CHAPTER 15

FORECASTING WITH SEASONAL MODELS

15.1 INTRODUCTION

Three families of models are presented in Part VI of the book for fitting to seasonal time series. In particular, SARIMA, deseasonalized and periodic models are described in Chapters 12 to 14, respectively. The objective of this chapter is to employ *forecasting experiments* for comparing the capabilities of these seasonal models to forecast accurately seasonal hydrological time series.

Forecasting can be utilized for *model discrimination* purposes. After fitting different types of models to one or more time series by following proper model construction procedures, the model or models which forecasts the best according to certain criteria can be selected for use in further practical applications. Because carrying out forecasting studies is a very time consuming undertaking, forecasting experiments cannot be used for discriminating among models in most applications. Nonetheless, if one finds, for example, in an extensive forecasting experiment, that a certain type of PAR model forecasts significantly better than its competitors when used with average monthly riverflow series, this would give one confidence in using PAR models in other applications involving average monthly riverflow series.

After explaining *how to calculate forecasts* for seasonal models in Section 15.2, two main forecasting studies are described in the next two major sections. In the *first set of forecasting experiments*, mean monthly flows from thirty rivers in North and South America are used to test the short-term forecasting ability of SARIMA, deseasonalized and PAR models. After splitting each series into two sections, the seasonal models are calibrated for the first portion of the data. The fitted models are then used to generate one-step ahead forecasts for the second portion of each time series. The forecasting performance of the models is compared using various measures of accuracy. The results suggest that PAR models identified using the sample periodic ACF and PACF provide the most accurate forecasts. The results of this study are also presented by Noakes et al. (1985) as well as Noakes (1984, Ch. V).

In the *second forecasting study*, the three quarter-monthly and three monthly riverflow series used in Sections 14.6 and 14.8.2 of the previous chapter, are used for comparing the forecasting accuracy of seasonal models. Besides the SARIMA (Chapter 12), deseasonalized (Chapter 13) and PAR models (Sections 14.2.2, 14.3, 14.4, 14.6 and 14.8), the PPAR models (Sections 14.5, 14.6 and 14.8) are also employed in this forecasting experiment. This second forecasting study was originally presented by Thompstone (1983, Ch. 4).

In both sets of forecasting experiments, *one step ahead forecasts* are used for comparing the forecasting abilities of the model. There are two reasons for doing this. Firstly, from a theoretical viewpoint one can show that for the families of seasonal models presented in

Chapters 12 to 14, the one-step ahead forecasts are independent of one another. This property allows one to use *statistical tests* based upon the independence assumption to ascertain whether or not one model forecasts significantly better than another. Secondly, in many practical applications the one step ahead forecasts are of most importance to decision makers. For example, when deciding upon the operating rules of a reservoir for generation of hydro-electric power, an accurate forecast for the inflows of the next month is crucial. After the real value of next month's flows is known, one can use this information in the seasonal forecasting model to produce the one step ahead forecast for the subsequent month and so on.

Because different kinds of time series models are not defined and calibrated in exactly the same way, it is not surprising that their forecasts for a given time series are not identical. In fact, a given type of model approaches forecasting from a unique perspective based upon its own particular strengths and weaknesses. To attempt to exploit the forecasting capabilities of each kind of model fitted to a time series, forecasts generated by individual models can be combined in an optional manner. Procedures for combining forecasts across models are presented in Section 15.5.2. Additionally, experimental results on combining forecasts for SARIMA and PAR models fitted to average monthly riverflows are given in Section 15.5.3, while findings on combining hydrological forecasts from transfer function-noise (TFN), PAR and conceptual models are described in Section 18.4.2.

Before the conclusions, a brief discussion is given in Section 15.6 on aggregating forecasts for the purpose of producing a forecast for a longer time interval. For instance, one can employ a monthly model to produce 12 monthly forecasts and then sum these 12 values to obtain the aggregated annual forecast.

For a summary of where material on forecasting is presented in the book, the reader can refer to Table 1.6.3. In particular, the table points out that forecasting with nonseasonal ARMA and TFN models is described in Chapters 8 and 18, respectively. Finally, for references on forecasting listed outside of this chapter, the reader may wish to refer to appropriate references given at the end of Chapter 1 as well as Chapters 8 and 18.

15.2 CALCULATING FORECASTS FOR SEASONAL MODELS

15.2.1 Introduction

Suppose that one fits an appropriate seasonal model to a seasonal time series and then wishes to forecast l steps ahead where $l \geq 1$. When using z_t to represent the value of the time series, as is done in Chapter 12 with SARIMA models, one can employ the calibrated seasonal model to forecast z_{t+l} given the observations up to and including time t. As explained in Section 8.2 for nonseasonal ARMA models, the *minimum mean square error (MMSE) forecast* $\hat{z}_t(l)$ for z_{t+l} can be obtained by minimizing $E[z_{t+l} - \hat{z}_t(l)]^2$. This minimization is equivalent to taking the conditional expectation of z_{t+l} at time t.

For the deseasonalized and periodic models of Chapters 13 and 14, respectively, it is convenient to let $z_{r,m}$ stand for the observation in year r and season m. Then $\hat{z}_{r,m}(l)$ represents the MMSE forecast for lead time $l \geq 1$ starting at $z_{r,m}$.

The general approach for calculating MMSE forecasts for the seasonal models of Part VI is very similar to that used for nonseasonal ARMA models in Section 8.2. The specific method for calculating MMSE forecasts for each of the seasonal models is described below.

15.2.2 Forecasting with SARIMA Models

The SARIMA model is defined in [12.2.7]. The most convenient format to employ when calculating MMSE forecasts is the generalized form of the SARIMA model given in [12.2.12]. More specifically, to calculate the conditional expectation of z_{t+l} at time t and, hence, the MMSE forecast $\hat{z}_t(l)$, one takes conditional expectations of [12.2.12] to obtain

$$[z_{t+l}] = \phi'_1[z_{t+l-1}] + \phi'_2[z_{t+l-2}] + \cdots + \phi'_{p+sP+d+sD}[z_{t+l-p-sP-d-sD}] + [a_{t+l}] - \theta'_1[a_{t+l-1}] - \theta'_2[a_{t+l-2}] - \cdots - \theta'_{q+sQ}[a_{t+l-q-sQ}] \quad [15.2.1]$$

where

$l = 1,2, \ldots,$ is the lead time for the forecast,

$[z_{t+l}]$ denotes the conditional expectation

$\underset{t}{E}[z_{t+l} | z_t, z_{t-1}, \cdots];$

ϕ'_i is the generalized AR parameter defined by

$\phi'(B) = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D$; and

θ'_i is the generalized MA parameter defined by

$\theta'(B) = \theta(B)\Theta(B^s)$.

The nonseasonal version of [15.2.1] is given in [8.2.22]. As explained in Section 8.2.4 for forecasting with a nonseasonal ARMA model, one can allow for a nonzero deterministic trend component by introducing the parameter θ_0 on the right hand side of [8.2.21] to obtain [8.2.23]. By taking conditional expectations of [8.2.23], one obtains [8.2.24] for calculating MMSE forecasts for a nonseasonal ARMA model containing the level parameter θ_0. In a similar fashion for a SARIMA model, one can introduce the parameter θ_0 on the right hand side of [12.2.12] and then take conditional expectations to obtain a formula for calculating MMSE forecasts. The resulting formula would be the same as [15.2.1] expect for the parameter θ_0 which would be added to the right hand side.

As is also done in Section 8.2.4 for nonseasonal ARMA models, the conditional expectations in [15.2.1] can be determined using the following four rules:

(1) $$\underset{t}{E}[z_{t-j}] = z_{t-j}, \; j = 0,1,2, \ldots, \quad [15.2.2]$$

(2) $$\underset{t}{E}[z_{t+j}] = \hat{z}_t(j), \; j = 0,1,2, \ldots, \quad [15.2.3]$$

is the MMSE forecast for lead time j,

(3) $$\underset{t}{E}[a_{t-j}] = a_{t-j},\ j = 0,1,2,\ldots, \tag{15.2.4}$$

and

(4) $$\underset{t}{E}[a_{t+j}] = 0,\ j = 1,2,\ldots . \tag{15.2.5}$$

If the series contains a level represented by θ_0, this can be added to the forecasts obtained using the above rules.

