m)}(k) - \theta_1^{(m)}\gamma_{za}^{(m-1)}(k-1) - \cdots - \theta_{q_m}^{(m)}\gamma_{za}^{(m-q_m)}(k-q_m) \qquad [14.2.18]$$

where $\gamma_k^{(m)}$ is the theoretical periodic autocovariance function in [14.2.7] and

$$\gamma_{za}^{(m)}(k) = E[(z_{r,m-k}^{(\lambda)} - \mu_{m-k})a_{r,m}] \qquad [14.2.19]$$

is the cross covariance function between $z_{r,m-k}^{(\lambda)} - \mu_{m-k}$ and $a_{r,m}$. Since $z_{r,m-k}^{(\lambda)}$ is only dependent upon shocks which have occurred up to time $(r,m-k)$, it follows that

$$\gamma_{za}^{(m)}(k) = 0, \quad k > 0$$
$$\gamma_{za}^{(m)}(k) \neq 0, \quad k \leq 0 \qquad [14.2.20]$$

Because of the $\gamma_{za}^{(m)}(k)$ terms in [14.2.18], one must derive other relationships before one can solve for the periodic autocovariances. This can be carried out by multiplying [14.2.15] by $a_{r,m-k}$ and taking expectations to obtain

$$\gamma_{za}^{(m-k)}(-k) - \phi_1^{(m)}\gamma_{za}^{(m-k)}(-k+1) - \phi_2^{(m)}\gamma_{za}^{(m-k)}(-k+2) - \cdots - \phi_{p_m}^{(m)}\gamma_{za}^{(m-k)}(-k+p_m)$$
$$= -[\theta_k^{(m)}]\sigma_m^2 \qquad [14.2.21]$$

where

$$[\theta_k^{(m)}] = \begin{cases} \theta_k^{(m)}, & k = 1,2,\ldots,q_m \\ -1, & k = 0 \\ 0, & otherwise \end{cases}$$

and $E[a_{r,m}a_{r,m-k}]$ is as defined in [14.2.2].

Equations [14.2.18] and [14.2.21] can be employed to solve for the theoretical *periodic autocovariance function* for a PARMA model for each season. For $k > q_m$, equation [14.2.18] reduces to

$$\gamma_k^{(m)} - \phi_1^{(m)}\gamma_{k-1}^{(m-1)} - \phi_2^{(m)}\gamma_{k-2}^{(m-2)} - \cdots - \phi_{p_m}^{(m)}\gamma_{k-p_m}^{(m-p_m)} = 0$$

or

$$\phi_m^{(m)}(B)\gamma_k^{(m)} = 0 \qquad [14.2.22]$$

where the differencing operator B operates on both the subscript and superscript in $\gamma_k^{(m)}$. If $k > \max(p_m,q_m)$, then [14.2.22] can be used to calculate the $\gamma_k^{(m)}$ directly from previous values. For $k = 0,1,2,\ldots,\max(p_m,q_m)$, equation [14.2.21] can be employed for solving for the periodic cross covariance function $\gamma_{za}^{(m)}(k)$ which can be substituted into [14.2.18] in order to solve for the periodic autocovariance function for the $z_{r,m}^{(\lambda)}$. By employing [14.2.8], one can easily calculate the theoretical periodic ACF after determining the theoretical periodic autocovariance function.

Recall that for a nonseasonal ARMA model in Section 3.4.2, the theoretical autocovariance function or theoretical ACF attenuates for increasing values of lag k. In a similar fashion, one can see from the form of [14.2.20] that the theoretical periodic autocovariance function dies off for a PARMA model in which $p_m \neq 0$ in season m.

Periodic Partial Autocorrelation Function

For season m, the PARMA model in [14.2.15] can be written as an infinite AR model by writing it as

$$a_{r,m} = \theta^{(m)}(B)^{-1}\phi^{(m)}(B)(z_{r,m}^{(\lambda)} - \mu_m) \tag{14.2.23}$$

where $\theta^{(m)}(B)^{-1}$ is an infinite series in B for $q_m \geq 1$. Because the definition of the theoretical *periodic PACF* is based upon an AR process, the periodic PACF is infinite in extent for a PARMA model and dies off with increasing lag. At higher lags, the behaviour of the periodic PACF depends upon the MA parameters and is dominated by a combination of damped exponentials and/or damped sine waves.

Three Formulations of a PARMA Model

In Section 3.4.3, it is explained how a nonseasonal ARMA model can be expressed in three equivalent forms. These same three formats can also be used with a PARMA model in season m. One formulation is to use the difference equation given in [14.2.15]. A second technique is to write the model as a pure MA model, which is also called the *random shock form*. Finally, by formulating the model as a pure AR model one obtains the inverted form for the model.

In random shock form, the PARMA model for season m is written as

$$\begin{aligned} z_{r,m}^{(\lambda)} - \mu_m &= \phi^{(m)}(B)^{-1}\theta^{(m)}(B)a_{r,m} \\ &= a_{r,m} + \psi_1^{(m)}a_{r,m-1} + \psi_2^{(m)}a_{r,m-2} + \cdots \\ &= a_{r,m} + \psi^{(m)}Ba_{r,m} + \psi_2^{(m)}B^2a_{r,m} + \cdots \\ &= (1 + \psi_1^{(m)}B + \psi_2^{(m)}B^2 + \cdots)a_{r,m} \\ &= \psi^{(m)}(B)a_{r,m} \end{aligned} \tag{14.2.24}$$

where $\psi^{(m)}(B) = 1 + \psi_1^{(m)}B + \psi_2^{(m)}B^2 + \cdots$, is the random shock or infinite MA operator for season m and $\psi_i^{(m)}$ is the ith parameter, coefficient or weight of $\psi^{(m)}(B)$. There are a variety of reasons for expressing a model in random shock form. For example, when forecasting in season m the $\psi_i^{(m)}$ weights are needed to calculate the variance of the forecasts (see [8.2.13] for the case of an ARMA model). When simulating in season m using a PARMA model, one way to simulate data is to write the model in random shock form and then to use this structure for producing the synthetic sequences (see Section 9.3 for the case of an ARMA model). Finally, by writing PARMA models in random shock form, the magnitude and sign of the $\psi_i^{(m)}$ parameters can be compared across models.

Following the arguments given in Section 3.4.3 to develop [3.4.21] for an ARMA model, one can obtain the $\psi_k^{(m)}$ weights from the $\phi_k^{(m)}$ and $\theta_k^{(m)}$ parameters for a PARMA model in season m by utilizing the expression

$$\phi^{(m)}(B)\psi_k^{(m)} = -\theta_k^{(m)} \qquad [14.2.25]$$

where B operates on k, $\psi_0^{(m)} = 1$, $\psi_k^{(m)} = 0$ for $k < 0$ and $\theta_k^{(m)} = 0$ if $k > q$. Rules for deciding upon how many random shock parameters to calculate and examples for determining these parameters are given in Section 3.4.3.

Because a PAR model is a special case of a PARMA model, one can, of course, write a PAR for season m in the random shock form given in [14.2.4]. As shown by Troutman (1979), a necessary and sufficient condition for periodic stationarity for a PAR model is

$$\sum_{i=0}^{\infty}(\psi_i^{(m)})^2 < \infty, \quad m = 1,2,\ldots,s \qquad [14.2.26]$$

To express the PARMA model in season m in *inverted form*, equation [14.2.15] is rewritten as

$$\begin{aligned} a_{r,m} &= \theta^{(m)}(B)^{-1}\phi^{(m)}(B)(z_{r,m}^{(\lambda)} - \mu_m) \\ &= (z_{r,m}^{(\lambda)} - \mu_m) - \Pi_1^{(m)}(z_{r,m-1}^{(\lambda)} - \mu_{m-1}) - \Pi_2^{(m)}(z_{r,m-2}^{(\lambda)} - \mu_{m-2}) - \cdots \\ &= (z_{r,m} - \mu_m) - \Pi_1^{(m)}B(z_{r,m}^{(\lambda)} - \mu_m) - \Pi_2^{(m)}B^2(z_{r,m}^{(\lambda)} - \mu_m) - \cdots \\ &= (1 - \Pi_1^{(m)}B - \Pi_2^{(m)}B^2 - \cdots)(z_{r,m}^{(\lambda)} - \mu_m) \\ &= \Pi^{(m)}(B)(z_{r,m}^{(\lambda)} - \mu_m) \end{aligned} \qquad [14.2.27]$$

where $\Pi^{(m)}(B) = 1 - \Pi_1^{(m)}B - \Pi_2^{(m)}B^2 - \cdots$, is the inverted or infinite AR operator for season m and $\Pi_i^{(m)}$ is the ith parameter, coefficient or weight of $\Pi^{(m)}(B)$. By comparing [14.2.26] and [14.2.24], one can see that

$$\psi^{(m)}(B)^{-1} = \Pi^{(m)}(B) \qquad [14.2.28]$$

Given the seasonal AR and MA parameters, one may wish to determine the inverted parameters. To achieve this, one can use the expression

$$\theta^{(m)}(B)\Pi_k^{(m)} = \phi_k^{(m)} \qquad [14.2.29]$$

where B operates on k, $\Pi_0^{(m)} = -1$, $\Pi_k^{(m)} = 0$ for $k < 0$, and $\phi_k^{(m)} = 0$ if $k > p$. Except for notational differences, [14.2.28] is the same as [3.4.27] which is used for obtaining the inverted weights for a nonseasonal ARMA model. Representative examples for calculating the inverted weights are presented in Section 3.4.3.

Example of a PARMA Model

If $p_m = q_m = 1$ for season m, a PARMA model for that season is written following [14.2.15] as

$$(1-\phi_1^{(m)}B)(z_{r,m}-\mu_m)=(1-\theta_1^{(m)}B)a_{r,m} \qquad [14.2.30]$$

To obtain the theoretical autocovariance function, one must solve [14.2.18] and [14.2.21] after setting all AR and MA parameters equal to zero except for $\phi_1^{(m)}$ and $\theta_1^{(m)}$. The reader can refer to Section 3.4.3 for examples of how to calculate the random shock and inverted parameters for a nonseasonal ARMA(1,1) model. The same approaches can be used for a PARMA model with $p_m = q_m = 1$ by replacing ϕ_i, θ_i, ψ_i, and Π_i by $\phi_i^{(m)}$, $\theta_i^{(m)}$, $\psi_i^{(m)}$ and $\Pi_i^{(m)}$ in Section 3.4.3.

14.3 CONSTRUCTING PAR MODELS

14.3.1 Introduction

As noted in Section 14.1, model construction techniques for PAR models are highly developed. Indeed, as demonstrated by research referenced in Section 14.1, PAR models can be conveniently used in practical applications and produce useful results. Consequently, this section concentrates upon how to construct PAR models by following the three stages of model construction. Applications for clearly illustrating how the construction techniques for PAR modelling are implemented in practice are presented in Sections 14.5 and 14.6 as well as Chapter 15. Finally, model construction methods for PPAR and PARMA models are given in Sections 14.5.3 and 14.7, respectively.

