# CHAPTER 13

# DESEASONALIZED MODELS

## 13.1 INTRODUCTION

*Deseasonalized models* are useful for describing time series in which the mean and variance within each season are stationary across the years. An example of a time series exhibiting these properties is the average monthly flows of the Saugeen River at Walkerton, Ontario, Canada, plotted in Figure VI.1. As can be seen from the sinusoidal shape of this graph, the mean and perhaps also the variance for each month change from season to season. Nevertheless, if one examines, for instance, the July flows across all of the years, these flows fluctuate around an overall mean level for the month of July and the variance is constant over time for the month of July. The deseasonalized models of this chapter and also the periodic models of Chapter 14 are ideally designed for capturing this type of statistical behaviour.

In addition to the *flexible design* of the deseasonalized models presented in Section 13.2, there are other distinct advantages for employing these models in practical applications. As explained in Section 13.3, a deseasonalized model can be easily identified and fitted to a given time series. Firstly, the seasonal component is removed from the series by subtracting from each observation the seasonal mean and perhaps also dividing this by the seasonal standard deviation. Subsequently, the most appropriate nonseasonal ARMA model is identified for fitting to the resulting nonseasonal series. Hence, the *model construction* tools presented in Part III for nonseasonal ARMA models, can also be used for building deseasonalized models.

To demonstrate clearly that the modelling methods can be easily used in practice, deseasonalized models are fitted to two *environmental time series* in Section 13.4. In the first application, a deseasonalized model is built for the average monthly flows of the Saugeen River displayed in Figure VI.1 while in the second example the most appropriate deseasonalized model is constructed for an average monthly ozone series.

Two other advantages for employing deseasonalized models are that they can be easily used for forecasting and simulation. As explained in Section 13.5, when *forecasting* with a deseasonalized model, the nonseasonal component of the series is forecasted using the procedure of Chapter 8 for nonseasonal ARMA models. Next, these forecasted values are converted to seasonal forecasts using the inverse of the deseasonalization procedure. Finally, if the original data were also transformed using a Box-Cox transformation, the inverse Box-Cox transformation is invoked to produce forecasts which possess the same units as the original untransformed series. A similar procedure is employed for *simulating* synthetic data using a deseasonalized model. The techniques of Chapter 9 are used to generate the simulated data from the nonseasonal ARMA model fitted to the deseasonalized series. Following this, the nonseasonal simulated values are filtered through the inverse deseasonalization method and inverse Box-Cox transformation to produce the untransformed synthetic sequence.

A drawback of using deseasonalized models is that they may require many model parameters for deseasonalizing the data. For instance, when dealing with a monthly time series, twelve monthly means and perhaps also twelve monthly standard deviations are needed for deseasonalization purposes. When modelling bimonthly and weekly time series a total of 48 and 104 deseasonalization parameters, respectively, may be required. To overcome the problem of *over-parameterization*, a *Fourier series approach* is described in Section 13.2.2. The applications in Section 13.4 are used for explaining how this method for reducing the number of model parameters is used in practice by following the overall model building stages explained in Section 13.3. In Section 13.5, approaches for forecasting and simulating with deseasonalized models are explained.

## 13.2 DEFINITIONS OF DESEASONALIZED MODELS

### 13.2.1 Introduction

As noted in the introduction, the deseasonalized model consists of the deseasonalization and ARMA model components. Because the basic design of the deseasonalized model reflects the inherent structure of many kinds of seasonal hydrological time series, this model has been used by hydrologists and environmental engineers for a long time for modelling, simulating and forecasting hydrologic phenomena (see, for example, Thomas and Fiering (1962), McMichael and Hunter (1972), McKerchar and Delleur (1974), Kavvas and Delleur (1975), Delleur et al. (1976), Tao and Delleur (1976), Croley and Rao (1977), Yevjevich and Harmancioglu (1989), and Jaywardena and Lai (1989), as well as the books on stochastic hydrology referred to in Section 1.6.3). In addition to defining this popular type of seasonal model in Sections 13.2.2 and 13.2.3, the AIC (Akaike information criterion) formula for the deseasonalized model is determined in Section 13.3.3 following the research of Hipel and McLeod (1979).

### 13.2.2 Deseasonalization

When considering data with $s$ seasons per year ($s = 12$ for monthly data) over a period of $n$ years, let $z_{r,m}$ represent a time series value in the $r$th year and $m$th season where $r = 1,2,\ldots,n$, and $m = 1,2,\ldots,s$. It is convenient to denote the $i$th previous value of $z_{r,m}$ by $z_{r,m-i}$, $i = 1,2,\ldots$. If, for example, one were dealing with monthly data, then $z_{9,12}$, $z_{10,0}$, and $z_{8,24}$ would all refer to the same observation.

If required, the given data may be transformed by the *Box-Cox transformation* (Box and Cox, 1964) to form the transformed series

$$z_{r,m}^{(\lambda)} = \begin{cases} \lambda^{-1}[(z_{r,m}+c)^{\lambda}-1], & \lambda \neq 0 \\ \ln(z_{r,m}+c), & \lambda = 0 \end{cases} \qquad [13.2.1]$$

for $r = 1,2,\ldots,n$, and $m = 1,2,\ldots,s$, where the constant $c$ is chosen just large enough to cause all entries in $z_{r,m}^{(\lambda)}$ to be positive, and $\lambda$ is the Box-Cox power transformation. Although one could have a separate Box-Cox transformation for each season of the year, in order to reduce the number of parameters it is assumed that the same $\lambda$ and $c$ are used for each season. The purpose of the Box-Cox transformation is to rectify anomalies such as heteroscedasticity and non-normality in the residuals of the ARMA model fitted to the deseasonalized time series.

The Box-Cox transformation given in [13.2.1] cannot remove seasonality from a time series. However, the following equations describe two *deseasonalization methods* that can be employed to deseasonalize the given data.

$$w_{r,m}^{(1)} = z_{r,m}^{(\lambda)} - \bar{\mu}_m \quad [13.2.2]$$

$$w_{r,m}^{(2)} = \frac{z_{r,m}^{(\lambda)} - \bar{\mu}_m}{\bar{\sigma}_m} \quad [13.2.3]$$

where $\bar{\mu}_m$ and $\bar{\sigma}_m$ are the fitted mean and standard deviation for the $m$th season.

The deseasonalization procedures given in [13.2.2] and [13.2.3], reflect the inherent statistical realities of many kinds of natural time series. For example, when considering average monthly river flow data, the observations for any particular month tend to fluctuate about some fixed mean level. Consequently, the deseasonalization method in [13.2.2] may be appropriate to employ if the monthly standard deviations of the $z_{r,m}^{(\lambda)}$ series are more or less constant throughout the year. When both the means and standard deviations of the $z_{r,m}^{(\lambda)}$ sequence are different from month to month, then the transformation in [13.2.3] should be utilized. In certain situations, a Box-Cox transformation may cause the standard deviations to become constant throughout the year for the $z_{r,m}^{(\lambda)}$ series and hence [13.2.2] can be used in preference to [13.2.3]. However, as pointed out in Chapter 12, for the type of natural time series just described, it is not recommended to difference the data to remove seasonality if the model is to be used for simulation.

