CHAPTER 18

FORECASTING WITH TRANSFER FUNCTION-NOISE MODELS

18.1 INTRODUCTION

A *transfer function-noise (TFN)* model can describe the dynamic relationship between a single output series and one or more input series. For example, a TFN model can formally specify the mathematical association existing between riverflows and the temperature and precipitation variables which caused the flows. Furthermore, the remaining noise component can be modelled using an ARMA model. Because its inherent flexible design reflects many types of physical situations that can take place in practice, the TFN model constitutes an important tool for use in water resources and environmental engineering plus many other fields.

In the previous chapter, the TFN model is defined and comprehensive model construction techniques are presented so that the model can be conveniently applied in practice. Moreover, practical applications are given in Chapter 17 to explain clearly how model building is carried out. If one is confronted with a situation where the direction of causality between two series is not clear, the residual cross-correlation function (CCF) of Section 16.2 can be utilized. Additionally, as explained in Section 17.3.1, after the type of causality is established, the results of a residual cross-correlation function study can be employed for deciding upon the parameters to include in a formal mathematical model to describe the relationship between the two series.

A particularly useful and common application of a calibrated TFN noise model is *forecasting*. For instance, forecasts of riverflows based upon other previous flows as well as other hydrological conditions are useful for optimizing the operation of multipurpose reservoir systems. Consequently, the objective of this chapter is to demonstrate the utility of TFN models in forecasting by employing practical applications in hydrology.

In the next section, it is explained how *minimum mean square error (MMSE) forecasts* can be generated using a TFN model. Then, *practical forecasting applications* are presented in the subsequent two sections. The forecasting experiments of Section 18.3 demonstrate that TFN models produce more accurate forecasts than other competing models, including what is called a conceptual hydrological model. The forecasting applications of Section 18.4 explain how forecasts from TFN and other models can be combined in an optimal fashion in an attempt to obtain improved forecasts. In particular, a TFN, periodic autoregressive (PAR) (see Chapter 14) and a conceptual model (see Section 18.3.3) are employed to forecast quarter monthly riverflows. These models all approach the modelling and forecasting problem from three different perspectives and each has its own particular strengths and weaknesses. The forecasts generated by the individual models are combined in an effort to exploit the strengths of each model. The results of this case study indicate that significantly better forecasts can be obtained when forecasts from different types of models are combined. The forecasting findings of Sections 18.3 and 18.4 are based upon research by Thompstone et al. (1985) and McLeod et al. (1987), respectively.

Because TFN models have been found to produce reliable forecasts in applications, they are becoming popular with practitioners. In addition to the forecasting studies described in this book, other documented results of TFN forecasting include contributions in hydrology (Anselmo and Ubertini, 1979; Baracos et al. 1981; Chow et al., 1983; Snorrason et al., 1984; Alley, 1985; Maidment et al., 1985; Olason and Watt, 1986; Fay et al., 1987; Haltiner and Salas, 1988), fish population studies (Stocker and Noakes, 1988; Noakes et al., 1990; Schweigert and Noakes, 1990) as well as many other fields. Moreover, as explained in Section 18.5, TFN models can also be employed for extending time series records, control and simulation.

For forecasting with nonseasonal ARMA and ARIMA models, the reader may wish to refer to Chapter 8. Forecasting experiments are presented in Section 15.5 on the three types of seasonal models from Part VI. These seasonal forecasting studies include experiments on combining forecasts from different seasonal models to try to procure better forecasts.

18.2 FORECASTING PROCEDURES FOR TFN MODELS

18.2.1 Overview

A TFN model describes mathematically how one or more inputs dynamically affect a single output or response variable. In Section 17.2, a TFN model having one input or covariate series is defined in [17.2.5]. Within Section 17.5.2, a TFN model with two or more covariate series is given in [17.5.3].

Intuitively, one would expect that forecasts for the response series should be considerably improved if one uses forecasting information coming from the covariate series. Consequently, the forecasts from a TFN model should be more accurate than those obtained from a separate time series model fitted only to the response series. In fact, the forecasting experiments of Section 18.3 demonstrate that a TFN model forecasts better than other competing time series models as well as a conceptual model. When a response variable can be anticipated by changes in the values of a covariate, economists refer to the covariate as a *leading indicator* for the response. The future net growth in a national economy, for instance, is often anticipated by leading indicators such as trade surplus or deficits, interest rates, unemployment and inflation.

Section 8.2 explains how to calculate minimum mean square error (MMSE) forecasts for nonseasonal ARMA and ARIMA models, while Section 15.2 describes how to compute MMSE forecasts for three types of seasonal models. The purpose of this section is to present procedures for determining MMSE forecasts for various types of TFN models. More specifically, in Section 18.2.2, formulae are given for calculating MMSE forecasts for TFN models having single or multiple inputs, ARMA or ARIMA noise and a deterministic trend component. Moreover, these kinds of TFN models can be fitted to yearly or deseasonalized data sets that may first be transformed using a Box-Cox transformation. In Section 18.2.3, an illustrative forecasting application is presented for clearly explaining how to calculate MMSE forecasts and for demonstrating that a TFN model forecasts more accurately than an ARMA model separately fitted to the response series.

18.2.2 Forecasting Formulae

For convenience of explanation, forecasting formulae are first developed for the case of a TFN model having a single covariate series. As explained below, these formulae can easily be extended for handling situations for which there are two or more input series. Other complications that are discussed in this subsection include how to handle seasonality, differencing and trends when forecasting with a TFN model.

Single Input TFN Model Having ARMA Noise

Derivation of MMSE Forecasts: As in Section 17.2, suppose that a variable X causes a variable Y. Let the observations for X and Y at time t be given by X_t and Y_t, respectively. If the given series are transformed using a transformation such as the Box-Cox transformation in [3.4.30], let the transformed series for X_t and Y_t be denoted as x_t and y_t, respectively. As in [17.2.5], a TFN model for mathematically describing the relationship between x_t and y_t as well as the noise, is written as

$$y_t - \mu_y = \nu(B)(x_t - \mu_x) + N_t \qquad [18.2.1]$$

where μ_y and μ_x are the theoretical means of y_t and x_t, respectively. In the above equation,

$$\nu(B) = \frac{\omega(B)}{\delta(B)} = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_m B^m)}{(1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r)} \qquad [18.2.2]$$

is the transfer function which models the dynamic effects of the input upon the output. If there is a delay time, b, (where b is a positive integer) for x_t to affect y_t, then x_t is replaced by x_{t-b} in [18.2.1]. The noise term, N_t, is assumed to follow an ARMA process as in [3.4.4] such that

$$\phi(B)N_t = \theta(B)a_t$$

or

$$N_t = \frac{\theta(B)}{\phi(B)}a_t = \frac{(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)}{(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)}a_t \qquad [18.2.3]$$

As pointed out later, N_t could also be an ARIMA model when the data are nonstationary.

As in [16.2.3], suppose that x_t can be described using an ARMA model such that

$$\phi_x(B)(x_t - \mu_x) = \theta_x(B)u_t$$

or

$$(x_t - \mu_x) = \frac{\theta_x(B)}{\phi_x(B)}u_t$$

$$= \frac{(1 - \theta_{x,1}B - \theta_{x,2}B^2 - \cdots - \theta_{x,q_x}B^{q_x})}{(1 - \phi_{x,1}B - \phi_{x,2}B^2 - \cdots - \phi_{x,p_x}B^{p_x})}u_t \qquad [18.2.4]$$

By substituting the above into [18.2.1], the TFN model becomes

$$y_t - \mu_y = \frac{\omega(B)\theta_x(B)}{\delta(B)\phi_x(B)}u_t + \frac{\theta(B)}{\phi_{(B)}}a_t$$

$$= \nu^*(B)u_t + \psi(B)a_t \qquad [18.2.5]$$

The transfer function for u_t in [18.2.5] is expanded as

$$\nu^*(B) = \nu_0^* + \nu_1^* B + \nu_2^* B^2 + \cdots$$

where the ν^* weights can be calculated by equating coefficients in the identity

$$\nu^*(B) = \frac{\omega(B)\theta_x(B)}{\delta(B)\phi_x(B)}$$

or

$$\delta(B)\phi_x(B)\nu^*(B) = \omega(B)\theta_x(B) \qquad [18.2.6]$$

As in [3.4.18], the random shock operator is

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \qquad [18.2.7]$$

where the ψ_i weights can be determined using the identity in [3.4.21].

By replacing t by $t+l$ in [18.2.5], the TFN model for the actual value of the response variable at time $t+l$ is

$$\begin{aligned} y_{t+l} - \mu_y = {} & \left(\nu_0^* u_{t+l} + \nu_1^* u_{t+l-1} + \nu_2^* u_{t+l-2} + \cdots \right. \\ & \left. + \nu_l^* u_t + \nu_{l+1}^* \mu_{t-1} + \nu_{l+2}^* u_{t-2} + \cdots \right) \\ & + \left(a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots \right. \\ & \left. + \psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots \right) \end{aligned} \qquad [18.2.8]$$

where l is a positive integer. Let $\hat{y}_t(l)$ be the forecast for y_{t+l} made at origin t. Keeping in mind that only information up to time t can be utilized, let this forecast be written as

$$\begin{aligned} \hat{y}_t(l) - \mu_y = {} & (\nu_l^o \mu_t + \nu_{l+1}^o u_{t-1} + \nu_{l+2}^o u_{t-2} + \cdots) \\ & + (\psi_l^o a_t + \psi_{l+1}^o a_{t-1} + \psi_{l+2}^o a_{t-2} + \cdots) \end{aligned} \qquad [18.2.9]$$

Then, using [18.2.8] and [18.2.9]

$$\begin{aligned} y_{t+l} - \hat{y}_t(l) = {} & \sum_{i=0}^{l-1} (\nu_i^* u_{t+l-i} + \psi_i a_{t+l-i}) \\ & + \sum_{j=0}^{\infty} \left[(\nu_{l+j}^* - \nu_{l+j}^o) u_{t-j} + (\psi_{l+j} - \psi_{l+j}^o) a_{t-j} \right] \end{aligned} \qquad [18.2.10]$$

where $\psi_0 = 1$. Following arguments put forward in Section 8.2.2 for forecasting with an ARMA

model, one can determine the MMSE forecast for the response variable. In particular, the mean square error for the forecast is calculated using [18.2.10] within the expected value given below as

$$E[y_{t+l} - \hat{y}_t(l)]^2 = (\nu_0^{*2} + \nu_1^{*2} + \nu_2^{*2} + \cdots + \nu_{l-1}^{*2})\sigma_u^2 + (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 + \sum_{j=0}^{\infty}[(\nu_{l+j}^{*} - \nu_{l+j}^{o})^2\sigma_u^2 + (\psi_{l+j} - \psi_{l+j}^{o})^2\sigma_a^2] \quad [18.2.11]$$

which is minimized only if $\nu_{l+j}^{o} = \nu_{l+j}^{*}$ and $\psi_{l+j}^{o} = \psi_{l+j}$. Consequently, the MMSE forecast $\hat{y}_t(l)$ of y_{t+l} at origin t is given by the *conditional expectation* of y_{t+l} at time t. Therefore, the MMSE forecast using the TFN model as written in [18.2.9] is simply

$$\hat{y}_t(l) - \mu_y = (\nu_l^{*} u_t + \nu_{l+1}^{*} u_{t-1} + \nu_{l+2}^{*} u_{t-2} + \cdots) + (\psi_l a_t + \psi_{l+1} a_{t-1} + \psi_{l+2} a_{t-2} + \cdots) \quad [18.2.12]$$