14.3.2 Identifying PAR Models

Introduction

Thomas and Fiering (1962) originally suggested that one could fit PAR models of order one for each season to monthly hydrological time series. 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. Based upon the results of an extensive forecasting study, Noakes et al. (1985) recommend that the best way to identify a PAR model is to employ the periodic ACF and PACF. Consequently, this approach to designing a PAR model is explained in this section. Another identification method which uses the AIC in conjunction with subset autoregression and the algorithm of Morgan and Tatar (1972) is outlined in Section 14.3.3. Moreover, two procedures for efficiently estimating the parameters of PAR models are described in Section 14.3.3 while diagnostic checks are discussed in Section 14.3.4. Finally, the results of the forecasting study of Noakes et al. (1985) are presented in Section 15.3 to demonstrate that PAR models identified using the periodic ACF and PACF forecast better than PAR models designed using other approaches as well as the deseasonalized and SARIMA models of Chapters 13 and 12, respectively.

Sample Periodic ACF: The theoretical periodic autocovariance function and ACF at lag k for the series $z_{r,m}^{(\lambda)}$ are defined in [14.2.7] and [14.2.8], respectively. In a practical application, the theoretical variables used in these equations are estimated using the sample time series $z_{r,m}$, where the years $r = 1,2,\ldots,n$, and the seasons $m = 1,2,\ldots,s$. To rectify problems with non-normality and/or heteroscedasticity in the residuals of the fitted PAR model, often the original series, $z_{r,m}$, is transformed using the Box-Cox transformation in [13.2.1] to obtain the transformed series $z_{r,m}^{(\lambda)}$. The theoretical variables in [14.2.7] and [14.2.8] are then estimated for the $z_{r,m}^{(\lambda)}$ series. More specifically, for the mth season, the mean, μ_m, is estimated using

$$\hat{\mu}_m = \frac{1}{n}\sum_{r=1}^{n} z_{r,m}^{(\lambda)} \qquad [14.3.1]$$

where $m = 1,2,\ldots,s$. To estimate the theoretical periodic autocovariance function, $\gamma_k^{(m)}$, in [14.2.7] for lag k and season m, the following formula is utilized:

$$c_k^{(m)} = \frac{1}{n}\sum_{r=1}^{n}(z_{r,m}^{(\lambda)} - \hat{\mu}_m)(z_{r,m-k}^{(\lambda)} - \mu_{m-k}) \qquad [14.3.2]$$

for $m = 1,2,\ldots,s$. When the lag k is zero, one obtains the estimate of the variance of the observations in season m, which is given as:

$$c_o^{(m)} = \frac{1}{n}\sum_{r=1}^{n}(z_{r,m}^{(\lambda)} - \hat{\mu}_m)^2, \quad m = 1,2,\ldots,s. \qquad [14.3.3]$$

The *sample or estimated theoretical periodic ACF* at lag k is determined for $\rho_k^{(m)}$ using

$$r_k^{(m)} = \frac{c_k^{(m)}}{\sqrt{c_o^m c_o^{m-k}}} \qquad [14.3.4]$$

where $m = 1,2,\ldots,s$.

Because the periodic ACF is symmetric about lag zero, it is only necessary to plot the sample ACF for season m from lag one to a maximum lag of about $n/4$. A separate sample ACF graph is made for each season of the year. To ascertain which values of the estimated ACF for period or season m are significantly different from zero, the approximate 95% confidence interval can be plotted. The sample ACF is asymptotically distributed as NID$(0,\frac{1}{n})$ at any lag. Consequently, the approximate 95% confidence interval is $\pm 1.96\sqrt{n}$.

As explained in Section 14.2.2, the theoretical ACF for a PAR model in season m, attenuates if AR parameters are in the model. Consequently, if the sample periodic ACF dies off for season m, this indicates that one or more AR parameters are needed in this season for the PAR model which is fitted to the series. If no values of the sample periodic ACF are significantly different from zero, this means that one can model this season using white noise by setting $p_m = 0$ in the PAR model in [14.2.3].

Sample Periodic PACF: For a given seasonal time series, the periodic PACF can be determined for each season of the year. The definition for the periodic PACF is derived from the definition of the PAR model. In particular, assuming that the AR model for season m is of order p_m, the PACF for that season is $\phi_{p_m}^{(m)}$. Be setting $p_m = 1,2,\ldots,$ the PACF is defined for lags 1,2,... .

For the case of a nonseasonal time series, one uses the Yule-Walker equations in [3.2.12] or [3.2.17] to estimate the PACF. Likewise, for the situation of a periodic or seasonal time series one can utilize the periodic Yule-Walker equations in [14.2.12] or [14.2.13] to estimate the periodic PACF.

To obtain Yule-Walker estimates for the AR parameters for season m in the PAR model in [14.2.1] or [14.2.3], simply replace each $\gamma_k^{(m)}$ in [14.2.12] or [14.2.13] by its estimate $c_k^{(m)}$ from [14.3.2]. This can be carried out for each of the s seasons by estimating the $\phi_i^{(m)}$ for $i = 1,2,\ldots,p_m$, for each season $m = 1,2,\ldots,s$. The resulting estimated PAR model is periodic stationary (Troutman, 1979) and the estimates are asymptotically efficient (Pagano, 1978). Furthermore, the estimates corresponding to different seasons are asymptotically independent. Sakai (1982) presents a practical computational algorithm for estimating the periodic AR parameters and, hence, also the periodic PACF from the periodic Yule-Walker equations in [14.2.12] or [14.2.13].

For period or season m, the correct order for a PAR model is given as p_m in [14.2.1]. Sakai (1982) shows that the sample PACF for a given season is asymptotically distributed as NID$(0,\frac{1}{n})$ at any lag greater than p_m. Therefore, the 95% confidence interval is $\pm 1.96\sqrt{n}$. The sample PACF and approximate 95% confidence interval can be plotted for each season up to a maximum lag of about $\frac{n}{4}$.

By definition the theoretical PACF in season m cuts off after lag p_m to a value of exactly zero. Consequently, if the sample periodic PACF is not significantly different from zero after lag p_m, this indicates that the order of the AR model fitted to the series in season m should be p_m. If none of the values of the sample periodic PACF in season are significantly different from zero, the model for season m within the overall PAR model should be white noise. In this case, p_m is set equal to zero.

Periodic IACF and IPACF

In Section 5.3, the sample IACF and IPACF are recommended as additional identification tools for determining the orders of the AR and MA operators in a nonseasonal ARMA model. One can define the periodic versions of these functions for use in identifying the order of the AR model for each season of a PAR model.

For season m, the *theoretical periodic IACF* of a PARMA model is defined to be the ACF of a PARMA model having the AR and MA components of orders q_m and p_m, respectively (i.e. the AR and MA operators are exchanged with one another). The PACF of this process for season m is defined to be the *theoretical periodic IPACF*.

For a PAR model having an AR operator of order p_m in season m, the IACF truncates after lag p_m. Thus, the behaviour of the periodic IACF is similar to that of the periodic PACF. Likewise, the periodic IPACF mimics the behaviour of the periodic ACF. For both of these latter functions, their values die off for increasing lags in season m when $p_m \neq 0$.

Further research is required for obtaining efficient estimates for the sample periodic IACF and IPACF. One could, for example, adopt estimation procedures similar to those developed for the nonseasonal versions of these functions in Chapter 5.

Tests for Periodic Correlation

The sample periodic ACF and PACF provide a means for detecting periodic correlation in seasonal time series and also information for designing a PAR model to fit to the series. Other approaches for finding periodic correlation in a data set include the statistical tests described by Hurd and Gerr (1991) and Vecchia and Ballerini (1991). When periodic correlation is present, one, should, of course, fit a periodic model such as a PAR or PARMA model to the time series under consideration. Tiao and Gruppe (1980) discuss the negative consequences of not using an appropriate periodic model when the data possesses periodic correlation.

14.3.3 Calibrating PAR Models

Introduction

A major advantage of using PAR models in practical applications is that two good algorithms are available for estimating the parameters of PAR models. In particular, the two estimation methods described in the next two subsections are the Yule-Walker estimator and multiple linear regression. These two estimator techniques are efficient both from statistical and computational viewpoints.

For deciding upon the order of the AR operator in each season, one can use plots of the sample periodic ACF and PACF, as explained in Section 14.3.2. Additionally, one can employ the AIC which is derived for the case of PAR models in this section. Finally, it is explained how the algorithm of Morgan and Tatar (1972) can be used in conjunction with the AIC to select the order of each AR operator in a PAR model.

Periodic Yule-Walker Estimator

The technique for obtaining Yule-Walker estimates for the parameters of a PAR model is explained in Section 14.3.2 under the subsection entitled Sample Periodic PACF. Even though this method is in fact a moment estimator, it is still efficient statistically for use with PAR model.

As discussed in Section 14.3.2, the periodic Yule-Walker equation in [14.2.12] or [14.2.13] can be used to obtain the estimates of the parameters for each season. Each theoretical ACF is replaced by its sample estimate from [14.3.2] and then the algorithm of Sakai (1982) is used to estimate the AR parameters for each season m using the periodic Yule-Walker equations. The parameters for each season can be estimated separately and the parameter estimates are asymptotically efficient (Pagano, 1978). Furthermore, the calibrated PAR model is periodic stationary (Troutman, 1979).

Multiple Linear Regression

Although λ could be estimated, assume that it is fixed at some value such as $\lambda = 0.5$ or $\lambda = 0$ for a square root or natural logarithmic transformation, respectively. For season m, the mean parameter μ_m is estimated by

$$\hat{\mu}_m = \frac{1}{n}\sum_{r=1}^{n} z_{r,m}^{(\lambda)}, \quad m = 1,2,\ldots,s \qquad [14.3.5]$$

For season or period (m), let $\beta_m = \phi_1^{(m)},\phi_2^{(m)},\ldots,\phi_{p_m}^{(m)}$, denote the vector of AR parameters for the PAR model and $\beta_m = \hat{\phi}_1^{(m)},\hat{\phi}_2^{(m)},\ldots,\hat{\phi}_{p_m}^{(m)}$, stand for the vector of estimated parameters. An efficient conditional maximum likelihood estimate $\hat{\beta}_m$ of $\hat{\beta}_m$ is are obtained directly from the multiple linear regression of $z_{r,m}^{(\lambda)}$ on $z_{r,m-1}^{(\lambda)}, z_{r,m-2}^{(\lambda)},\ldots,z_{r,m-p_m}^{(\lambda)}$.

The estimated innovations or residuals denoted as $\hat{a}_{r,m}$ are calculated from [14.2.3] by setting initial values to zero and the residual variance, σ_m^2 is then estimated by

$$\hat{\sigma}_m^2 = \frac{1}{n}\sum_{r=1}^{n} \hat{a}_{r,m}^2, \quad m = 1,2,\ldots,s \qquad [14.3.6]$$

Other Estimation Results

Pagano (1978) shows that $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normally distributed with mean zero and covariance matrix $\frac{1}{n}\mathbf{I}_m^{-1}$, where

$$\mathbf{I}_m = \frac{1}{\sigma_m^2}(\gamma_{i-j}^{(m)})$$

In practice, an estimate, $\hat{I}_m$ of I_m is obtained by replacing each $\gamma_k^{(m)}$ in [14.2.7] by its estimate $c_k^{(m)}$ in [14.3.2].

Pagano (1978) also demonstrates that the estimates for different periods are asymptotically uncorrelated. In other words, the joint information matrix of $\beta_1,\beta_2,\ldots,\beta_s$ is block diagonal. Consequently, the parameters for the mth season, can be estimated entirely independently of the parameters of any other season. Thus, for purposes of identification, estimation, and diagnostic checking, each season can be modelled independently of the other seasons.