The fitted means and standard deviations in [13.2.2] and [13.2.3] can be estimated using two approaches. One method is to estimate them using the standard formulae. Hence, the estimate, $\hat{\mu}_m$, for the mean of the mth season across all of the years, can be determined using

$$\hat{\mu}_m = \frac{1}{n}\sum_{r=1}^{n} z_{r,m}^{(\lambda)}, \quad m = 1,2,\ldots,s \quad [13.2.4]$$

The estimate of the standard deviation for the mth season, is calculated as

$$\hat{\sigma}_m = \left[\frac{1}{n}\sum_{r=1}^{n}(z_{r,m}^{(\lambda)} - \hat{\mu}_m)^2\right]^{0.5}, \quad m = 1,2,\ldots,s \quad [13.2.5]$$

For the case of monthly data, the deseasonalization transformation in [13.2.2] would require 12 means whereas the one in [13.2.3] needs 12 means and 12 standard deviations for a total of 24 deseasonalization parameters. When dealing with bimonthly, weekly and daily observations, a far greater number of deseasonalization parameters would be required

To reduce the total number of deseasonalization parameters that are needed, a *Fourier series approach to deseasonalization* can be utilized as another method for estimating the means and standard deviations. Let $F_\mu$ and $F_\sigma$ be the number of Fourier components to fit to the seasonal means in [13.2.4] and seasonal standard deviations in [13.2.5], respectively. Both $F_\mu$ and $F_\sigma$ can possess an integer value between 0 and $s/2$. The Fourier coefficients are determined from the equations.

$$A_k = \frac{2}{s}\sum_{m=1}^{s}\hat{\mu}_m \cos\frac{2\pi km}{s} \qquad [13.2.6]$$

$$B_k = \frac{2}{s}\sum_{m=1}^{s}\hat{\mu}_m \sin\frac{2\pi km}{s} \qquad [13.2.7]$$

$$C_h = \frac{2}{s}\sum_{m=1}^{s}\hat{\sigma}_m \cos\frac{2\pi hm}{s} \qquad [13.2.8]$$

$$D_h = \frac{2}{s}\sum_{m=1}^{s}\hat{\sigma}_m \sin\frac{2\pi hm}{s} \qquad [13.2.9]$$

where $k = 1,2,\ldots,F_\mu$, and $h = 1,2,\ldots,F_\sigma$ for $1 \le F_\mu$, $F_\sigma \le s/2$. To calculate the fitted means and standard deviations, set $\bar{\mu}_m = 0$ if $F_\mu = 0$ and let $\bar{\sigma}_m = 1$ if $F_\sigma = 0$. Otherwise, determine $\bar{\mu}_m$ and $\bar{\sigma}_m$ using

$$\bar{\mu}_m = A_0 + \sum_{k=1}^{F_\mu}\left(A_k\cos\frac{2\pi km}{s} + B_k\sin\frac{2\pi km}{s}\right), \quad m = 1,2,\ldots,s \qquad [13.2.10]$$

$$\bar{\sigma}_m = C_0 + \sum_{h=1}^{F_\sigma}\left(C_h\cos\frac{2\pi hm}{s} + D_h\sin\frac{2\pi hm}{s}\right), \quad m = 1,2,\ldots,s \qquad [13.2.11]$$

where $A_0 = \frac{1}{s}\sum_{m=1}^{s}\hat{\mu}_m$, and $C_0 = \frac{1}{s}\sum_{m=1}^{s}\hat{\sigma}_m$.

The deseasonalized series can be calculated using [13.2.2] if $F_\sigma = 0$ or otherwise [13.2.3]. When all of the Fourier components are used to calculate $\bar{\mu}_m$ (or $\bar{\sigma}_m$), then $\bar{\mu}_m = \hat{\mu}_m$ (and $\bar{\sigma}_m = \hat{\sigma}_m$) and, hence, $\bar{\mu}_m$ in [13.2.10] is equal to $\hat{\mu}_m$ in [13.2.4] and $\bar{\sigma}_m$ in [13.2.11] is the same as $\hat{\sigma}_m$ in [13.2.5]. Therefore, estimating the means and standard deviations using [13.2.4] and [13.2.5], respectively, can be considered a special case of the Fourier series approach.

### 13.2.3 ARMA Model Component

Subsequent to deseasonalization, the ARMA model defined in [3.4.4] can be fitted to the deseasonalized series represented by $w_{r,m}^{(1)}$ or $w_{r,m}^{(2)}$ in [13.2.2] or [13.2.3], respectively. For the case of the $w_{r,m}^{(2)}$ series in [13.2.3], the ARMA model would be written as

$$\phi(B)w_{r,m}^{(2)} = \theta(B)a_{r,m} \qquad [13.2.12]$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B - \cdots - \phi_p B^p$ is the AR operator of order $p$ for which $\phi_i$ is the ith AR parameter and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the MA operator of order $q$ for which $\theta_j$ is the jth MA parameter. The innovation series, $a_{r,m}$, is assumed to be distributed as IID$(0,\sigma_a^2)$. For estimating the model parameters, the normality assumption is invoked so that the innovations are required to be NID$(0,\sigma_a^2)$. If the residuals of the fitted model are not normally distributed and/or do not possess a constant variance, then an appropriate power transformation from [13.2.1] should be selected for correcting these problems.

## 13.3 CONSTRUCTING DESEASONALIZED MODELS

### 13.3.1 Introduction

When constructing a deseasonalized model, one follows in a general fashion the identification, estimation and diagnostic check stages of model building. During model construction, one must be able to identify and estimate the parameters contained in the deseasonalization and ARMA model components which form the overall deseasonalized model. In the next two sections, two basic approaches are presented for carrying out the model building process. In the first method, the time series under study is fully deseasonalized using [13.3.4] and [13.3.5] to estimate the deseasonalization parameters given in [13.2.2] and [13.2.3]. Then the best ARMA model is fitted to the resulting nonseasonal time series. The resulting model is then subjected to diagnostic check to ascertain if design modifications are needed to correct abnormalities in the ARMA model residuals. In the second approach, the AIC is used to determine which Fourier parameters are required in the deseasonalization step so that a more parsimonious model can be found. This second method is often needed when one is examining series, such as bimonthly or weekly sequences, for which the number of seasons per year is greater than twelve. Special AIC formulae are derived for deseasonalized models so one can employ the MAICE (minimum AIC estimation) procedure (see Section 6.3) for selecting the model having the fewest number of parameters. The AIC formulae are also valid for use with deseasonalized models having a Box-Cox transformation.

Figure 13.3.1 outlines the two procedures for fitting deseasonalized models to a time series. When following the fully deseasonalized approach of Section 13.3.2, one does not use the AIC to determine the Fourier components needed in the deseasonalization. Alternatively, when utilizing the AIC to determine which Fourier components are needed, one is constructing the deseasonalized model according to the procedure explained in Section 13.3.3. Notice that Figure 13.3.1 is similar to the general AIC model building methodology depicted in Figure 6.3.1. Moreover, most of the model building tools of Part III, can be employed for developing the ARMA component of the deseasonalized model of Sections 13.3.2 and 13.3.3.