The MMSE forecasts have a number of interesting properties which can be illustrated using the random shock form of the model in [12.2.15]. The forecast at time t for lead time l is

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots \tag{15.2.6}$$

Subtracting this from z_{t+l}, the forecast error is

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \tag{15.2.7}$$

Since $E[e_t(l)] = 0$, the variance of the forecast error is

$$V(l) = [Var\ e_t(l)] = [1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2]\sigma_a^2 \tag{15.2.8}$$

This variance can be utilized to estimate confidence intervals for forecasts at various lead times.

The one step ahead forecast error is

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1} \tag{15.2.9}$$

Although one step ahead forecast errors are statistically independent, forecast errors for lead times greater than one are correlated. For forecasts made from origin t, the correlation coefficient between forecast errors at lead times l and $l + j$ is given as (Box and Jenkins, 1976)

$$\rho[e_t(l), e_t(l+j)] = \frac{\sum_{i=0}^{l-1} \psi_i \psi_{j+i}}{\left\{ \sum_{h=0}^{l-1} \psi_h^2 \sum_{g=0}^{l+j-1} \psi_g^2 \right\}^{1/2}} \tag{15.2.10}$$

Inverse Box-Cox Transformation

Often the given series, z_t, is first transformed using the Box-Cox transformation in [12.2.1] to obtain the $z_t^{(\lambda)}$ series. The SARIMA model is then fitted to the $z_t^{(\lambda)}$ series as in [12.2.7]. The above calculations for obtaining MMSE forecasts are then carried out for the $z_t^{(\lambda)}$ series rather than z_t.

A naive approach for obtaining forecasts in the untransformed domain is to take the inverse Box-Cox transformation of the MMSE forecasts calculated in the transformed domain. However, in order to produce MMSE forecasts in the untransformed domain, a modified type of inverse Box-Cox transformation must be employed. More specifically, the exact MMSE forecast in the untransformed domain is determined from the fact that its transformed value follows a normal distribution with expected value $\hat{z}_t^{(\lambda)}(l)$ and variance $V(l)$. The expected value of the inverse Box-Cox transformed value is the desired MMSE forecast and it is determined numerically by Hermite polynomial integration (Granger and Newbold, 1976). In practice, it is found that the MMSE forecasts are slightly smaller than the naive forecasts. Moreover, studies with real data have shown that these MMSE forecasts do perform better than the naive forecasts. When data are transformed using a natural logarithmic transformation, as is often the case for seasonal hydrological time series, the MMSE forecast for the untransformed data is

$$\hat{z}_t(l) = exp[\hat{z}_t^{(\lambda)}(l) + \frac{V(l)}{2}] - c \ , \quad l = 1,2,\ldots, \qquad [15.2.11]$$

where $\hat{z}_t(l)$ is the MMSE forecast in the untransformed domain, $\hat{z}_t^{(\lambda)}(l)$ is the MMSE forecast produced by the model for the transformed logarithmic data, $V(l)$ is the variance of the forecast error given in [15.2.8], and c is the constant in the Box-Cox transformation required to make all entries be greater than zero.

For graphs of forecasts obtained using SARIMA models fitted to seasonal time series, the reader can refer to Section 12.5. In particular, Figures 12.5.1 and 12.5.2 display MMSE forecasts for monthly water demands and concentrations of atmospheric CO_2, respectively.

15.2.3 Forecasting with Deseasonalized Models

The main steps involved in forecasting with deseasonalized models are displayed in Figure 13.5.1. Firstly, one must calculate the MMSE forecasts for the ARMA model fitted to the deseasonalized series. This procedure is identical to that presented for the nonseasonal ARMA(p,q) model in Section 8.2. Let the deseasonalized series that is determined using either [13.2.2] or [13.2.3] be represented as $w_{r,m}$, where r and m stand for the year and season, respectively. By taking conditional expectations of the ARMA(p,q) model in [13.2.12], the MMSE forecasts for the deseasonalized series are calculated using

$$[w_{r,m+l}] = \phi_1[w_{r,m+l-1}]+\phi_2[w_{r,m+l-2}] + \cdots + \phi_p[w_{r,m+l-p}] + [a_{r,m+l}]$$
$$- \theta_1[a_{r,m+l-1}] - \theta_2[a_{r,m+l-2}] - \cdots - \theta_q[a_{r,m+l-q}] \qquad [15.2.12]$$

where

$l = 1,2,\ldots,$ is the lead time for the forecast, and

$[w_{r,m+l}]$ denotes the conditional expectation

$$\underset{t}{E}[w_{r,m+l} | w_{r,m}, w_{r,m-1}, \ldots,] .$$

Equation [15.2.12] can be used to calculate MMSE forecasts for the deseasonalized series by following the four rules given below for $l = 1,2,\ldots,$

(1) $$\underset{t}{E}[w_{r,m-j}] = w_{r,m-j}, \ j = 0,1,2,\ldots, \qquad [15.2.13]$$

(2) $$\underset{t}{E}[w_{r,m+j}] = \hat{w}_{r,m}(j), \ j = 1,2,\ldots, \qquad [15.2.14]$$

is the MMSE forecast for $w_{r,m+j}$,

(3)
$$\underset{t}{E}[a_{r,m-j}] = a_{r,m-j},\ j = 0,1,2,\ldots, \quad [15.2.15]$$

and

(4)
$$\underset{t}{E}[a_{r,m+j}] = 0,\ j = 1,2,\ldots . \quad [15.2.16]$$

When using the above rules, one should keep in mind that the time of occurrence of the deseasonalized series or the innovations can be written using a variety of equivalent subscripts. For instance, when there are s seasons per year $w_{r,m}$, $w_{r-1,m+s}$ and $w_{r+1,m-s}$ all stand for the same value.

Following the procedure described in Section 3.4.3, the random shock coefficient, ψ_i, $i = 1,2,\ldots,$ can be found for the ARMA(p,q) model describing the $w_{r,m}$ series. The variance of the forecast error for the deseasonalized series can then be determined as

$$V(l) = [1 + \psi_i^2 + \psi_2^2 + \cdots + \psi_{l-1}^2] \quad [15.2.17]$$

Finally, the one step ahead forecast error is

$$e_t(1) = w_{r,m+1} - \hat{w}_{r,m}(1) = a_{r,m+1} \quad [15.2.18]$$

As indicated in Figure 13.5.1, the next step is to route the MMSE forecasts through the inverse deseasonalization filter to obtain $z_{r,m}^{(\lambda)}$. The inverse deseasonalization for the two techniques given in [13.2.2] and [13.2.3] are

$$\hat{z}_{r,m}^{(\lambda)}(l) = \hat{w}_{r,m}(l) + \bar{\mu}_m \quad [15.2.19]$$

and

$$\hat{z}_{r,m}^{(\lambda)}(l) = \hat{w}_{r,m}(l)\bar{\sigma}_m + \bar{\mu}_m, \quad [15.2.20]$$

respectively. To obtain forecasts in the untransformed domain one must take the inverse Box-Cox transformation of $\hat{z}_{r,m}^{(\lambda)}(l)$. However, as noted in the previous subsection, if one wishes to have MMSE forecasts in the untransformed domain, one must make an appropriate adjustment before taking the inverse Box-Cox transformation. For the case of a logarithmic transformation, the MMSE forecast given in the same units as the original series is determined using

$$\hat{z}_{r,m}(l) = \exp[\hat{z}_{r,m}^{(\lambda)}(l) + \frac{1}{2}V(l)] - c \quad [15.2.21]$$

where $V(l)$ is the variance of the forecast error from [15.2.17].

15.2.4 Forecasting with Periodic Models

In Section 8.2, it is explained how to calculate MMSE forecasts for a nonseasonal ARMA model. A similar procedure is followed when forecasting with PAR, PPAR or PARMA models. For example, when calculating MMSE forecasts for a PAR model, one simply writes down the difference equation for season m in [14.2.1] and then determines the conditional expectations of the observations and innovations to arrive at the MMSE forecasts. Likewise, for a PARMA model, one uses the difference equation in [14.2.15] for the ARMA model in season m and then calculates the conditional expectations.