When estimating the parameters in a PAR model, the orders of the AR operators can be different across the seasons. Furthermore, *subset autoregression* (McClave, 1975) can be used for constraining AR parameters to zero. For example, in season m for a monthly time series, one may wish to estimate only the AR parameters $\phi_1^{(m)}$, $\phi_2^{(m)}$ and $\phi_{12}^{(m)}$. The parameters from $\phi_3^{(m)}$ to $\phi_{11}^{(m)}$, are omitted from the model and subset autoregression is used to estimate the remaining parameters.

Model Selection using the AIC

From Section 6.3, the general formula for the AIC is defined as

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

where ML stands for the maximized value of the likelihood function and k is the number of free parameters. When using the MAICE procedure, one selects the model which gives the minimum value of the AIC.

Assuming normality, the maximized log likelihood of the AR model for season m is derived as (McLeod and Hipel, 1978)

$$\log L_m = -n\ln(\hat{\sigma}_m) + (\lambda - 1)\sum_{r=1}^{n} z_{r,m} \qquad [14.3.7]$$

The summation term on the right hand side of [14.3.7] takes into account the Jacobian of the Box-Cox transformation. An explanation of how this is done is given in Section 13.3.3 for the deseasonalized model.

The AIC formula for the mth season is

$$AIC_m = -2\log L_m + 2p_m + 4 \qquad [14.3.8]$$

where p_m is the number of AR parameters in season m. Because the mean, μ_m, and the variance of the innovations are estimated, the last term on the right hand side of [14.3.8] is included in the seasonal AIC formula.

For each combination of AR parameters, the AIC_m can be calculated using [14.3.7] and [14.3.8]. The model which yields the minimum value of the AIC_m is selected for season m. This procedure is executed for choosing the models for all of the remaining seasons. Subsequently, the AIC for the overall PAR model is

$$AIC = \sum_{m=1}^{s} AIC_m + 2 \qquad [14.3.9]$$

where the constant 2 allows for the Box-Cox parameter λ. The calculations of AIC may be repeated for several values of λ such as $\lambda = 1, 0.75, 0.5, \ldots, -1$, and the transformation yielding the minimum value of the AIC is selected.

Exhaustive Enumeration for PAR Model Selection

As mentioned earlier in Section 14.3.2, the recommended procedure for identifying the most appropriate PAR model or set of models to fit to a seasonal series is to employ the sample periodic ACF and PACF. If there is more than one promising model, the MAICE procedure can then be used to select the best one.

Another approach for determining the best AR model for each season where the maximum value of p_m is specified, would be to examine all possible regressions for that season. An appropriate criterion, such as the AIC, could be invoked for choosing the most desirable model from the exhaustive set of models. This procedure could be carried out for each season and this would result in selecting the most suitable PAR model over all of the seasons. If, for the case of a monthly series, the maximum value of p_m were restricted to be 12 for each month, the AR model for the month of March, for example, may only have AR parameters, at lags 1, 2, 3 and 12 while the other parameters would be constrained to be zero.

A possible difficulty with the aforesaid procedure is the amount of computer time required for estimating the parameters for all possible regressions for each season. For a monthly model, for example, there are 4096 possible regression models for each month and 2^{144} possible orders of monthly AR models with $p_m \leq 12$, $m = 1, 2, \ldots, 12$. Fortunately, Morgan and Tatar (1972) have devised an efficient procedure for calculating the residual sum of squares for each regression. This method drastically reduces the computational effort involved when considering an exhaustive regression study.

Because the residual sum of squares can be calculated efficiently for each regression (Morgan and Tatar, 1972), the AIC can be employed for model discrimination. In particular the residual sum of squares is used in [14.3.6] to estimate σ_m^2 and then the value of the AIC in [14.3.8] can be calculated without having to estimate the AR parameters. By selecting the model with the minimum value of the AIC, this insures that the number of model parameters are kept to a minimum and also the PAR model provides a good statistical fit to the data. Using these techniques, the best fitting PAR model for monthly data can usually be selected in less than one minute of computer time.

Subsequent, to identifying the most desirable model for season m according to the exhaustive enumeration approach, the AR parameter for this model can be estimated using subset autoregression. This procedure is repeated for each of the seasons. The value of the AIC for the overall PAR model can then be determined using [14.3.9].

A possible drawback of this exhaustive enumeration approach is that models may be identified that cannot be justified from a physical viewpoint. For instance, is it reasonable in the month of July for an average monthly riverflow series to have AR parameters for lags 2, 5 and 8? On the other hand if there were AR parameters identified for lags 1, 2 and 12, this could be justifiable from a hydrological understanding of the physical phenomenon. Applications of the exhaustive enumeration approach to average monthly riverflow time series are presented by McLeod and Hipel (1978).

14.3.4 Checking PAR Models

The adequacy of a fitted model can be ascertained by examining the properties of the residuals for each season. In particular, the residuals should be uncorrelated, normally distributed and homoscedastic.

To ascertain if the residuals are white, one must estimate the *periodic RACF* (residual autocorrelation function). For season m, the RACF at lag k is estimated using

$$r_k^{(m)}(\hat{a}_{r,m}) = \frac{\frac{1}{n}\sum_{r=1}^{n}\hat{a}_{r,m}\hat{a}_{r,m-k}}{\hat{\sigma}_m\hat{\sigma}_{m-k}}, \quad k = 1,2,\cdots \qquad [14.3.10]$$

Note that is necessary to divide by $\hat{\sigma}_m\hat{\sigma}_{m-k}$ rather than $\hat{\sigma}_m^2$ since in general $\hat{\sigma}_m \neq \hat{\sigma}_{m-k}$. Use of the incorrect divisor, $\hat{\sigma}_m^2$, could result in correlation values greater than 1.

For each season, one can plot $r_k^{(m)}(\hat{a}_{r,m})$ up to about lag $\frac{n}{4}$. Because $r_k^{(m)}(\hat{a}_{r,m})$ is asymptotically distributed as NID(0, $\frac{1}{n}$), one can also draw the 95% confidence interval for each season. If the seasonal residuals are white, they should fall within the 95% confidence limits. Nonwhiteness indicate that additional AR parameters are needed in season m or perhaps another class of models should be considered.

As a single statistic for an overall test for whiteness of the residuals, one can use the Portmanteau statistic for season m given by

$$Q'^{(m)}_L = n\sum_{k=1}^{L}(r_k^{(m)})^2(\hat{a}_{r,m}) \qquad [14.3.11]$$

This statistic is χ^2 distributed with $L - p_m$ degrees of freedom (Box and Jenkins, 1976; Box and Pierce, 1970). A significantly large value of $Q_L^{(m)}$ indicates inadequacy of the model for season m. Hence, one can reject the null hypothesis that the data in season m are white if the calculated value of $Q_L^{(m)}$ in [14.3.11] is larger than the tabulated χ^2 value at a specified significance level. One can choose L to be large enough to cover lags at which correlation could be expected to occur. For example, for monthly data, one may wish to set $L = 12$ if sufficient data are available.

As shown by McLeod (1993), a modified Portmanteau test statistic improves the small sample properties. In particular, the following exact result holds for the periodic correlations fo white noise

$$Var(r_k^{(m)}(a_{r,m})) = \frac{n - \frac{k}{s}}{n(n+2)}, \text{ if } k \equiv 0 \text{ mod} s$$

$$= \frac{n - \left[\frac{k - m + s}{s}\right]}{n^2}, \text{ otherwise} \qquad [14.3.12]$$

where $[\cdot]$ denotes the integer part and $r_k^{(m)}(a_{r,m})$ is defined in [14.3.10] by replacing the residual, $\hat{a}_{r,m}$, by the theoretical innovation, $a_{r,m}$. The modified Portmanteau statistic is then defined as

$$Q''^{(m)}_L = \sum_{k=1}^{L} \frac{(r^{(m)})^2(\hat{a}_{r,m})}{\sqrt{Var(r_k^{(m)}(a_{r,m}))}} \qquad [14.3.13]$$

which is χ^2 distributed with $L - p_m$ degrees of freedom. The modified statistic in [14.3.13] reduces to that proposed for a nonseasonal ARMA model in [7.3.5] by Davies et al. (1977) and Ljung and Box (1978). One can demonstrate that

$$E\{Q''^{(m)}_L\} = L - p_m \qquad [14.3.14]$$

and

$$E\{Q'^{(m)}_L\} = n\sum_{k=1}^{L} Var(r_k^{(m)}(a_{r,m})) - p_m \qquad [14.3.15]$$

Across seasons the Portmanteau test statistics are asymptotically independent for $m = 1, 2, \ldots, s$. Consequently, for the case of the Portmanteau test statistic in [14.3.13] an overall check to test if the residuals across all the seasons are white is given by

$$Q''_L = \sum_{m=1}^{s} Q_L^{(m)} \qquad [14.3.16]$$

where Q''_L is χ^2 distributed on $\sum_{m=1}^{s}(L - p_m)$ degrees of freedom. The lag L used in [14.3.13] could be chosen to be different across the seasons but in most applications it is reasonable to use the same value of L for all seasons. One can also employ $Q'^{(m)}_L$ in place of $Q''^{(m)}_L$ in [14.3.16] to obtain Q'_L.

One can use the tests for normality and homoscedasticity presented in Sections 7.4 and 7.5, respectively, to check that these assumptions are satisfied for the residuals in each season. These tests could also be used to ensure the assumptions hold across all of the seasons. Heteroscedasticity and/or non-normality, can often be corrected using the Box-Cox transformation in [13.2.1].

14.4 PAR MODELLING APPLICATION

The model construction techniques of Section 14.3 are employed for determining the most appropriate PAR model to fit to the average monthly flows of the Saugeen River at Walkerton, Ontario, Canada, which are available from Environment Canada (1977) from January, 1916 to December, 1976. From the sinusoidal structure contained in the graph of the last ten years of the average monthly Saugeen riverflows shown in Figure VI.1, one can see that the observations are highly seasonal. As emphasized in Section 14.3.2 and Chapter 15, the recommended approach for identifying the AR parameters required in each season for the PAR model is to use the sample periodic ACF and PACF. Because it is known a priori that most average monthly riverflow series require a natural logarithmic transformation to avoid problems with the residuals of the fitted model, the logarithmic Saugeen flows are used right at the start of the identification stage.

Figure 14.4.1, displays the graph of the periodic ACF against lag k for the logarithmic monthly Saugeen riverflows. Notice that each period or month possesses an ACF which is plotted vertically. The two lines above a given period show the 95% confidence interval. To keep the graph simple, the zero line, which falls midway between the confidence interval, is not given. Opposite a particular lag, the estimated value of the ACF for a given period is plotted horizontally. If the line cuts the left or right line for the confidence interval, the value of the sample ACF is significantly different from zero. Notice in Figure 14.4.1 that the estimated periodic ACF at lag 1 is significantly different from zero for all periods or months except for March (period 3) where the value just touches the 95% confidence limits. Because flows in one month are usually correlated with flows in the previous month, this behaviour would be expected. In addition, for some of the months such as January, October, November and December, which are indicated by periods 1, 10, 11, and 12, respectively, it appears that the ACF may be attenuating.