### 13.3.2 Fully Deseasonalized Models

As just mentioned, the approach for fitting a *fully deseasonalized model* to a seasonal time series is found in Figure 13.3.1 by following the path which ignores using the AIC for determining the Fourier components needed in the deseasonalization method. Firstly, if it is known a priori that a Box-Cox transformation is needed, one can use [13.2.1] to transform the data. If the need for a Box-Cox transformation is not known in advance, it will be detected by testing the model residuals at the diagnostic check stage. Subsequent to implementing an appropriate transformation, the model parameters can be calibrated once again as shown in Figure 13.3.1.

Secondly, the transformed seasonal series is fully deseasonalized. If it is suspected based on graphs of the series or estimates of the seasonal standard deviations that the seasonal variance is about the same across the seasons, one can use [13.2.2] to deseasonalize the series. When using [13.2.2], each seasonal mean is estimated using the standard formula given in [13.2.4]. If, on the other hand, the variance changes across the seasons, [13.2.3] is used for obtaining the deseasonalized series where [13.2.4] and [13.2.5] are utilized for estimating the seasonal means and standard deviations, respectively.

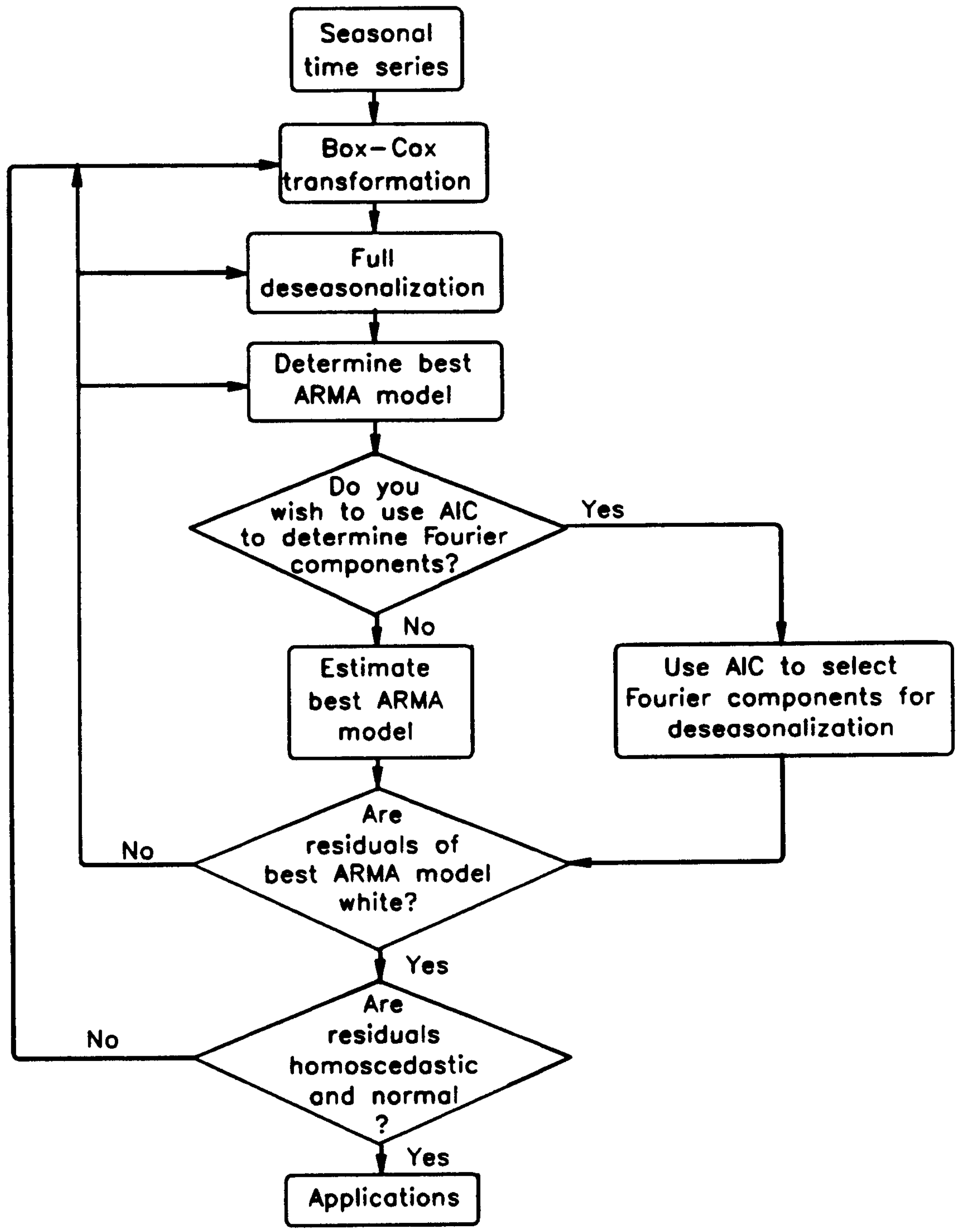


Figure 13.3.1. Constructing deseasonalized models.

Thirdly, the procedures of Part III are used to determine the best ARMA model to fit to the nonseasonal series found using [13.2.2] or [13.2.3]. If, for example, the best ARMA is not clearly identified using the identification graphs presented in Chapter 5, one may have to estimate the parameters for a few different ARMA models and select the one possessing the minimum value of the AIC (see Section 6.3). Whatever the case, in Figure 13.3.1, one omits using the AIC for determining the Fourier components and proceeds to estimating the parameters of the most appropriate ARMA model. Next, the residuals of the fitted ARMA model are checked to see if their underlying assumptions are satisfied. In particular, one should make sure that the residuals are uncorrelated, normally distributed and homoscedastic. Specific tests for checking that these properties are not violated are presented in detail in Chapter 7. Problems with heteroscedasticity and/or non-normality can be corrected by invoking a suitable Box-Cox transformation. If the residuals are correlated, a different ARMA model should be fitted to the fully deseasonalized series. After obtaining a satisfactory model, the overall deseasonalized model can be used for application purposes such as forecasting and simulation.

### 13.3.3 Fourier Approach to Deseasonalized Models

#### Overall Procedure

The overall procedure for carrying out the *Fourier approach to deseasonalized models* is explained first. Following this, the AIC formulae which are used in this procedure are derived.

The Fourier approach is traced in Figure 13.3.1 by following the path which uses the AIC for ascertaining the Fourier components needed in the deseasonalization. As can be seen, the first three steps consisting of a Box-Cox transformation, full deseasonalization, and determining the best ARMA model to fit to the fully deseasonalized data are identical to those explained for the fully deseasonalized model in Section 13.3.2. To cut down on the number of deseasonalization parameters that are needed in either [13.2.2] or [13.2.3], the MAICE procedure can be used to select the Fourier components that are required in [13.2.10] to [13.2.11]. When doing this, one can assume that the ARMA model identified for fitting to the fully deseasonalized data set requires the same number of AR and MA parameters. For each possible combination of Fourier components, one can estimate the AIC for a given type of ARMA model. The deseasonalized model possessing the minimum AIC value is then selected as the most satisfactory model. If one suspected that the type of ARMA model fitted to the deseasonalized data were dependent upon the number of Fourier components, one could also allow the number of AR and MA parameters to vary during the MAICE procedure. However, this would require a significant increase in the amount of computations and therefore is not shown in Figure 13.3.1. Whatever the case, subsequent to choosing the best deseasonalized model, one can subject the ARMA model residuals to the diagnostic checks given in Chapter 7. As shown in Figure 13.3.1, problems with the model residuals can be rectified by returning to an earlier step in the modelling procedure. The best overall deseasonalized model can then be used for application purposes.