The approach for calculating MMSE forecasts for PARMA models is explained first. Assuming that the observations and innovations are known up to the rth year and mth season, one takes the conditional expectation of [14.2.15] to obtain

$$[z_{r,m+l}^{(\lambda)}] = \phi_1^{(m)}[z_{r,m+l-1}^{(\lambda)}] + \phi_2^{(m)}[z_{r,m+l-2}^{(\lambda)}] + \cdots + \phi_{p_m}^{(m)}[z_{r,m+l-p_m}^{(\lambda)}] + [a_{r,m}]$$
$$- \theta_1^{(m)}[a_{r,m+l-1}] - \theta_2^{(m)}[a_{r,m+l-1}] - \cdots - \theta_{q_m}^{(m)}[a_{a_{r,m+l-q_m}}] \quad [15.2.22]$$

where

$l = 1,2, \ldots,$ is the lead time for the forecast, and

$[z_{r,m+l}^{(\lambda)}]$ denotes the conditional expectation

$$\underset{t}{E}[z_{r,m+l}^{(\lambda)} \mid z_{r,m}^{(\lambda)}, z_{r,m-1}^{(\lambda)}, \ldots]$$

By following the four rules listed below, equation [15.2.22] can be employed for calculating the MMSE forecasts for $z_{r,m}^{(\lambda)}$ for lead times $l = 1,2, \ldots,$.

(1) $$\underset{t}{E}[z_{r,m-j}^{(\lambda)}] = z_{r,m-j}^{(\lambda)}, \; j = 0,1,2, \ldots, \quad [15.2.23]$$

(2) $$\underset{t}{E}[z_{r,m+j}^{(\lambda)}] = \hat{z}_{r,m}^{(\lambda)}(j), \; j = 1,2, \ldots, \quad [15.2.24]$$

is the MMSE forecast for $z_{r,m+j}^{(\lambda)}$,

(3) $$\underset{t}{E}[a_{r,m-j}] = a_{r,m-j}, \; j = 0,1,2, \ldots, \text{ and} \quad [15.2.25]$$

(4) $$\underset{t}{E}[a_{r,m+j}] = 0, \; j = 1,2, \ldots, \; . \quad [15.2.26]$$

After calculating the forecasts for $l = 1,2, \ldots,$ the appropriate monthly mean μ_m must be added to each forecast when $\mu_m \neq 0$.

The procedure in Section 3.4.3 can be utilized to find the random shock coefficients $\psi_i^{(m)}$, $i = 1,2, \ldots,$ for the ARMA model in season m. To calculate the variance of the forecast error for $\hat{z}_{r,m}^{(\lambda)}(l)$ one uses

$$V^{(m)}(l) = [1 + \psi_1^{(m)^2} + \psi_2^{(m)^2} + \cdots + \psi_{l-1}^{(m)^2}]\sigma_m^2 \quad [15.2.27]$$

The one step ahead forecast error can be shown to be

$$e_t^{(m)}(1) = z_{r,m+1}^{(\lambda)} - \hat{z}_{r,m}^{(\lambda)}(1) = a_{r,m+1} \quad [15.2.28]$$

To obtain forecasts in the same units as the original series, one must take the inverse Box-Cox transformation of $\hat{z}_{r,m}^{(\lambda)}(l)$ for $l = 1,2, \ldots,$. When the data are transformed using natural logarithms (i.e., $\lambda = 0$), equation [15.2.21] can be utilized to calculate the MMSE forecasts in the untransformed domain where $V(l)$ is replaced by $V^{(m)}(l)$ from [15.2.27].

When determining forecasts for PAR or PPAR models, one can follow the approach explained for PARMA models. Consider, for example, the case of the PAR model. By taking conditional expectations of [14.2.1] or [14.2.3], the MMSE forecasts calculated after year r and season m are determined using

$$[z_{r,m+l}^{(\lambda)}] = \phi_1^{(m)}[z_{r,m+l-1}^{(\lambda)}] + \phi_2^{(m)}[z_{r,m+l-2}^{(\lambda)}] + \cdots + \phi_{p_m}^{(m)}[z_{r,m+l-p_m}^{(\lambda)}] + [a_{r,m}] \quad [15.2.29]$$

The four rules presented in [15.2.23] to [15.2.26] can then be used to calculate the MMSE forecasts for the transformed series. Additionally, when $\mu_m \neq 0$, one must add the appropriate mean level to each of the calculated forecasts. Finally, the modified version of the inverse Box-Cox transformation (see [15.2.21] for the case of $\lambda = 0$) must be taken to produce MMSE forecasts in the untransformed domain.

15.3 FORECASTING MONTHLY RIVERFLOW TIME SERIES

15.3.1 Introduction

To examine the efficacy of PAR models of Chapter 14, a comprehensive forecasting study is carried out by comparing their performance with that of several models used to model seasonal data. Using thirty monthly riverflow time series, the PAR models are compared to the SARIMA models of Chapter 12 as well as the deseasonalized models presented in Chapter 13. Methods of model order selection for the PAR models are also compared. The experiments described in this section, as well as by Noakes et al. (1985), are the most comprehensive yet reported in the hydrological literature. Other published comparisons have used only a few series and usually only two models [see, for example, Delleur et al. (1976)]. Also, the majority of the hydrological forecasting research to date has been concentrated on shorter time intervals in the order of a few hours or days [see, for example, the Proceedings of the Oxford Hydrological Forecasting Symposium, April 15-18 (International Association of Hydrological Sciences, 1980) and Thompstone et al. (1983)]. However, monthly riverflow forecasts are often used for operational planning of reservoir systems. Camacho (1990) considers both short term and long term forecasts in his riverflow forecasting study. Even modest improvements in the operation of large reservoir systems can result in multi-million dollar savings per year (see, for instance, Brocha (1978) as well as the comments on stochastic hydrology given in Section 1.1). Thus, the results of the forecasting study given in this section should be important to those concerned with the optimal medium and long-term operation of reservoir systems.

The performance of the forecasts from the different seasonal models are assessed using the *root mean square error (RMSE), mean absolute deviation (MAD), mean absolute percentage error (MAPE),* and *median absolute percentage error (MEDIAN APE),* criteria. Although these criteria give an indication as to which models seem to perform better, no statement concerning statistically significant differences can be made from such a comparison. To address this question, the nonparametric Wilcoxon signed rank test (Wilcoxon, 1945) is used to determine if a particular model produces significantly better forecasts when compared to another model. One could also employ Pitman's (1939) correlation test and the likelihood ratio test to check if one model forecasts significantly better than another. These latter two tests are described in Section 8.3.2 and used in the forecasting experiments with nonseasonal models presented in Section 8.3.4. The nonparametric Wilcoxon test is outlined in this section with the seasonal forecasting experiments and described in detail in Appendix A23.2. Noakes et al. (1983) and Noakes

(1984) present the results of the forecasting study of this section when Pitman's correlation study and the likelihood ratio tests are used. Finally, the overall procedure for carrying out the forecasting experiments in this section, Section 15.4 as well as Sections 8.3 and 15.3, is summarized in Figure 8.3.1.

15.3.2 Data Sets

The data used in this study comprise thirty monthly unregulated riverflow time series ranging in length from thirty-seven to sixty-eight years. The rivers are from a number of different physiographic regions and vary in size from a river with a mean annual flow of one cubic meter per second (m^3/s) to a river having a mean annual flow of almost 900 m^3/s. The data for the Canadian rivers were obtained from Water Survey of Canada records, the American riverflow series are from the United States Geological Survey, and the Brazilian data were kindly provided from Electrobras (the national electrical company of Brazil). The rivers and their mean annual flows for the water year from October to September are displayed in Table 15.3.1.

15.3.3 Seasonal Models

The last three years or 36 observations are omitted from each of the data sets in Table 15.3.1. After taking natural logarithms of the time series, SARIMA, deseasonalized and PAR models are fitted to the thirty truncated logarithmic series.

The most appropriate SARIMA models to fit to the series are identified using the graphical procedures of Section 12.3.2. All of the SARIMA models identified for fitting to the monthly riverflow series in Table 15.3.1 are determined to be of the form $(p,0,q)\times(0,1,Q)_{12}$ with $\lambda = 0$ and with typical values of p, q and Q being 1, 0 and 1.

Two types of deseasonalized models are used in the forecasting study. For the first kind of model, equation [13.2.2] is used to deseasonalize the logarithmic data after estimating each monthly mean of the logarithmic data using [13.2.4]. The most appropriate ARMA model is then fitted to this deseasonalized series using the model construction techniques of Part III. This overall deseasonalized model is referred to as DSM.