To identify more clearly the order of the AR model in each season, one must examine the sample periodic PACF plotted in Figure 14.4.2. Notice that the sample PACF for each period or season $m = 1,2,\ldots,12$, is plotted vertically along with the 95% confidence interval. There are significantly large values of the sample PACF at lag 1 across all 12 of the months, although in period 3 or March, the sample PACF is only just touching the 95% confidence interval. Furthermore, for all the months the sample PACF truncates and is not significantly different from zero after lag 1. Therefore, the identification plots indicate that for all months except possibly March, one should use an AR model of order 1 or a Markov model.

The parameters in the PAR(1,1,0,1,1,1,1,1,1,1,1,1) model are estimated using the periodic Yule-Walker equations in [4.2.12] for each season. The fitted model satisfies the tests for whiteness, heteroscedasticity and normality described in Section 14.3.4. For example, when the sample periodic RACF is plotted, the assumption of whiteness for the values of the RACF for each of the months is reasonably well satisfied. In particular, Figure 14.4.3 shows a graph of the periodic RACF calculated using [14.3.10] for the calibrated PAR model fitted to the logarithmic average monthly Saugeen riverflows. Notice that for all of the months, or periods, at most one value falls outside the 95% confidence limits which are calculated assuming that the RACF values are asymptotically NID$(0,\frac{1}{n})$. Moreover, at the crucial first few lags as well as lag 12, none of the RACF values are significantly different from zero for any of the seasons.

For both the periodic ACF and PACF graphs shown in Figures 14.4.1 and 14.4.2, respectively, at the fourth month or period there is a significantly large negative correlation at lag one. One way to interpret this behaviour from a physical viewpoint is that when spring flows in March cause large March flows due to the snowmelt runoff, the April flows tend to be substantially smaller.

Table 14.4.1 provides the parameter estimate and SE (standard error) for the AR parameter at lag one for each of the twelve seasons or periods for the PAR model fitted to the logarithmic average monthly Saugeen flows. One can see that the estimates reflect what is found in the periodic ACF and PACF plots in Figures 14.4.1 and 14.4.2, respectively. In particular, the AR parameter estimate for April is negative, as is also the case for the values of both the sample periodic ACF and PACF at lag one in period four.

Table 14.4.1. Parameter estimates and SE's for the PAR model having one AR parameter in each season, except for March, that is fitted to the logarithmic average monthly Saugeen riverflows.

Seasons or Periods	AR Parameter Estimates	SE's
1	0.6472	0.1037
2	0.4977	0.0886
3	0	0
4	-0.3124	0.0916
5	0.5300	0.1057
6	0.6091	0.0943
7	0.7087	0.1169
8	0.4228	0.0730
9	0.7039	0.1150
10	1.0598	0.1238
11	0.7699	0.0828
12	0.5901	0.1015

An advantage of employing the PAR model is that it can capture the type of varying seasonal correlation structure just described. Because of this, one would expect that the PAR model would more accurately and realistically describe the behaviour of the monthly Saugeen riverflows than competing types of seasonal models. This fact is confirmed by comparing the calculated value of the AIC in [14.3.9] for the Saugeen PAR model to those computed for the best SARIMA and deseasonalized models fitted to the average monthly Saugeen riverflow series in

Sections 12.4.4 and 13.4.2, respectively. The AIC value of 3357.82 for the PAR model is substantially less than those calculated for the other two models. Consequently, the PAR model is recommended over the SARIMA and deseasonalized models for fitting to the monthly Saugeen riverflows. Likewise, the forecasting experiments in Chapter 15, demonstrate that PAR models forecast average monthly riverflow series more accurately than its competitors and, therefore, is a better model to use with this type of seasonal data.

14.5 PARSIMONIOUS PERIODIC AUTOREGRESSIVE (PPAR) MODELS

14.5.1 Introduction

The PAR models described in the previous sections of this chapter attempt to preserve the seasonally-varying autocorrelation structure of a time series by fitting a separate AR model to each season of the year. However, one could reasonably question the necessity of going to the extreme of having a different model for each and every season. To decrease the number of model parameters required in a PAR model, one could combine individual AR models for various seasons in order to obtain a single model for all seasons in a given group. After grouping, the parameters of the more *parsimonious PAR* or *PPAR models* are estimated and diagnostically checked, and the PAR and PPAR models compared.

The approach for developing a PPAR model described in this section was originally presented by Thompstone et al. (1985a) and also Thompstone (1983). As an alternative procedure for reducing the number of parameters in PAR or PARMA models, Salas et al. (1980) propose a Fourier series approach. Recall that a Fourier series procedure is presented in Section 13.3.3 for reducing the number of deseasonalization parameters needed in the deseasonalized models of Chapter 13.

In the next subsection, the PPAR model is formally defined. Following this, flexible model construction techniques are given. In Section 14.6, all of the seasonal models of Part VI are compared by fitting them to six hydrological time series.

14.5.2 Definition of PPAR Models

As is also the case for the PAR model in [14.2.3], let the number of years and seasons be n and s, respectively, and let a transformed observation be given by $z_{r,m}^{(\lambda)}$, $r = 1,2, \ldots, n$, and $m = 1,2, \ldots, s$. Assuming the s seasons are grouped into G groups of one or more seasons with similar AR characteristics, the *parsimonious periodic autoregressive model (PPAR)* written as $(p_1,p_2, \ldots, p_G)$ may be defined in a manner analogous to the PAR model in [14.2.3] as

$$\phi^{(g)}(B)(z_{r,m}^{(\lambda)} - \mu_m) = a_{r,m} \qquad [14.5.1]$$

where $\phi^{(g)}(B) = 1 - \phi_1^{(g)}B - \phi_2^{(g)}B^2 - \cdots - \phi_{p_g}^{(g)}B^{p_g}$, is the AR operator of order p_g for group g, μ_m is the mean for season m, and $a_{r,m} \approx NID(0,\sigma_g^2)$. Notice from equation [14.5.1] that within a given group each seasonal mean is preserved by the parameter μ_m. However, for the observations in the seasons contained in the gth group, the AR parameters and the variance of the residuals are assumed to be the same.

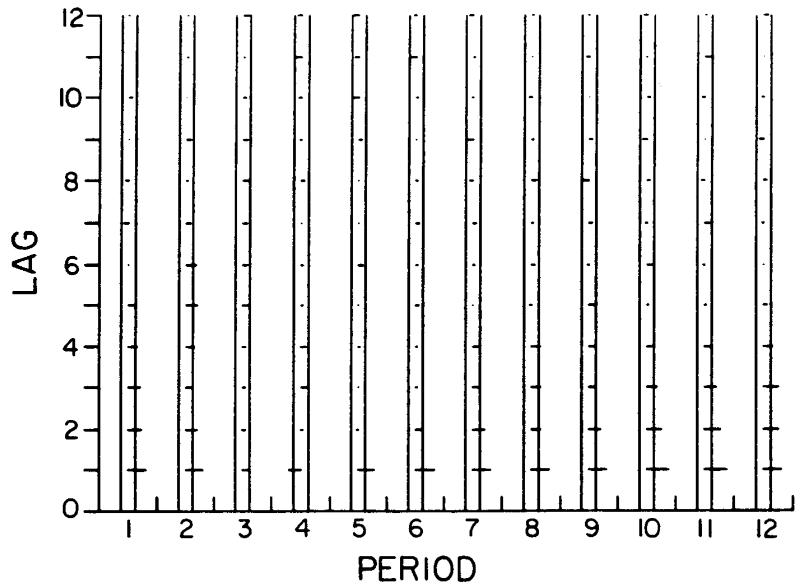

Figure 14.4.1. Sample periodic ACF for the logarithmic average monthly flows of the Saugeen River from January, 1916, until December, 1976, at Walkerton, Ontario, Canada.

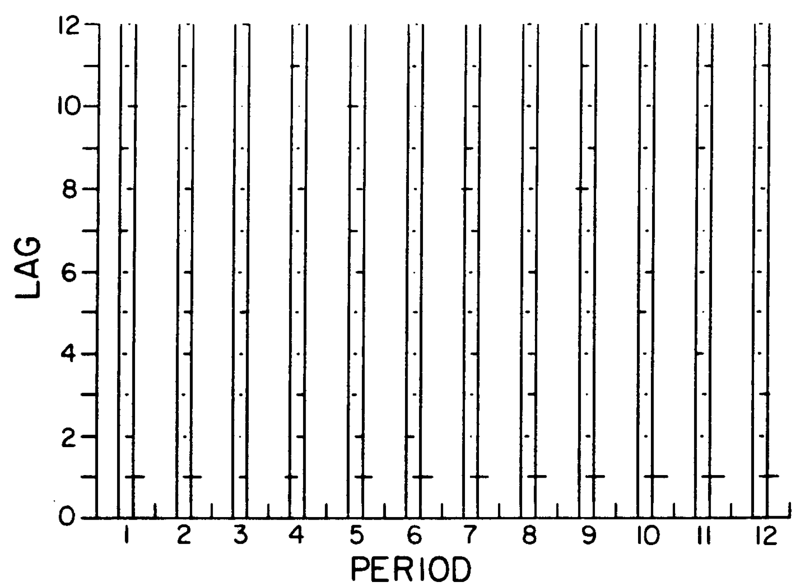

Figure 14.4.2. Sample periodic PACF for the logarithmic average monthly flows of the Saugeen River from January, 1916, until December, 1976, at Walkerton, Ontario, Canada.

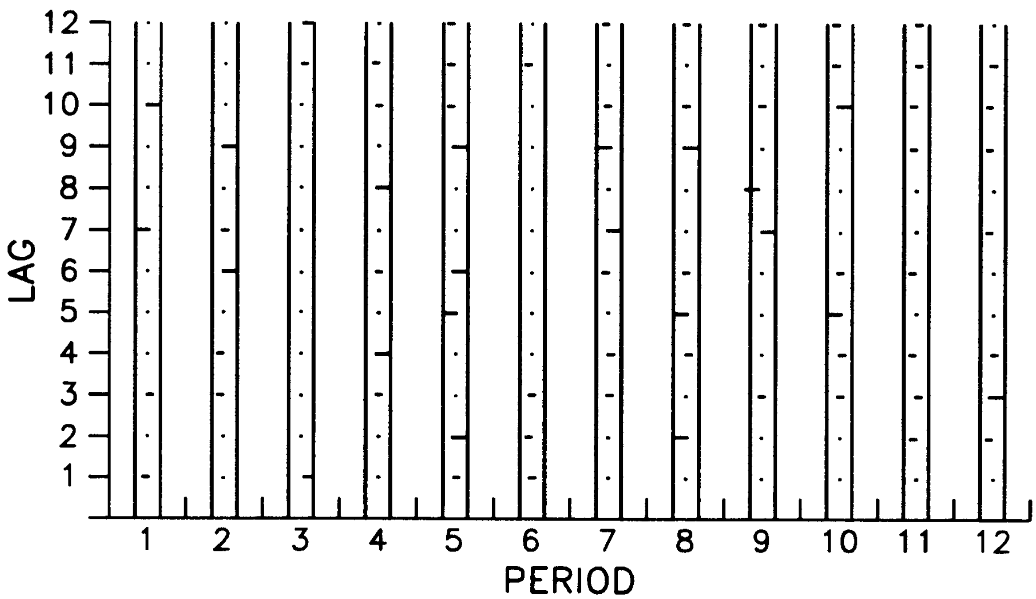

Figure 14.4.3. Periodic RACF for the PAR model having one AR parameter for each season, except for March, fitted to the average monthly riverflows of the Saugeen River from January, 1916, until December, 1976, at Walkerton, Ontario, Canada.