#### AIC Formulae for Deseasonalized Models

Recall from [6.3.1] Section 6.3, that the general formula for the AIC is given as (Akaike, 1974)

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

where $ML$ stands for the maximized value of the likelihood function and $k$ is the number of free parameters. The model which adequately fits a time series using a minimum number of parameters possesses the lowest AIC value. The procedure for determining the model having the minimum AIC value is referred to as MAICE.

When calculating the AIC formula for the deseasonalized model, one must consider the effects of a Box-Cox transformation and deseasonalization. Hence, one must determine the Jacobians of the deseasonalization transformations.

Following the derivation given by Hipel and McLeod (1979), first consider the deseasonalization procedure given in [13.2.2]. The Jacobian of the transformation from $z_{r,m}$ to $w_{r,m}^{(1)}$ is

$$J_1 = \left| \left( \frac{\partial w_{r,m}^{(1)}}{\partial z_{r,m}} \right) \right| = \prod_{r=1}^{n} \prod_{m=1}^{s} z_{r,m}^{\lambda-1} \qquad [13.3.1]$$

To explain in detail how [13.3.1] is derived, examine the Box-Cox transformation in [13.2.1]. Assuming that all of the $z_{r,m}$ are positive and, therefore, the constant is zero,

$$z_{r,m}^{(\lambda)} = \frac{z_{r,m}^{\lambda} - 1}{\lambda}$$

and from [13.2.2] the deseasonalized transformed series is

$$w_{r,m}^{(1)} = \frac{z_{r,m}^{\lambda} - 1}{\lambda} - \bar{\mu}_m$$

The entire matrix used in the Jacobian calculation in [13.3.1] is

$$\begin{pmatrix} \frac{\partial w_{11}^{(1)}}{\partial z_{11}} & \frac{\partial w_{11}^{(1)}}{\partial z_{12}} & \cdots & \frac{\partial w_{11}^{(1)}}{\partial z_{n,s}} \\ \frac{\partial w_{12}^{(1)}}{\partial z_{11}} & \frac{\partial w_{12}^{(1)}}{\partial z_{12}} & \cdots & \frac{\partial w_{12}^{(1)}}{\partial z_{n,s}} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial w_{r,s}^{(1)}}{\partial z_{11}} & \frac{\partial w_{r,s}^{(1)}}{\partial z_{12}} & \cdots & \frac{\partial w_{r,s}^{(1)}}{\partial z_{n,s}} \end{pmatrix}$$

In this matrix, all of the off-diagonal entries have a value of zero whereas each diagonal element is evaluated as

$$\frac{\partial w_{r,m}^{(1)}}{\partial z_{r,m}} = \frac{\lambda z_{r,m}^{\lambda-1}}{\lambda} = z_{r,m}^{\lambda-1}$$

Therefore, the value of the determinant of this matrix which constitutes the Jacobian $J_1$ is found by multiplying the diagonal entries together to obtain the result in [13.3.1].

The natural logarithm of $J_1$ is

$$\ln(J_1) = (\lambda - 1)\sum_{r=1}^{n}\sum_{m=1}^{s}\ln(z_{r,m}) \qquad [13.3.2]$$

The log-likelihood in terms of $z_{r,m}$ is then

$$\ln(L^{(1)}) \approx -ns\ln\left(\frac{MSS}{ns}\right) + \ln(J_1) \qquad [13.3.3]$$

where $MSS$ stands for the modified sum of squares described in Appendix A6.1.

The AIC for an ARMA model that is fit to the $w_{i,j}^{(1)}$ series is

$$AIC = -2\ln(L^{(1)}) + 2(p + q + 2 + \delta_1) + 4F_\mu - 2\delta_2 \qquad [13.3.4]$$

where $p$ is the number of AR parameters, $q$ is the number of MA parameters, $\delta_1 = 0$ when $\lambda = 1$ and $\delta_1 = 1$ when $\lambda \neq 1$, $F_\mu$ is the number of Fourier parameters in [13.2.10] used to estimate the seasonal means, and $\delta_2 = 1$ if $F_\mu = \frac{s}{2}$ while $\delta_2 = 0$ when $F_\mu < \frac{s}{2}$.

In the deseasonalization method given in [13.2.3], each entry in the $w_{r,m}^{(1)}$ series is divided by the appropriate standard deviation of $z_{r,m}^{(\lambda)}$ for the mth season. The Jacobian of the transformation from $w_{r,m}^{(1)}$ to $w_{r,m}^{(2)}$ is

$$J_2 = \left|\left(\frac{\partial w_{r,m}^{(2)}}{\partial w_{r,m}^{(1)}}\right)\right| = \left(\prod_{m=1}^{x}\hat{\sigma}_m^{-1}\right)^n \qquad [13.3.5]$$

To derive [13.3.5], consider the matrix used in the Jacobian formulae in [13.3.5] which is written as

$$\begin{pmatrix}
\frac{\partial w_{11}^{(2)}}{\partial w_{11}^{(1)}} & \frac{\partial w_{11}^{(2)}}{\partial w_{12}^{(1)}} & \cdots & \frac{\partial w_{11}^{(2)}}{\partial w_{n,s}^{(1)}} \\
\frac{\partial w_{12}^{(2)}}{\partial w_{11}^{(1)}} & \frac{\partial w_{12}^{(2)}}{\partial w_{12}^{(1)}} & \cdots & \frac{\partial w_{12}^{(2)}}{\partial w_{n,s}^{(1)}} \\
\cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & & \cdot \\
\frac{\partial w_{n,s}^{(2)}}{\partial w_{11}^{(1)}} & \frac{\partial w_{n,s}^{(2)}}{\partial w_{12}^{(1)}} & \cdots & \frac{\partial w_{n,s}^{(2)}}{\partial w_{n,s}^{(1)}}
\end{pmatrix}$$

From [13.2.2] and [13.2.3], the relationship between $w_{r,m}^{(1)}$ and $w_{r,m}^{(2)}$ determined as

$$w_{r,m}^{(2)} = \frac{w_{r,m}^{(1)}}{\bar{\sigma}_m}$$

Hence,

$$\frac{\partial w_{r,m}^{(2)}}{\partial w_{r,m}^{(1)}} = \frac{1}{\bar{\sigma}_m} = \bar{\sigma}_m^{-1}$$

Therefore, the diagonal elements in the matrix have a value of $\bar{\sigma}_m^{-1}$ while the off-diagonal elements are zero. The determinant of this matrix which is used to obtain the Jacobian $J_2$ is determined simply by multiplying together the values of the diagonal elements to obtain [13.3.5].