For the second type of deseasonalized model, equation [13.2.3] is used to deseasonalize the logarithmic series before fitting an ARMA model to the resulting nonseasonal series. In [13.2.3], the seasonal means and standard deviations are estimated using [13.2.4] and [13.2.5], respectively. This overall deseasonalized model is called DES.

Six types of PAR models are considered in this study. In the first model, a separate AR(1) model is fitted to each month (called PAR/1) using multiple linear regression. This model was originally suggested by Thomas and Fiering (1962) and has been used extensively by hydrologists.

The second and third PAR models are fitted to the data using the algorithm of Morgan and Tatar (1972) described in Section 14.3.3. This algorithm calculates the residual sum of squares of all possible regressions for each season. The AIC and BIC can thus be calculated for all possible models. The PAR model which gives the minimum value of the AIC in [14.3.8] or BIC in [6.3.5] (with $p_m \leq 12$) is selected as the most appropriate. This type of procedure has been called subset autoregression by McClave (1975), and thus is referred to as SUBSET/AIC or SUBSET/BIC modelling.

Table 15.3.1. Average monthly riverflow time series used in the forecasting experiments.

	River	Location	Period	Observations	Mean Flow (m^3/s)
1	American	Fair Oaks, California	1906-1960	660	106
2	Boise	Twin Springs, Idaho	1912-1960	588	33
3	Clearwater	Kamish, Idaho	1911-1960	600	231
4	Columbia	Nicholson, British Columbia	1933-1969	444	109
5	Current	Van Buren, Missouri	1922-1960	468	54
6	W.B. Delaware	Hale Eddy, New York	1916-1960	540	30
7	English	Sioux Lookout, Ontario	1922-1977	660	123
8	Feather	Oroville, California	1902-1977	708	167
9	James	Buchanan, Virginia	1911-1960	600	69
10	Judith	Utica, Montana	1920-1960	492	1
11	Mad	Springfield, Ohio	1915-1960	552	14
12	Madison	West Yellowstone, Montana	1923-1960	456	13
13	McKenzie	McKenzie Bridge, Oregon	1911-1960	600	47
14	Middle Boulder	Nederland, Colorado	1912-1960	588	2
15	Missinaibi	Mattice, Ontario	1921-1976	672	103
16	Namakan	Lac La Croix, Ontario	1923-1977	648	108
17	Neches	Rockland, Texas	1914-1960	564	69
18	N. Magnetawan	Burke Falls, Ontario	1916-1977	732	6
19	Oostanaula	Resaca, Georgia	1893-1960	816	78
20	Pigeon	Middle Falls, Ontario	1924-1977	636	14
21	Rappahannock	Fredericksburg, Virginia	1908-1971	768	45
22	Richelieu	Fryers Rapids, Quebec	1932-1977	468	331
23	Rio Grande	Furnas, Minas Gerais, Brazil	1931-1978	576	896
24	Saint Johns	Fort Kent, New Brunswick	1927-1977	600	30
25	Saugeen	Walkerton, Ontario	1915-1976	744	68
26	S.F. Skykomish	Index, Washington	1923-1960	456	278
27	S. Saskatchewan	Saskatoon, Saskatchewan	1911-1963	624	272
28	Trinity	Lewiston, California	1912-1960	588	47
29	Turtle	Mine Centre, Ontario	1921-1977	672	37
30	Wolf	New London, Wisconsin	1914-1960	564	49

The next PAR models are estimated by using the appropriate Yule-Walker equations (see Section 14.3.3). In the first case p_m is selected on the basis of the minimum value of the AIC or BIC. Unlike the previous case, however, intermediate parameters are not allowed to be constrained to zero. Thus, all of the parameters from $\phi_1^{(m)}$ to $\phi_{p_m}^{(m)}$ are estimated in this model for a given season to produce the PAR/AIC and PAR/BIC models.

The last PAR models are identified by examining plots of the sample periodic PACF, presented in Section 14.3.2. In general, an AR(p_m) model is fitted to month m, where p_m is the last lag for which the PACF is significantly different from zero. The adequacy of the selected model is checked by testing for significant residual correlation or non-normality. Thus, the PAR/PACF is the natural extension to PAR models of the modelling philosophy recommended by Box and Jenkins (1976) and adhered to in this book. Once again, no intermediate parameters are constrained to zero.

15.3.4 Forecasting Study

After omitting the last 36 values of each of the 30 average monthly riverflow series in Table 15.3.1, the nine seasonal models are fitted to the 30 truncated series. From Section 15.3.3, these nine models are labelled as the SARIMA, DSM, DES, SUBSET/AIC, SUBSET/BIC, PAR/AIC, PAR/BIC, PAR/1, and PAR/PACF models. The nine models are then used to generate thirty-six one-step-ahead forecasts for the logarithmic flows. Figure 15.3.1 shows a time series plot of the last five years of the logarithmic flows along with the forecasts for the last three years using the PAR/PACF method for river number 14 in Table 15.3.1. As can be seen from a visual viewpoint, the PAR/PACF model forecasts quite well.

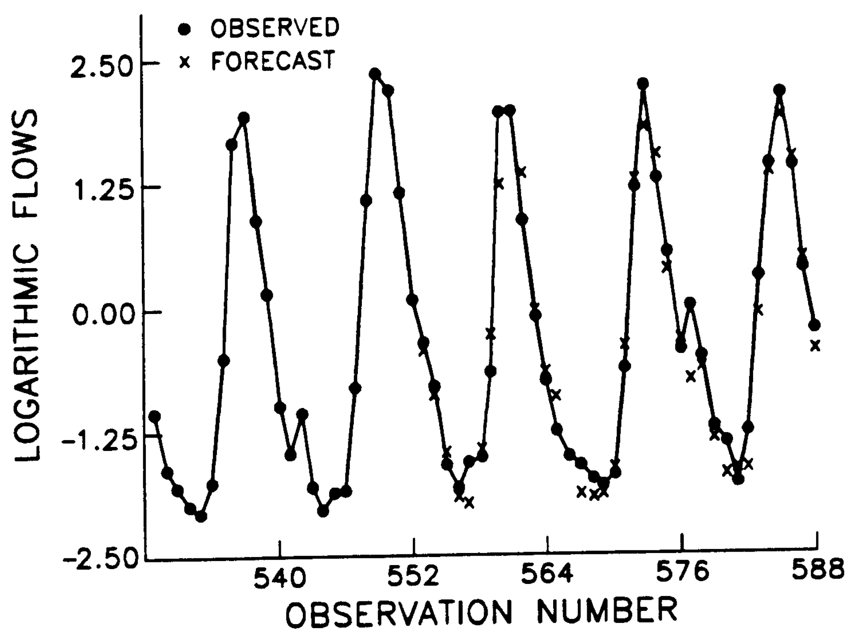

Figure 15.3.1. Logarithmic monthly flows and one step ahead PAR/PACF forecasts for the Middle Boulder Creek.

The monthly means of the logarithmic flows are also considered as forecasts and are referred to as MEANS. The logarithmic forecast errors associated with each of the ten forecasting models are then compared using the forecast performance measures RMSE, MAD, MAPE and MEDIAN APE, mentioned in Section 15.3.1.

RMSE results are given in Table 15.3.2 for each river. The results for each performance measure are summarized in Tables 15.3.3 to 15.3.6 where rank and rank-sum comparisons appear.

Table 15.3.2. RMSE × 1000 of the logarithmic forecast errors.

River	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	857	896	813	864	796	796	801	907	690	1240
2	280	279	273	280	307	289	264	289	273	248
3	323	330	334	331	346	330	359	339	367	544
4	183	190	180	198	204	211	184	181	182	209
5	426	418	445	410	464	423	389	408	390	357
6	658	642	628	666	681	664	689	690	698	775
7	191	218	187	203	218	201	209	205	440	633
8	337	338	394	338	415	335	354	347	358	481
9	516	495	536	544	562	548	489	489	488	579
10	470	469	463	469	500	471	582	427	576	746
11	435	428	416	431	481	440	426	431	424	539
12	98	91	120	90	125	98	98	118	107	127
13	175	175	208	176	254	221	167	171	169	186
14	273	273	272	274	281	296	290	290	302	365
15	619	614	604	618	634	626	707	639	752	961
16	242	244	238	243	248	238	253	259	261	515
17	909	909	930	910	1078	906	916	907	969	1147
18	407	407	416	407	419	407	408	411	419	440
19	424	418	425	420	427	425	448	447	446	487
20	600	591	592	604	627	618	673	707	694	1118
21	530	546	536	547	570	535	553	552	564	569
22	250	266	264	270	326	274	277	270	260	600
23	230	226	265	229	294	241	241	236	242	335
24	411	412	398	420	414	428	389	385	398	379
25	425	402	430	421	479	422	433	423	432	532
26	380	391	407	422	434	401	411	416	411	476
27	436	438	420	379	500	391	464	445	461	587
28	626	624	624	633	603	632	628	639	627	822
29	282	283	282	283	318	283	283	301	297	410
30	355	358	408	367	368	372	352	361	352	465

Table 15.3.3. RMSE of one-step MMSE forecasts of logged series (number of times each method has indicated rank).