14.5.3 Constructing PPAR Models

In order to identify an appropriate grouping of seasons, the approach examined herein involves first fitting PAR models to the time series in question as described in Section 14.3. One then attempts to find seasons for which the AR models are "compatible". The equation of season m_2 is said to be compatible with that of season m_1, if the residuals obtained when the equation fit to season m_2 is applied to season m_1 are not significantly different from the residuals obtained from the equation fit to season m_1. In order to test formally for compatibility, define $a_T(m_1,m_2)$ to be the residuals obtained when the model fit to season m_2 is applied to season m_1 using [14.2.3] with initial values set to zero. These residuals can be used to estimate $\sigma^2(m_1,m_2)$, the residual variance when the model for season m_2 is applied to season m_1.

Consider the null hypothesis

$$H_o:\sigma^2(m_1,m_2) = \sigma^2(m_1,m_1)$$

Assuming that $(a_{R,M}(m_1,m_2),a_{R,M}(m_1,m_1))$ are jointly normally distributed with mean zero and are independent for successive values, a test developed by Pitman (1939) can be used to test this null hypothesis. For a review of how to carry out a hypothesis test, the reader can refer to Section 23.2. Let

$$S_{R,M} = a_{R,M}(m_1,m_2) + a_{R,M}(m_1,m_1) \qquad [14.5.2]$$

and

$$D_{R,M} = a_{R,M}(m_1,m_2) - a_{R,M}(m_1,m_1) \qquad [14.5.3]$$

Pitman's test is then equivalent to testing if the correlation, ρ, between $S_{R,M}$ and $D_{R,M}$ is significantly different from zero. Thus, provided $n > 25$, H_0 would be accepted at the 5% level of significance if $|\rho| < 1.96/\sqrt{n}$.

In practice, the residuals may not satisfy exactly the assumptions of a joint normal distribution with mean zero and independence for successive values of the residuals. However, these assumptions are probably a sensible first approximation. The assumption of independence seems reasonable because, with annual periodicity, the residuals are chronologically one year apart. Furthermore, the mean of zero is assured for the case of $a_{R,M}(m_1,m_1)$ due to the method of fitting the model. Pitman's test has often been used for testing the equality of variances of paired samples (Snedecor and Chochran, 1980, p. 190). It was pointed out in Lehmann (1959, p. 208, problem 33) that in this situation the test is unbiased and uniformly most powerful.

The above definition of equation compatibility can be extended to mutual compatibility. In particular, equations for seasons m_1 and m_2 are mutually compatible, if, at a given level of significance, one would accept the following two hypotheses:

$$\sigma^2(m_2,m_1) = \sigma^2(m_1,m_1)$$

$$\sigma^2(m_1,m_2) = \sigma^2(m_2,m_2)$$

Thus, the criteria adopted herein for identifying seasons in the same group is that each pair of seasons in the group must be mutually compatible at a given level of significance and have the same order of AR model. In addition, seasons are not grouped together unless they are chronologically adjacent. Once the groups have been identified, the parameters are estimated using maximum likelihood estimation. Specifically, multiple linear regression can be used to estimate the AR parameters for each group of seasons, where the seasonal means are estimated using [14.3.5] and the estimated variance of the residuals for each season is calculated using the estimated residuals contained in the group of seasons. Diagnostic checking involves first calculating the residuals from [14.5.1] by setting initial values to zero, and then examining the seasonal RACF and related Portmanteau test statistics plus tests for normality and homoscedasticity.

For season m in a PAR model, the maximized log likelihood is presented in [14.3.7]. When considering a PPAR model, the maximized log likelihood for the gth group is

$$\log L_g = -n_g \ln(\hat{\sigma}_g) + (\lambda - 1) \sum_{z_{r,m} \in \, group \, g} z_{r,m} \qquad [14.5.4]$$

where n_g is the product of the number of seasons in group g and the number of years of data, n. Notice that the summation term on the right hand side of [14.5.4] is for all data points contained in the seasons in the gth group. The value of the maximized log likelihood can be obtained by summing [14.5.4] across all the seasons to obtain

$$L_{PPAR} = \sum_{g=1}^{G} \log L_g \qquad [14.5.5]$$

Likewise, for a PAR model the value of the maximized log likelihood across all s seasons is

$$L_{PAR} = \sum_{m=1}^{s} \log L_m \qquad [14.5.6]$$

where $\log L_m$ is defined in [14.3.7].

As was done for the PAR model, one can derive the AIC for a PPAR model for each group of seasons and also the overall model. In particular, for the gth group of seasons the AIC formula is

$$AIC_g = -2\log L_g + 2p_g + 2 + 2\,(\textit{number of means}) \qquad [14.5.7]$$

where p_g is the number of AR parameters in the gth group seasons. The other parameters are the variance of the residuals and the number of means in the gth group of seasons. The AIC for the overall PPAR model is determined as

$$AIC_G = \sum_{g=1}^{G} AIC_g + 2 \qquad [14.5.8]$$

where the constant 2 allows for the Box-Cox parameter λ. The overall AIC formula for the PAR model is presented in [14.3.9].

When both PAR and PPAR models are fitted to a given series, the log-likelihood ratio (Rao, 1973, p. 448) can be used to test the null hypothesis that there is no significant difference in the residuals of the two models. It may be expressed as

$$\tilde{R} = -2[L_{PPAR} - L_{PAR}] \qquad [14.5.9]$$

and, assuming the null hypothesis is true, $\tilde{R}$ follows a chi-squared distribution with the number of degrees of freedom equal to the difference in the number of free parameters in the PAR and PPAR models, respectively (i.e., the difference in the number of AR parameters and residual variances).

14.6 APPLICATIONS OF SEASONAL MODELS

All of the seasonal models presented in Part VI are fitted to three average monthly riverflow series and three average quarter-monthly riverflow time series and the resulting models are compared using the AIC. More specifically, the seasonal models fitted to the series are the SARIMA, deseasonalized, PAR and PPAR models defined in Sections 12.2, 13.2, 14.2.2 and 14.5.2, respectively. Grouping of seasons within the PPAR models is performed using three levels of significance in the Pitman test presented in Section 14.5.3, namely 50%, 20% and 5%. In general, as the level of significance decreases, fewer seasons are considered to have "incompatible" models and thus there is more grouping, or in other words, a smaller number of groups. A Box-Cox transformation with $\lambda = 0$ is used in all cases and, hence, the data are transformed by taking their natural logarithms. The above mentioned models are labelled as SARIMA, DES, PAR, PPAR/50, PPAR/20 and PPAR/05, respectively, in the upcoming tables. The results of this study were originally presented by Thompstone (1983, Section 3.5).

The six example hydrological time series consist of:

(1) inflows to reservoirs of the hydroelectric system operated by Alcan Smelters and Chemicals Ltd. in the Province of Quebec, Canada (Thompstone et al., 1980);

(2) flows of the Saugeen River measured at Walkerton, Ontario, Canada (Environment Canada, 1977);

(3) flows of the Rio Grande measured at Furnas, Minas Gerais, Brazil (supplied by Mr. Paulo Roberto de Holanda Sales at Eletrobras, (national electrical company of Brazil)).

The monthly series are comprised of Alcan system inflows from 1943 to 1979, Saugeen River flows, 1919-76, and Rio Grande flows, 1931-75; the quarter-monthly flows consist of Alcan system inflows, 1943-79, Saugeen riverflows, 1915-76, and Rio Grande flows, 1931-72. Note that the quarter-monthly data consists of flows in m^3/s averaged from the 1st to the 7th, from the 8th to the 15th, from the 16th to the 22nd, and from the 23rd to the end of the month, which constitute periods of approximately one week each.

For all six series, the order of the AR operator in a PAR or PPAR model for a season or group of seasons, respectively, is usually one while the highest order is three. Very few of the AR models for an individual season or group of seasons are white noise.

Table 14.6.1 summarizes the orders of the AR models contained within the PAR models fitted to the six series. Because there are 48 and 12 seasons for the quarter-monthly and monthly data, respectively, the number of AR models used in each quarter-monthly PAR model must equal 48 whereas the total for each PAR model is 12. For the case of the PAR model for the average monthly Saugeen riverflows, the order of the AR operator is one for 11 of the 12 months. As explained in Section 14.4, the month of March is white noise. Finally, the only other series which has white noise components in the PAR model is the Alcan system for monthly riverflows that contains four such months.

In order to illustrate the degree of grouping associated with various Pitman test significance levels, Table 14.6.2 shows the number of groups associated with the PAR, PPAR/50, PPAR/20 and PPAR/05 models for each series. For the case of quarter-monthly series, the highest degree of grouping is with the PPAR/05 model of Rio Grande flows: the 48 seasons are divided into 16 groups. The highest degree of grouping of monthly series is with the PPAR/05 model of the Saugeen riverflows: the 12 months are divided into 5 groups. Note that, in the case of the Alcan system monthly inflow series, no grouping is identified, even when using the 50% significance level.

Table 14.6.3 shows how all six of the seasonal models are ranked according to the AIC for each of the series. The model having the lowest AIC value is ranked first whereas the one with the highest value is ranked as 6. When fitting the deseasonalized model, the logarithmic series is fully deseasonalized using [13.2.3]. Although it isn't done in this study, one could reduce the number of deseasonalization parameters by implementing the Fourier series approach described in Section 13.3.3.

As shown in Table 14.6.3, the AIC always selects a PPAR model as the most desirable model. The only exception is the PAR model for the Alcan system for which no PPAR model is identified. As would be expected from the basic design of the SARIMA model, in all six cases the SARIMA model is the least desirable model, according to the AIC. This is because the

Table 14.6.1. Number of periods having a given order of an AR model for the PAR models fitted to six series.

Order of AR Model	Quarter-monthly Series			Monthly Series		
	Alcan System	Rio Grande	Saugeen	Alcan System	Rio Grande	Saugeen
0	0	0	0	4	0	1
1	36	42	43	6	9	11
2	10	6	5	2	2	0
3	2	0	0	0	1	0

Table 14.6.2. Number of groups in the PAR and PPAR models.

Model	Quarter-monthly Series			Monthly Series		
	Alcan System	Rio Grande	Saugeen	Alcan System	Rio Grande	Saugeen
PAR	48	48	48	12	12	12
PPAR/50	40	38	40	12	10	8
PPAR/20	29	27	23	12	8	7
PPAR/05	24	16	18	12	6	5

Table 14.6.3 Ranking of the seasonal models fitted to the six series according to the AIC.

Model	Quarter-monthly Series			Monthly Series		
	Alcan System	Rio Grande	Saugeen	Alcan System	Rio Grande	Saugeen
SARIMA	6	6	6	3	6	6
DES	5	5	5	2	5	5
PAR	4	3	3	1	4	4
PPAR/50	3	1	2	-	2	3
PPAR/20	1	2	1	-	1	1
PPAR/05	2	4	4	-	3	2

SARIMA model is not designed for describing stationarity within each season as well as a seasonally varying correlation structure. Because the deseasonalized model of Chapter 13 can account for a separate seasonal mean and variance within each season, the AIC results of Table 14.6.3 indicate that the deseasonalized model always performs better than the SARIMA in all six applications. Moreover, due to the fact that a periodic model can handle seasonally varying correlation, periodic models always do better than both deseasonalized and SARIMA models. Finally, forecasting experiments are carried out in Section 15.4.4 to compare the forecasting capabilities of the models listed in Table 14.6.3.