The natural logarithm of $J_2$ is

$$\ln(J_2) = -n \sum_{m=1}^{12} \ln(\bar{\sigma}_m) \qquad [13.3.6]$$

The log-likelihood for the $z_{r,m}$ series is

$$\ln(L^{(2)}) \approx -ns\ln\left(\frac{MSS}{ns}\right) + \ln(J_1) + \ln(J_2) \qquad [13.3.7]$$

The AIC formula for an ARMA model that is fitted to the $w_{r,m}^{(2)}$ sequence is

$$AIC = -2\ln(L^{(2)}) + 2(p + q + 2 + \delta_1) + 4(F_\mu + F_\sigma) - 2(\delta_2 + \delta_3) \qquad [13.3.8]$$

where $F_\sigma$ is the number of Fourier parameters in [13.2.11] used to estimate the seasonal standard deviations, $\delta_3 = 1$ when $F_\sigma = \frac{s}{2}$ while $\delta_3 = 0$ whenever $F_\sigma < \frac{s}{2}$.

For certain data sets, it may be known in advance what type of Box-Cox transformation should be used. Hence, it may be appropriate to set $\lambda$ at a specified value. In other situations, it may be desirable to obtain a maximum likelihood estimate for the Box-Cox exponent. This can be accomplished by maximizing the log-likelihood in [13.3.3] and [13.3.7] with respect to the model parameters.

As explained in Section 6.3.4, the difference in the values of the AIC for various models which are fit to the same data set, can be interpreted in different manners. For instance, if one model has an AIC value which is approximately $2k$ less than that for another model, this is analogous to the superior model having $k$ less parameters than the other model. An alternative approach for interpreting the difference in AIC values between two models, is to determine the plausibility of model $i$ versus model $j$ by using the formula

$$\text{Plausibility} = \exp[0.5(AIC_j - AIC_i)] \qquad [13.3.9]$$

where $AIC_i$ is the value of the AIC for the $i$th model, $AIC_j$ is the value of the AIC for the jth model and the $j$th model is assumed to be the model having the lower AIC value. This formula was suggested by H. Akaike in a private communication and is also written in [6.3.2].

## 13.4 APPLICATIONS OF DESEASONALIZED MODELS

### 13.4.1 Introduction

To demonstrate how the model construction approaches of Section 13.3 are used in practice, deseasonalized models are developed for two natural time series by adhering to the procedures outlined in Figure 13.3.1. In the first application, a deseasonalized model is fitted to the average monthly flows of the Saugeen River shown in Figure VI.1. The MAICE procedure is utilized in the second example for determining a deseasonalized model to describe a monthly ozone time series. Ozone is used as an indicator of pollution levels caused by exhausts from automobiles and other types of machinery driven by internal combustion engines. By using the MAICE procedure, very few deseasonalization parameters are needed in the fitted deseasonalized model.

### 13.4.2 Average Monthly Saugeen Riverflows

The average monthly flows of the Saugeen River at Walkerton, Ontario, Canada, are available from Environment Canada (1977) from January, 1915, until December, 1976, and the last ten years of this time series are plotted in Figure VI.1. This particular data set is suitable for modelling using a deseasonalized model because there are no major reservoirs on the river and also the land-use activities in the river basin have not changed significantly during the aforesaid time period. Consequently, the technique of intervention analysis of Part VIII does not have to be used to account for intervention effects.

First consider using the modelling construction approach of Section 13.3.2. This full deseasonalization procedure follows the vertical path on the left in Figure 13.3.1 which does not use the AIC for determining the optimum number of Fourier components needed in deseasonalization.

As pointed out in Section 12.4.4, in practice, it has been found necessary to first take natural logarithms of average monthly riverflow time series values. The logarithmic transformation often precludes problems with non-normality and/or heteroscedasticity in the residuals of the model which is fitted to the deseasonalized data. A common procedure in water resources engineering is to fully deseasonalize the transformed data by using the estimated means and standard deviations from [13.2.4] and [13.2.5], respectively, in place of the fitted means and standard deviations in [13.2.3]. This is equivalent to setting $F_{\mu} = F_{\sigma} = 6$ in [13.2.6] to [13.2.9]. Even though this may not constitute the most appropriate deseasonalization method for many time series, it is, however, useful for the model identification suggested in Figure 13.3.1. In addition, as shown by this example, most of the Fourier components are needed for deseasonalization. After taking natural logarithms of the monthly Saugeen River data which is given in cubic metres per second, the logarithmic data is deseasonalized following [13.2.3] by using the estimated means and standard deviations in [13.2.4] and [13.2.5], respectively. By following the identification procedures described in Chapter 5 for nonseasonal ARMA models, it is found that a model with one AR and one MA parameter [denoted as ARMA(1,1)] may adequately model the deseasonalized series.

Figure 13.4.1 displays the fully deseasonalized series for the last ten years of the average monthly riverflows of the Saugeen River. In order to be able to compare this figure to original Saugeen flows in Figure VI.1, the deseasonalized series in Figure 13.4.1 is determined for the situation where there is no Box-Cox transformation and hence $\lambda = 1.0$ in [13.2.1] as well as the deseasonalization calculation in [13.2.3]. As can be seen, the deseasonalized series in Figure 13.4.1 does not contain the distinct sinusoidal patterns reflecting seasonality shown in Figure VI.1. Finally, as noted above, the reader should keep in mind that the ARMA(1,1) model is identified for the fully deseasonalized series with $\lambda = 0$.

Table 13.4.1. AIC values for the ARMA(1,1) model fitted to the deseasonalized Saugeen River series.

| $F_\mu$ | $F_\sigma$ | AIC |
|---|---|---|
| 5 | 0 | 3383.15 |
| 5 | 1 | 3359.17 |
| 5 | 2 | 3365.98 |
| 5 | 3 | 3357.93 |
| 5 | 4 | 3355.82 |
| 5 | 5 | 3356.77 |
| 5 | 6 | 3361.81 |
| 6 | 0 | 3387.13 |
| 6 | 1 | 3363.15 |
| 6 | 2 | 3369.96 |
| 6 | 3 | 3361.91 |
| 6 | 4 | 3359.81 |
| 6 | 5 | 3360.75 |
| 6 | 6 | 3365.79 |

At the estimation stage of model development, the estimation procedure described in Appendix A6.1 is employed to obtain MLE's for the parameters of the ARMA(1,1) model which is fitted to the fully deseasonalized data in [13.2.3]. Using [13.3.8], the AIC for this fully deseasonalized model can be calculated for the case of full deseasonalization which is equivalent to $F_\mu = F_\sigma = 6$ in the Fourier series approach to deseasonalization. The value of the AIC for the fully deseasonalized model is given at the bottom of Table 13.4.1. This model passes the diagnostic checks suggested in Figure 13.3.1 and described in detail in Chapter 7. Consequently, the best ARMA model to fit to fully deseasonalized logarithmic monthly Saugeen flows is an ARMA(1,1) model.