Computing MMSE Forecasts: Equation [18.2.12] could be employed for calculating MMSE forecasts for a TFN model having a single input. However, a more convenient way to compute the forecasts is to use the TFN format from [18.2.1] which is written at time $t+l$ as

$$y_{t+l} - \mu_y = \frac{\omega(B)}{\delta(B)}(x_{t+l} - \mu_x) + \frac{\theta(B)}{\phi(B)}a_{t+l} \quad [18.2.13]$$

when the noise is modelled as an ARMA process. To eliminate the operators written in the denominators on the right hand side of the equation, one can multiply both sides of the equation by $\phi(B)\delta(B)$ to obtain

$$\phi(B)\delta(B)(y_{t+l} - \mu_y) = \phi(B)\omega(B)(x_{t+l} - \mu_x) + \delta(B)\theta(B)a_{t+l} \quad [18.2.14]$$

Subsequently, one can multiply together the operators in each term in [18.2.13] and then take conditional expectations to determine the MMSE forecasts. Specifically, in [18.2.13] let

$$\begin{aligned}
\delta^{*}(B) &= \phi(B)\delta(B) = 1 - \delta_1^{*}B - \delta_2^{*} - \cdots - \delta_{p+r}^{*}B^{p+r} \\
\omega^{*}(B) &= \phi(B)\omega(B) = 1 - \omega_0^{*} - \omega_1^{*} - \cdots - \omega_{p+s}^{*}B^{p+s} \\
\theta^{*}(B) &= \delta(B)\theta(B) = 1 - \theta_1^{*}B - \theta_2^{*}B^2 - \cdots - \theta_{q+r}^{*}B^{q+r}
\end{aligned} \quad [18.2.15]$$

One can see that δ_i^{*}, ω_i^{*} and θ_i^{*} coefficients can be easily computed by multiplying together the known operators as defined above. Then, employing square brackets to denote conditional expectations at time t, the MMSE forecast for lead time l is

$$\begin{aligned}
\hat{y}_t(l) - \mu_y &= [y_{t+l}] - \mu_y \\
&= \delta_1^{*}([y_{t+l-1}] - \mu_y) + \delta_2^{*}([y_{t+l-2}] - \mu_y) \\
&\quad + \cdots + \delta_{p+r}^{*}([y_{t+l-p-r}] - \mu_y)
\end{aligned}$$

$$+ \omega_o^*([x_{t+l}] - \mu_x) - \omega_1^*([x_{t+l-1}] - \mu_x)$$
$$- \omega_2^*([x_{t+l-2}] - \mu_x) - \cdots - \omega_{p+s}^*([x_{t+l-p-s}] - \mu_x) + [a_{t+l}]$$
$$- \theta_1^*[a_{t+l-1}] - \theta_2^*[a_{t+l-2}] - \cdots - \theta_{q+r}^*[a_{t+l-q-r}] \qquad [18.2.16]$$

In order to obtain the MMSE forecasts, the rules for iteratively calculating the conditional expectations in [18.2.5] for lead times $l = 1,2,\ldots,$ are as follows:

$$[y_{t+j}] = \begin{cases} y_{t+j} & \text{for } j \le 0 \\ \hat{y}_t(j) & \text{for } j > 0 \end{cases} \qquad [18.2.17a]$$

since y_{t+j} is a known observation for $j \le 0$ and unknown for $j > 0$.

$$[x_{t+j}] = \begin{cases} x_{t+j} & \text{for } j \le 0 \\ \hat{x}_t(j) & \text{for } j > 0 \end{cases} \qquad [18.2.17b]$$

where the forecasts for the input variable are determined using the ARMA model for the x_t series in [18.2.4] according to the forecasting rules laid out in Section 8.2.4 for an ARMA model.

$$[a_{t+j}] = \begin{cases} a_{t+j} & \text{for } j \le 0 \\ 0 & \text{for } j > 0 \end{cases} \qquad [18.2.17c]$$

because a_{t+j} is known for $j \le 0$ and has an expected value of zero for $j > 0$.

Variance of MMSE Forecasts: To obtain the ν_i^* and ψ_i weights for the TFN model as written in [18.2.5], one can employ the identities in [18.2.6] and [3.4.21], respectively. On the right hand side of [18.2.10], the forecast error is given by the first summation component. From the first two terms on the right hand side of [18.2.11], the variance of the forecast error for lead time l is written as

$$V(l) = E[y_{t+l} - \hat{y}_t(l)]^2$$
$$= \sigma_\mu^2 \sum_{j=0}^{l-1} \nu_j^{*2} + \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 \qquad [18.2.18]$$

where σ_μ^2 is the variance of the noise term for the ARMA model fitted to the x_t series in [18.2.4], σ_a^2 is the variance of the ARMA noise term for the TFN model in [18.2.1] and [18.2.12], and the ν_j^* and ψ_i coefficients are determined using the identities in [18.2.6] and [3.4.21], respectively. When the μ_t and a_t series are assumed to be NID(0,σ_μ^2) and NID(0,σ_a^2), respectively, one can conveniently calculate the probability limits for each MMSE forecast. For instance, the 95% probability limits for $\hat{y}_t(l)$ would be $\hat{y}_t(l) \pm 1.96\sqrt{\hat{V}(l)}$ where $\hat{V}(l)$ is estimated using [18.2.18] when the coefficient and noise estimates appropriately replace the theoretical values given on the right hand side of the equation.

Forecasts in the Original Domain: If the y_t or x_t are transformed using a Box-Cox transformation from [3.4.30], the MMSE forecasts calculated above are for the transformed domain. To obtain forecasts in the original units or untransformed domain, one would have to take an inverse Box-Cox transformation as is explained in Section 8.2.7 for the case of an ARMA model fitted to a single series. Keep in mind that both the forecasts and corresponding probability limits in the transformed domain can be determined for the untransformed domain.

Multiple Input TFN Model Having ARMA Noise

The TFN model in [17.2.5] and [18.2.1] has a single covariate or input series x_t. In general, one could have a TFN model with I input series which is written in [17.5.3] as

$$(y_t - \mu_y) = \frac{\omega_1(B)}{\delta_1(B)}(x_{t_1} - \mu_{x_1}) + \frac{\omega_2(B)}{\delta_2(B)}(x_{t_2} - \mu_{x_2}) + \cdots + \frac{\omega_I(B)}{\delta_I(B)}(x_{t_I} - \mu_{x_I}) + \frac{\theta(B)}{\phi(B)}a_t \qquad [18.2.19]$$

To employ this model for forecasting purposes, one follows a procedure similar to that carried out for the TFN model having a single covariate series. In particular, first one must write the TFN model so that there is no operator in the denominator of any term in [18.2.19]. This is accomplished by multiplying [18.2.19] by $\{\delta_1(B)\delta_2(B)\cdots\delta_I(B)\phi(B)\}$. Next, by separately fitting an ARMA model to each of the x_{t_i} series, one calculates the MMSE forecasts for each x_{t_i} series by following the procedure of Section 8.2.4. Thirdly, one iteratively calculates the MMSE forecasts for the response or output series for lead times $l = 1,2,\ldots,$ using the rules in [8.2.16]. Additionally, using a formula similar to that given in [18.2.18] for a TFN having one input, one can determine the probability limits for each forecast. Finally, if the response variable and other input covariates have been transformed using a Box-Cox transformation, one can, if desired, calculate the corresponding forecasts and probability limits in the untransformed domain.

Seasonal TFN Model

As noted in Section 17.2.1, a simple procedure is available for handling seasonal data. Firstly, the output series and each of the input series may be transformed using a Box-Cox transformation in order to cause each time series to be approximately normally distributed. Secondly, assuming that there is approximate stationarity within each season for a given series so that a graph of the series follows a shape similar to that in Figure VI.1 for an average monthly riverflow series, one can deseasonalize the series using a procedure from Section 13.2.2. Next, an appropriate TFN model is fitted to the set of deseasonalized series using the model construction techniques explained in Sections 17.3 and 17.5.3, and an ARMA model is separately developed for each deseasonalized input series by following the model building methods of Part III. Fourthly, by employing the forecasting formulae of Section 8.2.4, MMSE forecasts can be iteratively generated for each deseasonalized input series using the ARMA model fitted to the series. Next, by utilizing the forecasting formulae for TFN models presented in this section as well as the forecasts for the inputs, MMSE forecasts can be iteratively determined for the

response variable for lead times $l = 1,2,\ldots,$. Finally, to obtain forecasts in the untransformed domain, one must first take the inverse deseasonalization transformation of the forecasts and then invoke the inverse Box-Cox transformation. The procedure for forecasting seasonal data using a TFN model is depicted in Figure 18.2.1.

Another approach for handling seasonal data is to employ a periodic TFN model. The interested reader may wish to explore this possibility by answering problem 18.5.

TFN Model Having ARIMA Noise

Suppose that one wishes to fit a TFN model to a nonseasonal response series which has one nonseasonal input series and that these two series are nonstationary. One way to remove this nonstationarity is to introduce the differencing operator of Section 4.3.1 into the noise term of the TFN model so that the noise component follows an ARIMA rather than an ARMA process. Accordingly, the nonstationary version of the TFN model in [18.2.1] is

$$y_t = \frac{\omega(B)}{\delta(B)}x_t + \frac{\theta(B)}{\phi(B)\nabla^d}a_t \qquad [18.2.20]$$

where

$$\nabla^d = (1-B)^d$$

is the differencing operator defined for d taking on values of zero when the data are stationary and positive integers when the data are nonstationary. As exemplified by the examples in Section 4.3.3, usually $d = 1$ or 2 when an ARIMA model is fitted to a single yearly nonstationary time series. Because of the differencing operator in [18.2.20], both the y_t and x_t series are assumed not to have mean levels.

To obtain MMSE forecasts for the TFN model in [18.2.20], the procedure is similar to that for the stationary case. Firstly, one must eliminate operators in the denominator by multiplying [18.2.20] by $\delta(B)\phi(B)\nabla^d$ to obtain

$$\delta(B)\phi(B)\nabla^d y_t = \phi(B)\nabla^d\omega(B)x_t + \delta(B)\theta(B)a_t \qquad [18.2.21]$$

Next, after multiplying together the operators in each term in [18.2.21], one can iteratively calculate the MMSE forecasts by employing the rules in [18.2.17]. Finally, after taking into account the fact that there is a differencing operator, probability limits can be calculated for each forecast using a formula similar to that in [18.2.18].

If one were dealing with seasonal time series that follow graphs similar to those in Figures VI.2 or VI.3, one could possibly model a set of these time series using a TFN model by having a SARIMA noise term. As defined in Section 12.2.1, a SARIMA model contains nonseasonal and seasonal differencing operators to remove nonseasonal and seasonal nonstationarity, respectively. Moreover, the SARIMA model also has seasonal AR and MA operators in addition to the nonseasonal AR and MA operators.