The log-likelihood ratio test in [14.5.9] can be used to ascertain if the residuals of the fitted PPAR and PAR models differ significantly from each other. In the five cases for which PPAR models are identified (see Tables 14.6.2 or 14.6.3), residuals of none of the PPAR models are significantly different from those of the PAR model at the 5% level of significance. This reinforces the conclusion that even though PPAR models have fewer parameters and, hence, also the seasonal models, they still describe the data as well as the regular PAR model.

As noted in Section 14.5.1, another approach for reducing the number of AR parameters required in a PAR model is to use a Fourier series approach (Salas et al., 1980). However, this procedure assumes that a smooth sinusoidal type of curve is fitted to the AR parameters across the seasons or periods and, hence, also the sample periodic ACF. The question arises as to whether this assumption is reasonable. To investigate this, consider Figure 14.6.1, which shows a graph of the periodic ACF at lag one of the natural logarithms of the quarter-monthly flows of the Saugeen River. As can be seen, it would be impossible to fit a smooth cyclic curve through this plot. Notice, for example, the manner in which the first order correlation drops significantly downwards in the spring season (i.e., about the end of March in the 12th quarter-monthly period). Fortunately, both the PPAR and the PAR models are designed for modelling the type of behaviour exhibited in Figure 14.6.1. The approach to fitting PAR and PPAR models is sufficiently general to be applicable to series with or without a cyclic pattern in the seasonal correlations and AR parameters. In a similar fashion, one can see that it would be difficult to fit a Fourier series curve through the AR parameter estimates in Table 14.4.1 calculated for the PAR model fitted to the logarithmic average monthly Saugeen riverflows.

14.7 CONSTRUCTING PARMA MODELS

PARMA models can be fitted to seasonal series by following the identification, estimation and diagnostic check stages of model construction. Because model building procedures are highly developed for use with PAR models, this class of periodic models is focussed upon in this chapter. Nonetheless, there are now some good construction techniques available for fitting PARMA models to seasonal data sets. As is also the case for the PAR model, the ARMA model for each season of the year can be identified separately. The main area where further research is required for PARMA model building is the development of a maximum likelihood estimation technique which is computationally efficient. To obtain efficient estimates for a PARMA model, all parameters must be estimated simultaneously, including the innovation variances, and, moreover, it is necessary to use a nonlinear optimization technique since the likelihood function is nonlinear. Each evaluation of the likelihood function involves very lengthy computations when $s \geq 12$.

The sample periodic ACF and PACF described in Section 14.3.2 can be employed for identifying the orders of the AR and MA operators for the PARMA model in [14.2.15] to fit to each of the seasons in a given seasonal time series. If a pure MA model of order q_m is required, the sample periodic ACF for season m will not be significantly different from zero after lag q_m and the sample periodic PACF will die off. When a pure AR model of order p_m is needed to model season m, the sample periodic ACF attenuates while the sample periodic PACF is not significantly different from zero after lag p_m. When both AR and MA parameters should be included in the ARMA model to fit to the mth season, both the sample periodic ACF and PACF attenuate.

Assuming normality, Vecchia (1985a,b) developed a technique for obtaining MLE's of the parameters in a PARMA model. The approach that Vecchia (1985a,b) uses to write the likelihood function is the same as the one of Newbold (1974) for the univariate case and Hillmer and Tiao (1979) for the multivariate ARMA models presented in Chapter 20. Additionally, he proved that PARMA models and multivariate ARMA models are equivalent. From a computational point of view, his algorithm seems to be feasible for use in practical applications when the number of seasons is small (i.e., less than about 4 seasons per year). To overcome computational

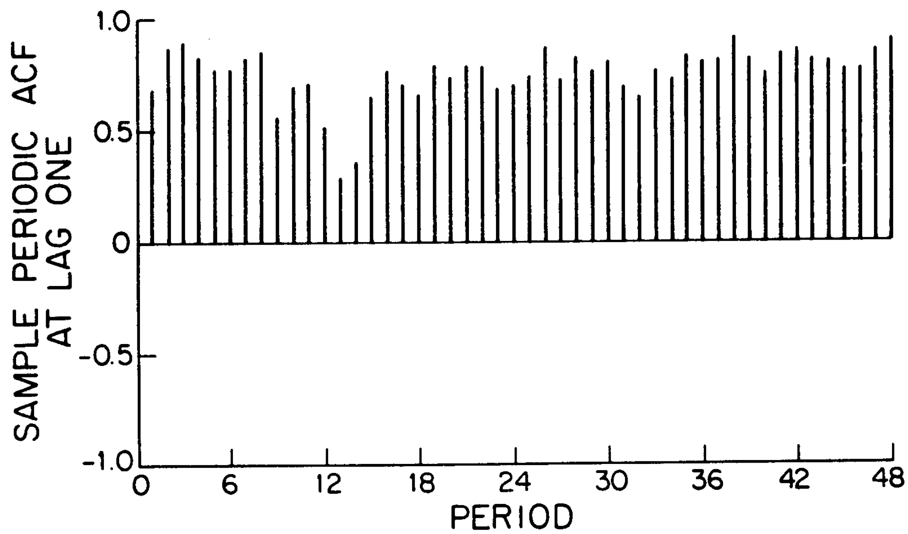

Figure 14.6.1. Periodic ACF at lag one of the logarithmic quarter-monthly riverflows of the Saugeen River at Walkerton, Ontario, Canada, from 1915 to 1976.

difficulties, Jimenez et al. (1989) propose a maximum likelihood parameter estimation technique which is implemented within a Kalman filtering framework. Finally, Li and Hui (1988) provide an algorithm for the exact likelihood of PARMA models.

As explained in Section 14.3.3, the Yule-Walker equations can be used as a moment estimation approach for obtaining efficient parameter estimates for the parameters of a PAR model. However, one should be cautious when using moment estimators with PARMA models, since the parameters estimates may not be efficient. Nonetheless, some research on moment estimation of PARMA model parameters has been completed. For example, Salas et al. (1982) derived Yule-Walker equations for PARMA models and showed how moment estimates can be calculated for PARMA models in which $p_m \geq 0$ and $q_m = 1$ in season m. Besides discussing moment estimation, Salas and Obeysekera (1992) described model identification and testing of model adequacy of PARMA models. Moreover, these authors proved a physical basis for PARMA models. In particular, based upon a conceptual-physical representation of a natural watershed, in which all inputs, storages, outputs and parameters are assumed to be periodic and the system is a linear reservoir, they demonstrated that the periodic groundwater storage and streamflow processes belong to the class of PARMA processes. Section 3.6 describes this kind of physical relationships for the case of nonseasonal ARMA models. Further results on how PARMA models can be used in physically-based modelling are provided by Claps et al. (1993).

When testing the adequacy of a calibrated PARMA model, one can use similar procedures to those suggested for PAR models in Section 14.3.4. The sample periodic RACF and related Portmanteau statistics can be employed to ascertain if the residuals are white. Other tests related to those presented in Chapter 7 for nonseasonal ARMA models can be used for testing if the normality and homoscedasticity assumptions are valid. Non-normality and/or heteroscedasticity can often be rectified by incorporating an appropriate Box-Cox transformation from [13.2.1].

In related research to PARMA modelling, Vecchia et al. (1983) investigated what happens when one aggregates across the seasons. Specifically, the aggregated time series resulting from summing over the seasons of a seasonal time series, which is assumed to be either AR(1) or ARMA(1,1) in each season, is shown to follow an ARMA(1,1) model at the annual level. Moreover, when the seasonal data and the model for each season are used rather than the annual data and the associated annual model, significant gain in parameter efficiency can be achieved. This, of course, further justifies the use of PAR and PARMA modelling in water resources and indicates that aggregation is preferable to disaggregation. A discussion of disaggregation and the controversy surrounding it is given in Section 20.5.2.

For most of the PARMA model construction techniques discussed thus far, it is assumed that the data or, equivalently, the model residuals are normally distributed. Fernandez and Salas (1986) studied PAR models having a Gamma marginal distribution. This Gamma or other kinds of distributional assumption could also be used with PARMA models. However, a substantial amount of theoretical research and development of flexible model building techniques are needed before these and other related models can be used in practice. Lewis (1985) and authors referenced therein, discuss non-Gaussian distributed innovations for use in nonseasonal and multivariate modelling. In Section 20.5.3, the employment of non-Gaussian marginal distributions in multivariate modelling is outlined.

14.8 SIMULATING AND FORECASTING WITH PERIODIC MODELS

14.8.1 Introduction

Subsequent to fitting a PAR or PARMA model to a seasonal time series, the calibrated model can be used for applications such as forecasting and simulation. In the next chapter, it is explained how minimum mean squared error forecasts from periodic models, as well as other kinds of seasonal models, can be calculated. Moreover, forecasting experiments with average monthly riverflow series demonstrate that PAR models forecast better than deseasonalized (Chapter 13) and SARIMA (Chapter 12) models.

In Chapter 9, two simulation procedures are presented for generating synthetic data from nonseasonal AR and ARMA models. The simulation techniques are designed so that random realizations of the underlying stochastic process are employed as starting values. Because fixed beginning values are not utilized, unwanted systematic bias is not introduced into the synthetic traces.

Because a PAR or PARMA model consists of having a separate AR or ARMA for each season of the year, simulation techniques similar to those presented in Chapter 9 for use with nonseasonal models can be employed with seasonal models. The technique labelled WASIM2, for example, in Section 9.4 exactly simulates an AR or ARMA process if the residuals are assumed to be normally distributed. Suppose, for example, one wishes to simulate using a PAR model. Let $k = \max(p_1, p_2 - 1, p_3 - 2, \ldots, p_s - (s-1))$ where s is the number of seasons. By utilizing the covariance matrix of $(z_{1,1}, z_{1,2}, \ldots, z_{1,k})$ to generate randomly the initial values, a technique very similar to WASIM2 can be used for producing synthetic traces from a PAR model. If deemed appropriate, parameter uncertainty can also be brought into a simulation study by following the WASIM3 procedure of Section 9.7. Salas and Abdelmohsen (1993) describe initialization techniques when simulation with single-site and multisite low-order PAR and PARMA models.

As explained in Chapter 9, simulation can be used for design purposes and investigating the theoretical properties of models. In the next subsection, it is shown using simulation that PAR models can preserve statistically the critical drought statistics defined by Hall et al. (1969).

Stedinger and Taylor (1982a,b) describe the steps involved in the development and use of a stochastic streamflow model. After properly fitting a time series model to a given nonseasonal or seasonal riverflow data set, these authors stress the importance of model verification and model validation. In model verification, one should demonstrate that a model has been implemented correctly and passes diagnostic checks. With respect to model validation, one should show that simulated sequences from the calibrated model produce reservoir system performance that is consistent with or statistically indistinguishable from that obtained utilizing the historical riverflows. Accordingly, the simulation experiments carried out in Section 14.8.2 as well as Section 10.6 can be considered to be model validations.