Rank	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	4	3	7	3	1	4	1	1	3	3
2	3	5	5	2	0	5	4	3	3	0
3	10	2	3	4	1	4	2	2	2	0
4	3	11	0	7	0	2	2	3	2	0
5	5	3	5	6	1	3	3	3	1	0
6	3	3	2	2	5	1	7	6	1	0
7	1	1	2	4	4	3	4	3	7	1
8	1	2	4	1	3	5	5	5	4	0
9	0	0	2	1	11	1	2	4	7	2
10	0	0	0	0	4	2	0	0	0	24
Rank-sum	110	119	127	134	230	145	166	173	178	268

Table 15.3.4. MAD of one-step MMSE forecasts of logged series (number of times each method has indicated rank).

Rank	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	4	4	4	1	1	5	1	2	5	3
2	6	4	4	4	1	4	3	3	1	0
3	5	8	5	3	1	4	2	2	0	0
4	6	6	2	8	1	4	1	1	1	0
5	6	2	4	6	0	2	5	1	4	0
6	2	3	3	2	6	3	4	4	2	1
7	0	1	2	5	3	3	6	6	4	0
8	1	2	4	1	3	2	6	3	8	0
9	0	0	2	0	10	3	1	8	5	1
10	0	0	0	0	4	0	1	0	0	25
Rank-sum	105	111	137	135	221	133	175	185	180	268

Table 15.3.5. MAPE of one-step MMSE forecasts of logged series (number of times each method has indicated rank).

Rank	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	3	5	3	1	3	5	1	1	5	3
2	5	4	3	5	2	3	2	5	1	0
3	4	7	4	4	1	3	5	1	0	1
4	7	2	5	7	0	3	3	1	2	0
5	7	6	4	2	1	3	1	2	4	0
6	2	2	1	5	1	4	6	5	4	0
7	1	1	2	5	2	3	5	7	3	1
8	1	3	6	1	3	2	6	2	4	2
9	0	0	2	0	11	4	0	6	7	0
10	0	0	0	0	6	0	1	0	0	23
Rank-sum	115	115	147	134	218	144	166	177	175	259

Table 15.3.6. MEDIAN APE of one-step MMSE forecasts of logged series (number of times each method has indicated rank).

Rank	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	5	1	3	1	6	4	2	1	3	4
2	3	3	5	4	4	2	1	3	4	1
3	4	5	6	2	0	3	3	4	2	1
4	6	4	3	6	2	3	2	3	0	1
5	4	5	5	6	2	1	3	2	2	0
6	3	3	1	3	4	6	3	2	4	1
7	3	6	2	2	2	6	2	5	2	0
8	1	1	2	3	3	2	7	4	5	2
9	1	1	3	1	4	2	4	5	8	1
10	0	1	0	2	3	1	3	1	0	19
Rank-sum	123	150	131	154	160	156	190	175	177	234

The rank-sums for the models are the sums of the product of the rank and the associated table entry. Thus, models with lower rank-sums perform better than those with larger rank-sums. The models PAR/PACF, PAR/1, PAR/AIC, PAR/BIC, and SUBSET/BIC fare very well on the basis of all performance criteria. As expected, using the MEANS proves unsatisfactory in most cases. The MEANS has the worst overall performance and produces the largest RMSE for twenty-four of the series. Table 15.3.2 shows that in the three cases (rivers 2, 5, and 24) where the MEANS has the smallest RMSE there is very little difference between any of the forecasting methods. Moreover, in these three cases all methods have low MAPEs and RMSEs. At the other extreme, the best alternative to MEANS for rivers 7, 16, and 22 has a RMSE less than half that of MEANS. Next to the PAR models mentioned above, the DSM, DES, and SARIMA models perform about equally as well. The SUBSET/AIC model performance is disappointing, although not totally surprising. The large number of parameters associated with the SUBSET/AIC model does not provide a sufficiently parsimonious and flexible model for producing accurate forecasts. The importance of parsimony in forecasting models is discussed by Ledolter and Abraham (1981).

For several of the rivers, there are large discrepancies between the MAPE and MEDIAN APE criteria. This is found to be due to a defect in the absolute percentage error when the observed value is small. For example, the observed logged flow for river 14 for November, 1959, is 0.0024 and the PAR/PACF forecast is -0.746. This creates an absolute percentage error of over 31,000!.

The forecasting results reported thus far are for the logarithmic flows. To compare results in the untransformed domain, one converts the forecasts using [15.2.21]. Table 15.3.7 shows the forecasting findings for the RMSE of one-step ahead MMSE forecasts for the untransformed time series. Once again, the same PAR models perform the best. However, there are some differences in the untransformed and transformed forecasting results. In particular, notice the improvement of the MEANS model and the poor performance of the DES model. The DSM and SARIMA models still perform reasonably well and the SUBSET/AIC improves slightly.

Table 15.3.7. RMSE of one-step MMSE forecasts of the flows
(number of times each method has indicated rank).

Rank	PAR/ PACF	PAR/1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	DSM	DES	SARIMA	MEANS
1	2	5	7	0	4	4	3	0	2	3
2	5	4	5	6	0	3	3	0	4	0
3	11	3	3	5	0	3	2	0	1	2
4	6	6	4	5	1	5	1	1	0	1
5	1	8	2	8	2	6	0	0	3	0
6	4	3	3	3	5	1	6	0	1	4
7	1	0	4	1	2	5	3	5	4	5
8	0	1	1	1	5	2	7	5	4	4
9	0	0	1	1	10	1	4	1	8	4
10	0	0	0	0	1	0	1	18	3	7
Rank-sum	105	112	115	129	202	135	178	268	196	210

The *Wilcoxon signed rank test* (Wilcoxon, 1945) for paired data is used to test for statistically significant differences in the forecasting ability of the various procedures. In this test, which is also described in Appendix A23.2, the differences in the squares of the logarithmic forecast errors are computed. These differences are ranked in ascending order, without regard to sign, and assigned ranks from one to thirty-six. The sum of the ranks of all positive differences are then computed as T in [A23.2.3] and compared to tabulated values in order to ascertain if the forecasts from one model are significantly better than the forecasts from a competing model. These results are then used to examine the performance of the models across all thirty series. In this test, the $P-$ value associated with each T value is calculated by estimating the area in the tail of the distribution. Then, the Fisher (1970, p. 99) method for *combining significance levels* for one-sided tests is

$$-2\sum_{i=1}^{k}\ln(p_i) \approx \chi^2_{2k}, \qquad [15.3.1]$$

where p is the calculated P-value associated with T and k is the number of series considered in the test. This combination technique generally has greater power than alternative methods such as simply summing the T's.

Fisher's test is employed to compare the overall performance of the PAR/PACF model to that of the other competing models. In addition, the PAR/1 parameters are also estimated using the Yule-Walker equations to provide an additional model for comparison (PAR/YW1). In this way, identical forecasts produced by the PAR/PACF and PAR/YW1 models could be ignored, ensuring that only the differences in the forecasting procedures are compared. The results of Fisher's test are presented in Table 15.3.8. The PAR/PACF model is significantly better than all of the models except the PAR/1 and the PAR/AIC at the five-percent level. Since different estimation procedures are employed for the PAR/PACF and PAR/1 models, there are several forecasts that are almost, but not quite, identical. These are all included in the analysis, thus masking the differences in the performance of the two models. The PAR/YW1 model, however, employs the same estimation procedure, thus resulting in identical forecasts when an AR(1) model is identified for a particular month for the PAR/PACF model. This allows ties to be dropped from consideration, and results in the testing of only the differences between the two models. All series with fewer than five untied forecasts are dropped from consideration in this test. The results of this comparison indicates that when ties are ignored, the PAR/PACF model is better than the PAR/YW1 model at the two-percent level of significance. Although the PAR/AIC compares quite favourably with the PAR/PACF when the significance levels are combined, detailed examination of the results reveal that for three rivers the PAR/PACF forecasts significantly better at the five-percent level than the PAR/AIC. However, in no case are the PAR/AIC forecasts significantly superior to those of the PAR/PACF. Additional details are given in the thesis of Noakes (1984).