Now consider how the Fourier approach to deseasonalization presented in Section 13.3.3 is used with the Saugeen data. As depicted in Figure 13.3.1, one must now fit the ARMA(1,1) model to each deseasonalized data set formed by all possible combination of Fourier components for estimating the seasonal means and standard deviations. Hence, the ARMA(1,1) model is fitted to each of the possible 49 deseasonalized data sets and [13.3.8] is used to calculate the AIC for each case. Table 13.4.1 lists the values of the AIC for some of the deseasonalized data sets. As can be seen from that table, the minimum value of the AIC occurs when 5 Fourier components are used for the means while the standard deviations require 4 Fourier components. Notice that the deseasonalization procedure which uses 6 Fourier components for both the means and standard deviations, is somewhat inferior to the best method. From [13.3.9], the plausibility

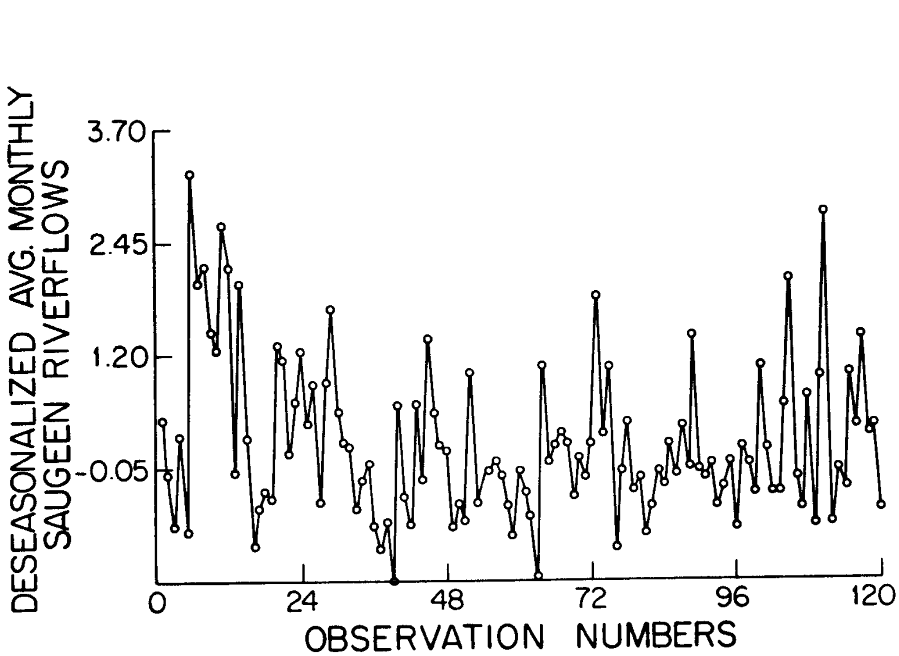


Figure 13.4.1. Fully deseasonalized series with $\lambda = 1$ in [13.2.3] for the last ten years of the average monthly flows of the Saugeen River at Walkerton, Ontario, Canada.

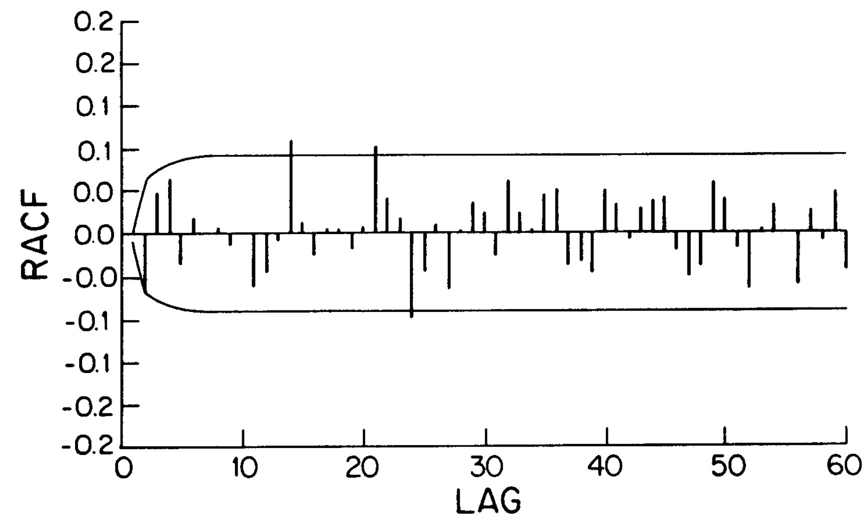


Figure 13.4.2. RACF for the ARMA(1,1) model fitted to the deseasonalized average monthly flows of the Saugeen River at Walkerton, Canada, with $\lambda = 0$, $F_{\mu} = 5$ and $F_{\sigma} = 4$.

of the model where the estimated means and standard deviations are used, versus the best seasonal model, is calculated to be 0.68%. Consequently, water resources engineers should not necessarily assume that the best deseasonalization method is to let $F_{\mu} = F_{\sigma} = 6$.

Examination of the residuals for the ARMA(1,1) model which is fitted to the deseasonalized data with $F_m = 5$ and $F_s = 4$, reveals that the whiteness assumption is satisfied. In particular, the graph of the RACF (residual autocorrelation function) in Figure 13.4.2 for this model shows that the values of the RACF fall within the 95% confidence limits. As discussed in Chapter 7 and elsewhere in this text, it has been found in practice that the important residual assumptions are usually fulfilled for the model which has been chosen using the MAICE procedure.

Diagnostic checks are given in Sections 7.4 and 7.5 to determine if the model residuals are approximately normally distributed and homoscedastic, respectively. If either of the aforesaid assumptions are not satisfied, the Box-Cox transformation in [13.2.1] can be invoked to rectify the situation. For the case of the Saugeen River data, it is assumed from the outset that a logarithmic transformation may be needed. Nevertheless, various values of $\lambda$ in [13.2.1] are examined and $\lambda = 0$ does indeed constitute a reasonable transformation.

As would be expected all the deseasonalized models in Table 13.4.1 possess lower values than the AIC for the best SARIMA model fitted to the average monthly Saugeen Riverflows in Section 12.4.4. This is because the design of the deseasonalized model allows for a stationary mean and standard deviation within each season. The reader should keep in mind that because different estimation procedures are used for obtaining estimates of the model parameters for different classes of models, one should entertain caution when comparing AIC values across families of models.

### 13.4.3 Ozone Data

Monthly values of the concentration of ozone (in parts per hundred million) at Azusa, California, are available from January 1956 until December 1970. Figure 13.4.3 shows a plot of the ozone data, which are available in a technical report by Tiao et al. (1973). Abraham and Box (1978) have analyzed this data and have suggested that a moving average model with one parameter [denoted ARMA(0,1)] and no transformation (i.e., $\lambda = 1$ in [13.2.1]) is appropriate to model the deseasonalized data in [13.2.2] with $F_{\mu} = 2$. However, by employing the MAICE procedure in conjunction with the AIC formulae developed in Section 13.3.3, an improved ozone model is obtained.