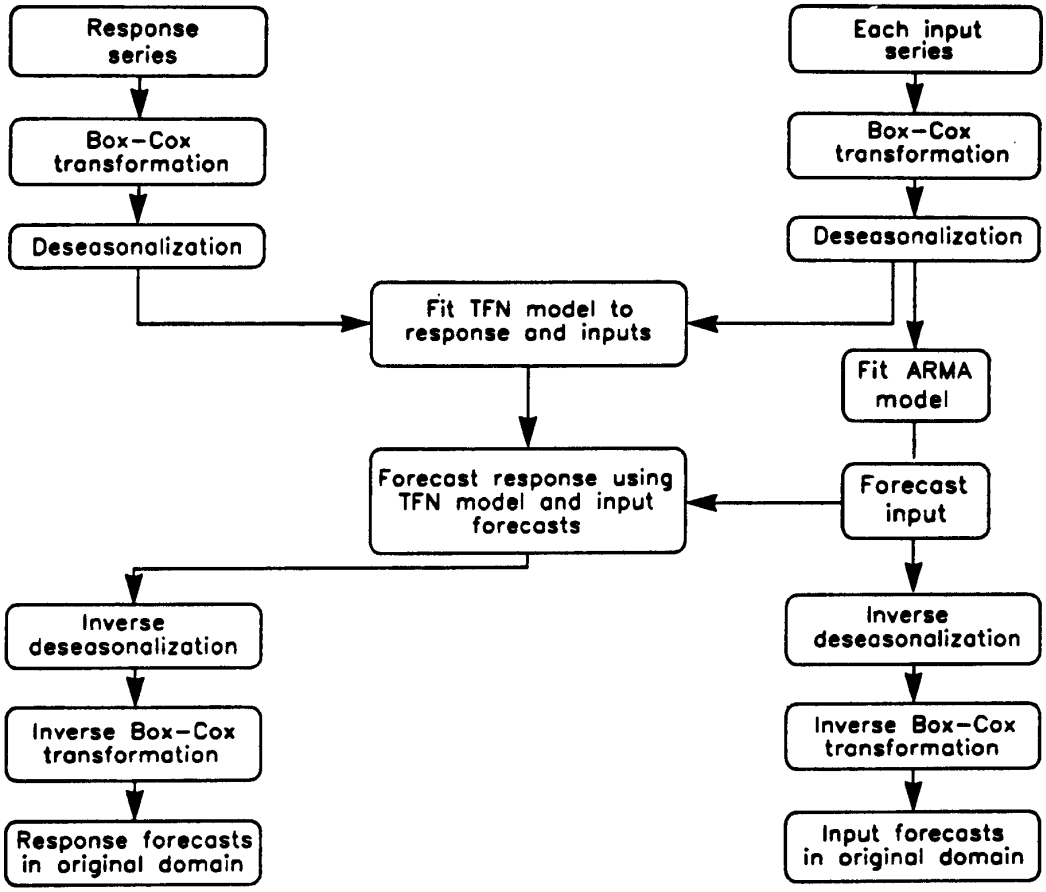

Figure 18.2.1. Forecasting seasonal series using a TFN model.

TFN Model Having a Deterministic Trend

As discussed in Sections 4.5 and 4.6, differencing is designed for removing stochastic trends in a time series. However, differencing may not eliminate a deterministic trend contained in a time series. To explain how a deterministic trend can be modelled and forecasted rewrite [18.2.20] as

$$\nabla^d y_t - \mu_w = \frac{\omega(B)}{\delta(B)}\nabla^d x_t + \frac{\theta(B)}{\phi(B)}a_t \qquad [18.2.22]$$

where μ_w is a level in the y_t series that the differencing cannot eliminate. By multiplying [18.2.22] by $\delta(B)\phi(B)$, one obtains

$$\delta(B)\phi(B)\nabla^d y_t = \theta_0 + \phi(B)\omega(B)\nabla^d x_t + \delta(B)\theta(B)a_t \qquad [18.2.23]$$

where

$$\theta_0 = \delta(B)\phi(B)\mu_w$$
$$= \delta(1)\phi(1)\mu_w$$

In the expression for θ_0, one replaces each B by unity in the two operators because $B^k\mu_w = \mu_w$ for $k = 0,1,2,\cdots$.

The procedure for calculating MMSE forecasts using [18.2.23] is the same as before except for the θ_0 term on the right hand side of [18.2.23]. Consequently, the rules for conditional expectations in [18.2.17] are employed to determine the MMSE forecasts for the response and input series, keeping in mind that $[\theta_0] = \theta_0$ in [18.2.23].

18.2.3 Application

The Red Deer River is a tributary of the South Saskatchewan (abbreviated as S.Sask.) River which flows eastwards from the Rocky Mountains across the Canadian prairies. In Section 17.4.2, a TFN model is constructed for describing the influence of the deseasonalized logarithmic Red Deer riverflows upon the deseasonalized logarithmic S.Sask. riverflows. From [17.4.1], this calibrated TFN model is written as

$$y_t = (0.572 + 0.238B)x_t + \frac{(1 - 0.494B)}{(1 - 0.856B)}a_t \qquad [18.2.24]$$

where y_t and x_t are the deseasonalized logarithmic S.Sask. and Red Deer riverflows, respectively, and $\hat{\sigma}_a^2 = 0.310$.

To write the model in [18.2.24] in a convenient form to calculate MMSE forecasts, first multiply the difference equation by the operator $(1 - 0.856B)$ to get

$$(1 - 0.856B)y_t = (1 - 0.856B)(0.572 + 0.238B)x_t + (1 - 0.494B)a_t$$

or

$$y_t = 0.856y_{t-1} + 0.572x_t - 0.252x_{t-1} - 0.204x_{t-2} + a_t - 0.494a_{t-1}$$

By replacing t by $t+l$ and taking conditional expectations in the above equations, the MMSE

forecast for lead time l is

$$\hat{y}_t(l) = 0.856[y_{t+l-1}] + 0.572\hat{x}_t(l) - 0.252[x_{t+l-1}] - 0.204[x_{t+l-2}] + [a_{t+l}] - 0.494[a_{t+l-1}] \quad [18.2.25]$$

For the case of a one-step-ahead forecast where $l = 1$ the above equation becomes

$$\hat{y}_t(1) = 0.856[y_t] + 0.572\hat{x}_t(1) - 0.252[x_t] - 0.204[x_{t-1}] + [a_{t+1}] - 0.494[a_t]$$

$$= 0.856y_t + 0.572\hat{x}_t(1) - 0.252x_t - 0.204x_{t-1} - 0.494a_t \quad [18.2.26]$$

Notice in [18.2.25] and [18.2.26], one requires MMSE forecasts for the x_t series. Consequently, one must separately fit an ARMA model to the x_t series and then use this model to generate MMSE forecasts for the x_t series. When an ARMA model is separately designed for describing the x_t series for the deseasonalized logarithmic Red Deer flows, the most appropriate ARMA model is found to be an ARMA(1,1). The estimated ARMA(1,1) model for the x_t series is given as

$$(1 - 0.845B)x_t = (1 - 0.292B)u_t \quad [18.2.27]$$

where u_t is the innovation series at time t and $\hat{\sigma}_u^2 = 0.482$. By substituting $t+l$ for t and taking conditional expectations in [18.2.27], the formula for iteratively generating MMSE forecasts for x_t is

$$\hat{x}_t(l) = 0.845[x_{t+l-1}] + [u_{t+l}] - 0.292[u_{t+l-1}] \quad [18.2.28]$$

To obtain the one-step-ahead MMSE forecast in [18.2.28] simply replace l by unity to obtain

$$\hat{x}_t(1) = 0.845[x_t] + [u_{t+1}] - 0.292[u_t]$$

$$= 0.845x_t - 0.292u_t \quad [18.2.29]$$

To calculate MMSE forecasts for the y_t series in the TFN model in [18.2.25], one can employ [18.2.28] to determine the MMSE forecasts for the x_t series which are needed as input forecasts in [18.2.25]. Consider the case where one wishes to find $\hat{y}_t(1)$ using [18.2.26]. Firstly, $\hat{x}_t(1)$ is found by utilizing [18.2.29] and then $\hat{x}_t(1)$ is substituted into [18.2.26] to get $\hat{y}_t(1)$. From [18.2.18], the variance of the one-step-ahead MMSE forecast error for $\hat{y}_t(1)$ is given theoretically as

$$V(1) = \omega_0^2\sigma_u^2 + \sigma_a^2 \quad [18.2.30]$$

where the estimate is calculated as

$$\hat{V}(1) = (0.572)^2 0.482 + 0.310 = 0.468$$

To calculate the lead one MMSE forecast for the original untransformed series, the deseasonalization and logarithmic transformations must be taken into account. Accordingly, from time t, the lead one MMSE forecast $\hat{y}_t(1)$ for the untransformed series is estimated using

$$\hat{y}_t(1) = \exp\left[\hat{y}_t(1)\hat{\sigma}_m + \hat{\mu}_m + \frac{1}{2}(0.468)\hat{\sigma}_m^2\right] \qquad [18.2.31]$$

where $\hat{\mu}_m$ and $\hat{\sigma}_m$ are the estimated mean and standard deviation calculated using [13.2.4] and [13.2.5], respectively, for the month that is currently connected with time t+1 of the y_t series. As explained in Section 8.2.7, the last term in the exponent in [18.2.31] is the correction required for producing the MMSE forecast in the untransformed domain.

When calculating MMSE forecasts for the y_t series using the TFN model in [18.2.25], information from the input series is used in the forecast calculation. Consequently, a priori, one would expect a TFN model to forecast more accurately than an ARMA model that is separately fitted to the response series. For the case of the y_t series representing the deseasonalized logarithmic flows of the S. Sask. River, the most appropriate model to fit to this series is an ARMA(1,1) model which is calibrated as

$$(1 - 0.819)y_t = (1 - 0.253)v_t \qquad [18.2.32]$$

where v_t is the innovation series at time t and $\hat{\sigma}_v^2 = 0.507$. Notice that the variance of the noise has a value of $\hat{\sigma}_v^2 = 0.507$ for the ARMA model in [18.2.32] and a magnitude of $\hat{\sigma}_a^2 = 0.310$ for the TFN model in [18.2.24]. Consequently, the TFN model provides a better fit to the available information than the single ARMA model and has a residual variance which is about 40% smaller. By replacing t by $t+l$ and taking conditional expectations in [18.2.32], the formula for iteratively determining MMSE forecasts for y_t using an ARMA model is

$$\hat{y}_t(l) = 0.819[y_{t+l-1}] + [v_{t+l}] - 0.253[v_{t+l-1}] \qquad [18.2.33]$$

To ascertain the one-step-ahead MMSE forecast in [18.2.33], simply assign l a value of one to get

$$\begin{aligned}\hat{y}_t(1) &= 0.819[y_t] + [v_{t+l}] - 0.253[v_t] \\ &= 0.819y_t - 0.253v_t \end{aligned} \qquad [18.2.34]$$

In the untransformed domain, the lead one MMSE forecast, $\hat{y}_t(1)$ is calculated using

$$\hat{y}_t(1) = \exp\left[\hat{y}_t(1)\hat{\sigma}_m + \hat{\mu}_m + \frac{1}{2}(0.507)\sigma_m^2\right] \qquad [18.2.35]$$

where $\hat{y}_t(1)$ is determined using [18.2.34], $\hat{\sigma}_v^2 = 0.507$ as in [18.2.32], and $\hat{\sigma}_m$ as well as $\hat{\mu}_m$ are the same as in [18.2.31].

As would be expected the TFN model for the response variable produces more accurate forecasts than an ARMA model separately fitted to the same series. More specifically, when the TFN model in [18.2.26] along with the inverse transformation in [18.2.31] are employed for obtaining lead one MMSE forecasts in the untransformed domain, the mean square error for all months in 1963 is about 20% less than for the forecasts obtained using the ARMA(1,1) model in [18.2.34] and [18.2.35] for the output series. Consequently, the input series in the TFN model acts as a leading indicator to significantly improve the accuracy of the MMSE forecasts of the output series.

18.3 FORECASTING QUARTER-MONTHLY RIVERFLOWS

18.3.1 Overview

A key problem in the operation of a water resources system is the forecasting of natural inflows to the various reservoirs in the system. It is increasingly recognized that time series analysis is of considerable practical use in dealing with this problem. The current section demonstrates the practical importance of this methodology by examining the use of the TFN models of Chapter 17 to forecast natural inflows in the Lac St. Jean reservoir, a major component of the multi-reservoir hydroelectric system operated by Alcan Smelters and Chemicals Ltd. in the province of Quebec, Canada.