14.8.2 Preservation of Critical Period Statistics

Introduction

Hall et al. (1969) discuss problems related to the design and operation of a reservoir when water shortages must be considered. They define the critical period as the period of time for which a given inflow series is most critical with respect to meeting water demands. Various statistics, which are closely related to the critical period, are defined and, by using simulation, Hall et al. (1969) conclude that the stochastic model they are investigating does not adequately preserve the historical critical period statistics. In a more exhaustive study, Askew et al. (1971) find that a large variety of stochastic models are not capable of retaining the critical period statistics. The purpose of the present section is to demonstrate that, for certain sample series, when the PAR and PPAR models are identified and fitted using the procedures described in this chapter, they adequately preserve the historical critical period statistics.

Critical Periodic Statistics for Water Supply

Hall et al. (1969) express the active reservoir storage as a ratio of the total volume of active storage in the reservoir to the volume of water due to the average annual inflow. The reservoir is operated to allow a *seasonal extraction* of X. It is assumed that the reservoir is full at the start and a value of X is determined which causes the reservoir storage to research zero at one point in time. The *length of the critical period* is denoted by L and is calculated as the time span from the zero storage point backward in time to the point when the reservoir was last full. The *percentage deficiency, D, for the critical period* is defined as

$$D = \frac{\sum_{CP}(\bar{V} - V_t)100}{\bar{V}L} \qquad [14.8.1]$$

where $\bar{V}$ is the average seasonal inflow volume, V_t is the seasonal inflow volume for period t, and the summation extends over the entire critical period, CP. As pointed out by Hall et al.

(1969), the aforesaid critical period statistics can be readily generalized to the case where the extraction is a function of time, the reservoir is at any level at the start of the calculations, and evaporation and other losses are considered. Note also that the critical period statistics are obviously a function of the length of the series for which they are defined. As illustrated by McMahon and Mein (1978, pp. 19-20), there may, in rare cases, be more than one critical period for a given inflow series.

Design of Simulation Experiments

In Section 14.6, PAR and PPAR models are fitted to three quarter-monthly and three-monthly time series. These same series are used in this section in split sample simulation experiments used to show that PAR and PPAR models preserve statistically the critical period statistics. These results were originally presented by Thompstone et al. (1987) and Thompstone (1983). McLeod and Hipel (1978) used simulation experiments to demonstrate that critical period statistics are preserved by PAR models but they did not use the split sample approach described herein. Recall that in Sections 9.8 and 10.6, simulation experiments are used to demonstrate that ARMA models preserve statistically the rescaled adjusted range and other statistics related to the Hurst phenomenon.

For each of the 6 seasonal series, PAR and PPAR models are identified and fit following the procedures of Sections 14.3 and 14.5.3, respectively. In each case, all but the last 20 years of available data are used to identify and fit the models. The PPAR model having the minimum value of the AIC is selected from three candidates, namely those with 50%, 20% and 5% levels of significance for the Pitman test grouping criterion described in Section 14.5.3.

As explained in Section 9.2, in order to generate synthetic sequences, it is first necessary to produce independent, normally distributed random numbers with a mean of zero and a variance of one. In the experiments described herein, an efficient and portable pseudo-random number generator, developed by Wickmann and Hill (1982), is used to produce numbers rectangularly distributed between zero and one, and these are then used in the algorithm of Box and Muller (1958) to produce the required random normal deviates. These innovations are then fed into the appropriately estimated PAR model in [14.2.3] or the PPAR model in [14.5.1] for a given series.

As pointed out in Chapter 9, an important consideration in the generation of synthetic hydrological sequences is the choice of initial values. Random realizations of the underlying stochastic process must be used to avoid introducing systematic bias into the simulation study. The approach to obtaining random realizations adopted in the original study by Thompstone et al. (1987) and Thompstone (1983) is to, in a preliminary study, set the required initial values to their expected values and then generate a full 40 years of synthetic data. The last few values of these 40 years of synthetic data provide the required initial values for the main simulation study.

For a given sample time series, the simulation experiment is conducted as follows. First, the remaining 20 years of the historical sample not used in model construction are used to calculate what are referred to as the historical critical period statistics. These are denoted as X(his), L(his) and D(his) for the historical extraction rate, historical length of the critical period, and historical deficiency, respectively. An active reservoir storage equal to the average volume of annual inflow is used. Next, 1,000 synthetic seasonal sequences of 40 years each are generated, and the first 20 years of each sample are dropped to provide 1,000 effectively independent sequences equal in length to the series used to calculate the historical critical period statistics.

It is important to note that almost all previous research concerning the preservation of statistics in synthetic hydrological sequence generation has not used the split-sample approach employed herein. In previous research, the same sample series was employed both to construct the model(s) being evaluated and to estimate the statistic(s) whose preservation is being studied. One would generally expect the split-sample design of the current research to be a more rigorous validation of the models under investigation.

In order to test if a given model preserves the critical period statistics, the P-values defined below are estimated:

$$P_X = Prob\{X(syn) < X(his)\} \qquad [14.8.2]$$

$$P_L = Prob\{L(syn) > L(his)\} \qquad [14.8.3]$$

$$P_D = Prob\{D(syn) > D(his)\} \qquad [14.8.4]$$

where *Prob* denotes probability, $X(syn)$ is the extraction rate in the synthetic series, $L(syn)$ is the length of the critical period in the synthetic series, $D(syn)$ is the percentage deficiency in the synthetic series, and other terms are as defined earlier.

The P-values are estimated separately for each series with the active reservoir storage equal to the volume of the average annual inflow for the 20-year historical sample not used to calibrate the models. This is done by counting the number of times the inequalities in [14.8.2] to [14.8.4] hold in each simulation run and dividing by 1,000. The P-values, as defined above, represent the probability of a critical period statistic in the synthetic sequence being more extreme than in the historical sequence. Thus a P-value of 0.05 indicates that there is only a 5% chance that the synthetic series will have a critical period statistic more extreme than the historical. Of course, this would happen 5% of the time even if the historical sequence were in fact generated by the corresponding fitted stochastic model. Nevertheless, P-values less than 5% do suggest possible model inadequacy, and hence, P-values can be used for diagnostic checking.

In Section 10.6.4, a χ^2 test is employed to ascertain, in an overall sense, if the Hurst statistics are preserved statistically by ARMA models fitted to 23 annual geophysical time series. In particular, when considering k time series for a given statistic, it can be shown (Fisher, 1970, p. 99)

$$-2\sum_{i=1}^{k}\ln(P_i) \approx \chi^2_{2k} \qquad [14.8.5]$$

where P_i can be the probability as defined in Equations [14.8.2] to [14.8.4] for the ith time series.

The Results of the Simulation Experiments

The results of the simulation experiments are summarized first for the PAR models, and then for the PPAR models. Table 14.8.1 shows the P-values for PAR models for the three critical period statistics and for the six example series, while Table 14.8.2 contains the chi-squared values calculated using [14.8.5] for the three critical period statistics with the series grouped according to their seasonal lengths. For a one-sided significance test, the chi-squared values with six degrees of freedom at the 5% and 1% significance levels are 12.592 and 16.812, respectively. For the monthly series, the critical statistics are preserved in each case at the 5% level, as can be seen in Table 14.8.1, and on an overall basis, also at the 5% level, as shown in Table 14.8.2.

For the quarter-monthly series, evidence of preservation of the critical period statistics by PAR models is not quite as strong. The extraction rate for the Saugeen series and the length of the critical period for the Alcan system inflows are preserved at the 1% level, but not at the 5% level. In all other cases, the statistics are preserved at the 5% level. According to the overall chi-squared test, the length of the critical period and deficiency percentage are preserved at the 5% level, but the extraction rate is preserved at only the 1% level.

Table 14.8.1. P-values for the PAR models.

Riverflow Series		Statistics		
		Extraction	Length of CP	Deficiency
Quarter-Monthly	Alcan System	0.233	0.040	0.730
	Rio Grande	0.350	0.267	0.654
	Saugeen	0.017	0.267	0.257
Monthly	Alcan System	0.423	0.118	0.774
	Rio Grande	0.521	0.090	0.825
	Saugeen	0.083	0.370	0.292

Table 14.8.2. Chi-squared values for the PAR models.

Seasonal Lengths	Statistics		
	Extraction	Length of CP	Deficiency
Quarter-Monthly	13.162	11.720	4.196
Monthly	8.003	11.079	3.359

Table 14.8.3 shows the P-values for PPAR models for the three critical period statistics and for the six example series, while Table 14.8.4 contains the chi-squared values for the three critical period statistics with the series grouped according to their seasonal length. For the case of the monthly series, there are two P-values which suggest that the critical period statistics are preserved at the 1% level, but not at the 5% level. These relate to the extraction rate and deficiency percentage for the Saugeen Series. All other cases indicate preservation at the 5% level. The overall chi-squared test indicates the length of the critical period and deficiency percentages are preserved at the 5% level, while the extraction is preserved at the 1% level.

For the quarter-monthly series, evidence of preservation of the critical period statistics by PPAR models is not quite as strong as for the monthly series. Again, the extraction rate and deficiency for the Saugeen series are preserved at the 1% level, but not at the 5% level. The length of the critical period is not preserved at the 1% level for the Alcan system inflow series. All other statistics are preserved at the 5% level. Nevertheless, the overall chi-squared statistics indicate that the deficiency and the length of the critical period are preserved at the 5% level, while the extraction rate is preserved at the 1% level (and very close to being preserved at the 5% level).

Table 14.8.3. P-values for the PPAR models.

Riverflow Series		Statistics		
		Extraction	Length of CP	Deficiency
Quarter-Monthly	Alcan System	0.395	0.007	0.886
	Rio Grande	0.414	0.684	0.307
	Saugeen	0.011	0.767	0.035
Monthly	Alcan System	0.673	0.293	0.800
	Rio Grande	0.175	0.373	0.400
	Saugeen	0.012	0.799	0.034

Table 14.8.4. Chi-squared values for the PPAR models.

Seasonal Lengths	Statistics		
	Extraction	Length of CP	Deficiency
Quarter-Monthly	12.641	11.214	9.309
Monthly	13.124	4.876	9.042

It should be noted that in the majority of these simulation experiments (22 out of 36 combinations of models, series and critical period statistics), the coefficient of skewness of the empirical distribution of the critical period statistics is different from zero at the 5% level. In fact, for the length of critical period statistic, the skewness coefficient is always significantly different from zero at the 0.1% level. In view of the significant skewness encountered in this study, the types of statistical tests used by Hall et al. (1969) and Askew et al. (1971) are not appropriate. Their tests are based on the assumption of normality, and this assumption is not valid for skewed statistics.

A further point that should be stressed is that the split sample approach to testing the preservation of critical period statistics is more exact than the approach in which an entire series is used for both model fitting and the calculation of the statistics to be preserved. This latter approach was used in the earlier studies of Hall et al. (1969), and Askew et al. (1971), as well as in Section 10.6.4 for the Hurst statistics.

14.9 CONCLUSIONS

Because the basic mathematical design of the periodic models described in this chapter closely reflects the statistical characteristics of many kinds of seasonal time series, especially those arising in the environmental sciences, periodic models are ideally suited for use in practical applications. Of particular import is the family of PAR models for which comprehensive model construction techniques have been developed. If the number of model parameters has to be reduced, one can employ the economical PPAR class of models. Although more research is required to devise estimation algorithms for PARMA models that are computationally efficient, good progress has been made in the practical development of this promising class of models.