Following Figure 13.3.1, a tentative model is identified to fit to the possible array of deseasonalized data sets by first examining the $w_{i,j}^{(2)}$ series in [13.2.3] with $F_{\mu} = F_{\sigma} = 6$. The fully deseasonalized series for the ozone data is displayed in Figure 13.4.3 and, as can be seen, deseasonalization removes the periodic seasonal component shown in Figure 13.4.3. A perusal of the sample ACF and PACF for this series, reveals that an AR model with one AR parameter [denoted ARMA(1,0)] may be suitable for modelling the data. The ARMA(1,0) model is fitted to the various types of deseasonalized time series which are obtained by simultaneously allowing $F_{\mu}$ and $F_{\sigma}$ to take on all possible values between 0 and 6. The values of the AIC obtained using [13.3.8] are listed in Table 13.4.2 for some of the models considered. It can be seen that an ARMA(1,0) with no Box-Cox transformation (i.e. $\lambda = 1$) and $F_{\mu} = 2$ and $F_{\sigma} = 1$ is preferable to

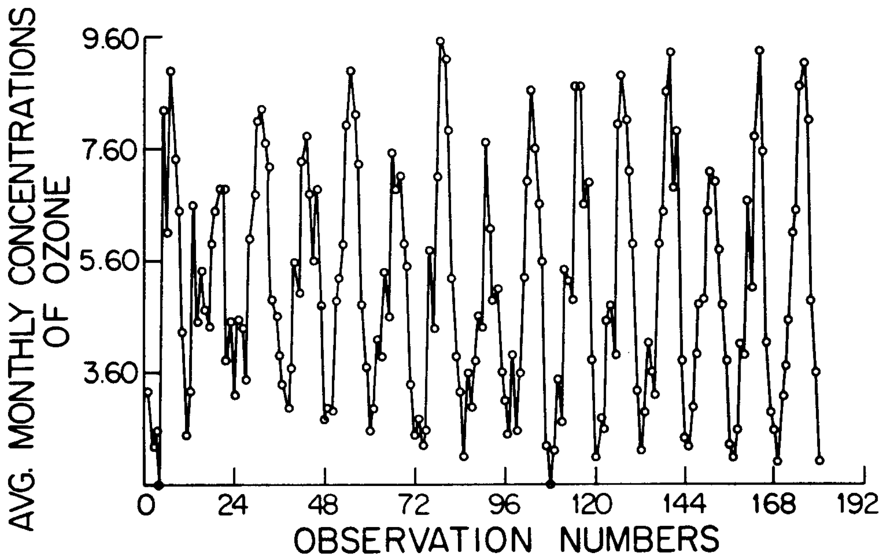


Figure 13.4.3. Average monthly concentrations of ozone (parts per hundred million) from January, 1956, to December, 1970, at Azusa, California.

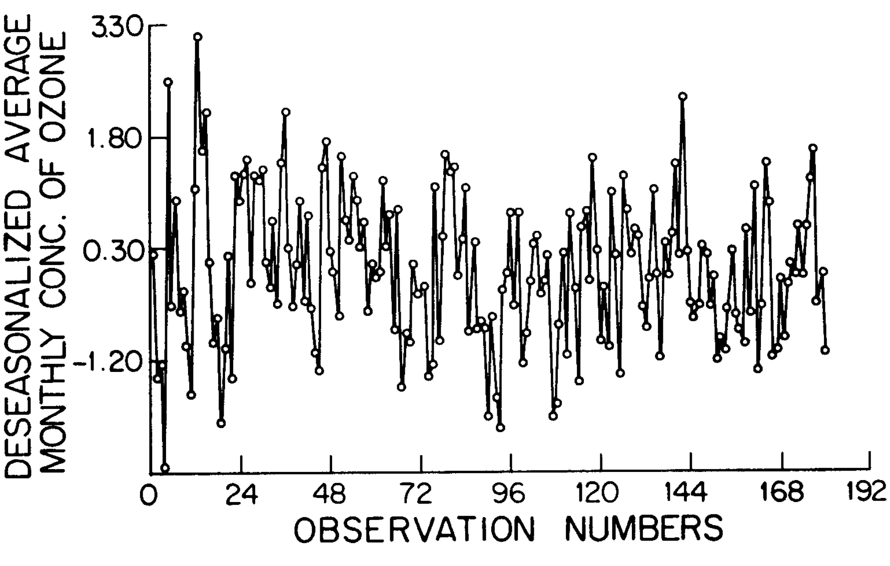


Figure 13.4.4. Fully deseasonalized series with $\lambda = 1$ in [13.2.3] for the last ten years of the average monthly concentrations of ozone (parts per hundred million) from January, 1956, to December, 1970, at Azusa, California.

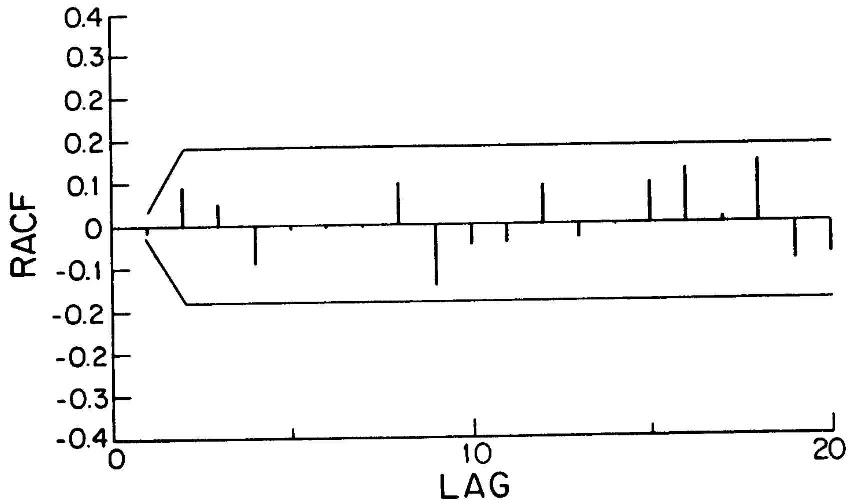


Figure 13.4.5. RACF for the ARMA(1,0) model fitted to the deseasonalized average monthly concentrations of ozone (parts per hundred million) at Azusa, California, with $\lambda = 0.5$, $F_\mu = 2$ and $F_\sigma = 1$.

the case where the same model is fitted to the deseasonalized data with $F_\mu = 2$ and $F_\sigma = 0$. In addition, as shown in Table 13.4.2, both of the aforesaid models possess lower AIC values than those obtained for ARMA(0,1) models which are fitted to the same deseasonalized series. However, an examination of the residuals for the most appropriate foregoing model, reveals that the residuals possess both significant skewness and heteroscedasticity. Consequently, various values of $\lambda$ are examined in order to determine a suitable Box-Cox transformation from [13.2.1]. A transformation with $\lambda = 0.5$ corrects the anomolies in the model residuals. The graph of the RACF in Figure 13.4.5 for the ARMA(1,0) model fitted to the deseasonalized series with $\lambda = 0.5$, $F_\mu = 2$ and $F_\sigma = 1$ demonstrates that the residuals are white.

In Table 13.4.2, the values of the AIC are listed for various models which are fitted to the data having a square root transformation. It is evident that the most desirable model is an ARMA(1,0) model with $\lambda = 0.5$, $F_\mu = 2$ and $F_\sigma = 0$. The plausibility of the process recommended by Abraham and Box (1978) versus the best model in Table 13.4.2, is calculated using [13.3.9] to be 0.95%.