15.4 FORECASTING QUARTER-MONTHLY AND MONTHLY RIVERFLOWS

15.4.1 Introduction

The results of the forecasting study of Section 15.3 indicate that certain types of PAR models work better than other competing seasonal models when forecasting average monthly riverflow time series. In particular, PAR models identified using the sample periodic PACF

Table 15.3.8. Results of Fisher's test for the Wilcoxon tests when each model is compared to the PAR/PACF model.

Model	PAR/1	PAR/ YW1	PAR/ AIC	PAR/ BIC	SUBSET/ AIC	SUBSET/ BIC	ARMA/ DSM	ARMA/ DES	SARIMA	MEANS
χ^2	65.2	60.5	64.8	57.8	116.0	87.8	99.4	101.2	113.3	276
DF	60	40	58	40	60	60	60	60	60	60
SL (%)	30	2	25	3	10^{-4}	1	0.1	0.05	10^{-5}	10^{-13}

provide the most accurate forecasts. Because PPAR models are not used in the forecasting experiments of the previous section, one of the objectives of the forecasting study presented in this section as well as by Thompstone (1983) and Thompstone et al. (1985) is to show a forecasting study involving PPAR models, as well as other types of seasonal models. A second goal is to perform forecasting experiments with both quarter-monthly and monthly time series.

15.4.2 Time Series

The data sets used in this forecasting study are identical to those utilized in the seasonal modelling applications of Section 14.6 and the simulation experiments with seasonal models in Section 14.8.2. In particular, the time series consist of both the quarter-monthly and monthly flows of the rivers called the Alcan system, Rio Grande and Saugeen. For all of these series, the last three years on record are not used when the seasonal models given in Table 14.6.3 were fitted to the six series.

15.4.3 Seasonal Models

The seasonal models used in the forecasting experiments are those listed in Table 14.6.3. The seasonal models consist of the SARIMA (Chapter 12), deseasonalized (Chapter 13), PAR (Sections 14.2.2 and 14.3) and three types of PPAR (Section 14.5) models. The deseasonalized model called DES refers to the situation when the most appropriate ARMA model is fitted to a series fully deseasonalized using [13.2.3], for which the seasonal means and standard deviations are estimated by utilizing [13.2.4] and [13.2.5], respectively. In all cases, the logarithmic series are used and the fitted models are identical to those described in Section 14.6.

15.4.4 Forecasting Experiments

For each time series and fitted model given in Table 14.6.3, one step ahead forecasts are calculated for the additional three year period in each series. Besides these models, forecast errors are also calculated for a model labelled MEANS, which simply entails using the seasonal means of the logarithmic series as the one step ahead forecasts.

For each fitted model and logarithmic time series, one can calculate the *MSE (mean squared error)* of the forecast errors. In Table 15.4.1, the models are ranked according to their MSE's for each of the six time series. The lowest value of MSE is ranked as 1 whereas the highest number refers to the model which produces the least accurate forecasts for the series. Because the forecast errors are approximately normally distributed, they can be employed in the Pitman test (Pitman, 1939) described in Section 8.3 to determine if there are significant differences in the MSE's of the forecasts between any two seasonal models. Some representative results using the Pitman test are given in Table 15.4.2.

Table 15.4.1. MSE's of the one step ahead forecasts and ranking of the models according to MSE's.

Model	Quarter-monthly Series			Monthly Series		
	Alcan System	Rio Grande	Saugeen	Alcan System	Rio Grande	Saugeen
MEANS	0.4314 (7)	0.2831 (7)	0.6177 (7)	0.3476 (4)	0.3359 (7)	0.5712 (7)
SARIMA	0.2871 (6)	0.2293 (6)	0.4274 (6)	0.3325 (2)	0.2399 (6)	0.4883 (6)
DES	0.2634 (5)	0.2222 (3)	0.4117 (5)	0.3011 (1)	0.2381 (4)	0.4788 (5)
PAR	0.2575 (3)	0.2213 (1)	0.4070 (1)	0.3457 (3)	0.2301 (2)	0.4189 (2)
PPAR/50	0.2567 (2)	0.2235 (5)	0.4075 (2)		0.2289 (1)	0.4187 (1)
PPAR/20	0.2583 (4)	0.2221 (2)	0.4096 (3)		0.2321 (3)	0.4265 (4)
PPAR/05	0.2552 (1)	0.2227 (4)	0.4098 (4)		0.2381 (4)	0.4217 (3)

Note: The parenthetical figure ranks the MSE's for a given series from the lowest (1) to the highest (7).

Table 15.4.2. Pitman's correlation test statistics for comparing MSE's of one step ahead forecasts for seasonal models fitted to the quarter-monthly Saugeen riverflows.

	MEANS	SARIMA	DES	PAR	PPAR/50	PPAR/20	PPAR/05
MEANS		0.4476(-)	0.4735(-)	0.4814(-)	0.4831(-)	0.4799(-)	0.4792(-)
SARIMA*	0.4476(+)		0.1460(=)	0.1358(=)	0.1320(=)	0.1205(=)	0.1239(=)
ARMA/DES*	0.4735(+)	0.1460(=)		0.0504(=)	0.0424(=)	0.0201(=)	0.0195(=)
PAR*	0.4814(+)	0.1358(=)	0.0504(=)		0.0227(=)	0.0633(=)	0.0657(=)
PPAR/50*	0.4831(+)	0.1320(=)	0.0424(=)	0.0227(=)		0.0625*(=)	0.0524(=)
PPAR/20*	0.4799(+)	0.1205(=)	0.0201(=)	0.0633(=)	0.0625(=)		0.0024(=)
PPAR/05*	0.4792(+)	0.1239(=)	0.0195(=)	0.0657(=)	0.0524(=)	0.0024(=)	

(1) Table shows $|r|$ for Pitman's correlation test statistic.

(2) Difference in MSE's of forecasts is significant at 5% level if $|r| > 0.163$.

(3) A parenthetical = indicates the difference is not significant, a + indicates the row model is "better" than the column model (significant difference and smaller MSE), and a - indicates the row model is "worse" than the column model.

(4) * indicates the model is better or equal to all other models.

Consider the results for the MSE's in Table 15.4.1 for comparing the forecasting capabilities of the seasonal models. The simplistic MEANS model consistently provides the worst forecasts, and this confirms that the methods of time series analysis provide meaningful improvements in forecasting ability. In five of the six cases, SARIMA models provide the second largest forecast errors. This could lead to some doubts regarding the appropriateness of SARIMA models for forecasting the inflow series considered herein. As noted in Chapter 12 and elsewhere in Part VI, from a physical viewpoint SARIMA models are not well designed for modelling riverflow time series, like the one in Figure VI.1, because they cannot explicitly model stationarity within each season as well as a seasonally varying correlation structure.

In three cases (one quarter-monthly series and two monthly series), PPAR models provide the smallest MSE's of forecasts, while in two other cases (both quarter-monthly), PAR models produce the best forecasts. More generally, one sees that in five of the six cases, the smallest and second smallest MSE's are furnished by PAR or PPAR models; in four of the six cases, the four smallest MSE's are provided by PAR or PPAR models. All this suggests that PAR and PPAR models have appealing forecasting abilities for the series considered herein. In only one case, the Alcan system monthly inflow series, the DES provides the smallest MSE.

Table 15.4.2 shows the results of Pitman's correlation test for the case of the quarter-monthly flows of the Saugeen River. The statistic, $|r|$, for comparing MSE's between one step ahead forecasts for two seasonal models is described in Section 8.3.2. A parenthetical equal sign, (=), indicates that, at the 5% level, the difference between the row model errors and the column model errors is not significant. A parenthetical plus sign, (+), indicates that the row model provides significantly better forecasts than the column model, and a parenthetical negative sign, (-), indicates the contrary. An asterisk beside the label of the row model indicates that it provides forecasts which are, at the 5% level, equal to or better than forecasts from all other models considered. As is the situation in Table 15.4.2 and the results for the other 5 series which are not shown, in no case does a model furnish forecasts which are significantly better than forecasts from all other models.