The electricity generated by this system is used at Alcan's aluminum smelter in Arvida, Quebec. In order to insure a constant and adequate supply of power, it is necessary to schedule releases from the reservoir in an optimum fashion. Thus, forecasts of the quarter monthly inflows into the reservoir are required so that the desired outflow and hydraulic head are available for power generation.

The forecasting experiments presented in this section were originally presented by Thompstone et al. (1985). The output for the TFN model used in the study are the quarter-monthly (i.e., near-weekly) natural inflows to the Lac St. Jean reservoir. The covariate series for the TFN model are rainfall and snowmelt, the latter being a novel derivation from daily rainfall, snowfall and temperature series. It is clearly demonstrated in Section 18.3.2 using the residual variance and the AIC (see Section 6.3) that modelling is improved as one starts with a deseasonalized ARMA model (Chapter 13) of the inflow series and successively adds transfer functions for the rainfall and snowmelt series. It is further demonstrated that the TFN model is better than a PAR model (Chapter 14) of the inflow series. The split-sample experiments are used in Section 18.3.4 to compare one-step-ahead forecasts from this TFN model with forecasts from other stochastic models as well as with forecasts from a so-called conceptual hydrological model described in Section 18.3.3 (i.e., a model which attempts to mathematically simulate the physical processes involved in the hydrological cycle). It is concluded that the TFN model is the preferred model for forecasting the quarter-monthly Lac St. Jean inflow series.

18.3.2 Constructing the Time Series Models

The application involves a series of quarter-monthly natural inflows in m^3/s to the Lac St. Jean reservoir in the Province of Quebec. One of the covariate series selected for possible incorporation in a dynamic model of the inflow was rainfall. The quarter-monthly rainfall series in mm/day represents the spatial average of rainfall over the entire 57,000 km^2 watershed (Thompstone, 1983). The second covariate series was a rather novel quarter-monthly snowmelt series in mm/day, and it was calculated using logic extracted directly from the conceptual hydrologic model which is described in the next subsection. Data were available for the years 1953-82 (Thompstone et al., 1980) but only the years 1953-79 were used in fitting the models described in this section. The other three years were reserved for the split-sample forecasting experiment described in Section 18.3.4.

Following Section 13.3, the identification, estimation and diagnostic checking stages of model construction were used to build a deseasonalized ARMA model for the Lac St. Jean inflow series. Several models were examined and, based on the AIC, the standard errors (SE's)

of estimation of the model parameters, and the results of diagnostic checking, the following ARMA (3,1) model was chosen:

$$(1 - 1.430B + 0.626B^2 - 0.113B^3)z_t^{(\lambda)} = (1 - 0.653B)a_t \qquad [18.3.1]$$

where $\lambda = 0.0$ indicates the given monthly series is transformed by taking natural logarithms as in [3.4.30], the z_t inflow series is deseasonalized by subtracting seasonal means and dividing by seasonal standard deviations, and a_t is the approximately normally distributed white noise innovation having a mean of zero and a variance of $\hat{\sigma}_a = 0.685$. All AR and MA parameters were more than two SE's from zero, and thus are statistically significant. Diagnostic checking of the residuals confirmed them to be uncorrelated, homoscedastic and approximately normally distributed (see Chapter 7). The AIC of the model was found to have a value of 13,771.24.

Both the empirical approach and the Box and Jenkins procedure were used to identify TFN models (see Section 17.3.1) for forecasting Lac St. Jean inflows using first the rainfall series, then the snowmelt series, and then both series together. The rainfall series was deseasonalized by subtracting the seasonal mean from each observation, and then dividing this by the seasonal standard deviation. The sample autocorrelation function (ACF) calculated using [2.5.9] showed the resulting series to be white noise. The sample CCF (cross correlation function) between the deseasonalized rainfall and the deseasonalized, logarithmic inflow series is shown in Figure 18.3.1. The 95% confidence limits in this figure are calculated under the assumption that the sample CCF values are NID(0,n^{-1}) where n is the length of the series (see Section 16.2.2). Because riverflows are caused by rainfall, the values of the sample CCF are significantly large for zero and negative values in Figure 18.3.1. As a result of the extra large value at lag -1, the order of the operator in the numerator of the transfer function in [18.2.2] is $m = 1$. The dying out effect for negative lags suggests that $r = 1$ for the operator in the denominator of the TFN in [18.2.2]. This form of model was fit to the data, and the resulting noise was identified as being ARMA(2,1). Consequently, the TFN model which gives the relationship between deseasonalized rainfall, x_{t1}, and deseasonalized logarithmic inflow, y_t, was selected to be:

$$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t1} + N_t \qquad [18.3.2]$$

where

$$N_t = \frac{(1 - \theta_1 B)}{(1 - \phi_1 B - \phi_2 B^2)}$$

Table 18.3.1 provides the MLE's (maximum likelihood estimates) of the parameters and their corresponding SE's. Diagnostic checking showed the residuals to be uncorrelated and approximately normally distributed. The AIC and residual standard deviation for the rainfall and inflow TFN model were found to be 13,159.76 and 0.583, respectively. These values compare with 13,771.24 and 0.685, respectively, for the deseasonalized inflow ARMA(3,1) model. Thus, the inclusion of the rainfall series into the modelling has improved the accuracy of the model for the inflow series.

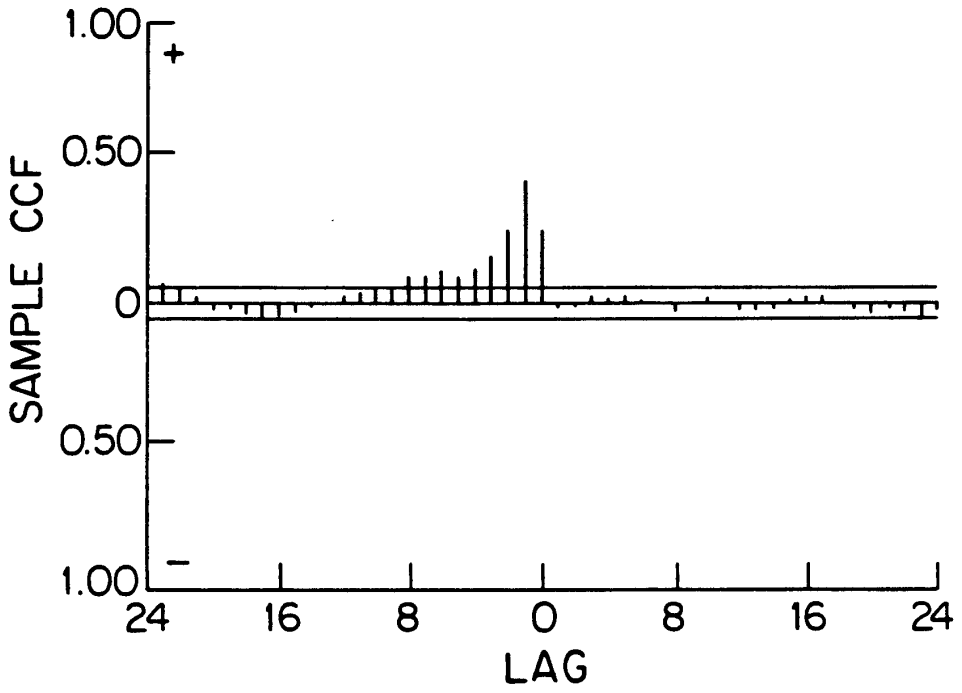

Figure 18.3.1. Sample CCF between deseasonalized, logarithmic inflow and deseasonalized rainfall series along with the 95% confidence limits.

Table 18.3.1. Parameter estimates and SE's for rainfall to inflow TFN model.

Parameters	MLE's	SE's
δ_1	0.608	0.034
ω_0	0.257	0.016
ω_1	-0.277	0.018
ϕ_1	1.410	0.173
ϕ_2	-0.472	0.127
θ_1	0.762	0.163

For the case of the snowmelt series, the selected deseasonalization involved only the subtraction of the seasonal mean from each observation. The following AR(2) model was identified and fitted to the deseasonalized series:

$$(1 - \phi_{1,2}B - \phi_{2,2}B^2)x_{t2} = a_{t2} \qquad [18.3.3]$$

where the estimates of the parameters and their corresponding of SE's are given in Table 18.3.2. Diagnostic checking showed that the residuals were uncorrelated and approximately normally

distributed.

Table 18.3.2. Parameter estimates and SE's for the AR(2) model of snowmelt.

Parameters	MLE's	SE's
$\phi_{1,2}$	0.267	0.027
$\phi_{2,2}$	-0.156	0.027

In accordance with the Box and Jenkins identification procedure of Section 17.3.1, the y_t output series was filtered, using [17.3.5] to obtain the estimated AR operator in [18.3.3], to produce the filtered output

$$\hat{\beta}_t = (1 - 0.267B + 0.156B^2)y_t \qquad [18.3.4]$$

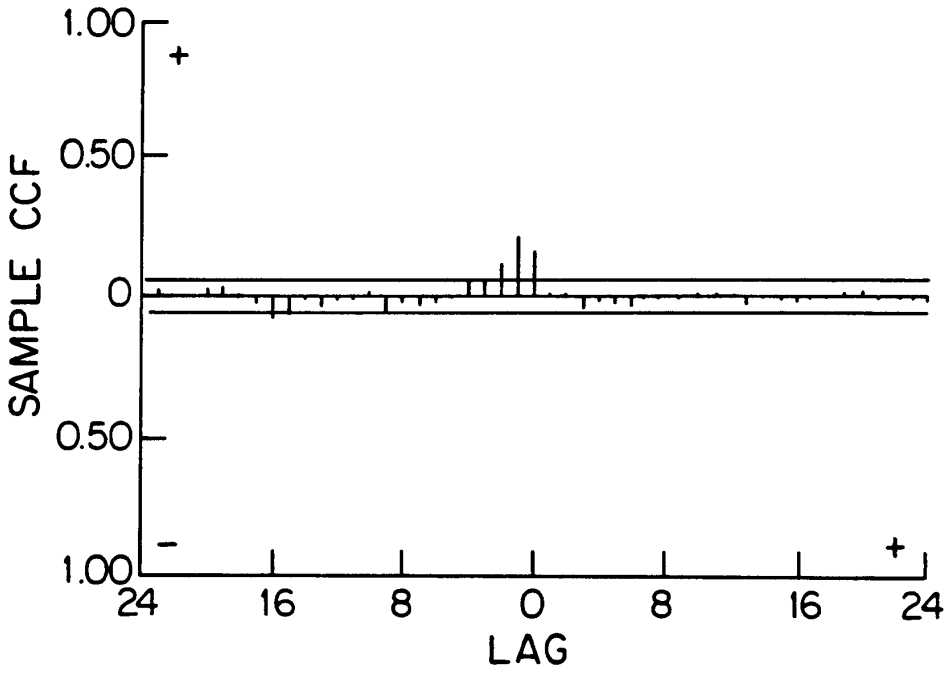

Figure 18.3.2. Sample CCF between the filtered inflow and prewhitened deseasonalized snowmelt.