As just noted, periodic models are well designed for use with natural time series. When dealing with seasonal socio-economic time series in which the mean level and possibly other statistics may change within each season over the years, one may wish to experiment with the following modelling approach. Firstly, one can model the seasonal series, such as monthly water demand, using a SARIMA model (Chapter 12). Secondly, one can model the residuals of the fitted SARIMA model using a PAR or other type of periodic model. In this way, one may be able to model a seasonally varying correlation structure which is not captured by the SARIMA model.

The simulation experiments of Section 14.8.2, demonstrate that properly fitted PAR and PPAR models can preserve statistically important historical statistics. In the next chapter, it is shown using forecasting experiments that these periodic models forecast seasonal riverflow series better than both deseasonalized (Chapter 13) and SARIMA (Chapter 12) models.

PROBLEMS

14.1 Complete the following:

(a) Assuming that there are four seasons per year, and the order of the AR model in each season is two, write down the complete set of equations.

(b) Develop the theoretical periodic autocovariance function for the PAR model in part (a).

(c) Determine the periodic Yule-Walker equations for this model.

14.2 The stationarity requirement for the PAR model in [14.2.4] having one AR parameter in each season is given in [14.2.6]. By referring to appropriate references, determine the stationarity condition for a general PAR model that is not restricted to being Markov.

14.3 Complete the following:

(a) Using the notation in [14.2.15], write down the complete set of equations for a PARMA model having four seasons where $p_m = q_m = 1$ in the first two seasons, and $p_m = 2$ and $q_m = 1$ for the second two seasons.

(b) Derive the theoretical periodic autocovariance function for the PARMA model in part (a).

(c) Ascertain the periodic Yule-Walker equations for the model.

14.4 The stationarity and invertibility conditions for a PARMA model having one AR and one MA parameter in each season are given in [14.2.6] and [14.2.17], respectively. Present and explain the conditions for stationarity and invertibility for the general PARMA model in [14.2.15] for which the number of AR and MA parameters for a specific season are not restricted in number.

14.5 Suppose that a PARMA model for season m is given as

$$(1 - 0.11B + 0.30B^2)z_{r,m} = (1 - 0.4B)a_{r,m}$$

where the mean of $z_{r,m}$ is zero.

(a) Obtain the random shock coefficients for at least eight terms and then write this model in random shock form as in [14.2.24].

(b) Also write the model in inverted form as in [14.2.26].

14.6 Carry out the instructions in problem 14.5 for the PARMA model in season m which is given as

$$(1 - 0.10B + 0.24B^2)z_{r,m} = (1 - 0.4B)a_{r,m}$$

where the mean of $z_{r,m}$ is assumed to be zero.

14.7 Select an average monthly riverflow series and then fit a PAR model to this series adhering to the following steps in model construction:

(a) Examine appropriate exploratory data analysis graphs as well as the sample periodic ACF and PACF plots to design the most appropriate set of PAR models.

(b) Estimate the model parameters for each model selected in part (a) and then use the MAICE procedure to find the best one. For the most appropriate model compare the estimates for the model parameters employing both multiple linear regression and the periodic Yule-Walker equations. Comment upon the results.

(c) Carry out diagnostic checks to ensure that the best PAR model from (b) satisfies the whiteness, normality, and constant variance assumptions. If there are any problems make suitable modifications based upon the diagnostic results and repeat steps (b) and (c). Whatever the case, be sure to employ the periodic RACF test for whiteness given in [14.3.10].

14.8 Develop PAR models for describing average monthly riverflows from three distinctly different geographical locations in the world. Using identification results such as the sample periodic ACF and the sample periodic PACF graphs as well as the structures of the calibrated PAR models, make comparisons among the fitted models. Wherever appropriate, provide physical explanations as to why certain modelling results vary across the regions.

14.9 Develop the most appropriate PAR and PPAR models to describe an average monthly hydrological time series. Explain why any groupings of months within the PPAR model make sense or else do not seem reasonable from both statistical and physical viewpoints.

14.10 Follow the instructions in problem 14.9 for an average weekly hydrological time series.

14.11 Select a hydrological data set for which you have both average weekly and average monthly observations. Carry out the studies put forward in the previous two questions for the monthly and weekly time series. Subsequently, compare the monthly and weekly modelling results for the PAR and PPAR models. Did you find, for instance, that there were more groupings of seasons for the fitted weekly PPAR model than with the monthly version?

14.12 In Section 14.7, model building procedures are discussed for PARMA models. Summarize and compare according to both advantages and disadvantages the PARMA estimation techniques given by Vecchia (1985a,b) and Jimenez et al. (1989).

14.13 After fitting a PAR model to an average monthly riverflow time series, execute a proper simulation study to ascertain whether or not the historical critical period statistics given in [14.8.1] are preserved.

14.14 Fit a PAR model to a seasonal hydrological time series of your choice. Then carry out simulation experiments to determine if the sample periodic ACF in [14.3.4] at lag one for each season is preserved statistically by the calibrated model.

REFERENCES

CRITICAL PERIOD STATISTICS

Askew, A. J., Yeh, W. W. G. and Hall, W. A. (1971). A comparative study of critical drought simulation. *Water Resources Research*, 7:52-62.

Hall, W. A., Askew, A. J. and Yeh, W. W. G. (1969). Use of the critical period in reservoir analysis. *Water Resources Research*, 5(6):1205-1215.

McMahon, T. A. and Mein, R. G. (1978). *Reservoir Capacity and Yield*. Elsevier, New York.

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.: Effect of parameter uncertainty. *Water Resources Research*, 18(4):919-924.

Thompstone, R. M., Hipel, K. W. and McLeod, A. I. (1987). Simulation of monthly hydrological time series. In MacNeill, I. B. and Umphrey, G. J., Editors, *Advances in Statistical Sciences*, Festschrift in Honor of Professor V. M. Joshi's 70th Birthday, volume IV, Stochastic Hydrology, pages 57-71, D. Reidel Publishing Co., Dordrecht, the Netherlands.

DATA SETS

Environment Canada (1977). Historical streamflow summary, Ontario. Technical report, Inland Waters Directorate, Water Resources Branch, Ottawa, Canada.

Thompstone, R. M., Poiré, A. and Valleé, A. (1980). A hydrometeorological information system for water resources management. *INFOR*, 18(3):258-274.

PERIODIC MODELLING

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

Cipra, T. (1985a). Periodic moving average processes. *Aplikace Matematily*, 30:218-229.

Cipra, T. (1985b). Statistical analysis of multiple moving average processes under periodicity. *Kybernetica*, 21:335-345.

Cipra, T. and Tlusty, P. (1987). Estimation of multiple autoregressive-moving average models using periodicity. *Journal of Time Series Analysis*, 8:293-301.

Claps, P., Rossi, F. and Vitale, C. (1993). Conceptual-stochastic modeling of seasonal runoff using autoregressive moving average models and different scales of aggregation. *Water Resources Research*, 29(8):2545-2559.

Croley II, T. E. and Rao, K. N. R. (1977). A manual for hydrologic time series deseasonalization and serial independence reduction. Report No. 199, Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, Iowa.

Dunsmuir, W. (1981). Estimation of periodically varying means and standard deviations in time series. *Journal of Time Series Analysis*, 2(3):129-153.

Fernandez, B. and Salas, J. D. (1986). Periodic gamma autoregressive processes for operational hydrology. *Water Resources Research*, 22(10):1385-1396.

Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics*, 2:385-388.

Gladyshev, E. G. (1963). Periodically and almost-periodically correlated random prrocesses with a continuous time parameter. *Theory of Probability and its Applications*, 8:173-177.

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.

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.

Hurd, H. L. and Gerr, N. L. (1991). Graphical methods for determining the presence of periodic correlation. *Journal of Time Series Analysis*, 12:337-350.

Jimenez, C., McLeod, A. I. and Hipel, K. W. (1989). Kalman filter estimation for periodic autoregressive-moving average models. *Stochastic Hydrology and Hydraulics*, 3:227-240.

Jones, R. H. and Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika*, 54:403-408.

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

Li, W. K. and Hui, Y. V. (1988). An algorithm for the exact likelihood of periodic autoregressive moving average models. *Communication in Statistics, Simulation and Computation*, 17(4):1483-1494.

McClave, J. T. (1978). Estimating the order of autoregressive models: The Max χ^2 method. *Journal of the American Statistical Association*, 73(363):122-128.

McLeod, A. I. and Hipel, K. W. (1978). Developments in monthly autoregressive modelling. Technical Report 45-XM-011178, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario.

Morgan, J. A. and Tatar, J. F. (1972). Calculation of the residual sum of squares for all possible regressions. *Technometrics*, 14(2):317-325.

Moss, M. E. and Bryson, M. C. (1974). Autocorrelation structure of monthly streamflows. *Water Resources Research*, 10:737-744.

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

Newton, H. J. (1982). Using periodic autoregressions for multiple spectral estimation. *Technometrics*, 24:109-116.

Noakes, D. J., McLeod, A. I. and Hipel, K. W. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*, 1:179-190.

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

Pagano, M. (1978). On periodic and multiple autoregressions. *The Annals of Statistics*, 6(6):1310-1317.

Rose, D. E. (1977). Forecasting aggregates of independent ARIMA processes. *Journal of Econometrics*, 5:323-345.

Sakai, H. (1982). Circular lattice filtering using Pagano's method. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-30(2):279-287.

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., Boes, D. C. and Smith, R. A. (1982). Estimation of ARMA models with seasonal parameters. *Water Resources Research*, 18(4):1006-1010.

Salas, J. D., Delleur, J. W., Yevjevich, V. and Lane, W. L. (1980). Applied modelling of hydrologic series. *Water Resources Publications*, Littleton, Colorado.

Salas, J. B. and Obeysekera, J. T. B. (1992). Conceptual basis of seasonal streamflow time series models. *Journal of Hydraulic Engineering*, ASCE, 118(8):1186-1194.

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).

Tao, P. C. and Delleur, J. W. (1976). Seasonal and nonseasonal ARMA models. *Journal of the Hydraulics Division, ASCE*, 102(HY10):1541-1559.

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*, pages 459-493. Harvard University Press.

Thompstone, R. M. (1983). *Topics in Hydrological Time Series Modelling*. PhD thesis, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

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

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

Tiao, G. C. and Gruppe, M. R. (1980). Hidden periodic autoregressive-moving average models. *Biometrika*, 67(2):365-373.

Troutman, B. M. (1979). Some results in periodic autoregression. *Biometrika*, 66(2):219-228.

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) modelling 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. *Technometric*, 27:375-384.

Vecchia, A. V. and Ballerini, R. (1991). Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, 78(1):53-63.

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.

PORTMANTEAU TEST STATISTIC

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, revised edition.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average models. *Journal of the American Statistical Association*, 65:1509-1526.

Davies, N., Triggs, C. M. and Newbold, P. (1977). Significance of the Box-Pierce Portmanteau statistics in finite samples. *Biometrika*, 64:517-522.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297-303.

McLeod, A. I. (1993). Diagnostic checking periodic autoregression models with application. *Journal of Time Series Analysis*.

STATISTICS

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29:610-611.

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburg, England.

Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburg, England.

Lehmann, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika*, 31:9-12.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley, New York, second edition.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*. The Iowa State University Press, Ames, Iowa.

Wickmann, B. A. and Hill, I. D. (1982). An efficient and portable pseudo-random number generator. *Applied Statistics*, 31:188-190.