In all three cases of quarter-monthly series, the DES, PAR, and PPAR/05 models give forecast errors which were statistically equivalent to or better than all other models. In two cases out of three, the SARIMA, PPAR/50 and PPAR/20 models are equal to or better than all other models with respect to their forecasting abilities. Only the MEANS model is, in all three cases, significantly worse than all other models.

Forecasting results for the monthly Alcan system inflows are inconclusive. No PPAR models are identified for this series, and there is no statistically significant difference in forecasts from the four other models. For the two other monthly series, forecasts are, at the 5% level of significance, indistinguishable for the SARIMA, DES, PAR, PPAR/50, PPAR/20 and PPAR/05 models. In the case of the monthly Rio Grande flows, the MEANS model is significantly worse than all other models, while for the monthly Saugeen riverflows, the MEANS model is indistinguishable from the SARIMA and DES models, but significantly worse than the others.

In summary, from the results of the Pitman test, it is difficult to conclude that, amongst the SARIMA, DES, PAR, and various PPAR models, one type of model is particularly outstanding with respect to its forecasting ability for the time series considered herein. However, it is interesting to recall that in three of the five cases for which they are identified, the PPAR models provide the smallest MSE's of forecasts (see Table 15.4.1), and in the other two cases it is the

PAR models which produces the best forecasts.

15.5 COMBINING FORECASTS ACROSS MODELS

15.5.1 Motivation

The selection of the "best" forecasting procedure is certainly a hopeful result of any forecasting study. Invariably, however, no one method will produce optimum forecasts in all cases. The task then becomes one of selecting the most appropriate forecasting procedure based upon the available information.

An alternative approach is to combine the forecasts from two or more procedures in accordance to their relative performances. In this way, it is hoped that the strengths of each method might be exploited. The successes achieved by combining economic forecasts are documented in several studies (Armstrong and Lusk, 1983; Bates and Granger, 1969; Bordley, 1982; Granger and Ramanathan, 1984; Makridakis et al., 1982; Newbold and Granger, 1974; Winkler and Makridakis, 1982). Within the field of water resources, McLeod et al. (1986) present experimental results on combining hydrologic forecasts which are also described in Sections 15.5.3 and 18.4.2 of this book.

In the next subsection, techniques for combining forecasts are given. Subsequently, in Section 15.5.3, seasonal riverflow forecasts generated using both PAR and SARIMA models are combined in an attempt to achieve improved forecasts. Within Section 18.4.2, seasonal riverflow forecasts from TFN, PAR and conceptual or physically based models are optimally combined in forecasting experiments.

15.5.2 Formulae for Combining Forecasts

There are certainly countless ways of combining forecasts from different forecasting procedures to arrive at a combined forecast. The simplest is probably to weight each forecast equally. If there are k forecasts available, the combined forecast f_c, would be

$$f_c = \sum_{i=1}^{k} w_i f_i \qquad [15.5.1]$$

where f_i is the forecast produced by the ith model, w_i is the weighting factor for the ith forecast and $w_i = w_j = 1/k$ for all i and j.

It would be expected that a better combination of forecasts could be obtained if the statistical properties of the forecast errors were considered. Winkler and Makridakis (1983) point out that if the covariance matrix of the forecast errors from k methods, Σ, is known, then the optimal weights are given by

$$w_i = \frac{\sum_{j=1}^{k} \alpha_{ij}}{\sum_{h=1}^{k}\sum_{j=1}^{k} \alpha_{hj}} \qquad [15.5.2]$$

where the α_{ij} terms are the elements of Σ^{-1}. In practice, Σ is not known and must be estimated. Estimates of the weights in [15.5.2] can be calculated from the inverse of $\hat{\Sigma}$ where

$$\hat{\Sigma}_{ij} = \mathrm{v}^{-1} \sum_{h=t-\mathrm{v}}^{t-1} e_h^{(i)} e_h^{(j)} \qquad [15.5.3]$$

$e_t^{(i)}$ is the percentage error for method i at time t and v is the number of previous forecast errors employed to calculate w_i.

In the study concerning the combination of economic forecasts by Winkler and Makridakis (1983), these authors found that estimating Σ^{-1} and calculating the weights using [15.5.2] gave the poorest results. One of the preferred procedures in their study was to ignore the correlation between the forecast errors. In this case, the forecast weights were calculated as

$$\hat{w}_i = \frac{\left[\sum_{h=t-\mathrm{v}}^{t-1} e_h^{(i)2}\right]^{-1}}{\sum_{j=1}^{p}\left[\sum_{h=t-\mathrm{v}}^{t-1} e_h^{(j)2}\right]^{-1}} \qquad [15.5.4]$$

where $e_t^{(i)}$ and v are as defined previously. This approach ensures that all of the estimated weights are greater than or equal to zero.

An alternative approach to calculating the combining weights when seasonal data are considered is developed by McLeod et al. (1986). In this procedure, the model residuals are employed to calculate the residual variance for each season. If two forecasts are to be combined, then the weights are calculated for each season such that

$$w_{1,j} = \frac{\sum_{k=1}^{n}[a_{j+(k-1)s}^{(1)}]^2}{\sum_{k=1}^{n}[a_{j+(k-1)s}^{(1)}]^2 + \sum_{k=1}^{n}[a_{j+(k-1)s}^{(2)}]^2} \qquad [15.5.5]$$

and

$$w_{2,j} = \frac{\sum_{k=1}^{n}[a_{j+(k-1)s}^{(2)}]^2}{\sum_{k=1}^{n}[a_{j+(k-1)s}^{(1)}]^2 + \sum_{k=1}^{n}[a_{j+(k-1)s}^{(2)}]^2} \qquad [15.5.6]$$

where $w_{1,j}$ is the weight assigned to forecasting procedure one for the jth season, $w_{2,j}$ is the weight assigned to forecasting procedure two for the jth season, $a_t^{(i)}$ is the residual at time t for the ith model, n is the number of years of data and s is the number of seasons per year. Since the data are seasonal, the forecast error variance might be expected to be seasonal and, hence, this procedure should account for this seasonality.

15.5.3 Combining Average Monthly Riverflow Forecasts

The thirty average monthly riverflow time series listed in Table 15.3.1 and referred to in Section 15.3.2 are the data sets employed in the experiments for combining forecasts among two models. As is also the situation in Section 15.3.3, the last three years or 36 observations are omitted from each of the data sets. Subsequently, after taking natural logarithms of each time series both PAR/PACF and SARIMA models are fitted to each of the truncated logarithmic sequences. Recall from Section 15.3.3 that PAR/PACF refers to a calibrated PAR model that is identified using the sample periodic PACF. The same 36 one-step-ahead forecasts calculated in Section 15.3.4 for each of these two models and each of the thirty series are employed in the combination experiments reported upon here.

The monthly logarithmic forecasts produced by the PAR/PACF and SARIMA models are combined using some of the procedures outlined previously in Section 15.5.2. Specifically, the combining weights are calculated using [15.5.4] with $v = 3, 6, 9$, and 12. In addition, seasonal combining weights are also determined employing [15.5.5] and [15.5.6]. The combined forecasts are then compared on the basis of MSE's. A summary of the results is presented in Table 15.5.1. The CMB-SEAS entries refer to the combined forecasts produced when separate weights are calculated for each season. The CMB-v entries represent the combined forecasts when the previous v forecast errors are employed to calculate the combining weights. The results show that, in general, the combined forecasts do not constitute an improvement over the PAR/PACF forecasts, regardless of the procedure utilized to calculate the combining weights. This is because the PAR family of models has a better mathematical design for forecasting an average monthly riverflow series like the one in Figure VI.1 while the SARIMA model is more suitable for forecasting series such as those in Figures VI.2 and VI.3. Accordingly, the PAR model forecasts better than the SARIMA model and attempting to combine inferior forecasts with better ones does not improve the situation for the PAR forecasts. Conversely, the SARIMA forecasts are almost always improved by combining them with PAR/PACF forecasts. Finally, a comparison of the various procedures for combining the forecasts seems to indicate that the more information employed to estimate the combining weights the better the forecasts.