The sample CCF between the prewhitened deseasonalized snowmelt series and the transformed output series is shown in Figure 18.3.2. This CCF suggested that the form of the transfer function be $r = 1$ and $m = 1$ in [18.2.2]. Such a model was fitted to the data, and the remaining noise was identified as AR(1). The TFN model chosen to relate deseasonalized snowmelt, x_{t2}, and deseasonalized logarithmic inflow, y_t, was therefore:

$$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t2} + N_t \qquad [18.3.5]$$

where

$$N_t = \frac{a_t}{(1 - \phi_1 B)}$$

and the estimates of the parameters and their SE's are as given in Table 18.3.3. The residuals were shown to be independent and approximately normally distributed. The AIC and residual standard deviation for this snowmelt to inflow TFN model were found to be 13,495.36 and 0.664, respectively. These results suggest that a model of inflows including the relationship with snowmelt is better than a model without snowmelt, but that the rainfall series is of more use than the snowmelt series in explaining inflow.

Table 18.3.3. Parameter estimates and SE's for snowmelt to the inflow TFN model.

Parameters	MLE's	SE's
δ_1	0.541	0.098
ω_0	0.113	0.015
ω_1	-0.083	0.018
ϕ_1	0.717	0.019

In order to further improve the modelling of the Lac St. Jean inflows, a TFN model including both the rainfall and snowmelt covariate series was constructed. The form of the transfer functions in [18.3.2] and [18.3.5] was conserved (i.e., $r = m = 1$), a model was estimated, and an ARMA(2,1) model was identified for the resulting noise series. The final model for explaining the deseasonalized logarithmic inflow series, y_t, as a function of the deseasonalized rainfall, x_{t1}), and snowmelt, x_{t2}, series was thus:

$$y_t = \frac{(\omega_{0,1} - \omega_{1,1} B)}{(1 - \delta_{1,1} B)} x_{t1} + \frac{(\omega_{0,2} - \omega_{1,2} B)}{(1 - \delta_{1,2} B)} x_{t2} + N_t \qquad [18.3.6]$$

where

$$N_t = \frac{(1 - \theta_1 B)}{(1 - \phi_1 B - \phi_2 B^2)} a_t$$

The estimates of the parameters and their SE's are given in Table 18.3.4 for the transfer functions, and in Table 18.3.5 for the noise term.

Table 18.3.4. Parameter estimates and SE's (in brackets) of estimates for the transfer functions in [18.3.6].

Series	j	$\delta_{1,j}$	$\omega_{0,j}$	$\omega_{1,j}$
Deseasonalized Rainfall	1	0.625 (0.033)	0.233 (0.016)	-0.269 (0.018)
Deseasonalized Snowmelt	2	0.579 (0.090)	0.102 (0.013)	-0.046 (0.017)

Note: Parenthetical figure is SE of estimation.

Diagnostic checking of the residuals from the fitted model in [18.3.6] suggested they were normally distributed. Figure 18.3.3 shows a plot of the values of the residual autocorrelation function (RACF) and their 95% confidence intervals, defined in Section 7.3.2. Because all of the values of RACF except one fall within the 95% confidence limits, the residuals are white. The large value at lag 26 is probably due to chance and not the lack of a suitable model. Further diagnostic checking involved cross correlation functions. Figure 18.3.4 shows the cross correlations between the deseasonalized rainfall series and the residuals for the TFN model in [18.3.6], while Figure 18.3.5 shows the values of the CCF between residuals of the AR(2) deseasonalized snowmelt series in [18.3.3] and the residuals of [18.3.6]. Because the values of the CCF in Figures 18.3.4 and 18.3.5 fall within the 95% confidence interval, the noise term in the TFN model is not correlated with the prewhitened input series.

The AIC for the TFN model in [18.3.6] is 13,074.37, and the residual standard deviation is 0.562. These two measures confirm that the use of both the rainfall and snowmelt covariate series better explains the inflow series than the employment of either of the series individually. Table 18.3.6 provides a summary comparison of the AIC values and the residual standard deviations of the four models of the Lac St. Jean uncontrolled inflows developed in this section. Note that it can be shown theoretically that the MMSE forecasts from the TFN model of [18.3.6] are more accurate than those from the deseasonalized ARMA model. This fact is confirmed by the forecasting experiment described in Section 18.3.4.

Finally, in Section 14.6 a PAR model was fitted to the Lac St. Jean quarter-monthly inflow series. The AIC of this model was calculated as 13,681.61, and this suggested it was preferable to the deseasonalized ARMA model, but not as good as any of the TFN models. Nevertheless it was retained for use in the forecasting experiment described in Section 18.3.4.

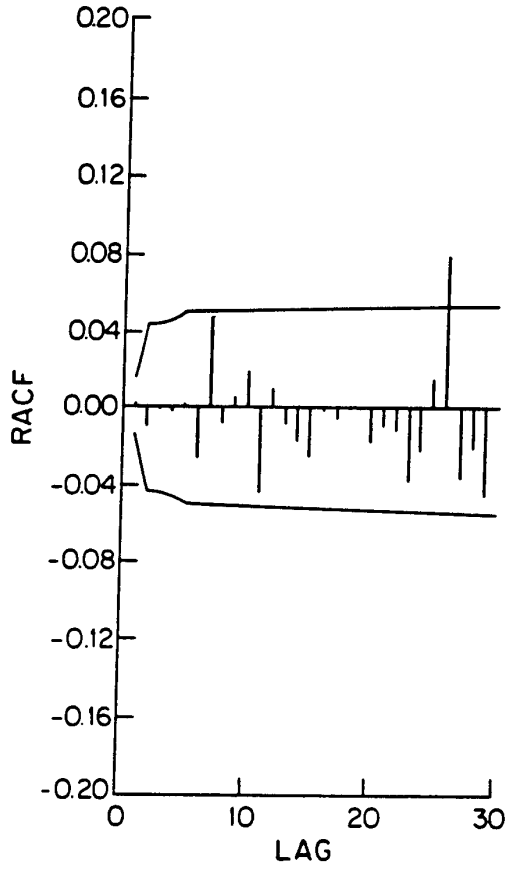

Figure 18.3.3. RACF and 95% confidence interval for the TFN model in [18.3.6].

Table 18.3.5. Parameter estimates and SE's for the noise model in [18.3.6].

Parameters	MLE's	SE's
ϕ_1	1.311	0.123
ϕ_2	-0.382	0.092
θ_1	0.712	0.113

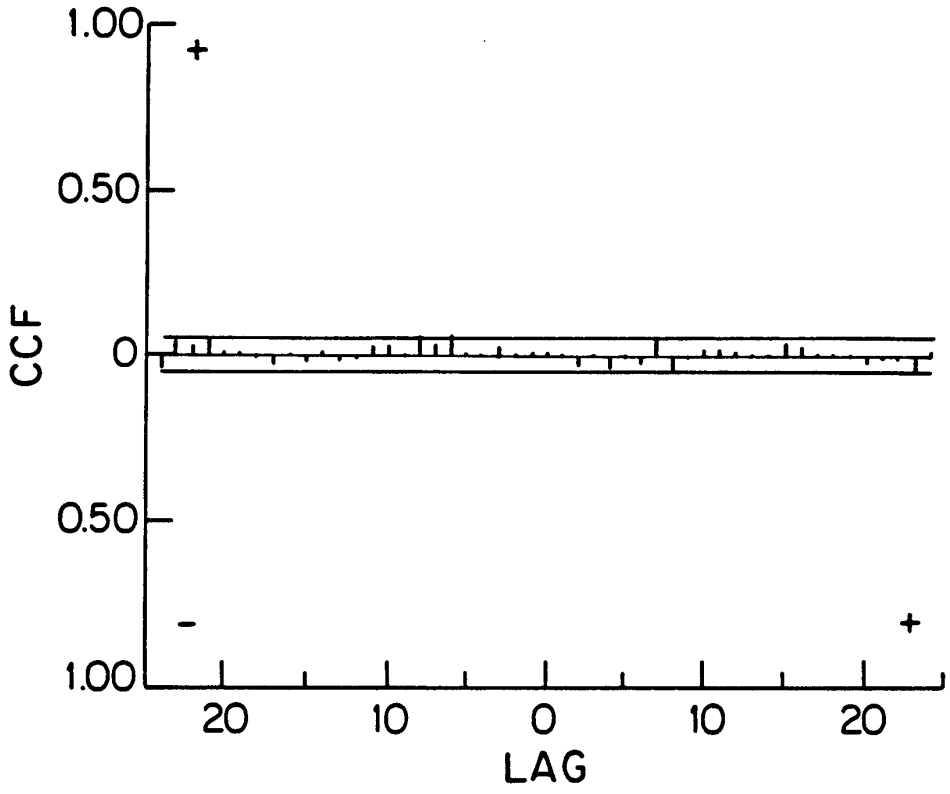

Figure 18.3.4. CCF between the deseasonalized rainfall and residuals of the TFN model in [18.3.6] along with the 95% confidence interval.

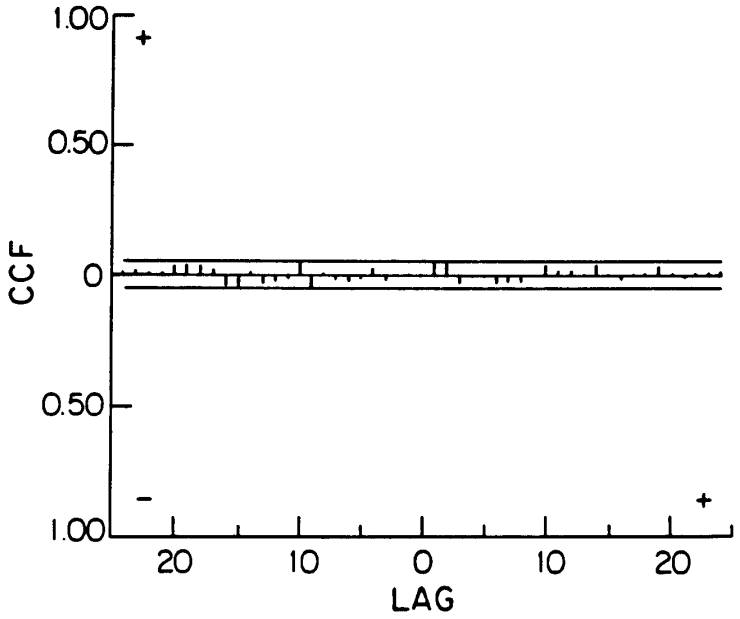

Figure 18.3.5. CCF between residuals of AR(2) deseasonalized snowmelt series and residuals of the TFN model in [18.3.6] along with the 95% confidence interval.

Table 18.3.6. Comparisons of AIC and $\hat{\sigma}_a$ values for the deseasonalized ARMA and TFN models.

Input Series	m	s	Noise	AIC	$\hat{\sigma}_a$
-	-	-	(3,1)	13,771.24	0.685
Deseasonalized Rainfall	1	1	(2,1)	13,159.76	0.583
Deseasonalized Snowmelt	1	1	(1,0)	13,495.36	0.664
Deseasonalized Rainfall Deseasonalized Snowmelt	1 1	1 1	(2,1) ·	13,074.37 ·	0.562 ·

18.3.3 Conceptual Hydrological Model

A realtime daily hydrological forecasting system has been developed for use in the operational management of the hydroelectric system operated by Alcan Smelters and Chemicals Ltd., in the Saguenay-Lac St. Jean region of Quebec. The forecasting system (Thompstone et al., 1981) uses a lumped parametric conceptual hydrological model to simulate the relationship

between daily meteorological conditions and natural inflows to various reservoirs. When the forecasting system is executed, recent meteorological conditions are represented using recent measurements at meteorological stations, and future meteorological conditions are represented using meteorological forecasts provided by the Atmospheric Environment Service of Environment Canada. The basic strategy in the selection of a conceptual hydrological model was to choose a simple and flexible model in preference to more elaborate models, provided no significant improvement in the accuracy of the forecasts could be obtained by the more complex models.