## 13.5 FORECASTING AND SIMULATING WITH DESEASONALIZED MODELS

As noted in the introduction in Section 13.1, deseasonalized models can be easily used for forecasting and simulation by slightly extending the approaches described in Chapters 8 and 9, respectively. The key difference in the procedures is that one has to take into account the effects of deseasonalization when forecasting or simulating with the deseasonalized model.

Table 13.4.2. AIC values for ARMA models fitted to the deseasonalized ozone data.

| ARMA Model | Box-Cox Transformation Parameter $\lambda$ | $F_\mu$ | $F_\sigma$ | AIC |
|---|---|---|---|---|
| (1,0) | 1 | 2 | 1 | -8.59 |
| (1,0) | 1 | 2 | 0 | -8.19 |
| (0,1) | 1 | 2 | 1 | -7.71 |
| (0,1) | 1 | 2 | 0 | -7.43 |
| (1,0) | 0.5 | 2 | 1 | -12.09 |
| (1,0) | 0.5 | 2 | 0 | -10.27 |
| (0,1) | 0.5 | 2 | 1 | -11.32 |
| (0,1) | 0.5 | 2 | 0 | -9.55 |

Figure 13.5.1 outlines the procedure for forecasting with a deseasonalized model which is fitted to a seasonal time series. Firstly, the ARMA model fitted to the deseasonalized transformed series is used to obtain minimum mean square error forecasts of this data set. The procedure of Section 8.2.4 can be used for accomplishing this. Next, the forecasts for $z_{r,m}^{(\lambda)}$ are calculated using the inverse deseasonalization method which can be obtained from [13.2.2] or [13.2.3]. Thirdly, [13.2.1] can be used to determine the inverse Box-Cox transformation in order to obtain forecasts for $z_{r,m}$. The forecasts for $z_{r,m}$ are now in the same units as the original series to which the deseasonalized model was fitted. In Section 8.2.7, it is explained how one should correctly calculate minimum mean square error forecasts in the untransformed domain when the data were transformed using a Box-Cox transformation before fitting the model.

A similar procedure to that given in Figure 13.5.1 for forecasting can be used when simulating with a deseasonalized model. Simply replace the word forecast by simulate and forecasts by simulated values. In the first step, one of the procedures of Section 9.3 or 9.4 can be used to obtain the simulated sequences for the ARMA model fitted to the deseasonalized series. The rest of the procedure is the same as that explained for forecasting.

## 13.6 CONCLUSIONS

As exemplified by the two applications in Section 13.4, the procedures of Section 13.3 can be used to fit conveniently deseasonalized models to seasonal environmental time series. The two basic approaches to model construction presented in Sections 13.3.2 and 13.3.3 are summarized in Figure 13.3.1. For the case of the average monthly riverflows of the Saugeen River at Walkerton, Ontario, most of the Fourier components were needed to fully deseasonalize the series before fitting the ARMA model to the data. However, as shown by the monthly ozone data application in Section 13.4.3, a time series which does not possess a strongly pronounced seasonal structure may require less Fourier parameters for deseasonalization.

Deseasonalized models are designed for preserving the mean and variance within each season of the year. If one also wishes to capture the seasonal correlation structure, the periodic models described in the next chapter can be employed.

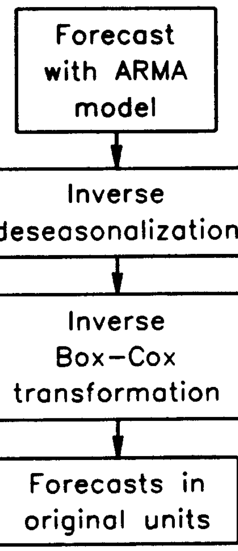


Figure 13.5.1. Forecasting with a deseasonalized model.

## PROBLEMS

**13.1** In the SARIMA modelling approach to seasonal time series modelling in Chapter 12, differencing is employed to remove periodicity or seasonality. Within this chapter, deseasonalization is utilized for modelling periodicity. Employing some of the results of Kavvas and Delleur (1975) as well as other authors, compare the advantages and drawbacks of these two procedures for describing periodicity in time series.

**13.2** Carry out an exploratory data analysis of an average monthly riverflow time series. Point out the main statistical characteristics of the data set. Of the three types of seasonal models presented in Part VI, which type of seasonal model do you think is most appropriate to fit to the data set?

**13.3** Using the same monthly riverflow time series as in problem 13.2, fit the most appropriate fully deseasonalized model to the data set.

**13.4** For the same time series used in problem 13.2, employ the approach of Section 13.3.3 to obtain a more parsimonious deseasonalized model. Comment upon your findings by comparing the results to those obtained in the two previous questions.

**13.5** Examine appropriate graphs from an exploratory data analysis study of an average weekly time series of your choice. After explaining your exploratory findings, follow the procedure of Section 13.3.3 to construct a parsimonious model for fitting to the weekly series.

**13.6** Execute problem 13.5 for an average daily time series that is of interest to you.

**13.7** Fit the most appropriate deseasonalized model to a monthly riverflow time series. Follow the procedures of Sections 13.5 and 8.2.4 to forecast 24 steps ahead from the last data point in the time series. Also, plot the 90% probability interval.

**13.8** Develop the most reasonable deseasonalized model to describe a monthly riverflow time series. Employing the techniques of Section 13.5 and Chapter 9, simulate and plot five sequences that are of the same length as the historical series. Compare the five simulated sequences and the historical data and then comment upon your findings.

## REFERENCES

### DATA SETS

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

Tiao, G. C., Box, G. E. P. and Hamming, W. J. (1973). Analysis of Los Angeles photochemical smog data: A statistical overview. Technical Report No. 331, Department of Statistics, University of Wisconsin, Madison, Wisconsin.

### DESEASONALIZED MODELS

Abraham, B. and Box, G. E. P. (1978). Deterministic and forecast adaptive time dependent models. *Applied Statistics,* 27:120-130.

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

Delleur, J. W., Tao, P. C. and Kavvas, M. L. (1976). An evaluation of the practicality and complexity of some rainfall and runoff time series models. *Water Resources Research,* 12(5):953-970.

Hipel, K. W. and McLeod, A. I. (1979). Modelling seasonal geophysical time series using deseasonalized models. Technical Report No. 52-XM-030579, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario.

Jayawardena, A. W. and Lai, F. (1989). Time series analysis of water quality data in Pearl River, China. *Journal of Environmental Engineering,* 115(3):590-607.

Kavvas, M. L. and Delleur, J. W. (1975). Removal of periodicities by differencing and monthly mean subtraction. *Journal of Hydrology,* 26:335-353.

McKerchar, A. I. and Delleur, J. W. (1974). Application of seasonal parametric linear stochastic models to monthly flow data. *Water Resources Research,* 10(2):246-255.

McMichael, F. C. and Hunter, J. S. (1972). Stochastic modeling of temperature and flow in rivers. *Water Resources Research*, 8(1):87-98.

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*, 459-493. Harvard University Press, Cambridge, Massachusetts.

Yevjevich, V. and Harmancioglu, N. B. (1989). Description of periodic variation in parameters of hydrologic time series. *Water Resources Research*, 25(3):421-428.

## TIME SERIES MODELLING

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19:716-723.

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