Table 15.5.1. Percentage of times model A gives better values for forecasting a series than model B.

Model A (1)	Model B						
	PAR/PACF (%) (2)	SARIMA (%) (3)	CMB-SEAS (%) (4)	CMB-3 (%) (5)	CMB-6 (%) (6)	CMB-9 (%) (7)	CMB-12 (%) (8)
PAR/PACF	0	70	56.7	60	60	56.7	56.7
SARIMA	30	0	20	16.7	20	20	20
CMB-SEAS	43.4	80	0	46.7	43.3	56.7	56.7
CMB-3	40	83.3	53.3	0	50	46.7	33.3
CMB-6	40	80	46.7	50	0	43.3	26.7
CMB-9	43.3	80	43.3	53.3	56.7	0	30
CMB-12	43.3	80	43.3	66.7	73.3	70	0

15.6 AGGREGATION OF FORECASTS

Suppose that one wishes to forecast future average annual riverflows for a given river for which the average monthly and hence also the annual values are known. One approach is to fit a nonseasonal time series model such as an ARMA model to the yearly data and then employ this model for forecasting annual values. Another procedure is to fit an appropriate seasonal model like the PAR model of Sections 14.2.2 and 14.3 to the average monthly series and then utilize this model to forecast the next 12 months. The sum of these 12 monthly forecasts would represent an *aggregated forecast* for the yearly value.

Noakes (1984, Ch. 6) carried out forecasting experiments with riverflow time series to ascertain if aggregated forecasts can improve the accuracy of forecasts determined for a larger time interval. For various yearly and seasonal time series models, Noakes found that for the data sets that he considered, the aggregated forecasts for annual values were generally not as good as those produced by the annual models.

For further research on aggregation of forecasts the reader can refer to Tiao (1972) and Tiao and Wei (1976). Moreover, a discussion on disaggregation and aggregation in time series modelling within the hydrological literature is given in Section 20.5.2.

15.7 CONCLUSIONS

As explained in Section 15.2, MMSE forecasts can be easily calculated for all the seasonal models presented in Part VI. The results of the forecasting experiments of Section 15.3 for 30 monthly riverflow series clearly indicate that PAR models identified using the sample periodic PACF forecast significantly better than SARIMA, deseasonalized, and PAR models identified using techniques other than the sample periodic PACF. When PPAR models are also considered, the forecasting studies of Section 15.4 show that PPAR models also forecast quite well. Finally, forecasts can be combined across models in an attempt to achieve improved forecasts by using procedures described in Section 15.5.2 and applied to monthly riverflow time series in Section 15.5.3.

PROBLEMS

15.1 Select a seasonal time series for which you think may be appropriate to fit a SARIMA model and then carry out the following tasks:

(a) Examine suitable exploratory data analysis graphs for discovering the key statistical characteristics of the time series.

(b) Remove the final year of observations from the time series and then by following the three stages of model construction, fit a SARIMA model to the remaining values.

(c) Calculate the MMSE forecasts and 90% probability limits for the last year of observations to which the model was not fitted. Clearly explain how you perform your calculations.

(d) Plot the MMSE forecasts and 90% probability limits on a graph with the historical observations for the final year. Determine the accuracy of the forecasts and comment upon any interesting findings.

15.2 Follow the instructions in problem 15.1 for a deseasonalized model.

15.3 Carry out the instructions in problem 15.1 for a PAR model.

15.4 Choose a time series to which it seems reasonable to fit SARIMA, deseasonalized and PAR models. For each of these models follow the instructions of question 15.1. Additionally, for the time series under study, compare the forecasting capabilities of the three seasonal models and ascertain if one model forecasts significantly better than another.

15.5 Makridakis et al. (1982) carry out forecasting experiments for a range of models fitted to 1001 time series consisting of yearly, monthly and quarter-monthly economic data sets. After reading their paper, respond to the following questions:

(a) Outline the major findings of their study.

(b) Describe the main steps these authors followed in executing their forecasting experiments and comparing the forecasting results for the various models and data sets.

(c) Explain the commonalities and differences between the procedures used by the authors of the forecasing paper for carrying out their forecasting experiments with those employed in this book.

15.6 Carry out the instructions of the previous question for the paper by Newbold and Granger (1974).

15.7 From your field of study, pick out a set of three or more seasonal time series that are of direct interest to you. After fitting appropriate time series models from Part VI to the first portion of each series, execute forecasting experiments to ascertain which class or classes of models provide the most accurate forecasts. A summary of how to perform a systematic forecasting study is given in Figure 8.3.1.

15.8 Employing procedures described in Section 15.5.2, combine forecasts among pairs of models used in problem 15.7 in order to ascertain if enhanced forecasts can be found. Comment upon any interesting discoveries that you may make.

15.9 Summarize the main research findings of Tiao (1972) as well as Tiao and Wei (1976) on the aggregation of forecasts.

15.10 The aggregation of forecasts is discussed in Section 15.6. Select an average monthly riverflow time series and then do the following:

(a) Fit a PAR model to all but the last three years of the monthly series. Employ this model to forecast the last 36 values. For each of the last 3 years, determine the aggregated forecast for each year.

(b) Fit an ARMA model to the average annual series for which the last 3 years are left out. Employ this calibrated model to forecast the next three years.

(c) Compare the accuracy of the annual forecasts obtained in points (a) and (b) and comment upon the results.

(d) Discuss the annual forecasting results when only one step ahead forecasts are employed in parts (a) and (b).

REFERENCES

In addition to the references listed here, the reader may also wish to refer to forecasting references given at the ends of Chapters 1, 8 and 18.

AGGREGATION OF FORECASTS

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregation of time series. *Biometrika*, 59:525-531.

Tiao, G. C. and Wei, W. S. (1976). Effect of temporal aggregation on the dynamic relationship of two time series. *Biometrika*, 63:513-523.

COMBINING FORECASTS

Armstrong, J. S. and Lusk, E. J. (1983). Commentary on Makridakis time series competition (m-competition). *Journal of Forecasting*, 2:259-311.

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly* (Journal of the Operational Research Society), 10:451-468.

Bordley, R. F. (1982). The combination of forecasts - a Bayesian approach. *Journal of the Operational Research Society*, 33:171-174.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3:197-204.

McLeod, A. I., Noakes, D. J., Hipel, K. W. and Thompstone, R. M. (1987). Combining hydrological forecasts. *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, 113(1):29-41.

Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, Series A*, 137:131-165.

Winkler, R. L. and Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, Series A*, 146:150-157.

FORECASTING

Brocha, M. (1978). Computerized rivers. *Science Dimension*, 10:18-20.

Camacho, F. (1990). Forecasts for Niagara River flows at Ontario Hydro. *Proceedings of the Great Lakes Water Level Forecasting and Statistics Symposium*, held May 17-18, 1990, in Windsor, Ontario, Canada, 73-80.

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.

Granger, C. W. J. and Newbold, P. (1976). Forecasting untransformed series. *Journal of the Royal Statistical Society, Series B*, 38(2):189-203.

International Association of Hydrological Sciences (1980). *Hydrological Forecasting Symposium, Proceedings of the Oxford Hydrological Forecasting Symposium*, held April 15-18, 1980. IAHS-AISH publication No. 129.

Ledolter, J. and Abraham, B. (1981). Parsimony and its importance in time series forecasting. *Technometrics*, 23(4):411-414.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1:111-153.

Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, Series A*, 137:131-165.

Noakes, D. J. (1984). Applied Time Series Modelling and Forecasting. PhD thesis, Dept. of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Noakes, D. J., McLeod, A. I., and Hipel, K. W. (1983). Forecasting experiments with seasonal hydrologic time series models. Technical Report 117-XM-220283, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

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

Thomas, H. A. and Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In Maas, 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, Cambridge, MA.

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. (1983). Transfer function-noise modelling for powerhouse inflow forecasting. *INFOR*, 21:259-269.

Thompstone, R. M., Hipel, K. W. and McLeod, A. I. (1985). Forecasting quarter-monthly riverflow. *Water Resources Bulletin* 21(5):731-744.

STATISTICS AND TIME SERIES ANALYSIS

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Revised Edition.

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

McClave, J. T. (1975). Subset autoregression. *Technometrics*, 17(2):213-220.

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

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

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1:80-83.