There exists a multitude of conceptual models which have been used in operational hydrological forecasting, each model having its particular strengths and weaknesses (World Meteorological Organization, 1975). The conceptual model chosen for the Alcan forecasting system was originally developed by S.I. Solomon and Associates (1974), and subsequently modified by Kite (1978), the modified model being called the Water Resources Branch model. It has undergone further modifications since inclusion in the Alcan system. A detailed description of the model and the reasons it was chosen are contained in Thompstone (1983) and references therein.

The realtime daily hydrological forecasting system which uses the conceptual hydrological model provides hydrological forecasts based on meteorological forecasts and long term daily meteorological statistics (Thompstone et al., 1981). This system, referred to as PREVIS, has been operational since March, 1979, and it can be executed on a daily basis to provide hydrological forecasts for seven days into the future. The meteorological forecasts have been obtained, interpreted and entered into the forecasting system only on weekdays. Consequently, meteorological forecasts were not available for use in the proposed forecasting study.

In order to provide a basis for comparison of forecasts from the conceptual hydrological model, it was decided that observed meteorological conditions would be used in place of the meteorological forecasts and long term statistics. In other words, the conceptual hydrological model was used in the simulation mode rather than the forecasting mode. Thus, results of the forecasting study are biased in favour of the forecasts from the PREVIS system.

In using the PREVIS system, it has been recognized that the model generally follows the trends of inflows, but during certain periods is consistently higher or lower than the observed inflows. Consequently, an ad hoc smoothing of the raw hydrological forecasts was introduced into the system. The inflow forecast for the next seven days is adjusted by adding the average error of simulated versus observed inflows for the previous seven days. During the spring period, since inflows vary relatively rapidly, the smoothing period is reduced to the previous three days. In order to approximate this crude smoothing, a second set of so called forecasts from the PREVIS system was developed by adjusting the inflow forecast for the next quarter-month period by the error for the previous quarter-month period. These forecasts are labelled herein as PREVIS/S.

Note that in order to compare forecasts from the PREVIS and PREVIS/S models in the same domain as forecasts from the other models, these former forecasts are transformed using natural logarithms. This is necessary since the Pitman (1939) correlation test (see Section 8.3.2) used to compare mean squared errors of forecasts is based on the forecast errors being approximately normally distributed.

18.3.4 Forecasting Experiments

In order to compare the forecasting abilities of the deseasonalized ARMA models, TFN model, PAR model and conceptual hydrological model, a split-sample approach was adopted whereby one-step-ahead quarter-monthly forecasts were generated for three years of data, from the beginning of 1980 to the end of 1982. Data from these years were not used in either building the time series models or in calibrating the conceptual hydrological model.

Until recently, a great deal of effort had been devoted to the advancement of forecasting procedures while relatively little research had been devoted to developing methods for evaluating the relative accuracy of the forecasts produced by the different procedures (Thompstone et al., 1985; Noakes et al., 1985, 1988). Granger and Newbold (1973, 1977) have provided useful comments concerning the evaluation of forecasts and the costs of errors. The mean square error (MSE) is a cost function which is intuitively simple to understand and has been widely used in previous forecasting studies. It is the MSE and its square root, the standard error of forecast, which are used herein to compare the competing forecasting models. Various forecasting tests are discussed in detail in Section 8.3.2 and utilized in forecasting experiments carried out in Chapters 8, 15 and 18.

Noakes et al. (1985, 1988) have underlined the importance of not simply ranking models according to the MSE's of competing procedures. In their study, they used the test of Pitman (1939) and a likelihood ratio test as well as a nonparametric test to compare the one-step-ahead forecasts from different models (see also Sections 8.3.4, 15.3 and 15.4). Since the tests led to essentially the same conclusions, and the Pitman test is computationally less demanding, it has been adopted for the current research.

In order to describe the Pitman (1939) test, which is also presented in Section 8.3.2, let $e_{1,t}$ and $e_{2,t}$ ($t = 1,2,\ldots,L$) denote the one-step-ahead forecast errors for models 1 and 2 respectively. Then, the null hypothesis from [8.3.2] is

$$H_o: MSE(e_{1,t}) = MSE(e_{2,t}) \qquad [18.3.7]$$

where $MSE(e) = \langle e^2 \rangle$, and $\langle . \rangle$ denotes expectation. The alternative hypothesis, H_1, is the negation of H_0.

As explained in Section 8.3.2 just after [8.3.2], for Pitman's test, let $S_t = e_{1,t} + e_{2,t}$ and $D_t = e_{1,t} - e_{2,t}$. Pitman's test is equivalent to testing if the correlation, r, between S_t and D_t is significantly different from zero. Therefore, provided $L > 25$, H_0 is significant at the 5% level if $|r| > 1.96/\sqrt{L}$.

The results of the forecasting study are summarized in Tables 18.3.7 and 18.3.8. Table 18.3.8 shows the root mean squared errors (RMSE's) of the forecasts of the logarithmic series for the five different models. The model with the smallest RMSE is the TFN model, while the second best model is the deseasonalized ARMA model. The worst forecasts are provided by the PREVIS model, while the PREVIS/S model is second worst.

Table 18.3.7. RMSE's of forecasts for the logarithmic quarter-monthly Lac St. Jean uncontrolled inflows from 1980 to 1982.

Models	RMSE's
ARMA/DES	0.298
PAR	0.301
PREVIS	0.389
PREVIS/S	0.354
TFN	0.278

Table 18.3.8. Correlation test statistics and forecast errors for the forecasts for the quarter-monthly logged Lac St. Jean uncontrolled inflows from 1980 to 1982.

Models	ARMA/DES	PAR	PREVIS	PREVIS/S	TFN
ARMA/DES*		0.0296 (=)	0.2675 (+)	0.1814 (+)	0.0902 (=)
PAR*	0.0296 (=)		0.2561 (+)	0.1704 (+)	0.1000 (=)
PREVIS	0.2675 (-)	0.2561 (-)		0.0995 (=)	0.3225 (-)
PREVIS/S	0.1814 (-)	0.1704 (-)	0.995 (=)		0.2421 (-)
TFN	0.902 (=)	0.1000 (+)	0.2421 (+)	0.2421(+)	

(1) Table shows $|r|$.

(2) Difference in MSE's of forecasts 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.

Table 18.3.8 examines the statistical significance of differences in the mean squared errors of forecasts from the various models. Using a 5% significance level, it is concluded that each of the time series models is better than or equal to the PREVIS and PREVIS/S models. There is no significant difference in forecasts from the ARMA/DES, PAR, and TFN models. However, since the TFN model has the smallest RMSE of forecasts and is favoured with respect to the AIC and residual variance, it is recommended that it be adopted for forecasting the Lac St. Jean inflow on a quarter-monthly basis. The physical relationship known to exist between rainfall, snowmelt and inflow reinforces this recommendation. Furthermore, a comparison of the RMSE's of forecasts in the inflow domain for which the flows are not logarithmic confirms that forecasts from the TFN model are preferable to forecasts from the conceptual model (RMSE of 512.30 as opposed to 625.85).

18.3.5 Conclusions

The TFN model described in [18.3.6] provides an effective means of forecasting quarter-monthly inflows to the Lac St. Jean reservoir based on rainfall and snowmelt. The most recent statistical techniques and an understanding of the physical processes involved are used to identify, estimate and verify a reasonable model. Both the empirical approach and the Box and Jenkins approach are useful in model identification (see Section 17.3.1). The MAICE procedure (Section 6.3) indicates the TFN model with both covariate series is better than a deseasonalized ARMA model, PAR model or TFN model with only one or another of the covariate series. The split-sample forecasting experiments of Section 18.3.4 demonstrate that the full TFN model provides better forecasts than a particular conceptual hydrological model. Consequently, the TFN model is the preferred model for forecasting the quarter-monthly Lac St. Jean inflow series. It is interesting to note that Chow et al. (1983) also found flood forecasts from a TFN model to be as reliable as forecasts generated from a complex conceptual model. Hence, they concluded that TFN models provide an attractive alternative to conceptual models for use in realtime flood forecasting.

18.4 COMBINING HYDROLOGICAL FORECASTS

18.4.1 Overview

Often a variety of models can be fitted to a given data set. For example, in Section 18.3, time series models consisting of TFN, PAR (Chapter 14) and deseasonalized ARMA (Chapter 13) models, plus two related conceptual models, are fitted to a hydrological time series. Each of these calibrated models can then be employed for generating forecasts for the series. Although one model may produce more accurate forecasts than others in the long run, it may not do so in every instance. Consequently, one may wish to improve the forecasts by combining forecasts from two or more models in accordance to their relative performances.

The objective of this section is to show how better forecasts can be obtained when TFN forecasts are combined with other types of forecasts. In particular, a TFN, PAR and two similar conceptual models are employed to forecast quarter monthly riverflows, as is done in Section 18.3. These models all approach the modelling and forecasting problem from three different perspectives and each has its own particular strengths and weaknesses. The forecasts generated by the individual models are combined in an effort to exploit the strengths of each model. The results of this case study indicate that significantly better forecasts can be obtained when forecasts from different types of models are combined. In particular, the best forecasts are obtained when TFN and PAR forecasts are optimally combined. These forecasting experiments are also reported by McLeod et al. (1987).

Formulae for combining forecasts in an optimal manner from competing models are presented in Section 15.5.2. Additionally, forecasting experiments are presented in Section 15.5.3 for combining forecasts for monthly riverflows using SARIMA (Chapter 12) and PAR (Chapter 14) models. Because the SARIMA model is not well designed for modelling monthly riverflows for which there is stationarity within each season (see the introduction to Part VI and Section 12.1), combining forecasts from this model with the better forecasts from the PAR model does not produce improved forecasts.

18.4.2 Combination Forecasting Experiments

The data used in this section are identical to those employed in the forecasting experiments of Section 18.3. More specifically, the quarter-monthly inflows for the Lac St. Jean reservoir are utilized. Recall from Section 18.3.1, that accurate quarterly-monthly forecasts for riverflows are required so that Alcan can optimally generate hydro-electrical power for use in its aluminum smelters.

Thirty years of quarter-monthly riverflows are available from 1953 to 1982, inclusive. As is done in Sections 18.3.2 and 18.3.3, models are fitted to the first twenty-seven years of the data and then used to forecast the one-step-ahead forecasts for the last three years. Prior to fitting models to the riverflows, the data are first transformed using natural logarithms.

The calibrated models used in the study are already described in Sections 18.3.2 and 18.3.3. In particular, the finite difference equation for the best TFN model is given in [18.3.6] while its parameter estimates are listed in Tables 18.3.4 and 18.3.5. The most appropriate PAR model is identified using graphs of the sample periodic ACF and PACF (defined in Section 14.3.2). The two versions of the conceptual model used in the combination forecasting study are the PREVIS and PREVIS/S conceptual models described in Section 18.3.3.

The RMSE's of the logarithmic forecast errors are presented in Table 18.3.7. As can be seen, the TFN model has the smallest RMSE of all the models considered. As such, this value will be used as a basis for comparison of the various techniques employed to combine the individual forecasts.

Notice that the deseasonalized ARMA and PAR models have almost the same RMSE's in Table 18.3.7. Because the PAR model is generally better to use than the deseasonalized model for modelling seasonal riverflows for which there are sufficient data (see discussion in Part VI), the deseasonalized ARMA model is not employed in the combination experiments of this section.

The equations for combining forecasts are given in Section 15.5.2. In this study, the weights for combining the individual forecasts were calculated using both [15.5.2] and [15.5.4] with $\upsilon = 4$, 8 and 12. Since the model residuals were not employed, the first υ forecasts were combined using equal weights. The weights were then recalculated for each subsequent forecast using the previous υ forecast errors.

The forecasts from the four models were combined in a pairwise fashion with the exception of the two conceptual models (PREVIS and PREVIS/S). The resulting RMSE's of the combined forecasts using [15.5.2] to calculate the combining weights are given in Table 18.4.1. The subscripts associated with the RMSE's indicate the number of previous forecast errors that were employed to calculate the weights. For example, when the previous four forecast errors were used to combined the TFN and PAR forecasts, the resulting RMSE was 0.142. In most cases, the greater the number of previous forecast errors employed to calculate the weights, the smaller the resulting combined RMSE.

Table 18.4.1. RMSE's of the combined quarter monthly forecasts with combining weights calculated using [15.6.2].

Model Combinations	$RMSE_4$	$RMSE_8$	$RMSE_{12}$
TFN - PAR	0.142	0.120	0.119
TFN - PREVIS	0.787	0.524	0.418
TFN - PREVIS/S	0.994	0.318	0.271
PAR - PREVIS	0.243	0.229	0.222
PAR - PREVIS/S	0.217	0.186	0.187

The smallest RMSE was obtained when the TFN and PAR forecasts were combined using the previous 12 forecast errors to calculate the weights. The resulting RMSE was less than half the value of the smallest RMSE for the individual models suggesting that significant benefits can be obtained by combining the forecasts from these two models. Conversely, the largest RMSE's were found when the TFN forecasts were combined with the PREVIS or PREVIS/S forecasts. Only when the previous 12 forecast errors were employed to calculate the weights did the combined TFN and PREVIS/S forecasts yield a smaller RMSE than the best individual model. Even then, the difference was only in the third decimal place.

The resulting RMSE's of the combined forecasts when [15.5.4] was employed to calculate the combining weights are given in Table 18.4.2. In this case, only one combination had a larger RMSE than the best individual model. Once again, the smallest RMSE was found when the TFN and PAR forecasts were combined using the previous 12 forecast errors to calculate the combining weights. The largest RMSE's were found when the TFN forecasts were combined with the forecasts from the two conceptual models. These RMSE's did, however, represent a significant improvement when compared to the RMSE's obtained when [15.5.2] was used to calculate the combining weights. In the previous case, poor estimates of Σ in [15.5.3] resulted in the calculation of one negative weight and one weight greater than one. As a result, the corresponding RMSE's were more than three times as large as the RMSE of the best individual model. It is therefore recommended that, unless reasonably good estimates of Σ can be obtained, the suboptimal estimates of the combining weights calculated using [15.5.2] be employed.

As a test of combining forecasts from more than two models, the forecasts produced by the TFN, PAR and PREVIS/S models were combined using equal weights. The resulting RMSE was 0.136. Although this does not represent the lowest RMSE, even this naive combination of forecasts produced a RMSE which was less than half the RMSE of the best individual model.

Table 18.4.2. RMSE's of the combined quarter-monthly forecasts with combining weights calculated using [15.5.4].

Model Combinations		$RMSE_4$	$RMSE_8$	$RMSE_{12}$
TFN	- PAR/PACF	0.146	0.124	0.122
TFN	- PREVIS	0.275	0.251	0.247
TFN	- PREVIS/S	0.283*	0.252	0.250
PAR/PACF	- PREVIS	0.244	0.230	0.222
PAR/PACF	- PREVIS/S	0.214	0.187	0.188

*Larger RMSE than TFN forecast error in Table 18.3.7.

18.4.3 Conclusions

Combining economic forecasts from various models has become fairly common practice. However, the case studies presented in Sections 18.4.2 and 15.5.2 as well as by McLeod et al. (1987) represent the first reported experiments dealing with the combination of riverflow forecasts. Combining forecasts from conceptual models, a TFN model and a PAR model resulted in a significant reduction in the RMSE's of the forecasts. These three models approach the modelling problem from three distinctly different perspectives. The relative strengths of each model were enhanced by combining the individual forecasts. Thus, based upon the results of this case study, it would appear that significant improvements in forecasting performance can be obtained when the forecasts from different types of models are combined.

18.5 RECORD EXTENSIONS, CONTROL AND SIMULATION

18.5.1 Overview

The main objectives of this chapter are to explain how reliable forecasts can be calculated using TFN models and to demonstrate how forecasting can be conveniently carried out in practice using the hydrological forecasting experiments of Sections 18.2.3, 18.3 and 18.4. The purpose of this section is to outline how TFN models can be employed for three other kinds of applications: extensions of hydrologic records, control and simulation.

18.5.2 Record Extensions

Using natural time series records from the Arctic, Baracos et al. (1981) explain how hydrometric records can be extended using TFN models. In particular, weather records have been kept in the Arctic for a much longer period of time than have hydrometric or riverflow measurements. Based on a knowledge of the dynamic relationship between riverflow series and meteorologic series, it is possible to give an estimate of the values the hydrometric series is likely to have taken during the period when weather data are available, but before flow records were kept. This may be thought of as an artificial extension of the hydrometric record and can be considered to be a type of *back forecasting*. The true values of the unmeasured flows can of course never be obtained by this method, but likely values, given the covariate meteorologic input series, can be calculated. These estimates are simply the output of the TFN model with the noise term set to its conditional expectation of zero.

Baracos et al. (1981) develop ARMA, TFN and intervention models (see Chapter 19) for modelling 16 average monthly riverflow series as well as precipitation and temperature series from the Northwest Territories in the Canadian Arctic. The data sets are available from the Water Survey of Canada which is part of Environment Canada in Ottawa. To explain how riverflow records can be extended using meteorological inputs, consider the TFN model developed for the flows of the Tree River. Average monthly flows for the Tree River are available for 8 years from the start of 1969 to the end of 1976. However, the two meteorologic input series consisting of precipitation and temperatures from the Coppermine weather station are 44 years in length and span the years from the start of 1933 to the end of 1976. For the years in which the riverflows overlap with the meteorologic data, a TFN model can be developed to model how the meteorologic inputs dynamically affect the riverflow output series. The TFN model can then be employed for extending or back forecasting the riverflow series for the years during which there are only meteorological records.

The calibrated TFN model for the Tree River is written as

$$y_t = 0.0012x_{t1} + 0.04x_{t2} - 0.031Bx_{t3} + \frac{1 - 0.32B + 0.25B^8}{1 + 0.57B}a_t \qquad [18.5.1]$$

where

y_t is the Tree River series which is first transformed by taking natural logarithms and then deseasonalized by removing the monthly means for the logarithmic series using [13.2.2].

x_{t1} is the Coppermine rainfall series which is deseasonalized by subtracting the appropriate monthly mean from each observation. Snowmelt is included as part of the rainfall series. In order to produce a plausible representation of snowmelt input to a riverflow series, the monthly snowfalls are summed over each winter, and then the total snowfall for the winter is introduced as a pulse input to the rainfall series during the first month that the mean temperature rises above zero Celsius for each year. Snowfalls that occur during months when the mean temperature is above zero Celsius are assumed to have melted immediately, and are added to the rainfall series rather than to the winter's snow accumulation.

x_{t2} is the Coppermine temperature series which is deseasonalized by removing monthly means. Because the temperature is below zero in the winter and hence does not melt the snow, the values from January, February, March, November and December are set to zero.

x_{t3} is input series containing the deseasonalized temperature only for the month of April. All other months are set equal to zero. The reason for including the x_{t3} series in the third term on the right hand side of [18.5.1] is because for the month of April there is a large negative cross correlation at lag one between the prewhitened Tree riverflows and the Coppermine temperature series.

a_t is the noise term for the TFN model which is NID$(0,\sigma_a^2)$.

To employ the calibrated TFN model for extending the riverflows, the conditional expectation of the noise is assumed to be zero and hence one uses only the dynamic component on the right hand side of [18.5.1] to calculate y_t as

$$y_t = 0.0012x_{t1} + 0.04x_{t2} - 0.031Bx_{t3} \qquad [18.5.2]$$

By substituting in known values of x_{t1}, x_{t2} and x_{t3} in [18.5.2], one can determine y_t's for any desired values of t. Subsequently, to find the values of the flows in the original untransformed domain one simply takes the inverse deseasonalization and logarithmic transformation of the generated y_t's from [18.5.2].

Using the above procedure, the average monthly flows for the Tree River can be predicted for any period during the years for which meteorological records exist from 1933 to 1976. By utilizing graphical and numerical results, Baracos et al. (1981) demonstrate that the predicted flows using [18.5.2] produce reasonable results. In particular, during the time period for which the flows are known, from 1969 to 1976, the predicted flows are close to the known historical flows.

Following a similar procedure to the one described in this section, Beauchamp et al. (1989) extend daily riverflow records of a downstream station based upon a TFN noise model that connects the downstream flows to a longer upstream time series of daily riverflows. They also employ regression analysis for extending the same riverflows. However, they point out that the regression model was found to have a significant amount of correlation in the residuals which the TFN could eliminate, since the noise in a TFN model can be modelled as an ARIMA model.

Snorrason (1986) employs a TFN model to extend seasonal riverflow records for a river in Iceland. A longer temperature series constitutes the input to the TFN model which has the riverflows as the output.

18.5.3 Control

This chapter deals mainly with employing TFN models for forecasting or predicting the future values of the response variable. As pointed out by Young (1984, p. 104), another important application area of TFN models is designing control and management schemes for the system that is currently being studied. In the chemical industry, for example, TFN models are employed extensively for scientifically controlling processes for optimally producing a wide range of chemical products. The key reason why TFN models are ideally suited for control purposes is that they mathematically describe how the inputs dynamically affect the response in the presence of correlated noise.

In a control problem, one often wishes to keep a response variable as close as possible to a target value in a system subject to the inputs and noise. One could attempt to design control schemes which minimize an overall measure of error at the output such as the mean square error. As explained by Box and Jenkins (1976, Chapter 12), one can categorize control procedures into three main domains - feedforward control, feedback control and a mixture of these two. In *feedforward control*, one or more sources of disturbances (inputs) are measured and these observations can be employed for compensating for potential deviations in the output. Because input into the system is used to control the output of the system, this is referred to as feedforward control. On the other hand, in some applications the only information available about the existence of the input disturbances is the deviation from the target which they cause in the response. If only this deviation is utilized for deciding upon how to adjust the system, the action is called *feedback control*. A combination of the aforementioned two methods of control is referred to as *feedforward-feedback control*.

For a detailed discussion of discrete control schemes, the reader may wish to refer to Part IV of Box and Jenkins (1976). Certainly, the design of control schemes has many potential applications in water resources and environmental engineering. For example, to maximize the hydroelectrical output of a system of reservoirs, good control and management plans are required. The efficient operation of a sewage treatment facility that handles both industrial and residential liquid wastes poses many interesting control problems.

18.5.4 Simulation

Besides forecasting, a TFN model can, of course, also be employed for *simulation* purposes. To simulate with a TFN model, it is most convenient to use the model as given in [18.2.13] or [18.2.20] where appropriate multiplications have been made so that no operators appear in the denominator in any term on both sides of the equation. For explanation purposes, consider the TFN model having one input series and ARMA noise which is written in [18.2.13] for time t as

$$\phi(B)\delta(B)(y_t - \mu_y) = \phi(B)\omega(B)(x_t - \mu_x) + \delta(B)\theta(B)a_t \qquad [18.5.3]$$

The main steps to follow in simulating with a TFN model are:

1. By employing the ARMA model that is separately fitted to the x_t series in [18.2.4], use the simulation techniques of Section 9.3 or 9.4 to simulate the x_t's.
2. To simulate the a_t's needed in the second term on the right hand side of [18.5.3], employ an appropriate method from Section 9.2.3 to simulate the a_t's which are NID$(0,\sigma_a^2)$.
3. If starting values are needed for the y_t's in [18.5.3], these can be generated using a separate ARMA model fitted to the y_t series in conjunction with a simulation technique from Section 9.3 or 9.4.
4. Use the simulated x_t and a_t series from steps 1 and 2, respectively, as well as the starting values for y_t from step 3, in the TFN model in [18.5.3] to simulate the y_t series.

18.6 CONCLUSIONS

The TFN model of Chapter 17 is particularly well designed for use in the natural sciences such as hydrology and water quality modelling. This is because the TFN model in [18.2.18] and [17.5.3] can formally describe, using a finite difference equation, the dynamic relationships existing between a single output series and one or more input series. For instance, the TFN model in [18.3.6] describes how the input or covariate series consisting of rainfall and snowmelt cause riverflows. Furthermore, the correlated noise in the model can be modelled using an ARMA(2,1) model.

Because the structure of the TFN model in an equation such as [18.3.6] realistically reflects the physical relationships among the variables, one would expect the model to provide good forecasts. In addition, since the TFN model incorporates more information into its structure by means of the input series, one would think that better forecasts should be obtained using this model. Indeed this is exactly what happens. The forecasting experiments of Section 18.3 demonstrate that the TFN noise model forecasts seasonal riverflows better than its competitors. In particular, the TFN model of [18.3.6] provides more accurate forecasts of the quarter-monthly

riverflows into Lac St. Jean than the deseasonalized ARMA, PAR or either of the two conceptual models. Moreover, as shown by the forecasting results in Section 18.4, even better forecasts can be obtained when the TFN forecasts are optimally combined with those provided by the PAR model.

Forecasting experiments with a range of nonseasonal models are furnished in Section 8.3. For a description of forecasting experiments with the seasonal models of Part VII, the reader can turn to Sections 15.3 and 15.4. Experiments with combinations of forecasts from seasonal models are also given in Section 15.5.3.

In addition to handling multiple input series, the TFN model of Part VII can be expanded to take care of other situations that arise in practice. More specifically, the intervention model of Part VIII constitutes a general type of TFN model that can be used to model the effects of external interventions upon the mean level of a series, estimate missing observations and also to describe the dynamic relationships between multiple input series and a single output. Besides Part VIII, further interesting applications of intervention and TFN modelling are presented in Chapter 22.

PROBLEMS

18.1 Suppose that a TFN model is written as

$$y_t - \mu_y = \frac{\omega_0 B}{(1 - \delta_1 B)}(x_t - \mu_x) + \frac{(1 - \theta_1 B)}{(1 - \phi_1 B)}a_t$$

For this model, carry out the following tasks:

(a) Using formulae, clearly explain how to iteratively calculate MMSE forecasts for lead times $l = 1,2, \ldots,$.

(b) Derive the formula for determining the variance of the forecast error for $\hat{y}_t(l)$.

18.2 Carry out the instructions of Problem 18.1 for the following TFN model having three input series.

$$(y_t - \mu_y) = \frac{\omega_1(B)}{\delta_1(B)}(x_{t1} - \mu_{x1}) + \frac{\omega_2(B)}{\delta_2(B)}(x_{t2} - \mu_{x2})$$

$$+ \frac{\omega_3(B)}{\delta_3(B)}(x_{t3} - \mu_{x3}) + \frac{\theta(B)}{\phi(B)}a_t$$

18.3 For the TFN model written below, execute the instructions given in Problem 18.1.

$$y_t = \frac{(\omega_0 - \omega_1 B)B^2}{(1 - \delta_1 B - \delta_2 B^2)}x_t + \frac{(1 - \theta_1 B)}{(1 - \phi_1 B)(1 - B)}a_t$$

18.4 Consider the situation where one has monthly observations for both a response variable, Y_t, and an input series, X_t. Each series is first transformed by taking natural logarithms and then deseasonalized using [13.2.3]. Next, the TFN model fitted to the resulting nonseasonal series is

$$y_t = \frac{\omega_0 - \omega_1 B}{(1 - \delta_1 B)} x_t + \frac{1}{(1 - \phi_1 B)} a_t$$

where y_t and x_t are the deseasonalized response and covariate series, respectively. By employing suitable equations, explain how to calculate MMSE forecasts for y_t and x_t as well as forecasts for the original Y_t and X_t series.

18.5 In Chapter 14, periodic models are defined for application to a single seasonal time series for which there are s seasons per year. Assuming one input series and an ARMA noise term, write down the difference equations to define a periodic TFN model. Explain the drawbacks of this type of model and how these disadvantages could be overcome.

18.6 The field of econometrics deals with the development of statistical and stochastic methods for application to economic data. Find an article in the econometrics literature where one or more leading indicators are used to forecast some aspect of the economy. Outline the procedure that is employed and explain how you think it could be improved.

18.7 Within the water resources literature, locate a paper where TFN modelling is used for forecasting. Briefly describe how the forecasting study was carried out and point out any interesting facts that you discover.

18.8 Often an overly complex model does not forecast as accurately as a much simpler time series model such as a TFN model. Explain why you think this could happen. Find a paper in a field which is of interest to you where a TFN model provides better forecasts than a more complicated model, such as a conceptual model. Describe the main findings of the paper and emphasize the most interesting results.

18.9 Fit a TFN model to a nonseasonal data set where you have a response series and one input series. Employ the calibrated model to calculate MMSE forecasts for lead times from 1 to 12. Plot the forecasts along with the 95% confidence limits.

18.10 Carry out the instructions of the previous question for two monthly time series.

18.11 Find two seasonal series designated by y_t and x_t for the response and input series, respectively. Omit the last three years of the data set and then fit SARIMA (Chapter 12), deseasonalized ARMA (Chapter 13) and PAR (Chapter 14) models to the y_t series. Also, fit a TFN model to the y_t and x_t series for which the last three years of the data are not used for calibration purposes. Following the approach of Section 18.3, determine which of the four models produces the best one step ahead MMSE forecasts of the last three years of the series.

18.12 Employ the combination methods of Section 18.4 to determine if the accuracy of the forecasts obtained using the four models in Problem 18.11 can be improved by optimally combining the forecasts. Clearly explain your findings.

18.13 Select a response series for which you have a longer record for one or more input series. Fit a TFN model to the data for the time period during which the response and input series overlap. By following the procedure of Section 18.5.2, use the TFN model to extend the response series for the time interval for which only the input data are known.

18.14 Snorrason (1986) employs a TFN model to extend seasonal riverflow data from a glaciated basin in Iceland. A longer temperature record is used as the input to the TFN model while the output is the riverflows. His record extension technique is a slightly different variation of the one presented in Section 18.5.2. Describe the data extension approach of Snorrason and compare it to the one presented in Section 18.5.2.

18.15 Using equations and diagrams, explain the feedforward, feedback and mixed control schemes put forward by Box and Jenkins (1976, Chapter 12). Describe how each of these schemes could be possibly employed for modelling a water resources or environmental system.

18.16 Fit a TFN model to a data set for which you have one input series and, of course, a single response series. Follow the procedure of Section 18.5.4 to simulate a sequence of values that has the same length as the historical series. Clearly explain all of the steps that you follow and compare a graph of the simulated y_t sequence to a plot of the historical response series.

REFERENCES

For references about causality and how to fit TFN models, the reader may wish to refer to the references in Chapters 16 and 17, respectively. Moreover, further references on forecasting are listed at the ends of Chapters 1, 8 and 15.

CONTROL

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

Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis: An Introduction.* Springer-Verlag, Berlin.

CONCEPTUAL MODELS

Kite, G. W. (1978). Development of a hydrologic model for a Canadian watershed. *Canadian Journal of Civil Engineering,* 5(1):126-134.

Solomon, S. I. and Associates Limited (1974). *Preliminary Analysis of Potential Remote Sensing Applications in Hydrology.* Report to Environment Canada.

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., Bouchard, S., Pilon, P. J. and Bergeron, R. (1981). Real-time daily hydrological forecasting for a multi-reservoir hydroelectric system. In *Proceedings of the Canadian Society for Civil Engineers Fifth National Hydrotechnical Conference*, pages 37-57, Fredericton, New Brunswick.

World Meteorological Organization (1975). Intercomparison of conceptual models used in operational hydrological forecasting. Operational Hydrology Report 7, Secretariat of the World Meteorological Organization, Geneva, Switzerland.

COMBINING FORECASTS

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

DATA SETS

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

FORECASTING TESTS

Granger, C. W. J. and Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, 5:35-47.

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

Noakes, D. J., Hipel, K. W., McLeod, A. I., Jimenez, J. and Yakowitz, S. (1988). Forecasting annual geophysical time series. *International Journal of Forecasting*, 4:103-115.

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

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

RECORD EXTENSION

Baracos, P. C., Hipel, K. W. and McLeod, A. I. (1981). Modelling hydrologic time series from the Arctic. *Water Resources Bulletin*, 17(3):414-422.

Beauchamp, J. J., Downing, D. J. and Railsback, S. F. (1989). Comparison of regression and time-series methods for synthesizing missing streamflow records. *Water Resources Bulletin* 25(5):961-975.

Snorrason, A. (1986). Analysis of river flow data from glaciated basin in Iceland. In: *Analysis of Hydrologic Processes, Proceedings of the Fourth International Hydrology Symposium*, Eds. Shen, H. W., Obeysekera, J. T. B., Yevjevich, V. and DeCoursey, D. G., held at Colorado State University, Fort Collins, Colorado, July 15-17, 1985. Published by the Engineering Research Center, Colorado State University, 651-663.

TRANSFER FUNCTION-NOISE FORECASTING

Alley, W. M. (1985). Water balance models in one-month-ahead streamflow forecasting. *Water Resources Research*, 21(4):597-606.

Anselmo, V. and Ubertini, L. (1979). Transfer function-noise model applied to flow forecasting. *Hydrological Sciences Bulletin*, 24:353-359.

Chow, K. C. A., Watt, W. E. and Watts, D. G. (1983). A stochastic-dynamic model for real time flood forecasting. *Water Resources Research*, 19(3):746-752.

Fay, D. M., Watt, W. E. and Watts, D. G. (1987). A stochastic real-time spring flood forecasting system for Carman, Manitoba. *Canadian Journal of Civil Engineering*, 14(1):87-96.

Haltiner, J. P. and Salas, J. D. (1988). Short-term forecasting of snowmelt runoff using ARMAX models. *Water Resources Research* 24(5):1083-1089.

Maidment, D. R., Miaou, S. P. and Crawford, M. M. (1985). Transfer function models of daily urban water use. *Water Resources Research*, 21(4):425-432.

Noakes, D. J., Welch, D. W., Henderson, M. and Mansfield, E. (1990). A comparison of preseason forecasting methods for returns of two British Columbia sockeye salmon stocks. *North American Journal of Fisheries Management*, 10(1):46-57.

Olason, T. and Watt, W. E. (1986). Multivariate transfer function-noise model of river flow for hydropower operation. *Nordic Hydrology*, 17:185-202.

Schweigert, J. K. and Noakes, D. J. (1990). Forecasting Pacific herring (clupea harengus pallasi) recruitment from spawner abundance and environmental information. *Proceedings of the International Herring Symposium*, Anchorage Alaska, 373-387.

Snorrason, A., Newbold, P. and Maxwell, W. H. C. (1984). Multiple input transfer function-noise modeling of river flow. In Maxwell, W. H. C. and Beard, L. R., Editors, *Frontiers in Hydrology*, pages 111-126. Water Resources Publications, Littleton, Colorado.

Stocker, M. and Noakes, D. J. (1988). Evaluating forecasting procedures for predicting Pacific herring (clupea harengus pallasi) recruitment in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(6):928-935.

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